Available online at www.sciencedirect.com

ScienceDirect

Current Opinion in Chemical Biology

Recent advances in *N*- and *C*-terminus cysteine protein bioconjugation

Richard J. Spears and Vijay Chudasama

Abstract

Advances in the site-specific chemical modification of proteins, also referred to as protein bioconjugation, have proved instrumental in revolutionary approaches to designing new protein-based therapeutics. Of the sites available for protein modification, cysteine residues or the termini of proteins have proved especially popular owing to their favorable properties for site-specific modification. Strategies that, therefore, specifically target cysteine at the termini offer a combination of these favorable properties of cysteine and termini bioconjugation. In this review, we discuss these strategies with a particular focus on those reported recently and provide our opinion on the future direction of the field.

Addresses

Department of Chemistry, University College London, 20 Gordon Street, London, UK

Corresponding author: Chudasama, Vijay (v.chudasama@ucl.ac.uk)

Current Opinion in Chemical Biology 2023, **75**:102306

This review comes from a themed issue on **Sulfur and Selenium**

Edited by **Kate Carroll**

For a complete overview see the Issue and the Editorial

Available online xxx

https://doi.org/10.1016/j.cbpa.2023.102306

Keywords

Cysteine conjugation, N-terminal modification, C-terminal modification, Termini modification, Protein modification, Peptide modification.

Introduction

The chemical modification of proteins, or protein bioconjugation [1], has revolutionized approaches to the construction of novel therapeutics and studying biological processes [2,3]. In particular, strategies that offer site-specific modification are becoming increasingly utilized when preparing bioconjugates. The resulting homogenous protein bioconjugates tend to display more desirable properties compared to their heterogeneous counterparts, with one of the most notable examples being that of antibody–drug conjugates (ADCs), a novel class of therapeutics revolutionizing approaches to targeted therapies [4,5]. Out of the bioconjugation methods described in the last three decades, those focused on cysteine as a handle for site-specific bioconjugation have proven especially popular [6] owing to its favorable properties such as low natural abundance and reactivity of the thiol side chain at neutral pH ($pK_a \approx 8$) [7]. In addition to cysteine, the termini of proteins, that is, the *N*- and *C*-terminus are of great interest; as these sites are usually solvent accessible and, in the case of single-chain proteins, there is only one *N*- and *C*-terminus. They also typically represent unique reactivity sites in a protein; for example, the *N*-terminal amino group (pK_a 6–8) differs in reactivity to the lysine ε side chain amino group (pK_a 10), which can allow for specific protein bioconjugation outcomes at the *N*-terminus [8]. For the *C*-terminus, differences in oxidative potentials of the *C*-terminus carboxyl group vs. internal carboxylates of aspartate and glutamate can be exploited for bioconjugation *via* photoredox-mediated decarboxylation strategies, with recent examples showcased by Bloom, Liu et al. in 2018 [9], Garreau et al. in 2019 [10], and Zhang, Floyd et al. in 2021 [11]. Different outcomes can also be achieved depending on the terminal amino acid side chain. This additional layer of specificity has enabled strategies that can lead to stable conjugates at the terminus over any internal amino acid [8,12]. This has proven to be especially true for cysteine at the termini of proteins, which has seen multiple methods developed for this handle with at least five novel bioconjugation strategies reported within just the past three years [6,13,14].

Herein, we highlight and describe these termini bioconjugation methods for protein modification of a cysteine *N*-terminus (denoted as NCys in line with the previous literature) and a cysteine *C*-terminus (denoted as CCys in this review), focusing on recent developments in the field and highlighting their applications in bioconjugation and chemical biology more broadly.

NCys modifications

Before bioconjugation, the NCys residue must first be incorporated into the target protein. This is typically carried out by incorporating a peptide tag at the *N*-terminus, which can undergo proteolytic cleavage to reveal the desired NCys. In recombinant protein production, the cysteine can be placed after the initiator methionine of the peptide tag; the methionine residue is then

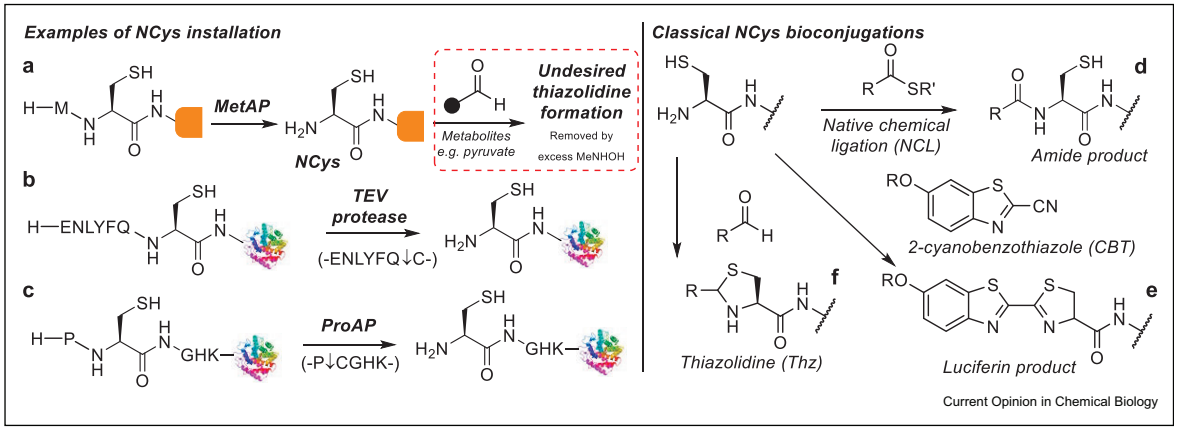
removed by methionine aminopeptidases (MetAPs) *in vivo* (thought to occur cotranslationally), yielding an NCys-containing protein [31]. Reactive aldehydes metabolites, for example, pyruvate can react with the cysteine to give an unreactive *N*-terminus thiazolidine (Thz); this can, however, be removed using an excess of methylhydroxylamine (Figure 1a) [14]. More commonly, the cysteine is placed after a sequence that can be cleaved *in vitro* by an enzyme. For example, the tobacco etch virus (TEV) protease can cleave the peptide sequence ENLYFQ↓C (where ↓ indicates the site of cleavage) to yield the desired NCys-containing protein, as reported by Tolbert and Wong (Figure. 1b) [32]. Other similar strategies include using different peptide sequences which are cleaved by proteases, such as Factor Xa [33], thrombin [34], and foot-and-mouth disease virus 3C protease (FMDV $3C^{pro}$protease) [35]. Very recently, Hempfling et al. have described a new procedure that involves the production of recombinant fusion proteins containing a MPCGHK or MPCGHKPGSSGSS peptide sequence at the N-terminus; in this strategy, the initiator methionine is removed *in vivo*, yielding a recombinant protein containing a proline at the *N*-terminus which can be cleaved *in vitro* by prolyl aminopeptidase (ProAP) from *Aeromonas sobria* (*A. sobria*) to generate NCys (Figure 1c). Although a short peptide tag remains in the final NCys protein, the specificity of ProAP for proline at the N-terminus offers the potential to avoid nonspecific cleavage that can be observed in other protease-based NCys production strategies [36]. NCys-containing proteins can be also be produced using intein splicing [37–39] or *via* expression systems that incorporate a leader peptide sequence (required for protein secretion) that is cleaved to reveal the desired cysteine [19,40].

