Selective N-terminal cysteine protein modification with cyclopropenones

Alena Istrate,[1] Claudio D. Navo,[2] Barbara B. Sousa,[3] Marta C. Marques,[3] Michael J. Deery,[4] Andrew D. Bond,[1] Francisco Corzana,[5] Gonzalo Jiménez-Osés,*[,2] and Gonçalo J. L. Bernardes*[,1,3]

[1]Department of Chemistry, University of Cambridge, Lensfield Road, CB2 1EW Cambridge, United Kingdom.

[2] Center for Cooperative Research in Biosciences (CIC bioGUNE), Basque Research and Technology Alliance (BRTA), Bizkaia Technology Park, Building 801A, 48160 Derio, Spain

[3]Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisboa, Portugal.

[4]Cambridge Centre for Proteomics, Cambridge Systems Biology Centre, Department of Biochemistry, University of Cambridge, Tennis Court Road, Cambridge CB2 1QW, United Kingdom.

[5]Departamento de Química, Universidad de La Rioja, Centro de Investigación en Síntesis Química, 26006 Logroño, Spain.

*e-mail: gb453@cam.ac.uk (G.J.L.B.) and gjoses@cicbiogune.es (G.J.-O.)

ABSTRACT

Protein conjugates are valuable tools to create therapeutics, such as antibody-drug conjugates, or to study biological processes. Despite a number of protein conjugation strategies having been developed over recent years, the ability to modify one specific amino acid on a protein in the presence of other side chains with similar reactivity remains a challenge. We used the reaction between a monosubstituted cyclopropenone (CPO) probe and the 1,2-aminothiol of an N-terminal cysteine to give a stable 1,4-thiazepa-5-none linkage under mild, biocompatible conditions. The method relies on the ability of cyclopropenones to ring-open after sequential sulfhydryl and α-amine conjugation to be truly specific for N-terminal cysteine. We show that our CPO probes selectively label N-terminal cysteine containing peptides and proteins

even in the presence of internal, solvent-exposed cysteines, which can be subsequently modified by using conventional cysteine modification reagents. The ability to distinguish and specifically target N-terminal cysteine residues on a protein will facilitate the construction of elaborate multi-labelled bioconjugates.

INTRODUCTION

Protein conjugates are important tools to create valuable therapeutics, such as antibody-drug conjugates (ADCs)[1-2] and PEGylated proteins,[3] build new functionalised materials,[4] and study biological processes.[5-6] Among various strategies for protein conjugation, the modification of naturally occurring amino acids remains the method of choice because it offers the advantage of straightforward accessibility without the need for sequence alterations by means of genetic methods.[7] Ideally, conjugation reactions should proceed with complete chemo- and regioselectivity to generate well-defined protein constructs, which is a crucial requirement for many applications such as ADCs.[1-2] Similarly, such transformations should occur rapidly in mild aqueous solutions at room temperature and physiological pH. Although many protein conjugation strategies have been developed over recent years, the ability to modify one specific amino acid on a protein in the presence of other side chains with similar reactivity remains a challenge.[7] Protection of particularly reactive amino acids such as cysteine[8] or extensive sequence engineering with the introduction of specific tags for enhanced reactivity[9] is often required in order to achieve selectivity for the intended residue.

Among proteinogenic amino acids, lysine[10-11] and cysteine[9, 12-13] are the two most commonly targeted for bioconjugation because they are highly nucleophilic under physiological conditions. Native lysine residues are very convenient targets, but they are abundant on protein surfaces, so it is difficult to achieve a high degree of selectivity.[10] Conversely, cysteines are less abundant in proteins (<2%) and commonly less solvent-exposed, which makes them an excellent target for site-selective conjugation.[14] However, when relying on cysteine for protein modification, there are several factors that must be taken into account. Specifically, cysteines often form disulfide bonds that are critical for a protein's structure, and modification of such residues can lead to loss of protein function. Moreover, many surface-exposed endogenous cysteines are directly involved in the catalytic activity of the protein, and

thus cannot be exploited for modification. Therefore, methods for targeted, selective cysteine modification could further advance the field of protein conjugation and enable the construction of functional and well-defined biomolecule conjugates without the need for genetic encoding of unnatural amino acids.

The most reliable strategy to differentiate one cysteine in the presence of other thiol groups is to target the N-terminal cysteine residue (NCys). Several methods for selective NCys modification have been developed, including reaction with thioesters [native chemical ligation (NCL)], condensation with aromatic aldehydes, and condensation with 2-cyanobenzothiazole derivatives (**Figure 1**).[15] NCL provides a means to link protein or peptide fragments in a robust and chemoselective manner through trans-thioesterification and S-to-N acyl exchange (**Figure 1a**).[16-17] However, this method is rarely used to produce protein conjugates because of difficulties in preparation and lack of stability of suitable thioester reagents. The reaction of 1,2-aminothiols with aldehydes to form thiazolidine has also been explored as a strategy for NCys modification (**Figure 1b**).[18] Neri and co-workers have successfully applied this approach for site-specific coupling of cytotoxic aldehydes to tumor-targeting antibodies, which produced homogenous conjugates that were then used for targeted delivery and slow release of the cytotoxic component.[19] However, this reaction requires long incubation times (≈48 h), occurs under acidic conditions (pH 4–5) and is typically performed with a large excess of the aldehyde derivative. These limitations can be addressed by using formyl benzeno boronic acids (FBBA) that stabilize thiazolidine formation through N→B coordination (**Figure 1c**). Recently, FBBA reagents have been used to rapidly modify several model NCys-containing peptides at neutral pH.[20-21] This reaction is reversible and the product is not stable in an acidic environment (pH < 6) or in the presence of nucleophiles (e.g., free cysteine or benzyl hydroxylamine). It is however possible to use a thiazolidino boronate intermediate which undergoes an intramolecular acyl transfer to give more stable N-acyl thiazolidines.[22]

Another N-terminal Cys labelling reaction was inspired by the final step of chemical synthesis of D-luciferin,[23] and is based on the condensation of free cysteine with 2-cyanobenzothiazol (CBT) (**Figure 1d**). After Rao and co-workers first demonstrated

the potential of this reaction for NCys modification,[24] the approach has been widely used in site-selective protein labelling and molecular imaging. This click reaction has major advantages for bioconjugation because of its efficiency, biocompatibility and stability of the resulting luciferin linkage.[25] However, 2-cyanobenzothiazol derivatives also react quickly (although reversibly) with simple thiols. As a result, when using excess CBT to ensure complete conjugation, other free-protein thiols must be protected. Therefore, alternative bioconjugation reagents for fast and selective labelling of N-terminal cysteines are still required to enable construction of complex protein conjugates of well-defined structures.

In this work, we explore the possibility of using cyclopropenone (CPO) for site-selective labelling of N-terminal cysteines on peptides and proteins. Cyclopropenone is a small unit with unique reactivity, which makes it an attractive component for several biological applications, for example as a key part of selective protease inhibitors[26] or reagents for bioorthogonal protein ligation.[27-28] Indeed, the aromatic character of the ring[29] contributes to the cyclopropenone's stability, whereas significant angle strain and large dipole moment allow these moieties to participate in cycloaddition and ring-opening reactions.[30-31] As α,β-unsaturated ketones, cyclopropenones also act as electrophiles in 1,2- and 1,4-nucleophilic addition reactions. Several natural products contain cyclopropenone substituents, which highlights its biocompatibility and potential stability under physiological conditions.[32-33] Previous work has demonstrated that CPOs react with thiols under basic conditions,[28] yet the reaction mechanism or products are not well established. Here, we describe the selective and efficient reaction of monosubstituted cyclopropenones with 1,2-aminothiol of the N-terminal cysteine and their use for site-selective bioconjugation of peptides and proteins (**Figure 1e**). Importantly, the use of cyclopropenones to efficiently modify an N-terminal cysteine in the presence of solvent-exposed cysteines either on the same sequence, or in a mixture of proteins is demonstrated for the first time.

