Contemporary approaches to site-selective protein modification

Emily A. Hoyt,[1] Pedro M. S. D. Cal,[2] Bruno L. Oliveira[1,2] and Gonçalo J. L. Bernardes[1,2*]

[1]Department of Chemistry, University of Cambridge, Lensfield Road, Cambridge CB2 1EW (UK)

[2]Instituto de Medicina Molecular, Faculdade de Medicina, Universidade de Lisboa, Avenida Professor Egas Moniz, 1649-028, Lisboa (Portugal)

*email: gb453@cam.ac.uk; gbernardes@medicina.ulisboa.pt

Proteins constitute the majority of nature's worker biomolecules. Designed for specific functions, complex tertiary structures make proteins ideal candidates for analyzing natural systems and creating novel biological tools. Due to both large size and the need for proper folding, *de novo* synthesis of proteins has been quite a challenge, leading scientists to focus on modifying protein templates already provided by nature. Recently developed methods for protein modification fall into two broad categories: those that can modify the natural protein template directly and those that require genetic manipulation of the amino acid sequence prior to modification. The goal of this review is to provide not only a window through which to view the many opportunities created by novel protein modification techniques, but also to act as an initial guide to help scientists find direction and form ideas in an ever-growing field. In addition to the highlighting methods reported in the past five years, we aim to provide a broader sense of the goals and outcomes of protein modification and bioconjugation in general. While the main body of the paper comprises reactions directly

involving proteins as a starting material, some further functionalization strategies as well as biological applications are also acknowledged. The discussion concludes by speculating what trends and discoveries will most likely come next in the field.

[H1] **Introduction**

Over time, proteins have become the most valuable biomolecules among the vast variety of cellular components. Due to their multifaceted nature, performing many roles both within and outside of the cell cycle, amino acid-based (AA-based) protein modification can lead to a myriad of diverse applications.[1,2] This is particularly apparent in the field of bioconjugates where therapeutic potential in the treatment of cancer and other medical problems has led to them infiltrating the pharmaceutical market (in addition to an already established presence in smaller markets such as biomaterials, textile manufacturing, and food processing). With huge commercial value resulting from such medical promise, it is not surprising that the toolbox available to chemists is ever-expanding. [3] Enhanced bioavailability, fluorescent tracking, post- translational modification insertions, and targeted delivery are just a few of the numerous possible applications of protein conjugates.[1,4] In order for these powerful applications to be realized, however, the protein modification must avoid interfering with protein function. Arguably, the ability to retain protein function is primarily affected by the site and size of the protein modification. Site-selective methods that result in homogeneous products are thus in demand because they maximize the chance of success in the desired application.

Starting from classical methods that are unable to deliver site-specificity, chemical biologists have been able to improve protein modification procedures and identify routes that circumvent the inherent obstacles of bioconjugation (Figure 1).[5] One of the most challenging

criteria is the need for modifications to occur under mild reaction conditions, in an aqueous environment, and in the presence of multiple unprotected, chemical entities that can promote cross-reactions. Moreover, promoting such reactions under natural biological conditions while also maintaining structural and functional integrity adds an extra level of difficulty. Nevertheless, different methods have been developed that take advantage of reactive, endogenous AA sidechains.. The nucleophilicity, solvent accessibility, and relative abundance of lysine (Lys) and cysteine (Cys) residues have encouraged scientists to target these sidechains using maleimides, *N*-hydroxysuccinimide (NHS) esters, and α-halocarbonyls as electrophiles for modification.[6,7] Michael addition, activated ester amidation, and reductive amination have become particularly popular (Figure 1).[8] Each method presents particular advantages and disadvantages, but common motivations for the continued search for optimized protein modification methods centre on improving reaction rate and product homogeneity.

Given that the available chemical functional groups are naturally limited to the canonical AAs, different strategies have been pursued to increase selectivity and improve kinetics.[9] To do so, researchers have employed creative solutions that take advantage of strategies within the realm of nature (for example, enzymatic tags/recognition sites and acknowledgement of the various microenvironments within a protein's structure), genetic engineering for the introduction of natural or abiotic functional groups (e.g. genetic sequence insertions and subsequent chemical reactions), or even previously unexplored chemistry or reaction optimizations (e.g. controlled reaction conditions or metal-catalyzed/directed reactions) (Figure 1).[1,2,5,7,8,10,11] Among the most successful methods to achieve homogenous products, genetic engineering to incorporate a new amino acid (either canonical or non-canonical) within the structure of a protein followed by modification is currently the method of choice.[9] Direct native protein modification is arguably the ideal, however,

(avoiding the need for any prior protein modification), and seems to be a promising emerging strategy with many more examples being described. [12] This review focuses on appraising modification methodologies from the last five years as well as novel examples of downstream functionalization of these modifications and therapeutic biological applications present in this ever-growing field of site-selective protein modification.