For the chemical bioconjugation strategies discussed in this review, the modification of the NCys 1,2 aminothiol functionality typically occurs *via* a two-step pathway. The first step involves the reversible addition of either the amino or thiol nucleophile to the conjugation partner, followed by a second step involving the addition of the second nucleophile to generate the final conjugate. In cases where thiol addition occurs first, the reversible nature of this pathway leaves cysteines without the pendant amino group (i.e. internal/CCys residues) unmodified in the final conjugates, enabling site-specificity of the reported NCys bioconjugation strategies.

Classical NCys bioconjugation: native chemical ligation, cyanobenzothiazoles, thiazolidines

For NCys, the traditional methods of bioconjugation involve either using thioester reagents in native chemical ligation (NCL), first reported by Dawson et al. [41], to form amide bonds (Figure. 1d), [15,16], or using 2-cyanobenzothiazole (CBT) reagents to form luciferin type products, first reported by Ren et al. (Figure 1e) [17]. Previous reviews in the literature have covered NCL and CBT bioconjugation, with select examples of these reactions for NCys bioconjugation as shown in Table 1, Entries 1–4. Additionally, outside of being an unwanted by-product during NCys protein production, the reaction of the 1,2 aminothiol group of cysteine and an aldehyde to form a Thz group (Figure 1f) has been well-documented for its uses in chemical biology [42,43]. NCL remains a powerful tool for the semi-synthesis of peptides/proteins [44], and CBT bioconjugation is well established with several *in vitro* and *in vivo* applications reported [17,45–51] Furthermore, Thz formation has been reported by Casi et al. for the formation of antibody fragment drug conjugates

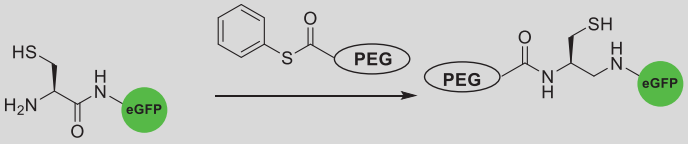
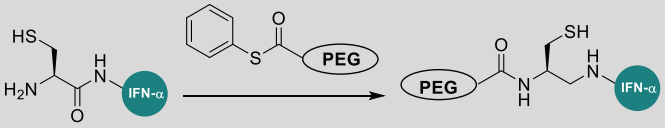
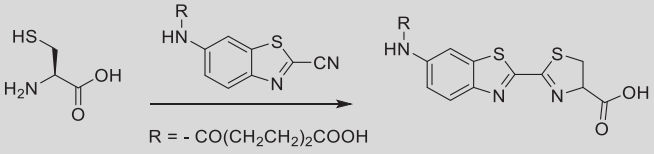
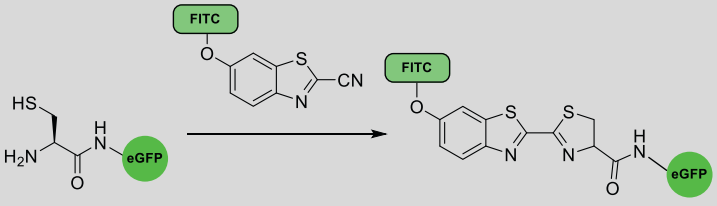
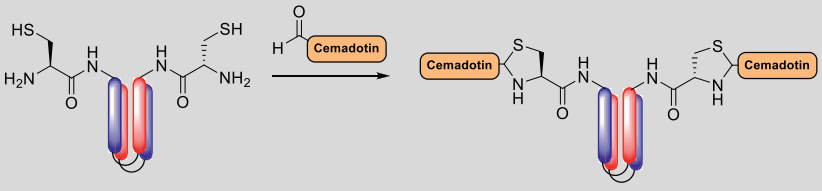
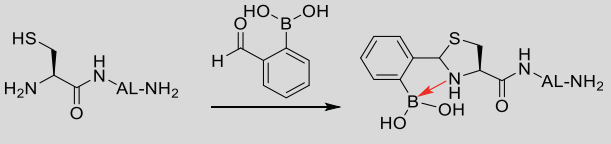
Figure 1



a) NCys installation *via* cleavage a methionine–cysteine bond with methionine aminopeptidases (MetAPs), which can undergo reaction with aldehyde metabolites *in vivo*. (**b**) NCys installation *via* cleavage of the tobacco etch virus (TEV) protease recognition sequence using TEV. (**c**) NCys installation *via* cleavage of a proline–cysteine bond with prolyl aminopeptidase (ProAP). (**d**) Native chemical ligation (NCL) of NCys. (**e**) 2-cyanobenzothiazole (CBT) bioconjugation of NCys. (**f**) Thiazolidine (Thz) formation at NCys.

Table 1

Selected examples of NCys and CCys bioconjugations, associated reaction conditions, and rate of reactions (or conversion/yield, if applicable).

Entry	NCys/CCys reaction and conditions	Rate (M^{-1} s^{-1})[a]	Ref
1	NCys Native chemical ligation  164 μM eGFP, 3–5 equiv. P(OEG_3-Glu)$_{n'}$-SPh, 2 mM TCEP, Tris.HCl pH 6.5, RT, (n' = 7 or 20).	10 h, 70–76% conversion,[b] 50–65% isolated yield[c]	[15]
2	NCys Native chemical ligation  3.87 mM IFN-α, 3 equiv. N-acetylated P(EG3-Glu)-SPh (*ca.* 82 kDa), Tris.HCl pH 7.0, RT	8 h, 35% conversion[d]	[16]
3	NCys Cyanobenzothiazole  150 μM L-cysteine, 150 μM CBT-COOH, 300 μM TCEP, PBS pH 7.4, 23 °C	9.19	[17]
4	NCys Cyanobenzothiazole  20 μM eGFP, 50 μM CBT-FITC, 2 mM TCEP, PBS pH 7.4	1 h, product formation confirmed[e]	[18]
5	NCys Thiazolidine  20 μM diabody, 1–4 mM CHO-CEM, 1 mM DTT, AcOK pH 4.5, 4 °C	60–96 h, >90% conversion[f]	[19]
6	NCys Thiazolidino boronate  10 μM H-CAL-NH_2, 10 μM 2-FPBA/2FBBA, pH 7.0	5500	[20]

(*continued on next page*)

Table 1 (*continued*)