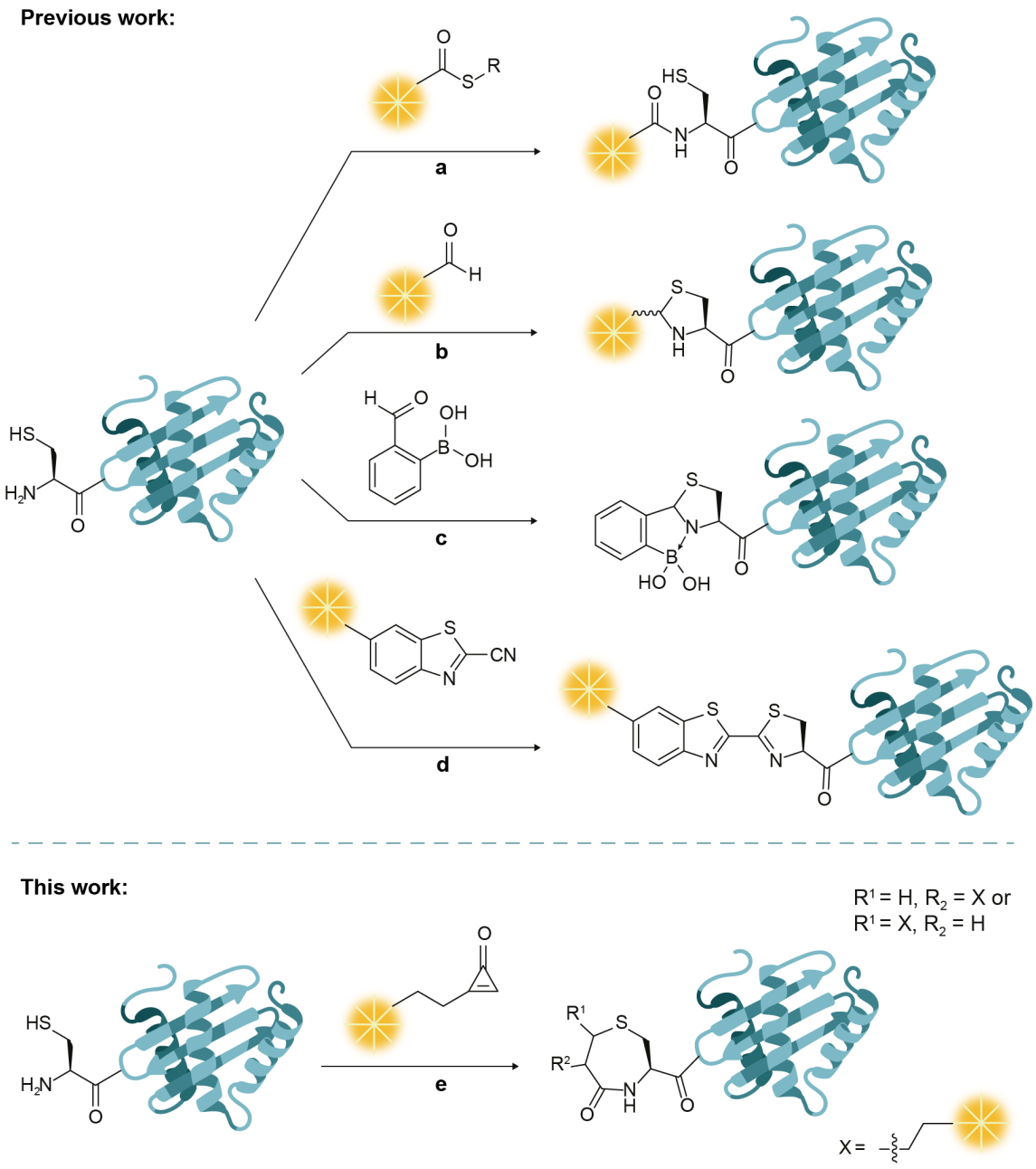

Fig. 1 | Protein modification at N-terminal cysteine. a. Modification with thioesters through native chemical ligation; **b.** Condensation with aromatic aldehydes to form thiazolidines; **c.** Modification using formyl benzeno boronic acids; **d.** Condensation with 2-cyanobenzothiazole derivatives resulting in the formation of luciferin linkage; **e.** Reaction with monosubstituted cyclopropenones to generate 1,4-thiazepa-5-none linkage.

RESULTS & DISCUSSION

Monosubstituted cyclopropenones are efficient reagents for N-terminal cysteine labelling.

To start our investigation, we synthesized 2-phenethylcyclopropenone (**1**) as a model cyclopropenone and tested its stability in aqueous buffer solutions. Compound **1** was prepared from commercially available 4-phenyl-1-butyne by a simple, two-step procedure. Specifically, 4-phenyl-1-butyne was treated with difluorocarbene generated from $TMSCF_3$ to afford a *gem*-difluorocyclopropen derivative,[34] which was hydrolyzed on wet silica gel to afford desired cyclopropenone **1**.[28, 35] Importantly, cyclopropenone **1** showed excellent stability after incubation in phosphate buffers (50 mM, pH 7–8) at 37 °C for 7 days (**Figure S6** and **S7**). We next studied the reaction of cyclopropenone **1** with L-cysteine in the presence of base (Na_2CO_3) at 4 ºC. The exceptionally high reactivity of cyclopropenone **1** required assaying the labelling reactions with small, highly accessible nucleophiles (i.e. amino acids, peptides, solvent-exposed sidechains of proteins) at lower temperatures than usual (4 ºC). This allowed proper reaction monitoring and measuring labelling kinetics without interference of undesired side reactions. The reaction resulted in four isomeric compounds in 87% yield after for 30 min in aqueous solution. Liquid chromatography (LC)–mass spectrometry (MS) and NMR spectroscopic analysis showed that the products are two pairs of diastereomeric regioisomers of 1,4-thiazepa-5-none derivative, a stable 7-membered ring adduct (**Figure 2a**). Isomer **2a** was isolated and characterised by X-ray crystallography, which revealed its stereochemistry and confirmed its structure (**Figure 2b**). To study the reaction further and confirm its outcome on a simpler, non-chiral model, we incubated cyclopropenone **1** with cysteamine (**CA**). As expected, the reaction yielded two pairs of enantiomeric regioisomers (**Figure 2a**) that were separated by column chromatography and characterized by NMR spectroscopy.

Cyclopropenone ring-opening after double nucleophilic attack makes the reaction with 1,2-aminothiols irreversible.

The whole reaction mechanism was inspected through quantum mechanical (QM) calculations using an abbreviated cyclopropenone model (**1'**) and **CA** as reagents (**Figure 2c**). In the first reaction step, cysteamine thiolate undergoes conjugate addition to either the di-substituted (*pathway A*) or tri-substituted (*pathway B*) alkene