[H1] Direct modification of native proteins

With only a select number of reactive groups represented and often repeated throughout the protein sequence, modification of endogenous AAs has been difficult to accomplish with a high level of selectivity. Although challenging, the potential for the higher yields available from a single-step process has inspired efforts to develop such methods. Difficulties with conventional strategies for modifying native protein sequences result from the lack of reaction site-specificity (heterogeneous products result from repeated functionalities being modified because the methods are site-selective rather than site-specific).[1] Site-specificity can avoid the modification of reactive residues that are critical to protein function (for example, catalytic Cys residues). However, functionalization of less reactive sidechains often requires harsh reaction conditions that can be detrimental to protein activity.[7] Thus, more creative methods are required, and the last five years of research in this area have highlighted two general strategies. The first relies on targeting unique and accessible N-/C-terminal chemical environments present in single-chain proteins.[17] The second focuses instead on protein tertiary structures that create more reactive microenvironments that enable selective modification (for example, ligand- and metal-binding sites, hyperreactive sidechains, and disulfide bonds).[18,19]

[H2] *N-/C-terminus Modification*

Terminus modifications have the potential to be more generally applicable —most protein termini are accessible and have chemical environments distinct from the remainder of the protein. Several recent discoveries have been made that improve method versatility and terminal AA tolerance while addressing past problems.[17] For C-terminal modifications, targeting relies on the difference in oxidation potential between terminal carboxylic acids and in-chain glutamic acid (Glu)/Asp residues. MacMillan and coworkers have reported a method based on visible-light-mediated single-electron transfer (SET) to perform decarboxylative alkylation at C-terminal residues (3 equiv. photocatalyst, 10 equiv. Michael acceptor, 8 h, r.t., pH 3.5, 41–49% conversion, Figure 2i).[16] Unlike prior amide coupling and esterification strategies, this SET reaction favors the C-terminus over carboxylic acid moieties in Glu and Asp residues.[20–22] Such selectivity originates from increased stability of the C-terminal carbon-centered radical.[16] The aforementioned conditions allow for the selective alkylation of human insulin at the A chain C-terminus. However, using alternate decarboxylative conditions, originally studied for targeting tryptophan residues at the peptide level, selective modification of the B chain C-terminus was observed (0.5-1 equiv. photocatalyst, 10-30 equiv. Michael acceptor, 10 equiv. K_2HPO_4, 2.5-3 h, DMF, Merck Photoreactor (450 nm), 15-38% conversion).[23] Though the precise reasons for this selectivity remain as yet undetermined, the divergence in these C-terminal modification products showcases the number of factors at play when performing protein conjugation.

When considering N-terminal modification, the first point to consider is that the N-terminal amine has pK_a ~ 8 while that of an in chain Lys's ε-amine has pK_a ~ 10. Thus, not only will the N-terminus be charged at physiological pH, encouraging solvent accessibility, but also in slightly basic environments, the N-terminal amine will be deprotonated and more nucleophilic than those in Lys residues.[24,25] Although the N-terminal amine pK_a is similar to that of Cys (pK_a ~ 8), Cys

residues are lower in abundance and may not be as accessible as the N-terminus. Thus, Cys-modification can be avoided by using protection steps or by careful, amine-reactive reagent selection.[25,26] Therefore, targeting the N-terminus in a single chain protein can lead to site-specific modification. Even with such promise, the earliest reported methods targeting the N-terminus required high concentrations of modifying reagents and long reaction times.[27,28] Numerous methods that address this, involving one step, low reagent concentrations, and reasonable reaction times have been reported: oxidative coupling with *o*-aminophenols (5 equiv. *o*-aminophenol, 250 equiv. oxidant, 30 min, room temperature (r.t.), pH 7.5, Figure 2ii), addition of 2-pyridinecarboxaldehydes (2-PCA, 400 equiv., 16 h, 37 ºC, pH 7.5, 33-95% conversion, Figure 2iii), and reductive alkylation with aldehydes (2 equiv. aldehyde, 5 equiv. reductant, 6-48 h, r.t., pH 6.1, 30-70% conversion, Figure 2iv).[24–26] While some N-terminal residue types show higher conversion values than others, only a few are found to be incompatible (for example, an N-terminal Cys leads to reductive alkylation).[24,25] Such versatility can be even further extended by enzyme-mediated modification at the N-terminus. Butelase 1 has been reported as an asparagine/aspartic acid (Asn/Asp) specific ligase to efficiently conjugate thiodepsipeptide substrates to N-terminal residues (0.001 equiv. butelase 1, 5 equiv. substrate, 2.5 h, 42 ºC, pH 6.5, 70-82% conversion, Figure 2v). Butelase 1 has a high substrate tolerance, and can successfully perform conjugation with any N-terminal AA other than proline. Reaction tends to be favouredby the presence of a hydrophobic residue in the second position.[29] Having several conjugation strategies available allows navigation of the complex combinations of disadvantages and advantages in order to choose the optimal method for a diverse range of proteins and applications.