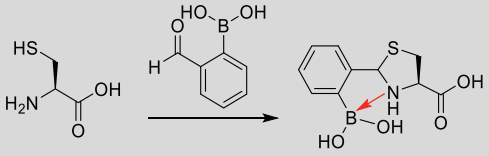
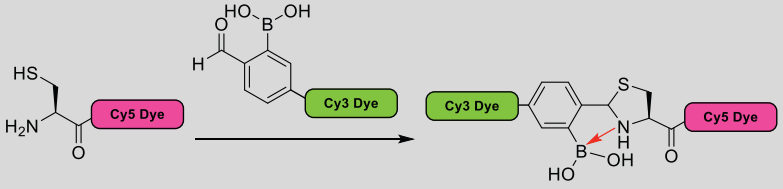
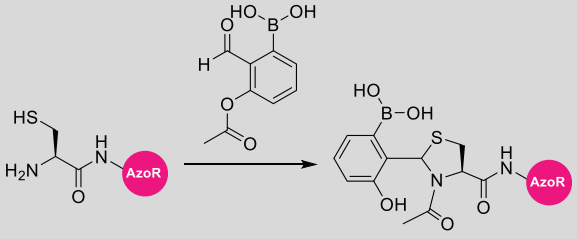
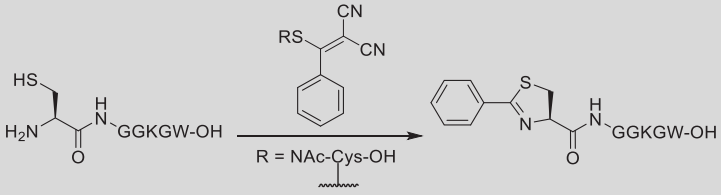
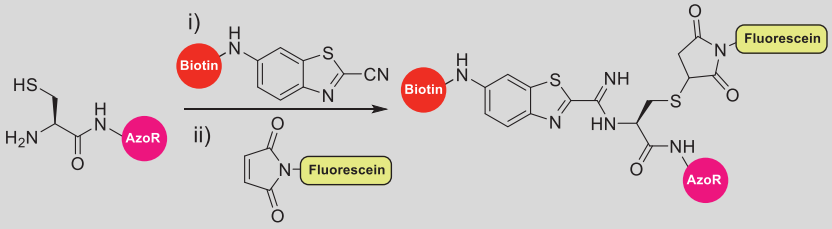
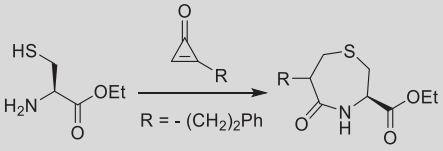
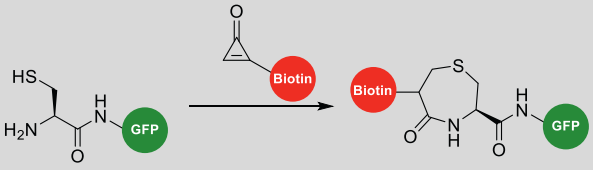
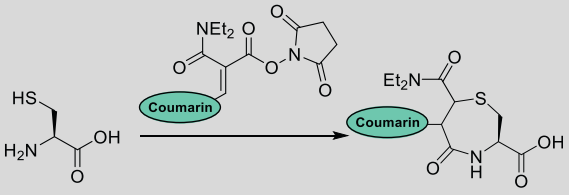
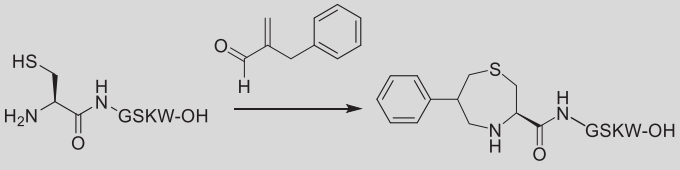
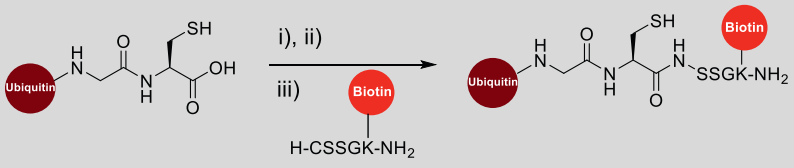
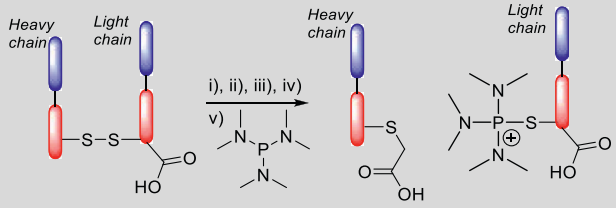
Entry	NCys/CCys reaction and conditions	Rate (M^{-1} s^{-1})[a]	Ref
7	NCys Thiazolidino boronate  500 μM L-cysteine, 200 μM 2-FPBA/2FBBA, NH_4OAc pH 7.4, 10% DMF, 23 °C	238	[21]
8	NCys Thiazolidino boronate  2.5 μM NCys Cy5 dye, 2.5 μM *ortho* boronic acid Cy3 dye, pH 6.0, 7.4 or 8.0, 23 °C	250 (pH 6.0)613 (pH 7.4) 21,261 (pH 8.0)	[22]
9	NCys Thiazolidino boronate  10 μM AzoR, 50 μM KL42, pH 6.0, RT	2 h > 99% conversion[f]	[23]
10	NCys 2-((alkylthio) (aryl)methylene)malononitrile  100 μM H-CGGGKGW-OH, 100 μM TAMM-NacCys, PB pH 7.4, 25 °C	4.2	[24]
11	NCys Cyanobenzothiazole + Maleimide, *N*, *S* double labeling  17 μM CIS-AzoR, (i) 238 μM biotin-CBT, 1 mM TCEP (ii) 363 μM maleimide, PBS pH 8.7, RT	20 min CBT label, then 1 h maleimide. *N,S*-dual label major product (>90%).[f]	[25]

Table 1 (*continued*)

Entry	NCys/CCys reaction and conditions	Rate (M^{-1} s^{-1})[a]	Ref
12	NCys Cyclopropenone  2 mM L-cysteine ethyl ester, 2 mM CPO-Ph,3 mM Na_2CO_3, H_2O:MeCN (1:1), 4 °C	2.98	[26]
13	NCys Cyclopropenone  5 μM GFP, 500 μM CPO-Biotin, 2.5 mM DTT, PB pH 7.0, 5% MeCN, 25 °C	2 h, >99% conversion[f]	[26]
14	NCys NHS acrylate  50 μM L-cysteine, 50 μM NHS-acrylate 7-diethylamino coumarin, PB pH 7.0, 23 °C	4.53	[27]
15	NCys BAA  5 μM H-CGSKW-OH, 5 μM BAA, 10 μM TCEP, 100 μM $NaBH_3CN$, PB pH 7.4	2707	[28]
16	CCys Hydrazinolysis  (i) 90–180 μM ubiquitin, 609 mM MESNa, 17 mM TCEP, 5% w/v (543 mM) hydrazine acetate, PB pH 5.8, 45 °C, (then exchanged to pH 4.0), (ii) 3 mM $NaNO_2$, PB pH 4.0, 0 °C, then(iii) 100 mM MPAA, 0.71 mM H-CSSGK(Biotin)-NH_2, PB pH 7.0, RT	(i) 48 h(ii) 20 min(iii) 1 hProduct assessed as major species[f]	[29]

(*continued on next page*)

Table 1 (*continued*)

Entry	NCys/CCys reaction and conditions	Rate (M^{-1} s^{-1})[a]	Ref
17	CCys Thiophosphonium  (i) 150 μM Fab,1.5 mM TCEP, PB pH 7.4, 6 mM EDTA, 37 °C, (ii) 225 μM α-chloro thioester PB pH 7.4, 5 mM EDTA, 22 °C, (iii) BBS pH 8.5, 5 mM EDTA, 22 °C, (iv) 80 μM protein, 1.6 mM Ellman's reagent, BBS pH 8.5, 5 mM EDTA, 10% DMF, 37 °C, then (v) 40 μM protein, 5 mM HMPT, PB pH 8.0, 10% MeCN, 37 °C	(i) 1 h (ii) 30 min (iii) 24 h (iv) 30 min (v) 1 h>99% conversion for each step, light chain thiophosphonium major product[f]	[30]

[a] If no rate studies were performed then, alternatively, reaction time + conversion/yield is quoted.