carbon atoms of the cyclopropenone. The initial activation barriers for these competitive pathways ultimately leading to the two observed regioisomeric products are quite similar and easily accessible at the reaction temperature despite the different steric hindrance at both positions ($\Delta G^{\ddagger}_{\mathbf{TS1}\ A}$ = 14.6 and $\Delta G^{\ddagger}_{\mathbf{TS1}\ B}$ = 15.6 kcal mol^{-1}, respectively), indicating that both pathways are feasible. Thio-enolate adducts are quite unstable ($\Delta G_{\mathbf{CP}^-}$ ~11 kcal mol^{-1}), as commonly calculated for such intermediates in S-Michael-type reactions.[36] However, and opposing the common trend calculated for non-cyclopropenone electrophiles, protonation of the enolates does not stabilize the adducts ($\Delta G_{\mathbf{CP}}$ = 7–8 kcal mol^{-1}). As a consequence, this first step as well as the addition of simple thiolates (see below) are endergonic and reversible. Two racemic pairs of cyclopropanone diastereomers (**CP** $A_{syn,anti}$) are generated upon protonation of enolates in *pathway A*, and a racemic cyclopropanone regioisomer (**CP** B) in *pathway B*. Of note, transition states for spontaneous ring-opening leading to the ketene-ylide intermediates proposed to mediate in the bioorthogonal ligation of cyclopropenones assisted by triarylphosphines[27] could not be located from these thio-cyclopropanone adducts. Conversely, intramolecular nucleophilic attack of the pendant amino group to the cyclopropanone carbonyl by the same face where the thioether is located proceeds with easily accessible activation barriers ($\Delta G^{\ddagger}_{\mathbf{TS2}}$ = 14–17 kcal mol^{-1}) to give first zwitterionic ($\mathbf{HA_{zw}}$) then neutral (**HA**) hemiaminal bicyclic adducts. As shown previously,[37] such transformations are strongly favored for cyclopropanones due to the significant strain release exerted by the carbonyl $sp^2 \rightarrow sp^3$ rehybridization, despite the formation of 2-thia-5-azabicyclo[4.1.0]heptane structures. The OH-deprotonated form of these bicyclic intermediates ($\mathbf{HA}^-$) undergo fast ring-expansion reactions resembling those occurring in Favorskii rearrangements[37-38] ($\Delta G^{\ddagger}_{\mathbf{TS3}}$ = 6–8 kcal mol^{-1}). The resulting di-substituted carbanion in pathway A ($\Delta G_{\mathbf{TZ}^-\mathbf{A}}$ = -6 kcal mol^{-1}), is more stable that the tri-substituted regioisomer generated in pathway B ($\Delta G_{\mathbf{TZ}^-\mathbf{B}}$ = –4 kcal mol^{-1}), and negative hyperconjugation by neighboring sulfur atoms[39] can play an important role in the stabilization of both ring-expansion products. Finally, these species are protonated to irreversibly form the two experimentally observed 1,4-thiazepan-5-ones as racemic regioisomers with nearly identical energies ($\Delta G_{\mathbf{TZ}}$ = –42 kcal mol^{-1}). Competitive but slower E_1cB-type elimination reactions were calculated from the same alkoxide intermediates ($\mathbf{HA}^-$) ($\Delta G^{\ddagger}_{\mathbf{TS4}}$ = 10–15 kcal mol^{-1}, **Figure S4**), in agreement with the observed preference for the ring-expansion reactions. In fact, this elimination

mechanism would yield the type of α,β-unsaturated amides ((*Z*)- and (*E*)-crotonamides through pathway A and methacrylamide through pathway B) observed as ligation adducts with 1,2-aminophosphines, albeit through a different mechanism.[27]

The activation barriers for the conjugate addition of 2-methylcyclopropenone **1'** and methanethiolate, an abbreviated model of internal cysteine residues and simple biological/unnatural thiols, are very similar to those calculated for the addition of cysteamine ($\Delta G^{\ddagger}_{\mathbf{TS1'}}$ = 14–15 kcal mol^{-1}, **Figure S3**). However, the lack of adjacent free amino groups precludes cyclization and subsequent ring-expansion of the bicyclic intermediates and the associated ring strain release. Therefore, although internal cysteines and other thiols may compete with N-terminal cysteines for reacting with cyclopropenone derivatives, in the absence of such thermodynamic driving force the corresponding adducts would not be observed due to the reversibility of the intrinsically endergonic and reversible S-Michael reaction. Methylamine was also examined as a model nucleophile for the conjugate addition of cysteamine and N-terminal cysteine amino groups or lysine sidechains, to 2-methylcyclopropenone **1'**. The much higher activation barriers calculated for this reaction ($\Delta G^{\ddagger}_{\mathbf{TS1'N}}$ ~ 20 kcal mol^{-1}, **Figure S3b**) demonstrate the high preference for reactions with thiolates (and 1,2-aminothiolates) over primary amines.

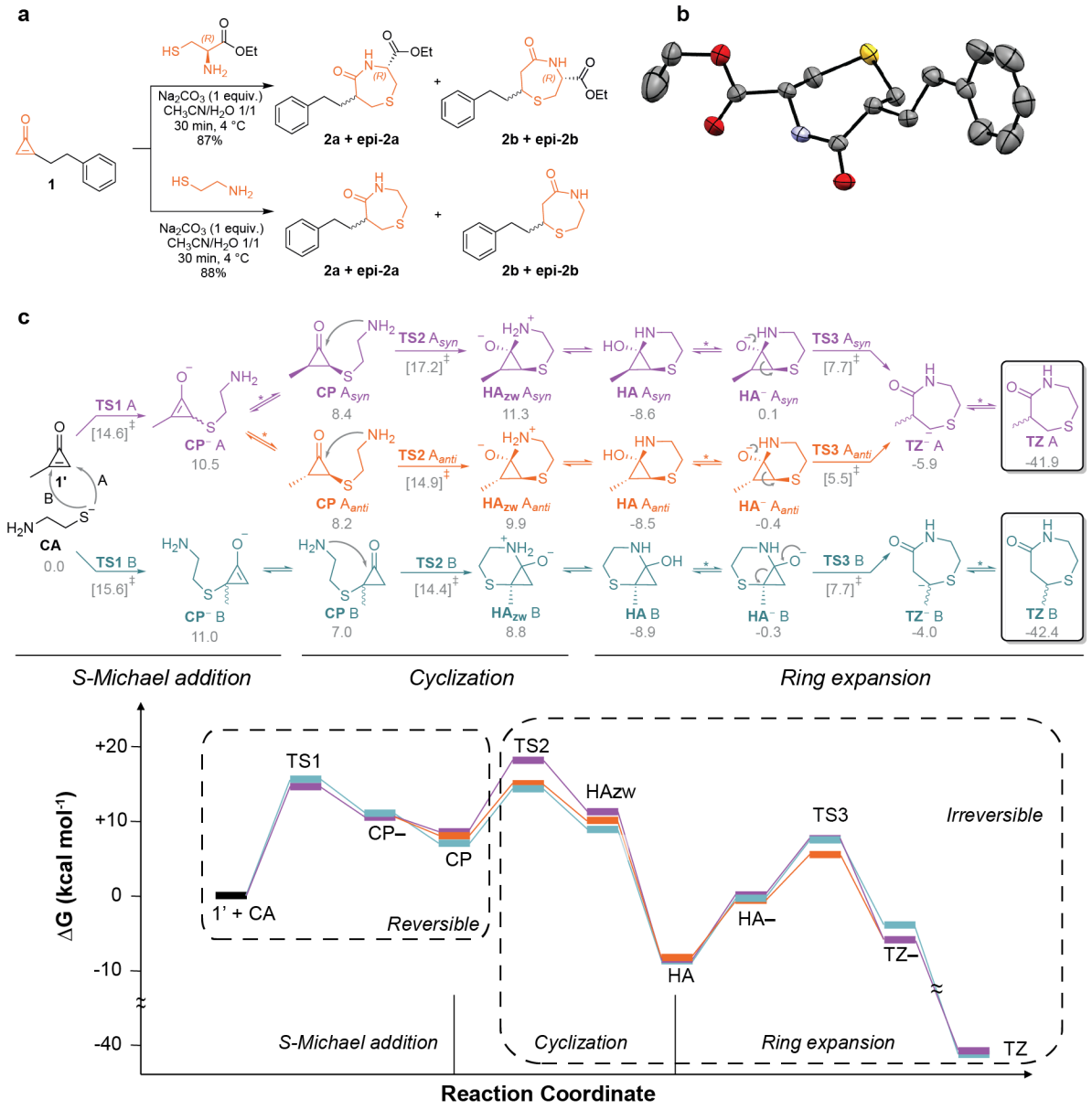

Fig. 2 | a. Reaction of cyclopropenone **1** with L-cysteine ethyl ester hydrochloride and cysteamine; **b.** ORTEP plot (50% probability level) of compound **2a**. Hydrogen atoms are omitted for clarity; **c.** Proposed mechanism and minimum energy pathways for the reaction between 2-methylcyclopropenone (**1'**) and cysteamine thiolate (**CA**) calculated with PCM(H_2O)/M06-2X/6-31+G(d,p). Note that each calculated structure represents a racemic compound constituted by two enantiomers with identical energies. *: protonation and deprotonation equilibria were calculated using tri-hydrated bicarbonate ($HCO_3^{2-} \cdot 3\ H_2O$) and carbonate ($CO_3^{2-} \cdot 3\ H_2O$) anions as an acid and a base, respectively. Given the intrinsic inaccuracy of such estimations, relative energies of charges/neutral species must be considered with caution. Relative free energies (ΔG) are given in kcal mol^{-1}. Numbers in brackets represent the relative

energies of transition structures. Three-dimensional structures for all lowest-energy stationary points are available in the **Supplementary Information**.