While some of the aforementioned methods cannot modify proteins with certain terminal residues, the methods complement each other such that there exists at least one possible method

for every terminal residue type. Further advantages and limitations stem from specific reagents or conditions used. With *o*-aminophenol oxidative coupling, double modification can be seen at basic pH either due to Lys modification or dimerization.[26] In the case of 2-PCA, after 12 h at 37 ºC, 20-30% of the modified protein is lost.[25] For the method relying on butelase 1, the short half-life of the thiodepsipeptide reagents must be acknowledged and reaction conditions adjusted.[29] These limitations highlight why method diversity is needed: to allow for the choice of a modification method with limitations that will not affect the conjugate application. Such diversity also applies to the various advantages of methods: the use of reductive alkylation maintains the charge on the N-terminal amine which may be necessary for protein function, and the use of 2-PCA or *o*-aminophenol can facilitate multiple, site-specific modifications.[24–26] Optimization of older methods has also led to the progression of N-/C-terminal modification strategies. As an example, Rapoport's Salt (RS) for oxidation at the N-terminus shows an expanded tolerance for different N-terminal residues in higher pH environments (~ pH 8.5).[30] The important advantages to incorporate and limitations to avoid are depend on the application of interest. However, the structure and makeup of the protein, the presence of post translational modifications (PTMs), or the need for the termini to be available for protein function, mean that other protein modification strategies (those that target in-chain sites) are also in high demand.[17,31]

[H2] *In-chain Residue Modification*

Protein modification at in-chain AAs is critical for certain biological applications (e.g. profiling inhibition and modulation of enzymatic active sites).[32] Careful selection of reaction type, conditions, and reagents, have enabled such modifications have been achieved using endogenous AAs.[7] One such strategy is the selective targeting of Trp residues (200–300 equiv. sodium

trifluoromethanesulfinate, 25 equiv. *tert*-butyl hydroperoxide, 25 equiv. methionine (Met), 5–10 min, 0 ºC, pH 6, 65–80% conversion, Figure 3i). Modification of Trp occurs at a 30-fold faster reaction rate than Cys the next most reactive. Trp is a low-abundance residue and is highly likely to be found in the hydrophobic core of the protein. The ability of this method to trifluromethylate Trp selectively is a notable achievement.[33] However, when targeting more common endogenous AAs for modification, strategies may rely on the modification of the most accessible and reactive copy of a repeated residue to achieve site-specificity.

Due to their innate nucleophilicity, Lys residues represent one of the more abundant AAs (~6%) commonly targeted for protein modification.[18] While Lys abundance can lead to extreme product heterogeneity with conventional modification methods (for example amidation with succinimide-based esters), the tertiary structure of proteins can enable higher levels of specificity as a result of unique, microenvironment-driven pK_a changes.[1,18,19] A recent study by Cravatt and coworkers observed and quantified the reactivities of 9,000 Lys residues in the human proteome, finding several hundred hyperreactive cases.[18] Therefore, specific reagents or conditions have been developed to modify the most reactive Lys residues in different proteins. By adjusting the electrophilicity of Michael acceptors, our group has targeted the most reactive Lys residues in proteins while leaving other Lys and Cys residues untouched (1 equiv., 1-2 h, 25-37 ºC, pH 8.0, >95% conversion, Figure 3ii). The design of sulfonyl acrylate reagents was assisted by computational calculations that determined transition states facilitating the desired bond formation.[12] Similarly, a multicomponent reaction involving formaldehyde, acetylene, and a Cu-ligand complex as a catalyst was found to modify a single Lys residue in a handful of different proteins (100 equiv. aldehyde, alkyne, and CuI, 400 equiv. ligand, 72 h, r.t., pH 7.8, 40-99% conversion, Figure 3iii). Conveniently, the N-terminus is reversibly protected by the

formaldehyde, allowing for the reactive Lys residue to be targeted over the N-terminal amine. However, the kinetics for the multicomponent reaction are slower in addition to the suboptimal use of a Cu(I) catalyst that introduces potential toxicity and requires difficult purification steps.[34]