[b] Obtained using fluorescence intensity/Typhoon™ FLA 9500 laser scanner (SDS-PAGE).

[c] Obtained using a NanoPhotometer® (A_{280}).

[d] Obtained using Typhoon™ FLA 9500 laser scanner (SDS-PAGE).

[e] By MALDI-TOF/SDS-PAGE.

[f] As judged by ESI-MS.

(Table 1, Entry 5) [19,52] which show cytotoxicity in cell lines *via* drug release, with the NCys Thz linkage acting as a cleavable linker [19]. For future applications of NCys bioconjugation, it is likely other more recent methodologies that offer superior kinetics or greater NCys selectivity (or both) will begin to see greater use over the coming years.

Thiazolidino boronate linkages

The synthesis of bioconjugates through Thz linkages requires acidic conditions (pH 4.5) and long reaction times (days). In 2016, the synthesis of bioconjugates *via* thiazolidino boronate (TzB) complexes was described by both Bandyopadhyay et al. and Faustino et al. for an analogous but significantly faster method of NCys bioconjugation (Figure 2 a, Table Entries 6–7) [20,21]. Here, aromatic aldehydes containing an *ortho* boronic acid, such as 2-formylphenylboronic acid/2-formylbenzeneboronic acid (2-FPBA/2FBBA), were used as the aldehyde substrates. Similarly, for hydrazone/oxime ligation [53], the boronic acid moiety facilitates the activation of the imine intermediate, expediting thiazolidine formation and thus greatly enhancing the rate of product formation. The rate of TzB formation is pH dependant as recently highlighted by Rose et al.; here, the rate of conjugation between 1,2 aminothiol and *ortho* boronic acid cyanine dyes was shown to be enhanced *ca.* 30 fold if the conjugation was performed at pH 7.4 vs. pH 6.0. Alternatively, a 100-fold increase in rate was noted if the conjugation was performed at pH 8.0 vs. pH 6.0 [22]. In addition to NCys bioconjugation, TzB formation using 2-FPBA can be used to facilitate *C*-terminal protein ligations involving asparaginyl endopeptidase (AEP); here, 2-FPBA was used to quench the NCys by-product resulting from AEP-mediated ligation, thus driving the forward reaction to yield enzymatically labeled conjugates [54,55].

TzB formation has been demonstrated in the fluorescent labeling of peptides [20], and it enables rapid NCys bioconjugations [21,22]. The same is true for the reverse reaction, however, with the conjugates shown by Bandyopadhyay et al. to undergo dissociation within hours in the presence of L-cysteine or at acidic pH (Figure 2b) [20]. To overcome this, a thiazolidino boronate-mediated acyl transfer strategy was reported by Li et al. in 2020 (Figure 2c) [23]. In this example, derivatives of 2-FPBA that contain an acyl group capable of transferring to the Thz N atom during TzB formation were utilized, leading to the synthesis of more stable *N*-acyl Thz conjugates while maintaining rapid kinetics of formation. This strategy was subsequently used for NCys bioconjugation of azoreductase (AzoR) with a model or biotin-containing acetyl ester of 2-FPBA (Figure 2d, Table 1, Entry 9) [23].

TAMM

Reported in 2020 by Zheng et al., the reaction of 2-((alkylthio) (aryl)methylene)malononitrile (TAMM) with cysteine gives the condensation product 2-aryl-4,5-dihydrothiazole (ADT, Figure 2e, Table 1, Entry 10) [24]. In model peptide reactions, TAMM was found to selectively modify NCys in the presence of lysine. For internal cysteines, thiol-vinyl addition was observed; this could, however, be mitigated in the presence of an external thiol such as *N*-acetyl cysteine. In a similar

manner previously described for CBT labeling [17], TAMM reagents could also be used in cell surface labeling [24]. HEK cells could be transfected with a plasmid expressing a fusion protein containing the murine immunoglobulin (Ig) κ-chain leader sequence, the ENLYFQ↓C TEV recognition sequence, mCherry protein, and platelet-derived growth factor receptor (PDGFR) transmembrane domain. Labeling of the extracellular mCherry proteins could then be achieved within 30 min through incubation with TEV protease (revealing the NCys) and fluorescent TAMM-BODIPY (ADT formation) in a one-pot manner. Live cell fluorescence could then be used to confirm both the presence of mCherry and site-specific labeling with TAMM BODIPY on the HEK cell surface (Figure 2f). NCys modification with TAMM was further exemplified through the cyclization of phage-displayed peptides using a TAMM-chloroacetyl reagent [24].

N,S-Dual labeling

Outside of specific examples [56,57], CBT bioconjugation generally leads to reversible biomolecule modification at internal cysteines and irreversible conjugation at NCys [46,48]. In 2020, Wang and Gao described the use of CBT reagents to modify the NCys *via* an alternative pathway that leads to amidine products through opening of the 2-aminothiazolidine ring during CBT addition [25]. Through the screening of peptide libraries, it was found that a Cys-Ile-Ser tripeptide motif significantly favored amidine formation over the luciferin product typically observed in CBT bioconjugation; it was hypothesized this was due to hydrogen bonding interactions between the Cys and Ser side chains. The liberated thiol could then be modified through a different cysteine bioconjugation strategy, such as with maleimides (Figure 2g), to afford *N*,*S* dually labeled bioconjugates such as biotinylated, fluorescently labeled AzoR (Figure 2h, Table 1, Entry 11) [25].

1,4-Thiazepan-5-one linkages activated Michael acceptors

Reported by Istrate, Geeson et al. in 2020, cyclopropenone (CPO) reagents have been shown to react with NCys to give 1,4-thiazepan-5-one products (Figure 2i, Table 1, Entry 12) [26]. The CPO motif shows selectivity for the 1,2 aminothiol group, does not react with other nucleophilic amino acids, and shows no reactivity toward glutathione (one of the most abundant thiols found in cells). In model peptide reactions, the chemoselective modification of the NCys of linear vasopressin with CPO in the presence of an internal Cys and tyrosine residues was observed. CPO reagent selectivity for NCys was rationalized using quantum mechanical (QM) calculations, with irreversible 1,4-thiazepan-5-one formation only occurring when another nucleophile was in close proximity, for example, the amino group of the 1,2-aminothiol. NCys-eGFP could undergo selective modification with CPO reagents to synthesize fluorescently labeled or biotinylated conjugates (Table 1, Entry 13). Furthermore, the dimerization of NCys-IL-2 mimic to synthesize protein−protein conjugates (Figure 2j) could be achieved through bioconjugation with azide/strained alkyne containing CPO probes, followed by strain-promoted azide−alkyne cycloaddition (SPAAC) [26]. In addition to CPO reagents, 1,4-thiazepan-5-one peptide conjugates can also be prepared using reagents containing both an NHS ester and acrylate segment to selectively target NCys as reported in 2021 Silva et al. [58]. Multivalent NHS-activated acrylates were also reported for NCys modification by Djaló et al. in 2022 (Table 1, Entry 14) [27]. Another example of using activated Michael acceptors reported by Wu, Li et al. in 2021 for rapid NCys modification is the use of 2-benzylacrylaldehyde (BAA) reagents with $NaBH_3CN$ to give 1,4-thiazepane products (Table 1, Entry 15) [28].