CPO-cysteine reaction is selective, and proceeds with fast kinetics in mild conditions.

Next, we examined the kinetics and selectivity of the CPO-cysteine reaction. To evaluate the reaction kinetics, we followed the HPLC chromatogram peak area of starting cyclopropenone **1** at different time points during the reaction (**Figure S5**). The second-order rate constant for this reaction was determined to be 3.0 $M^{-1}s^{-1}$ at 4 ºC, which is comparable to the value reported for the CBT-cysteine reaction (9.19 $M^{-1}s^{-1}$) and strain-promoted azide-alkyne cycloaddition reactions (10^{-2} – 1 $M^{-1}s^{-1}$) performed at higher temperatures (37 ºC).[40] In fact, the rate constant for the reaction between cyclopropenone **1** and cysteine extrapolated to 37 ºC would be 67 M^{-1} s^{-1}, which supersedes the aforementioned values. These results are in good agreement with the low activation barrier calculated for the reaction with cysteamine and confirm the high reactivity of monosubstituted cyclopropenones, making them ideal for protein modification with reduced amounts of labelling reagent and/or at particularly unreactive positions. Then, we tested the reactivity of probe **1** with other biological nucleophiles, including lysine, serine, threonine, tyrosine, glutathione, and cysteine methyl ester with tert-butoxycarbonyl protected amino group (Boc-Cys-OMe). As shown by LC–MS, compound **1** did not react with lysine, serine, threonine or tyrosine, but showed excellent selectivity for cysteine. In addition, these amino acids did not interfere with the reaction between CPO **1** and cysteine (**Figure S8-S12**). Boc-Cys-OMe reacted with compound **1** to form a complex mixture of products, although the reaction was slower relative to the one with cysteine that contains an unprotected amino group (**Figure S14**). In fact, when stoichiometric amounts of both cysteines (with free and Boc-protected amino groups) were treated with CPO probe **1**, we observed high selectivity towards 1,2-aminothiol resulting in almost exclusive formation of 1,4-thiazepan-5-one products **2a–2d** (**Figure S14**). Importantly, glutathione, the most abundant low molecular-weight thiol in cells, did not react with CPO probe **1**, and did not interfere with the reaction between CPO **1** and cysteine (**Figure S13**), which confirms the computationally predicted high selectivity of cyclopropenones towards unprotected 1,2-aminothiols over simple thiols.

Cyclopropenone probes have a clear advantage over CBT probes that react with glutathione and other thiol nucleophiles[24] for the reaction with N-terminal cysteine.

Straightforward synthesis from easily accessible alkynes affords CPO probes with a range of payloads.

Before testing the applicability of the reaction in peptide and protein models, we designed and synthesised several cyclopropenone-based probes with different functionalities. We started with commercially available 5-hexynoic acid **4**, which was converted into corresponding pentafluorophenol (PFP) ester **5**. Ester **5** was subjected to cyclopropenation reaction to afford corresponding CPO-modified activated ester **CPO-PFP**. **CPO-PFP** is an easy-to-handle, stable (> 6 months) white solid that can be used to link the cyclopropenone unit to a molecule of interest through a primary amino group in a fast and efficient manner. With this intermediate, we prepared several labelling moieties of interest for peptide and protein bioconjugation. We optimised the reaction conditions by using a simple amine, benzylamine, to produce a model cyclopropenone with benzyl group (**CPO-BN**) in high yield. Next, we applied the optimised conditions to prepare cyclopropenone-based probes bearing polyethylene glycol (**CPO-PEG**), alkyne (**CPO-PEG-Alkyne**) and a fluorescent dye (**CPO-EDANS**). Overall, simple incubation of intermediate **CPO-PFP** with amine-bearing molecules of interest at room temperature for 20 min afforded CPO-based probes for protein and peptide labelling in high yields.

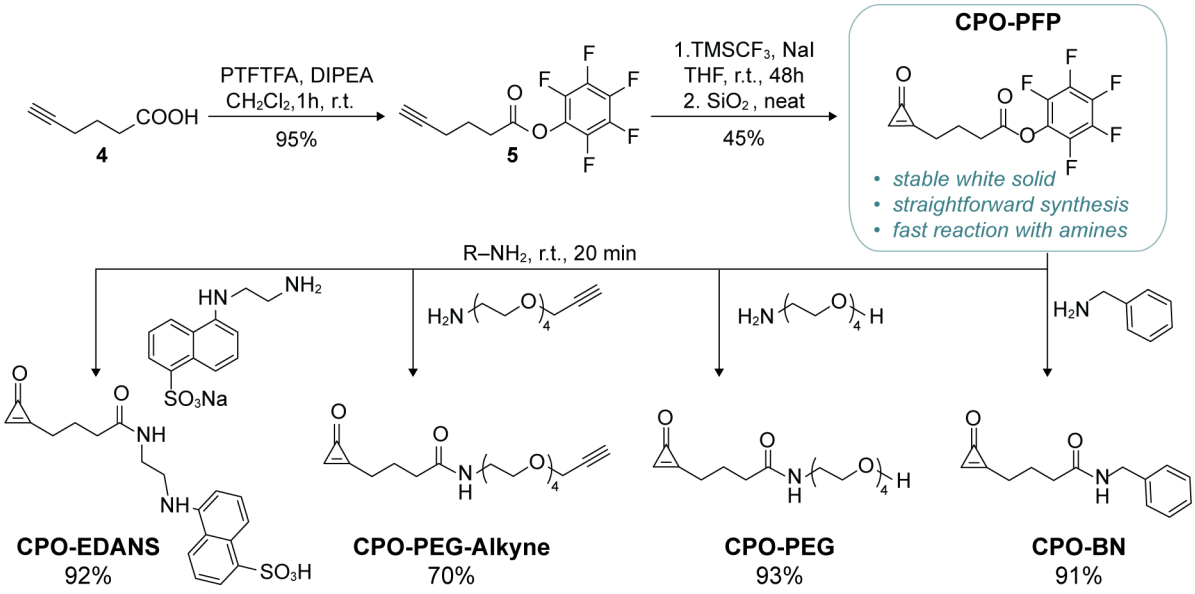

Fig. 3 | Synthesis of cyclopropenone-based probes with different functionalities.

CPO-probes selectively modify N-terminal cysteine on peptides, allowing dual cysteine labelling.

The high reactivity and selectivity of compound **1** towards 1,2-aminothiols prompted us to investigate the ability of cyclopropenone probes to modify N-terminal cysteine in peptides. Our investigation began with two unprotected 5-mer model peptides with N-terminal cysteines: CAIAI (**P1**) and CAIKI (**P2**). Notably, the **P2** contained a lysine residue to test the selectivity of the CPO reagent for cysteines over lysines. Treatment of both peptides (2 mM) with **CBO-BN** linker (2 equiv.) in NaP_i buffer (20 mM, pH 7)/acetonitrile resulted in complete conversion into the expected products after 1 h at 4 ºC (**Figure 4a, S15**, and **S17**). Similarly, modification of peptides **P1** and **P2** with **CPO-PEG** linker resulted in formation of the expected PEGylated products, as confirmed by LC–MS (**Figure S16** and **S18**).