Also reliant on the tertiary structure of target proteins are methods mediated by native metal- and ligand-binding sites. Such strategies manipulate the spatial arrangement of reagents and residues such that proximity promotes site-specific labeling. However, these methods rely on two factors: the native presence of a metal- or ligand- binding site and a nearby, reactive residue. If these conditions are met, then a substoichiometric amount of the catalytic targeting component can be used.[35] For metal-binding sites, the most recent advances have relied on the use of Pd(II) for the arylation of Cys residues as exhibited with the site-specific modification of mannosyltransferase (80 equiv. Pd(II) reagent, 500 equiv. aryl halide, 4 h, 65 ºC, pH 7.6, >85% conversion, Figure 3iv).[32] Further arylation techniques and products are discussed in this review, but a more in-depth and comprehensive review of this type of arylation has been published recently.[36] Meanwhile, for ligand-binding sites, polyproline peptides targeting SH3 domain proteins were used to facilitate a reaction between Asp residues and aryldiazonium reagents (5 equiv., 2.5 h, r.t., pH 7.4, 30% conversion, Figure 3v).[19] Furthermore, targeting ligand-binding sites has also enabled the use of imidazole-1-sulfonyl diazotransfer reagents to convert surrounding amino groups into azide moieties for downstream bioorthogonal functionalization (1 equiv. diazotransfer reagent, 100 equiv. Cu(II) catalyst, 1 h, r.t., pH 7.4, Figure 3vi).[37] Methods have also been reported combining the use of metal reactivity and peptide binding sites for selective modification in SH3 domain proteins and antibodies. These metallopeptides consist of dirhodium(II) cores and either SH3 domain-binding (0.5 equiv. metallopeptide, 50 equiv. diazo compounds, 5 h, r.t., pH 6.2-7.4, >95%) or Fc-binding peptides (2 equiv. metallopeptide, 125

equiv. alkyne-diazo, 8 h, 10 ºC, pH 6.4, >90% conversion) for the production of functionalized protein and antibody conjugates (Figure 3vii).[38,39] Other binding sites on antibodies have also been taken advantage of for proximity-induced site-specific modification. To avoid disrupting the Fc-binding area (an important part of the antibody for receptor interaction), peptides that bind to the junction between two of the heavy chains in a full length antibody were mutated to contain 4-fluorophenyl carbamate moieties and used to modify a lysine residue proximal to the junction (8 equiv. binding protein, 48 h, 37 ºC, pH 8.5, >90% conversion).[40] However, proximity-induced reactions can also be facilitated by strategies like a recently reported linchpin-directed method (LDM, 25 equiv. LDM reagent, 6-24 h, r.t. or 37 ºC, pH 7.0, 34-57% conversion).[41] This LDM relies on a Lys-reactive group attached to a His-reactive group via a spacer. First, an intermolecular reaction allows for the Lys-reactive component of the LDM reagent to attach non-specifically and reversibly to the Lys residues in the target protein. Once the reagent is bound to Lys residues, the His-reactive component now has the opportunity to bind irreversibly and intramolecularly to a His in proximity of a bound Lys. The lysine residues are then released by the addition of an aminooxy reagent, leaving only the specific His residues modified.[41]

Endogenous AA modification has also been achieved by disulfide rebridging, and several general techniques have been established. Conventional native Cys modification, often relies on the reduction of interchain disulfides followed by modification of the free Cys, but has the potential to cause protein instability due to the disruption of structure-stabilizing disulfide crosslinks.[12] In disulfide rebridging methodologies, the disulfide bonds are reformed, and thus their structural function retained after modification. Additionally, disulfides that are selectively modified in these approaches tend to be found toward the exterior of the protein and have structure stabilizing functions, allowing internal disulfides that are vital for activity to remain protected.[42] However,