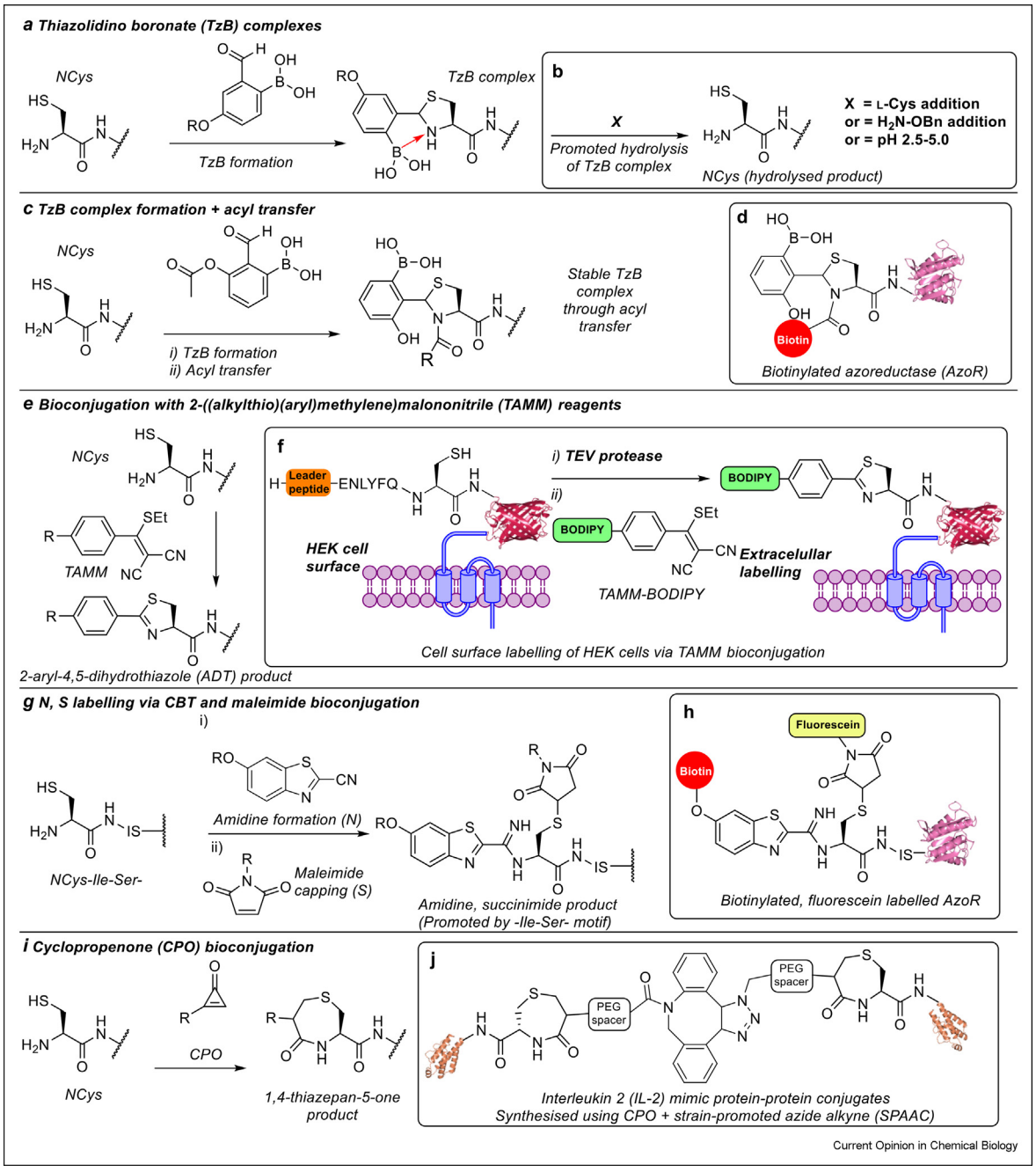
CCys modifications

Cysteines at the *C*-terminus (CCys, Figure 3a) can be introduced by site-directed mutagenesis or occur naturally in specific cases [59]. Alternatively, the addition of a synthetic di-cysteine linker to thioester/intein-containing proteins generates a 1,2 aminothiol functionality at the *C*-terminus; these motifs can then undergo bioconjugations akin to that of NCys modification [18,60]. Akin to specifically focusing on the 1,2 aminothiol group of NCys in the previous section, this section will focus specifically on protein bioconjugation at the CCys 3-mercaptopropionic acid functionality, discussing two strategies that directly involve or give a unique outcome at the CCys protein position. For wider reading, we would like to direct readers to a recent review by Arbour et al. that covers the synthesis of CCys-containing peptides (which pose unique challenges in peptide synthesis) [61] and work by Zhang, Floyd et al. [11] which shows the photoredox-catalyzed decarboxylative alkynation methodology first reported by Bloom, Liu et al. [9] can be achieved on CCys peptides provided the thiol is alkylated first by iodoacetamide [11].

Hydrazinolysis

Peptides containing a *C*-terminal hydrazide have been reported as more stable thioester equivalents for use in NCL. These hydrazides can be prepared synthetically through Boc or Fmoc solid-phase peptide synthesis (SPPS), or recombinantly through hydrazine-mediated cleavage of intein fusion proteins [62−64]. This strategy has also been applied in CCys protein modification, reported in 2013 by Adams et al. [29]. In this case, *C*-terminal protein hydrazides could be prepared *via* the addition of a hydrazine source to intercept transient thioesters (resulting from *N*- > *S* acyl transfer) across His−Cys and Gly−Cys motifs [65−67]. The *C*-terminal