We then assayed the method with more complex peptides containing multiple nucleophilic side chains. To begin with, we focused on laminin-derived synthetic peptide **P3**, which inhibits tumor growth.[41] This peptide contains nine amino acids, including N-terminal cysteine, tyrosine, serine, asparagine and arginine. Application of our optimized cysteine modification protocol resulted in full conversion of the starting peptide after 1 h, as confirmed by LC–MS (**Figure 4a** and **S19**). Then, we targeted the more challenging GTP-Binding Protein Fragment **P4**, which is a 16-mer peptide with multiple nucleophilic amino acids (one N-terminal cysteine, three lysines, two serines, one threonine, one methionine). The reaction of peptide **P4** with **CPO-BN** linker (2 equiv.) resulted in full conversion into products **P4-BNa** and **P4-BNb** within 1 h at 4 ºC (**Figure 4a** and **S20**). No mass signal for double chemical modification was observed, which further highlights the chemoselectivity of the protocol.

As described above, N-terminal cysteine could be modified selectively even in the presence of other thiol nucleophiles. To test whether the same behaviour will be observed on the peptide level, we used **CPO-BN** to modify a peptide that contains both the N-terminal and internal cysteine residues. Vasopressin (**P5**), a 9-mer cyclic peptide, upon the reduction of the disulfide bridge generates cysteine residues in positions 1 and 6. In control experiments without the addition of the reducing agent, vasopressin showed no reaction with **CPO-BN** probe, or another common reagent for cysteine conjugation, *N*-methylmaleimide (NEM). We next reduced the disulfide bridge

with TCEP (**Figure 4b** and **S21**) and incubated the reduced peptide **P5-SH** with N-methylmaleimide or CPO-BN probes. Cyclopropenones readily react with phosphines[27-28] so care was taken to ensure that excess TCEP was removed from the reaction mixture before the addition of cyclopropenone (e.g., by using immobilised TCEP or limiting the amount used to 1 equiv.). As expected, NEM does not show any selectivity for the position of cysteine residue within the peptide, and incubation of **P5-SH** with NEM quickly resulted in modification of both cysteines (**Figure S22**). In contrast, incubation of the reduced vasopressin with **CPO-BN** probe resulted in selective modification of the N-terminal cysteine residue (**Figure 4b** and **S24**). The same result was obtained even if an excess of CPO probe (2 equiv.) was used. In further tests, we confirmed that the internal cysteine was still available for modification after selective labelling of the N-terminal cysteine with **CPO-BN**. Treatment of CPO-modified vasopressin with NEM allowed quantitative modification of the internal cysteine. In line with previous experiments, the internal cysteine could be efficiently conjugated to this reagent (**Figure 4b** and **S25**). LC–MS/MS analysis confirmed the sites of both modifications (**Figure S26**). Next, we examined whether CBT-based probes offer similar selectivity for the N-terminal cysteine of vasopressin. Under the conditions used for the reaction between **CPO-BN** and **P5-SH**, 2-cyano-6-hydroxybenzothiazole (**CBT)** probe afforded a mixture of single- and double-modified peptides (**Figure 4b** and **S23**), which highlights the higher selectivity of the cyclopropenone-based reagents. This experiment demonstrates that CPO-labels can selectively modify N-terminal cysteine functionality even in the presence of other thiol groups, which can allow selective dual-cysteine labelling.

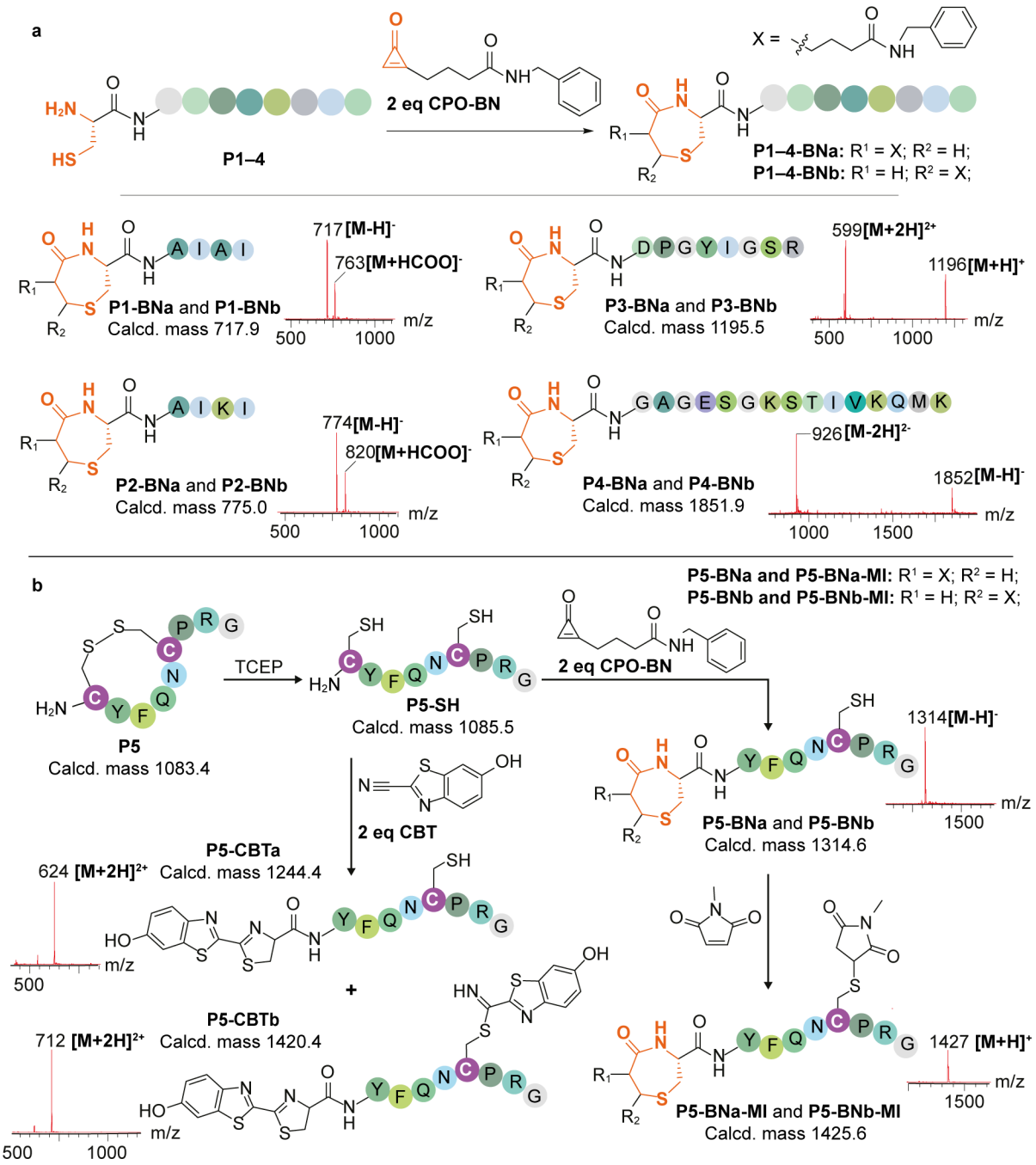

Fig. 4 | Chemo- and regioselective modification of 1,2-aminothiols on peptides. **a.** Site-selective modification of peptides **P1–P4** by **CPO-BN** reagent. Modification occurs only on N-terminal cysteine residue; **b. CPO-BN** probe selectively modifies N-terminal cysteine of vasopressin (**P5**) and leaves the internal cysteine unmodified and accessible for further functionalization.

CPO-probes react exclusively with 1,2-aminothiols on proteins.