the size and bulkiness of rebridging reagents must be limited to avoid disruption of the structure of the target protein. Recently, commercially available oxetane reagents have demonstrated the ideal distance for disulfide rebridging while also improving the stability and activity of therapeutic proteins and antibodies (20-60 equiv. oxetane, 6-12 equiv. tris(2-carboxyethyl)phosphine (TCEP), 24-48 h, 25-37 ºC, pH 8.5, >95% conversion, Figure 3viii).[43] Alternatively, unlike their bis-sulfone counterparts[44], allyl sulfones have been proposed as viable disulfide rebridging reagents with high aqueous solubility and reactivity (2 equiv. allyl sulfone, ~1 equiv. TCEP, 24 h, r.t., pH 7.8, 19% and 28% isolated yield with insulin and lysozyme as model proteins, Figure 3ix).[45] Meanwhile, divinylpyrimidine (DVP) was just reported as a stable rebridging agent for antibody and protein conjugates (10-15 equiv. DVP, 5 equiv. TCEP, 1-2 h, 37 ºC, pH 8.0, >95% conversion).[46] A type of disulfide rebridging reagent for obtaining highly homogenous antibody drug conjugates (ADCs) with drug-to-antibody ratios (DARs) of 2 was reported: dibromopyridazinedione (dibromo-PBD) derivatives (16 equiv., 80 equiv. TCEP, 16 h, 4 ºC, pH 8.0, Figure 3x). To achieve this, one compound containing two dibromo-PBD derivatives connected by a linker crosslinks two disulfide bridges. Since there are four interchain disulfide bridges available in IgG antibodies, two such dibromo-PBD-linker compounds are used, each with one payload incorporated.[47] Further reports of dibromo-PBD rebridging reagents for a "plug and play approach" for the production of ADCs and for the attachment of antibody fragments in a specific, favorable orientation to nanoparticles have been published recently.[48,49] Even photomediated disulfide bridging has been explored using a one-pot thiol-yne conjugation strategy on a reduced antibody fragment (0.8 equiv. initiator, 1 equiv. alkyne reagent, 4 h, 0 ºC, UV radiation (365 nm), 40% conversion, Figure 3xi). The method has proven applicable for a variety of alkyne reagents and has exhibited the potential of photomediated disulfide rebridging for therapeutic applications.[50]

Despite the potential of these discoveries, drawbacks remain prevalent: not being able to achieve high selectivity, homogeneity, efficiency, reagent stability, or conversion; not having control over the position of the conjugation site; and needing a multiple step synthesis to produce the reagents needed for the modification.[32–34,38–42,45] The diversity between and within the classes of modification strategies (e.g. reactive-residue targeting, metal-/ligand-binding sites, disulfide rebridging, etc.) signifies that efforts to modulate and add functionality to proteins via endogenous AAs are becoming more pronounced and promising. When yields are optimized, such strategies have the potential for remarkable academic and industrial relevance as the protein of interest does not require prior sequence engineering.

[H1] Modification via Genetic Manipulation

The genetic engineering of proteins has proven to be an indispensable methodology for site-selective protein modification and the production of homogenous bioconjugates. Rather than working only with a scaffold of endogenous AAs, genetic engineering allows for the controlled introduction of a variety of abiotic and biotic chemical handles at designated sites.[1] The potential versatility and generality inherent in the genetic engineering platform make these strategies invaluable for the exploration and modulation of protein function. The discussion of the various, genetically-based strategies that follows, focuses on the diversity and complementarity of the different method components and potential applications. Such diversity can be evaluated based on certain attributes: multiple modifications; biologically compatible conditions; reaction selectivity; directly added functionality vs. added bioorthogonal handle for subsequent modification; reaction efficiency and stability; method applicability; and modification reversibility. Additionally, the strategies presented here have been categorized into three broader insertion groups: canonical AAs, non-canonical (ncAAs), and motifs or enzymatic tags.

[H2] *Canonical Amino Acid Insertion*

While direct modification of endogenous AAs can be performed using the same chemical transformations discussed earlier in this section, these methods can often be more broadly applied and successful when the target residue is genetically inserted at specific positions in protein sequences. Although the selectivity is based primarily on chemical functionality, the insertion site in the protein is also of great importance: the site must be accessible to the modifying reagents and the protein must retain its structure and function. In addition, the insertion and expression of canonical AAs avoid disadvantages and complexities inherent to ncAA and non-AA motif insertion. As Cys has been the focus of many recently discovered methods due to its broad reactivity profile, this section has been split into two: methods targeting inserted Cys residues and methods targeting other low-abundance, endogenous residues.[6]