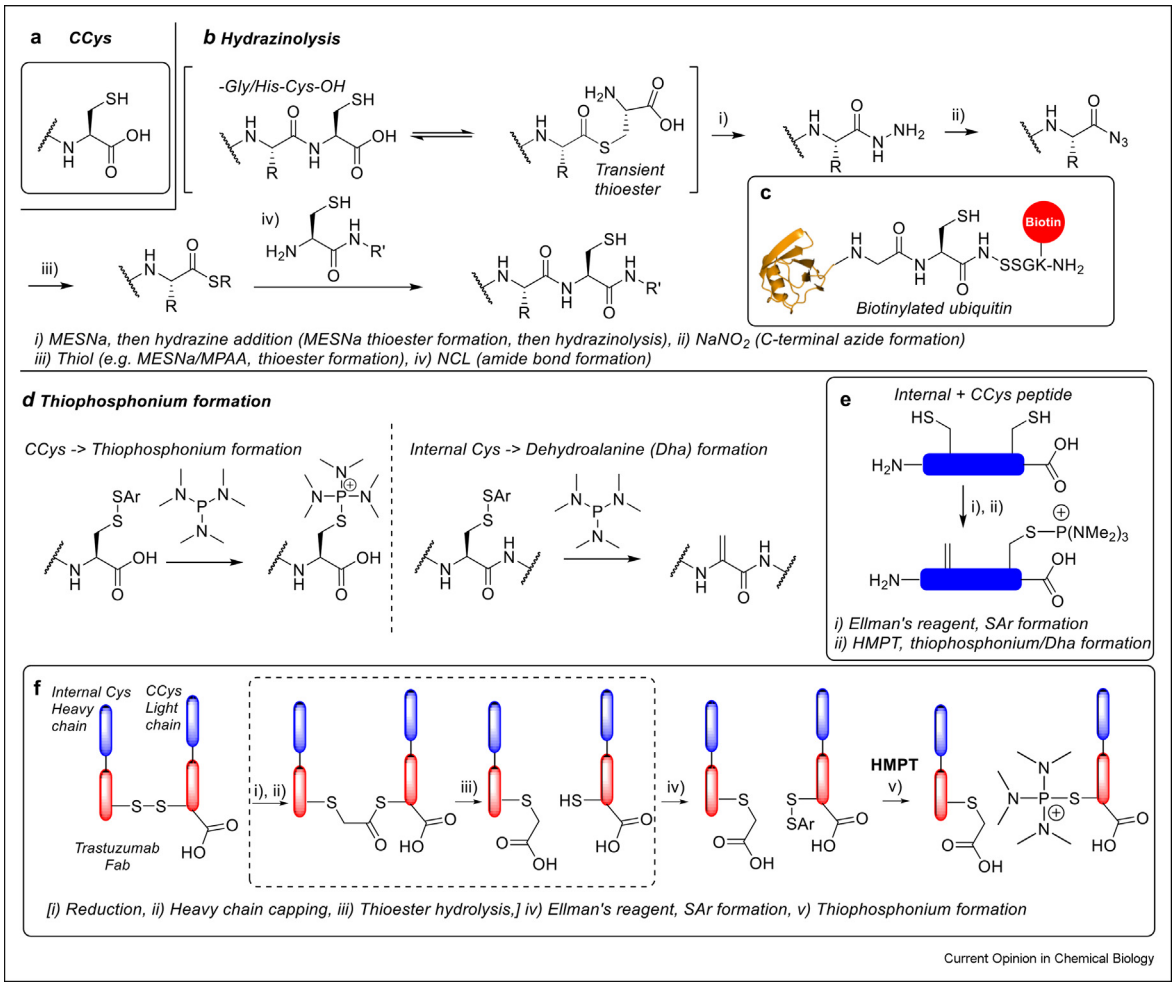
Figure 2



a) Thiazolidino boronate (TzB) complex formation at NCys. (**b**) Hydrolysis of the TzB complex. (**c**) TzB complex formation and stabilization through acyl transfer. (**d**) Biotinylated azoreductase (AzoR) *via* TzB complex formation at NCys. (**e**) Bioconjugation of NCys using 2-((alkylthio) (aryl)methylene) malononitrile (TAMM) reagents. (**f**) Extracellular labeling of HEK cells using TAMM reagents. (**g**) *N, S* labeling of NCys through sequential bioconjugation of CBT and maleimide reagents. (**h**) Biotinylated, BODIPY labeled AzoR *via* CBT + maleimide *N,S* labeling. (**i**) Bioconjugation of NCys using cyclopropenone (CPO) reagents. (**j**) Interleukin 2 (IL-2) mimic protein–protein conjugates constructed with CPO and SPAAC bioconjugation.

protein hydrazides could then be converted to acyl azides through activation with sodium nitrite, followed by thioester formation with sodium 2-mercaptoethanesulfonate (MESNa) or 4-mercaptophenylacetic acid (MPAA). The resulting thioesters could then undergo NCL with CCys-containing peptides (Figure 3b). For example, ubiquitin G76C could be converted to the corresponding ubiquitin *C*-terminal hydrazide through the addition of 5% w/v hydrazide acetate in the presence of TCEP and MESNa.

Figure 3



a) 3-Mercaptopropionic acid functionality of CCys. (**b**) CCys bioconjugation *via* a hydrazinolysis and NCL strategy. (**c**) Biotinylated ubiquitin is prepared *via* hydrazinolysis and NCL. (**d**) Thiophosphonium formation at CCys S-SAr (alkyl-aryl) disulfides and dehydroalanine (Dha) formation at internal cysteine S-SAr disulfides. (**e**) Thiophosphonium and Dha formation on a bis-cysteine peptide. (**f**) Thiophosphonium formation of the light chain of Trastuzumab Fab.

Treatment with sodium nitrite, followed by MPAA and an H-CSSK(biotin)-NH_2 peptide, led to the *C*-terminal biotinylated ubiquitin (Figure 3c, Table 1 Entry 16). Additionally, recombinantly produced erythropoietin 1-161 (A160G) could undergo hydrazinolysis across the A160G-C161 *C*-terminus. Although hydrazinolysis was, in some cases, observed across a His–Cys internal bond, successful NCL of the peptide H-CRTGDRC-OH at the *C*-terminus of the EPO fragment *via* hydrazinolysis/thioester intermediates could be achieved [29].

Thiophosphonium linkage

The reaction of disulfides with phosphine reagents typically leads to the reduction of the disulfide bond *via* a thiophosphonium intermediate [68]. This intermediate is, outside of a few select examples such as salt formation [69] or rotaxane-mediated stabilization [70], typically viewed as highly reactive and unstable, rapidly forming decomposition products such as dehydroalanine (Dha) [71]. In the case of unsymmetrical alkyl-aryl (S-SAr) disulfides involving a CCys, however, unusual stability of the thiophosphonium intermediate was reported by Spears et al. in 2022 when utilizing tris(alkylamino)phosphines (Figure 3d), with other alkyl phosphines such as TCEP leading to disulfide reduction [30]. This observation was further confirmed on model peptides; CCys S-SAr disulfides treated with tri(dimethyl)aminophosphine (hexamethylphosphorous triamide, HMPT) rapidly led to the formation of stable thiophosphonium adducts, whereas internal cysteine S-SAr disulfides underwent decomposition to Dha under identical conditions (Figure 3e). In the case of NCys, S-SAr disulfides gave a mixture of thiophosphonium adduct and decomposed species upon treatment with

HMPT; this decomposition product underwent further transformation to an *N*-terminus ketone (likely through an enamine-imine tautomerization and imine hydrolysis pathway), which was confirmed through oxime ligation. Furthermore, chemoselective modification on a bis Cys S-SAr peptide was observed, whereby a thiophosphonium adduct was generated at the *C*-terminus and Dha at an internal position; the Dha residue could then be modified by thiols. The chemistry could also be used to install a thiophosphonium adduct at the CCys cysteine of Trastuzumab Fab light chain [30]. In this example, Trastuzumab Fab was reduced with TCEP, followed by liberation of the light chain thiol and capping of the heavy chain thiol (as reported by Chrzastek et al. in 2022) [72]. The addition of Ellman's reagent resulted in a mixed disulfide S-SAr-containing light chain, followed by the formation of the thiophosphonium-containing light chain *via* the addition of HMPT (Figure 3f, Table 1 Entry 17) [30].