To provide proof of concept for our CPO-reagent to be used for protein modification, we produced recombinant enhanced green fluorescent protein containing *N*-terminal

cysteine (**Cys-GFP**). Recombinant proteins with *N*-terminal cysteines are widely used for native chemical ligation technology, so various approaches have been developed for their direct production.[42-44] To produce **Cys-GFP**, we engineered GFP protein variant with the tobacco etch virus (TEV) protease recognition sequence (ENLYFQ↓C; arrow indicates the cleavage site) introduced after the hexa-His purification tag at the *N*-terminus. In this way, the cleavage of the expressed protein by TEV protease simultaneously removed the His-tag and generated GFP protein with the N-terminal cysteine. Incubation of **Cys-GFP** with **CPO-BN** (100 equiv.) in NaP_i buffer (20 mM, pH 7.0) at room temperature for 2 h gave desired conjugate **GFP-CPO-BN** with high efficiency, as confirmed by LC–MS (**Figure 5a, 5b,** and **S33**). The high reactivity of GFP terminal cysteine required the use of DTT (25 equiv.) to avoid protein dimerization and thus an excess of cyclopropenone labelling reagent and a higher temperature compared to previous modification experiments was required in order to achieve complete conversion. Tryptic digestion of the conjugate and subsequent LC–MS/MS analysis confirmed the site of modification (**Figure S36**). Next, we evaluated the stability of **GFP-CPO-BN** conjugate with an excess of glutathione. After 24 h incubation at 37 ºC with 5 mM glutathione, bioconjugate **GFP-CPO-BN** was shown to be stable with no de-conjugation observed (**Figure S34** and **S35**). 0.5 μs Molecular dynamic simulations were performed on the four possible conjugates of GFP-CPO-BN (SI and **Figure S49**). The root-mean-square deviation values of the peptide backbone in all complexes range from 1.21 to 2.67 Å. Hence, these data suggest that the installation of CPO-BN scaffold does not cause relevant structural modifications on the protein, which is required for biological activity.

After demonstrating chemoselectivity and efficiency of **CPO-BN** for *N*-terminal cysteine bioconjugation, we expanded the scope of the reaction to other CPO-based linkers. Testing **CPO-PEG** and **CPO-EDANS** under the same reaction conditions as used for **CPO-BN** resulted in successful conversion of **Cys-GFP** into the expected products (**Figure 5c, 5d, S37** and **S38**). For **CPO-EDANS**, lower conversion (70%) is likely explained by the poor solubility of **CPO-EDANS**. We next assessed whether pre-purification of CPO-reagents is necessary for the conjugation reaction. For this purpose, we mixed **CPO-PFP** reagent with commercial amine-functionalised biotin to obtain **CPO-biotin** linker (**Figure S42**). Then, without any additional purification steps, the resulting solution was directly added to **Cys-GFP** (100 equiv. of **CPO-biotin** to 1

equiv. of **Cys-GFP**), and the reaction was monitored by LC–MS. To our delight, conversion of **Cys-GFP** into the expected conjugate was complete after incubation for 2 h (**Figure 5e** and **S36**). Thus, there is no need to purify the conjugation reagent before the labelling step, which is beneficial for the practical preparation of bioconjugates with payloads (eg. dyes, oligonucleotides, or drugs).

Notably, during our work on reaction optimization with **Cys-GFP,** addition of dithiothreitol (DTT) to protein stock solution was necessary to maintain the reactivity of cysteine residue, probably due to high susceptibility of the *N*-terminal cysteine to undergo oxidation and protein cross-linking. Nevertheless, an excess of DTT (500 equiv.) in the reaction mixture did not hinder the bioconjugation reaction of CPO probes with *N*-terminal cysteine. However, with similar NEM-based probes for **Cys-GFP** bioconjugation under the same reaction conditions, very low conversion rates were observed (0–15%; **Figure S39** and **S41**). This can be explained by the incompatibility of maleimides with excess DTT. Again, this highlights the selectivity of the cyclopropenone-based reagents towards 1,2-aminothiols for which 100 equiv. of the probe was enough to afford protein labelling with a range of probes even in the presence of excess of DTT (**Figure 5b** and **5e**).

Finally, we decided to explore whether CPO-based probes can selectively label a NCys-containing protein in the presence of other proteins bearing internal cysteines. We prepared a mixture of four proteins: **Cys-GFP** (3 free cysteines including N-terminal cysteine), engineered variant of the C2A domain of Synaptotagmin-I (**C2Am**, 1 free cysteine), Annexin V (**AnxV**, 1 free cysteine), engineered variant of a nanobody (**DesAB-HET**, 1 free cysteine), and incubated it with **CPO-biotin** linker or a similar maleimide-based linker (**MI-biotin**, **Figure S43**). **CPO-biotin** successfully conjugated to **Cys-GFP**, whereas **C2Am, AnxV,** and **DesAB-HET** proteins were unchanged. Similar results were observed regardless of whether or not DTT (500 equiv.) was present in the reaction mixture (**Figure 5f, S41** and **S43**). In contrast, **MI-biotin** probe fully modified 3 out of four proteins (**Cys-GFP**, **C2Am**, and **DesAB-HET**) when DTT was absent from the reaction mixture (**Figure 5f** and **S42**). **AnxV** was not modified because it usually requires larger excess of reagents, higher temperatures or longer reaction times for the cysteine modification to proceed.[45] As expected, in the presence of DTT, only minor modification of proteins (0–15%) with **MI-biotin** was observed

(**Figure 5f, S44**). Overall, these data confirm the high selectivity, orthogonality and highlight the suitability of CPO-reagents to label specific proteins in complex mixtures.