[H3] *Cysteine insertion*

Out of the two most nucleophilic canonical residues (Cys and Lys), Cys remains the residue of interest largely due to its relative low abundance (~1.9%), high nucleophilicity, and ability to react in environments closer to neutral pH.[17] Whether targeted as an endogenous AA or a mutated sidechain, Cys's broad scope of reactivity from transfer (i.e. atom, electron, or hydride) and metal-binding to exchange reactions indicates the incredible number of already determined Cys modification methods. Several recent reviews have focused on Cys-targeted conjugation.[6,51] Furthermore, Cys residues can be reduced to dehydroalanine (Dha) to extend the reach of an already versatile reactivity profile.[6] The standard Cys reaction for efficiency and selectivity has until recently been maleimide-based reactions. However, due to the observation of retro-Michael additions under basic conditions and thiol exchange, *in vivo* therapeutic applications for maleimide linkages are slowly becoming replaced by more robust linkages. neglected.[51] Even with the

development of methods to stabilize the maleimide linkage by hydrolyzing the thiosuccinimide ring, these strategies prove to be less efficient and result in a mixture of hydrolyzed and non-hydrolyzed products.[52–58] Therefore, novel reagents with thiol-specific reactivity have been pursued that lead to irreversible, stable conjugation while attempting to reach efficiencies comparable to maleimide reactions.

While some of these recently discovered reagents add a unique chemical handle at the site of the inserted Cys residue for downstream functionalization, other methods manage to add functionality directly in a single step. Strategies for the efficient insertion of chemical handles include the addition of isobutylene (50-1000 equiv., 50 equiv. TCEP, 1-6 h, 4 ºC-r.t., pH 8.0-9.0, Figure 4i) and oxetane bromo-derivatives as electrophilic handles (1500 equiv. oxetane, 440 equiv. TCEP, 2-5 h, 37 ºC, pH 8.0-11.0, >95% conversion, Figure 4ii). Both electrophilic handles rely on a bromide leaving group for further modification by alkylation with small molecule nucleophiles. While the isobutylene handle facilitates conjugation under more biocompatible conditions, the oxetane linkage adds advantageous attributes to the conjugate by increasing aqueous solubility and metabolic stability.[59,60] As electrophiles do not exist endogenously in proteins, these handles create unique sites on the protein for downstream functionalization. Meanwhile, other strategies focus on synthesizing modifying reagents that incorporate the desired functionality upon reaction with the protein.

Even with the variety of reactive partners available for installed chemical handles, high-yielding, simple, one-step processes for conjugation are hotly pursued.. Sodium 4-((4-(cyanoethynyl)benzoyl)oxy)-2,3,5,6-tetrafluorobenzenesulfonate (CBTF) is one reagent that fulfils these criteria. CBTF contains both an activated ester for amine functionalization prior to conjugation and a 3-arylpropiolonitrile moiety that can react with a Cys residue. CBTF displays

a high potential for rapid reactions and forms highly-stable products (12 equiv. CBTF, 1.1-2.2 equiv. TCEP, 12 h, r.t., pH 6.8, Figure 4iii).[61,62] Alternatively, by decorating the aromatic ring with a variety of electron withdrawing groups (EWGs) and fluorine substituents, fluorobenzene type reagents have also been explored for the stable and selective S-arylation of Cys residues (2-67 equiv., 16 h, 37 ºC, pH 8.3, Figure 4iv).[63] Other reagents that form stable, irreversible linkages while also introducing desired functionality are: carbonylacrylic, allenamide, and cyclopropenyl ketone derivatives.[15,64,65] Carbonylacrylic reagents succeed in performing the desired modification via a rapid Michael addition in a stoichiometric manner (1-50 equiv., 1-2 h, r.t.-37 ºC, pH 8.0, >95% conversion, Figure 4v).[15] Allenamide reagents, being less electron deficient than ketones and esters, react specifically with Cys and avoid modification by common biological nucleophiles (10-25 equiv. allenamide, 10 equiv. dithiothreitol (DTT), 30 min, 37 ºC, pH 8.0, Figure 4vi).[64] Unlike the carbonylacrylic and allenamide reagents, the cyclopropenyl ketone derivatives, functionalized via amide connections formed by NHS-ester promoted reactions, rely on ring strain to promote irreversible, rapid formation of stable products (50 equiv. cylcopropenyl ketone, 5 equiv. tris(hydroxypropyl)phosphine, 10 min, r.t., pH 6.0, Figure 4vii).[65] Furthermore, dichlorotetrazine (10 equiv., 1 h, r.t., pH 5.2, Figure 4viii) and 2-azidoacrylate reagents (10 equiv., 12 h, r.t., pH 7.4, >85% conversion, Figure 4ix) have been proposed as both being able to directly add functionality in one step in addition to introducing bioorthogonal sites for further functionalization.[66,67] These dual functionalization strategies circumvent the need for two ncAA insertions and are very promising for improved theranostic applications.[67]