Conclusions and outlooks

The last three decades have seen major progress in the strategies available for protein bioconjugation, with bioconjugation at the *N*-terminus and *C*-terminus being no exception. Methods available for targeting NCys have greatly expanded in recent years owing to new and innovative approaches offering rapid, site-specific labeling under mild conditions to yield functional bioconjugates such as dually labeled bioconjugates and protein−protein bioconjugates. This does come with the caveat of some of the more bespoke labeling reagents being expensive or more difficult to synthesize and thus less accessible to non-chemists, but the next few years will likely see even further development of these methods in terms of accessibility, optimization, and application. For CCys, which represents a unique but significantly challenging handle to target, bioconjugation strategies for its site-specific modification have also started to emerge. These current methods available offer a more limited scope with regards to the mildness of conditions, site-specificity, and requirement of multi-step bioconjugation protocols to achieve modification in comparison to NCys bioconjugation strategies. This is unsurprising given the significant challenges faced for *C*-terminus modification in general, such as distinguishing between the carboxyl groups of the *C*-terminus and Asp/Glu. Current methods for *C*-terminal/terminus protein modification are largely dominated by enzymatic strategies [73−75]. While the reactive profile of the *C*-terminus carboxyl group could also lend itself to facilitate specific outcomes at this position, as observed by Spears et al. for CCys and Bloom, Liu et al. for other *C*-terminus amino acids, it will also likely pose the primary obstacle when attempting to develop bioconjugation strategies, especially for CCys. Indeed, Boll and Raines have very recently reported that the reactivity of CCys is less than that of internal cysteine and NCys [76]. Nevertheless, given that several native peptides and therapeutic proteins such as full antibodies naturally contain a CCys [59], endeavors toward further advancements in this area will likely prove highly valuable within the field of protein bioconjugation.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Data availability

No data was used for the research described in the article.

Acknowledgements

We gratefully acknowledge the Leverhulme Trust for funding R. J. S. (RPG-2020-010).

References

Papers of particular interest, published within the period of review, have been highlighted as:

* of special interest
** of outstanding interest

1. Stephanopoulos N, Francis MB: *Nat Chem Biol* 2011, **7**:876−884.
2. Hoyt EA, Cal PMSD, Oliveira BL, Bernardes GJL: *Nat Rev Chem* 2019, **3**:147−171.
3. Sornay C, Vaur V, Wagner A, Chaubet G: *R Soc Open Sci* 2022, **9**:211563.
4. Walsh SJ, Bargh JD, Dannheim FM, Hanby AR, Seki H, Counsell AJ, Ou X, Fowler E, Ashman N, Takada Y, Isidro-Llobet A, Parker JS, Carroll JS, Spring DR: *Chem Soc Rev* 2021, **50**:1305−1353.
5. Fu Z, Li S, Han S, Shi C, Zhang Y: *Signal Transduct Targeted Ther* 2022, **7**:93.
6. Chen F-J, Gao J: *Chem Eur J* 2022, e202201843.
† This review provides an in-depth discussion of cysteine bioconjugation (including NCys bioconjugation) and associated rates of reaction.
7. Spicer CD, Davis BG: *Nat Commun* 2014, **5**:4740.
8. Rosen CB, Francis MB: *Nat Chem Biol* 2017, **13**:697−705.
9. Bloom S, Liu C, Kölmel DK, Qiao JX, Zhang Y, Poss MA, Ewing WR, MacMillan DWCC: *Nat Chem* 2018, **10**:205−211.
10. Garreau M, Le Vaillant F, Waser J: *Angew Chem Int Ed* 2019, **58**: 8182−8186.
11. Zhang L, Floyd BM, Chilamari M, Mapes J, Swaminathan J, Bloom S, Marcotte EM, V Anslyn E: *ACS Chem Biol* 2021, **16**: 2595−2603.
12. De Rosa L, Di Stasi R, Romanelli A, D'Andrea LD: *Molecules* 2021:26.
13. Jiang H, Chen W, Wang J, Zhang R: *Chin Chem Lett* 2022, **33**: 80−88.
14. Asiimwe N, Al Mazid MF, Murale DP, Kim YK, Lee J-S: *Pept Sci* 2022, **114**, e24235.
15. Hou Y, Yuan J, Zhou Y, Yu J, Lu H: *J Am Chem Soc* 2016, **138**: 10995−11000.
16. Yuan J, Zhang Y, Li Z, Wang Y, Lu H: *ACS Macro Lett* 2018, **7**: 892−897.

17. Ren H, Xiao F, Zhan K, Kim YP, Xie H, Xia Z, Rao J: *Angew Chem Int Ed* 2009, **48**:9658–9662.

18. Wang H-C, Yu C-C, Liang C-F, Huang L-D, Hwu J-R, Lin C-C: *Chembiochem* 2014, **15**:829–835.

19. Casi G, Huguenin-Dezot N, Zuberbühler K, Scheuermann J, Neri D: *J Am Chem Soc* 2012, **134**:5887–5892.

20. Bandyopadhyay A, Cambray S, Gao J: *Chem Sci* 2016, **7**: 4589–4593.

21. Faustino H, Silva MJSA, Veiros LF, Bernardes GJL, Gois PMP: *Chem Sci* 2016, **7**:5052–5058.

22. Rose NC, V Sanchez A, Tipple EF, Lynam JM, Spicer CD: *Chem Sci* 2022, **13**:12791–12798.

23. Li K, Wang W, Gao J: *Angew Chem Int Ed* 2020, **59**:14246–14250.
** This paper demonstrates the first use of an acryl transfer strategy to rapidly generate stable thiazolidino–boronate complexes for NCys bioconjugation.

24. Zheng X, Li Z, Gao W, Meng X, Li X, Luk LYP, Zhao Y, Tsai Y-H, Wu C: *J Am Chem Soc* 2020, **142**:5097–5103.
** This paper demonstrates the first use of TAMM reagents as NCys bioconjugation reagents, with applications in chemical biology such as labelling of cell surfaces and phage display.

25. Wang W, Gao J: *J Org Chem* 2020, **85**:1756–1763.
** This paper demonstrates a novel NCys bioconjugation strategy which can generate NCys dually labelled bioconjugates using established CBT and maleimide chemistry, which is assisted by interactions from a CIS tripeptide motif.

26. Istrate A, Geeson MB, Navo CD, Sousa BB, Marques MC, Taylor RJ, Journeaux T, Oehler SR, Mortensen MR, Deery MJ, Bond AD, Corzana F, Jiménez-Osés G, Bernardes GJL: *J Am Chem Soc* 2022, **144**:10396–10406.
** This paper describes the first use of cyclopropenone reagents in NCys bioconjugation, generating bioconjugates including protein–protein conjugates

27. Djaló M, Silva MJSA, Faustino H, Pinto SN, Mendonça R, Gois PMP: *Chem Commun* 2022, **58**:7928–7931.

28. Wu Y, Li C, Fan S, Zhao Y, Wu C: *Bioconjugate Chem* 2021, **32**: 2065–2072.

29. Adams AL, Cowper B, Morgan RE, Premdjee B, Caddick S, Macmillan D: *Angew Chem Int Ed* 2013, **52**:13062–13066.

30. Spears RJ, Chrzastek A, Yap SY, Karu K, Aliev AE, Baker JR, Chudasama V: *Chem Commun* 2022, **58**:5359–5362.
** This paper describes a novel approach to CCys bioconjugation involving model peptides and trastuzumab Fab reactions with model tris(dialkylamino)phopshine reagents to generate thiophosphonium adducts, which prove to be unusually stable in this context

31. Gentle IE, De Souza DP, Baca M: *Bioconjugate Chem* 2004, **15**: 658–663.

32. Tolbert TJ, Wong C-H: *Angew Chem Int Ed* 2002, **41**:2171–2174.

33. Fushman D, Xu R, Cowburn D: *Biochemistry* 1999, **38**: 10225–10230.

34. Liu D, Xu R, Dutta K, Cowburn D: *FEBS Lett* 2008, **582**: 1163–1167.

35. Busch GK, Tate EW, Gaffney PRJ, Rosivatz E, Woscholski R, Leatherbarrow RJ: *Chem Commun* 2008:3369–3371.

36. Hempfling JP, Sekera ER, Sarkar A, Hummon AB, Pei D: *J Am Chem Soc* 2022, **144**:21763–21771.
* This paper describe a novel strategy for recombinant production of NCys proteins using prolyl aminopeptidase *in vitro*.