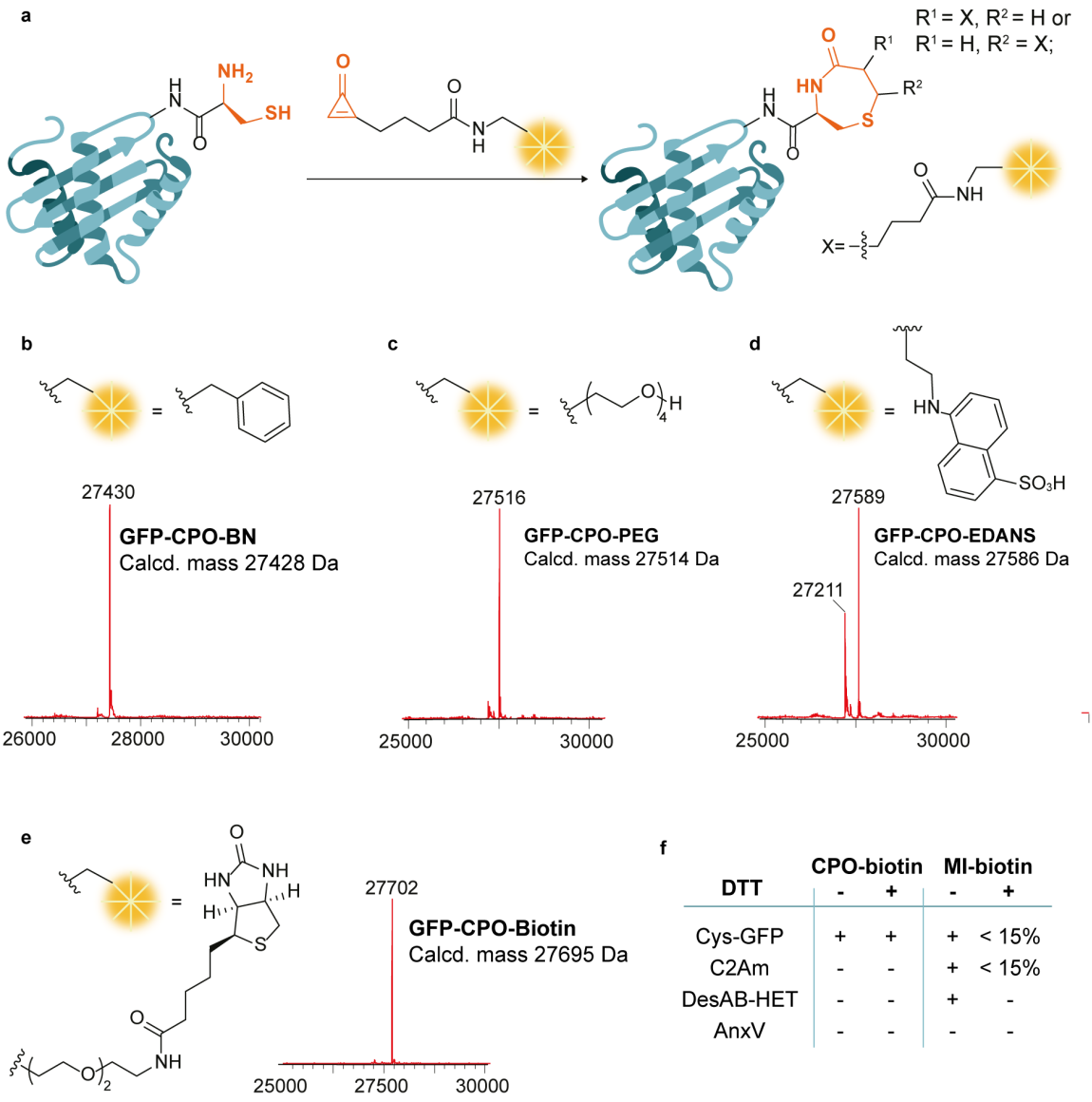

Fig. 5 | Chemo- and regioselective modification of 1,2-aminothiols on proteins. **a.** Site-selective bioconjugation of Cys-GFP with CPO-BN **(b)**, CPO-PEG **(c)**, CPO-EDANS **(d)**, and CPO-biotin **(e)** probes; **f.** Modification of **Cys-GFP**, **C2Am**, **AnxV**, and **DesAB-HET** protein mixture (5 μM each) with **CPO-biotin** or **MI-biotin** probes (50 equiv.) in NaP_i buffer (20 mM, pH 7.0) with or without DTT (500 equiv.).

CONCLUSIONS

We have developed an efficient method for N-terminal cysteine modification on peptide and proteins. The method is based on a reaction of monosubstituted

cyclopropenones with 1,2-aminothiols, which results in formation of a stable 1,4-thiazepa-5-none linkage. Synthesis of cyclopropenone-based conjugation reagents that bear various functional groups can be easily achieved in three steps starting from commercially available 5-hexynoic acid. The CPO-cysteine reaction proceeds with high efficiency in mild conditions (aqueous buffer, pH 7, 4–25 °C). The requirement for a nucleophile adjacent to the cysteine's thiol for the reaction to be irreversible, makes this transformation ideal for the selective labelling N-terminal cysteines. We have demonstrated that this approach can be used to modify N-terminal cysteine residues on a peptide or protein of interest even in the presence of remote nucleophilic residues, other biological thiols and reagents incompatible with common labelling reagents. Finally, we show the CPO-based probes can achieve selective labelling of NCys-containing protein in the presence of other proteins bearing internal cysteines. Further work is ongoing on the selective protein labelling in complex proteomes and in living cells. Overall, this ability to target specific cysteine residue on a protein will allow the straightforward construction of complex bioconjugates with well-defined structure, which makes this approach an important addition to the protein conjugation toolbox.

ASSOCIATED CONTENT

Supporting information

Detailed methods, characterization data and additional figures (PDF)
CCDC 2010531 contains the supplementary crystallographic data for this paper

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

This project has received funding from Fundação para a Ciência e a Tecnologia, Portugal (Investigator IF/00624/2015 to G.J.L.B), the Spanish Ministry of Science and Innovation (MCI) co-financed with FEDER funds (RTI2018-099592-B-C21 to F.C and RTI2018-099592-B-C22 to G.J.O) and Swiss National Science Foundation (Early Postdoc.Mobility P2BEP3_175253 to A.I). G.J.L.B. is a Royal Society University Research Fellow (URF\R\180019). The authors thank Dr Vikki Cantrill and Claudia Flandoli (draw.science) for their help with the editing of this manuscript and its figures,

respectively. We also thank Prof. Michele Vendruscolo and Phil Lindstedt for the generous gift of DesAB-HET antibody, Dr. André Neves and Prof. Kevin Brindle for providing C2Am.

REFERENCES

1. Beck, A.; Goetsch, L.; Dumontet, C.; Corvaïa, N., Strategies and challenges for the next generation of antibody–drug conjugates. *Nat. Rev. Drug Discov.* **2017,** *16* (5), 315–337.
2. Chudasama, V.; Maruani, A.; Caddick, S., Recent advances in the construction of antibody–drug conjugates. *Nat. Chem.* **2016,** *8* (2), 114–119.
3. Pelegri-O'Day, E. M.; Lin, E.-W.; Maynard, H. D., Therapeutic Protein–Polymer Conjugates: Advancing Beyond PEGylation. *J. Am. Chem. Soc.* **2014,** *136* (41), 14323–14332.
4. Spicer, C. D.; Pashuck, E. T.; Stevens, M. M., Achieving Controlled Biomolecule–Biomaterial Conjugation. *Chem. Rev.* **2018,** *118* (16), 7702–7743.
5. Zhang, G.; Zheng, S.; Liu, H.; Chen, P. R., Illuminating biological processes through site-specific protein labeling. *Chem. Soc. Rev.* **2015,** *44* (11), 3405–3417.
6. Xue, L.; Karpenko, I. A.; Hiblot, J.; Johnsson, K., Imaging and manipulating proteins in live cells through covalent labeling. *Nat. Chem. Biol.* **2015,** *11* (12), 917–923.
7. Hoyt, E. A.; Cal, P. M. S. D.; Oliveira, B. L.; Bernardes, G. J. L., Contemporary approaches to site-selective protein modification. *Nat. Rev. Chem.* **2019,** *3* (3), 147–171.
8. Taylor, M. T.; Nelson, J. E.; Suero, M. G.; Gaunt, M. J., A protein functionalization platform based on selective reactions at methionine residues. *Nature* **2018,** *562* (7728), 563–568.
9. Zhang, C.; Welborn, M.; Zhu, T.; Yang, N. J.; Santos, M. S.; Van Voorhis, T.; Pentelute, B. L., π-Clamp-mediated cysteine conjugation. *Nat. Chem.* **2016,** *8* (2), 120–128.
10. Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Guerreiro, A.; Cal, P. M. S. D.; Bertoldo, J.; Maneiro, M.; Perkins, E.; Howard, J.; Deery, M. J.; Chalker, J. M.; Corzana, F.; Jiménez-Osés, G.; Bernardes, G. J. L., Chemo- and Regioselective

Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **2018,** *140* (11), 4004–4017.

11. Adusumalli, S. R.; Rawale, D. G.; Singh, U.; Tripathi, P.; Paul, R.; Kalra, N.; Mishra, R. K.; Shukla, S.; Rai, V., Single-Site Labeling of Native Proteins Enabled by a Chemoselective and Site-Selective Chemical Technology. *J. Am. Chem. Soc.* **2018,** *140* (44), 15114–15123.

12. Bernardim, B.; Cal, P. M. S. D.; Matos, M. J.; Oliveira, B. L.; Martínez-Sáez, N.; Albuquerque, I. S.; Perkins, E.; Corzana, F.; Burtoloso, A. C. B.; Jiménez-Osés, G.; Bernardes, G. J. L., Stoichiometric and irreversible cysteine-selective protein modification using carbonylacrylic reagents. *Nat. Commun.* **2016,** *7* (1), 13128.

13. Vinogradova, E. V.; Zhang, C.; Spokoyny, A. M.; Pentelute, B. L.; Buchwald, S. L., Organometallic palladium reagents for cysteine bioconjugation. *Nature* **2015,** *526* (7575), 687–691.

14. Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G., Chemical Modification of Proteins at Cysteine: Opportunities in Chemistry and Biology. *Chem. Asian J.* **2009,** *4* (5), 630–640.