In addition to the more conventional type Michael addition and alkylation reactions, metal-based reactions have also been explored for the modification of Cys. However, transition metal-based reactions have not been pursued to the same extent as metal-free methods due to common

difficulties with complex biomolecules: side reactions with endogenous protic, basic, or thiol-containing moieties; heterogeneous product mixtures; catalyst deactivation; and incompatible protein modification conditions.[68,69] Only recently has the specificity, efficiency, and versatility of transition-metal based chemistry been harnessed thanks to judicious metal and ligand choices.[70] Au(III) (15-20 equiv., 30 min, r.t., pH 8.0, Figure 4x) and Pd(II) complexes (1.1-10 equiv., 24 h, 37 ºC, pH 8.5, >94% conversion, Figure 4xi) have been used for the production of stable S-arylated products by targeting Cys residues. The Au(III) and Pd(II) catalyzed systems are based on two-electron strategies. The S-arylation products are stable and the conjugation irreversible, making these methods useful for many potential therapeutic applications. [68,70–72]

Reversible conjugation processes can, in the right context, be useful. These methods include the efficient Michael addition of either 5-methylene pyrrolones (5MPs, 200-500 equiv., 2 h, 37 ºC, pH 7.5, Figure 4xii) or 4-acetoxy cyclopentenones (50 equiv., 1-2 h, r.t., pH 7.0-7.4, 22-95% conversion, Figure 4xiii) with controlled release by increasing pH/thiol exchange or Michael donor addition respectively; the formation of a thiazolidino boronate (TzB) product at N-terminal Cys residues by 2-formyl phenylboronic acid (2-FPBA) that dissociates in slightly acidic environments (this particular case shows modification of native N-terminal Cys, 1 equiv., 30 min, r.t., pH 7.0, >95%, Figure 4xiv); and the formation of a thioether bond by a fast reaction driven by irradiation at 350 nm of 3-(hydroxymethyl)-2-napthols (napthoquinone methide precursors, NQMPs) that can be reversed by irradiation of a dilute solution of labeled conjugate or when mixed with vinyl ether (8-9 equiv., 2-6 min, r.t., pH 7.4, 350 nm irradiation, Figure 4xv).[73–77] Such variable processes for controlled release of a Cys residue can lead to information on critical epigenetic roles or reversible modulation of protein function.[73,76] Additionally, the acidic environment-driven release of 2-FPBA could be used for endosomal release and delivery of

cytotoxic drugs from an antibody drug conjugate (ADC) construct. Furthermore, combinatorial approaches using these methods have an impact in the field. For example, NQMP-Cys conjugation is orthogonal to standard azide–alkyne click chemistry, allowing for many nuanced, complex reversible and release/catch applications.[74,75]

The methods discussed here only cover a fraction of the applications possible through targeted Cys conjugation alone. Even with risks of disulfide disruption or shuffling due to reduction steps required to produce free Cys thiols, the Cys-based modification strategies pursued by scientists represent a window into the future goals of the site-selective protein modification field. The ultimate goal is to obtain a complete toolkit of strategies that cover everything from transition metal-mediated to photoinitiated processes that can allow for dual or reversible modification that can be used in concert with other (orthogonal) methods. Cys modification symbolizes the key concept of site-selective protein modification: the strength and power of the technique is a result of the complementarity inherent in a substantial variety and number of methods.

[H3] *Other Low-Abundance Canonical Amino Acids*

To supplement engineered Cys residue modification, methods targeting alternative, low-abundance AAs have been explored. Genetically inserted residues that have captured particular interest are: Trp, Tyr, Met and His. Apart from all four being of relative low-abundance to other canonical AAs, endogenous Met and Trp residues are also most commonly found in the hydrophobic interior of the protein.[78,79] Both low abundance and positioning can increase the likelihood that a genetically inserted copy is the only instance of that residue available for modification. In other words, genetic manipulation can allow for the formation of a unique chemical handle. However, due to the lower reactivity of these four sidechains in comparison to

residues such as Cys and Lys, more extreme or biologically incompatible conditions were initially relied upon.[79,80] While not all of the methods for modifying these less reactive, low-abundance residues require metal mediation, it is a common approach that can ensure specificity and avoid the need for highly reactive reagents which may lead to unwanted modification elsewhere on the protein..[81,82]