37. Muir TW, Sondhi D, Cole PA: *Proc Natl Acad Sci USA* 1998, **95**: 6705 LP–6710.

38. Evans Jr TC, Benner J, Xu M-Q: *Protein Sci* 1998, **7**:2256–2264.

39. Evans TC, Benner J, Xu M-Q: *J Biol Chem* 1999, **274**: 18359–18363.

40. Hauser PS, Ryan RO: *Protein Expr Purif* 2007, **54**:227–233.

41. Dawson PE, Muir TW, Clark-Lewis I, Kent SB: *Science* 1994, **266**: 776–779.

42. Bi X, Pasunooti KK, Tareq AH, Takyi-Williams J, Liu C-F: *Org Biomol Chem* 2016, **14**:5282–5285.

43. Laps S, Satish G, Brik A: *Chem Soc Rev* 2021, **50**:2367–2387.

44. Agouridas V, El Mahdi O, Diemer V, Cargoët M, Monbaliu J-CM, Melnyk O: *Chem Rev* 2019, **119**:7328–7443.

45. Jeon J, Shen B, Xiong L, Miao Z, Lee KH, Rao J, Chin FT: *Bioconjugate Chem* 2012, **23**:1902–1908.

46. Xu Y, Qi C, Wang C: *Comput Theor Chem* 2020, **1185**:112874.

47. He Y, Kang T-R, Liu Q-Z, Chen L-M, Tu Y-L, Liu Y-J, Chen T-B, Wang Z-Q, Liu J, Xie Y-M, Yang J-L, He L: *Org Lett* 2013, **15**: 4054–4057.

48. Zheng Z, Chen P, Li G, Zhu Y, Shi Z, Luo Y, Zhao C, Fu Z, Cui X, Ji C, Wang F, Huang G, Liang G: *Chem Sci* 2017, **8**:214–222.

49. Zhu Y, Zhang X, You Q, Jiang Z: *Bioorg Med Chem* 2022, **68**: 116881.

50. Kilpatrick LE, Friedman-Ohana R, Alcobia DC, Riching K, Peach CJ, Wheal AJ, Briddon SJ, Robers MB, Zimmerman K, Machleidt T, V Wood K, Woolard J, Hill SJ: *Biochem Pharmacol* 2017, **136**:62–75.

51. Friedman Ohana R, Hurst R, Rosenblatt M, Levin S, Machleidt T, Kirkland TA, Encell LP, Robers MB, V Wood K: *Sci Rep* 2019, **9**: 7046.

52. Bernardes GJL, Steiner M, Hartmann I, Neri D, Casi G: *Nat Protoc* 2013, **8**:2079–2089.

53. Cambray S, Gao J: *Acc Chem Res* 2018, **51**:2198–2206.

54. Tang TMS, Cardella D, Lander AJ, Li X, Escudero JS, Tsai Y-H, Luk LYP: *Chem Sci* 2020, **11**:5881–5888.

55. Tang TMS, Luk LYP: *Org Biomol Chem* 2021, **19**:5048–5062.

56. Ramil CP, An P, Yu Z, Lin Q: *J Am Chem Soc* 2016, **138**: 5499–5502.

57. Keyser SGL, Utz A, Bertozzi CR: *J Org Chem* 2018, **83**: 7467–7479.

58. Silva MJSA, Faustino H, Coelho JAS, V Pinto M, Fernandes A, Compañón I, Corzana F, Gasser G, Gois PMP: *Angew Chem Int Ed* 2021, **60**:10850–10857.

59. Shen Y, Zeng L, Zhu A, Blanc T, Patel D, Pennello A, Bari A, Ng S, Persaud K, Kang YK, Balderes P, Surguladze D, Hindi S, Zhou Q, Ludwig DL, Snavely M: *mAbs* 2013, **5**:418–431.

60. Ziaco B, Pensato S, D'Andrea LD, Benedetti E, Romanelli A: *Org Lett* 2008, **10**:1955–1958.

61. Arbour CA, Mendoza LG, Stockdill JL: *Org Biomol Chem* 2020, **18**:7253–7272.

62. Fang G-M, Li Y-M, Shen F, Huang Y-C, Li J-B, Lin Y, Cui H-K, Liu L: *Angew Chem Int Ed* 2011, **50**:7645–7649.

63. Fang G-M, Wang J-X, Liu L: *Angew Chem Int Ed* 2012, **51**: 10347–10350.

64. Mo Z, Lin S, Chen W, He C: *Angew Chem Int Ed* 2022, **61**, e202115377.

65. Richardson JP, Chan C-H, Blanc J, Saadi M, Macmillan D: *Org Biomol Chem* 2010, **8**:1351–1360.

66. Kang J, Reynolds NL, Tyrrell C, Dorin JR, Macmillan D: *Org Biomol Chem* 2009, **7**:4918–4923.

67. Kang J, Richardson JP, Macmillan D: *Chem Commun* 2009: 407–409.

68. Burns JA, Butler JC, Moran J, Whitesides GM: *J Org Chem* 1991, **56**:2648–2650.

69. Krafft GA, Siddall TL: *Tetrahedron Lett* 1985, **26**:4867–4870.

70. Oku T, Furusho Y, Takata T: *Org Lett* 2003, **5**:4923–4925.

71. Chalker JM, Gunnoo SB, Boutureira O, Gerstberger SC, Fernández-González M, Bernardes GJL, Griffin L, Hailu H, Schofield CJ, Davis BG: *Chem Sci* 2011, **2**: 1666–1676.

72. Chrzastek A, Thanasi IA, Irving JA, Chudasama V, Baker JR: *Chem Sci* 2022, **13**:11533–11539.

73. Schumacher D, Helma J, Mann FA, Pichler G, Natale F, Krause E, Cardoso MC, Hackenberger CPR, Leonhardt H: *Angew Chem Int Ed* 2015, **54**:13787–13791.

74. Rehm FBH, Tyler TJ, Yap K, de Veer SJ, Craik DJ, Durek T: *J Am Chem Soc* 2021, **143**:19498–19504.

75. Morgan HE, Turnbull WB, Webb ME: *Chem Soc Rev* 2022, **51**: 4121–4145.

76. Boll LB, Raines RT: *Chembiochem* 2022, **23**, e202200258.

*This paper investigates how the reactivity of NCys, internal cysteines and CCys small molecules with maleimides can give significantly different outcomes due to effects of thiol location and thus pK_a.