15. Rosen, C. B.; Francis, M. B., Targeting the N terminus for site-selective protein modification. *Nat. Chem. Biol.* **2017,** *13* (7), 697–705.

16. Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B., Synthesis of proteins by native chemical ligation. *Science* **1994,** *266* (5186), 776.

17. Agouridas, V.; El Mahdi, O.; Diemer, V.; Cargoët, M.; Monbaliu, J.-C. M.; Melnyk, O., Native Chemical Ligation and Extended Methods: Mechanisms, Catalysis, Scope, and Limitations. *Chem. Rev.* **2019,** *119* (12), 7328–7443.

18. Zhang, L.; Tam, J. P., Thiazolidine Formation as a General and Site-Specific Conjugation Method for Synthetic Peptides and Proteins. *Anal. Biochem.* **1996,** *233* (1), 87–93.

19. Casi, G.; Huguenin-Dezot, N.; Zuberbühler, K.; Scheuermann, J.; Neri, D., Site-Specific Traceless Coupling of Potent Cytotoxic Drugs to Recombinant Antibodies for Pharmacodelivery. *J. Am. Chem. Soc.* **2012,** *134* (13), 5887–5892.

20. Faustino, H.; Silva, M. J. S. A.; Veiros, L. F.; Bernardes, G. J. L.; Gois, P. M. P., Iminoboronates are efficient intermediates for selective, rapid and reversible N-terminal cysteine functionalisation. *Chem. Sci.* **2016,** *7* (8), 5052–5058.

21. Bandyopadhyay, A.; Cambray, S.; Gao, J., Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation. *Chem. Sci.* **2016,** *7* (7), 4589–4593.
22. Gao, J.; Li, K.; Wang, W., Fast and Stable N-Terminal Cysteine Modification via Thiazolidino Boronate Mediated Acyl Transfer. *Angew. Chem. Int. Ed.* **2020,** doi:10.1002/anie.202000837.
23. Podsiadły, R.; Grzelakowska, A.; Modrzejewska, J.; Siarkiewicz, P.; Słowiński, D.; Szala, M.; Świerczyńska, M., Recent progress in the synthesis of firefly luciferin derivatives. *Dyes Pigm.* **2019,** *170*, 107627.
24. Ren, H.; Xiao, F.; Zhan, K.; Kim, Y.-P.; Xie, H.; Xia, Z.; Rao, J., A Biocompatible Condensation Reaction for the Labeling of Terminal Cysteine Residues on Proteins. *Angew. Chem. Int. Ed.* **2009,** *48* (51), 9658–9662.
25. Yuan, Y.; Liang, G., A biocompatible, highly efficient click reaction and its applications. *Org. Biomol. Chem.* **2014,** *12* (6), 865–871.
26. Cohen, M.; Bretler, U.; Albeck, A., Peptidyl cyclopropenones: Reversible inhibitors, irreversible inhibitors, or substrates of cysteine proteases? *Protein Sci.* **2013,** *22* (6), 788–799.
27. Shih, H.-W.; Prescher, J. A., A Bioorthogonal Ligation of Cyclopropenones Mediated by Triarylphosphines. *J. Am. Chem. Soc.* **2015,** *137* (32), 10036–10039.
28. Row, R. D.; Shih, H.-W.; Alexander, A. T.; Mehl, R. A.; Prescher, J. A., Cyclopropenones for Metabolic Targeting and Sequential Bioorthogonal Labeling. *J. Am. Chem. Soc.* **2017,** *139* (21), 7370–7375.
29. Experiments show cyclopropenone is aromatic. *Chem. Eng. News* **1983,** *61* (38), 33.
30. Potts, K. T.; Baum, J. S., Chemistry of cyclopropenones. *Chem. Rev.* **1974,** *74* (2), 189–213.
31. Krebs, A. W., Cyclopropenylium Compounds and Cyclopropenones. *Angew. Chem. Int. Ed.* **1965,** *4* (1), 10–22.
32. Kogen, H.; Kiho, T.; Tago, K.; Miyamoto, S.; Fujioka, T.; Otsuka, N.; Suzuki-Konagai, K.; Ogita, T., Alutacenoic Acids A and B, Rare Naturally Occurring Cyclopropenone Derivatives Isolated from Fungi: Potent Non-Peptide Factor XIIIa Inhibitors. *J. Am. Chem. Soc.* **2000,** *122* (8), 1842–1843.

33. Reber, K. P.; Gilbert, I. W.; Strassfeld, D. A.; Sorensen, E. J., Synthesis of (+)-Lineariifolianone and Related Cyclopropenone-Containing Sesquiterpenoids. *J. Org. Chem.* **2019,** *84* (9), 5524–5534.
34. Wang, F.; Luo, T.; Hu, J.; Wang, Y.; Krishnan, H. S.; Jog, P. V.; Ganesh, S. K.; Prakash, G. K. S.; Olah, G. A., Synthesis of gem-Difluorinated Cyclopropanes and Cyclopropenes: Trifluoromethyltrimethylsilane as a Difluorocarbene Source. *Angew. Chem. Int. Ed.* **2011,** *50* (31), 7153–7157.
35. Rullière, P.; Cyr, P.; Charette, A. B., Difluorocarbene Addition to Alkenes and Alkynes in Continuous Flow. *Org. Lett.* **2016,** *18* (9), 1988–1991.
36. Krenske, E. H.; Petter, R. C.; Houk, K. N., Kinetics and Thermodynamics of Reversible Thiol Additions to Mono- and Diactivated Michael Acceptors: Implications for the Design of Drugs That Bind Covalently to Cysteines. *J. Org. Chem.* **2016,** *81* (23), 11726–11733.
37. Salaun, J., Cyclopropanone hemiacetals. *Chem. Rev.* **1983,** *83* (6), 619–632.
38. Guijarro, D.; Yus, M., The Favorskii Rearrangement: Synthetic Applications. *Curr. Org. Chem.* **2005,** *9* (17), 1713–1735.
39. Bernasconi, C. F.; Kittredge, K. W., Carbanion Stabilization by Adjacent Sulfur: Polarizability, Resonance, or Negative Hyperconjugation? Experimental Distinction Based on Intrinsic Rate Constants of Proton Transfer from (Phenylthio)nitromethane and 1-Nitro-2-phenylethane. *J. Org. Chem.* **1998,** *63* (6), 1944–1953.
40. Oliveira, B. L.; Guo, Z.; Bernardes, G. J. L., Inverse electron demand Diels–Alder reactions in chemical biology. *Chem. Soc. Rev.* **2017,** *46* (16), 4895–4950.
41. Sakamoto, N.; Iwahana, M.; Tanaka, N. G.; Osada, Y., Inhibition of Angiogenesis and Tumor Growth by a Synthetic Laminin Peptide, CDPGYIGSR-NH_2. *Cancer Res.* **1991,** *51* (3), 903–906.
42. Gentle, I. E.; De Souza, D. P.; Baca, M., Direct Production of Proteins with N-Terminal Cysteine for Site-Specific Conjugation. *Bioconjugate Chem.* **2004,** *15* (3), 658–663.
43. Tolbert, T. J.; Wong, C.-H., New Methods for Proteomic Research: Preparation of Proteins with N-Terminal Cysteines for Labeling and Conjugation. *Angew. Chem. Int. Ed.* **2002,** *41* (12), 2171–2174.
44. Liu, D.; Xu, R.; Dutta, K.; Cowburn, D., N-terminal cysteinyl proteins can be prepared using thrombin cleavage. *FEBS Lett.* **2008,** *582* (7), 1163–1167.

45. Cal, P. M. S. D.; Sieglitz, F.; Santos, F. M. F.; Parente Carvalho, C.; Guerreiro, A.; Bertoldo, J. B.; Pischel, U.; Gois, P. M. P.; Bernardes, G. J. L., Site-selective installation of BASHY fluorescent dyes to Annexin V for targeted detection of apoptotic cells. *Chem. Commun.* **2017,** *53* (2), 368–371.