Two methods reported in the last five years, one metal-mediated and the other organoradical-based, target Trp residues.[79,83] The metal-mediated method for Trp uses 1-[(triisopropylsilyl)ethynyl]-1,2-benziodoxol-3(1H)-one (TIPS-EBX). Catalyzed by [AuCl(SMe_2)], the hypervalent iodide TIPS-EBX adds regioselectively to the Trp indole (10 equiv. TIPS-EBX, 5 equiv. catalyst, overnight, r.t., ~90% conversion, Figure 5i), rendering this the first Trp modification method both selective for Trp and able to modify a specific site on Trp residues (i.e. modification occurs at Trp C2).[83] Meanwhile, the reagents used for the organoradical, selective modification of Trp are derivatives of 9-azabicyclo[3.3.1]nonane-3-one-N-oxyl (keto-ABNO). Stabilized on the oxygen bonded to the nitrogen, the radical adds to the indole ring of Trp to achieve a highly homogenous product (1-5 equiv. keto-ABNO, 0.6-3 equiv. $NaNO_2$, 30 min, r.t., pH 7.4, 11-64% conversion, Figure 5ii).[79] Based on some drawbacks to these methods, including the acidic conditions required for high conversion with keto-ABNO conjugation, metal-mediated strategies act as valuable alternatives for proteins and applications requiring alternate conditions.[79]

Metal-mediation is also required for certain Tyr-selective modification strategies. As the more polar Tyr tends to be at the protein surface, site-specificity is somewhat harder to achieve. However, due to low abundance, even in the case of multiple Tyr modifications, high product homogeneity is still likely.[84] In one instance, iron-containing hemin was reported to catalyze the

addition of *N*-methylated luminol derivatives to the *ortho* Tyr position in the presence of H_2O_2 (1 equiv. hemin, 100 equiv. peroxide, 100 equiv. luminol derivative, 1 h, r.t., pH 7.4, Figure 5iii). The production of a reactive cyclic diazodicarboxyamide intermediate *in situ* drives the reaction forward.[84] A similar intermediate has been used as the starting reagent for Tyr modification, but the hemin-mediated method avoids the need to store unstable reagents.[84,85] Whether through coordination or covalent binding, some methods require direct metalation of the protein rather than metal mediation of the modification. A method for Tyr modification has been established recently that uses rhodium(III) chloride and boronic acid to link arene complexes to the *ortho* position in Tyr (50 equiv. rhodium(III) chloride and boronic acid, overnight, r.t., pH 9.4, Figure 5iv). The Rh(III) Tyr complexes maintain both a metastability and controlled reversibility (via DTT or H_2O_2) due to the inorganic linkage.[86] Although still not as highly regarded as metal-free methods, advantages of metal-based reactions have become more apparent and have led to further exploration of novel, specific interactions.

Although they are not yet able to augment protein functionality (aside from facilitating transition metal complex interactions to induce luminescence), methods selectively targeting His with Pt(II) (5 equiv. complex, 1 h, r.t., pH 7.0, Figure 5v) and Ru(II) (excess complex, 30 min, r.t., pH 7.0, Figure 5vi) complexes have been developed for protein labeling and staining. While the interaction between the complexes and His residues has not yet been identified as covalent or noncovalent as a result of conflicting analysis results, the staining or "switch-on" probe protocols developed with these complexes only require 30 min – 1 h to reach completion.[81,82] As transition metal complexes usually bind nonspecifically to proteins, these complexes may signify new interactions inspire novel site-selective modification methods. Aside from Trp-, Tyr-, and His-selective methods, a redox activated tagging (ReACT) Met-targeted reaction was also reported

using oxaziridine reagents (1.1-10 equiv., 10 min, r.t., pH 7.4, >95% conversion, Figure 5vii). Oxaziridine reagents have demonstrated selective oxidation of Met residues to sulfimides, which can then be used as a chemical handle for installing payloads.[78] Similarly targeting Met, hypervalent iodine species were used to create a high energy sulfonium protein synthon (500-1667 equiv. iodonium salt, 200-667 equiv. thiourea, 50-167 equiv. TEMPO, 50-167 equiv. formic acid (~pH 3), <5 min, 0-20 ºC, 84-95% conversion). While functionalization is possible upon oxidation of Met by incorporating the desired functionality into the iodonium salt, the resulting sulfonium product also has an electrophilic diazo group that allows for further modification.[87]

Outside of the proteinogenic, canonical AAs, selenocysteine (Sec) has also made an appearance quite recently in reports of site-specific protein modification (oxidation of Sec: 20 equiv. 2,2'-dithiobis(5-nitropyridine), 15 min, r.t., acidic conditions, not isolated; arene addition: 10 equiv. arene, 5 h, 37 ºC, pH 8.0, 23% isolated yield).[88] This report acknowledges the ability for electron-rich arenes (e.g. vancomycin) to attach at the site of an oxidized Sec residue, containing an electrophilic Se-S bond. While modification of the oxidized Sec was shown directly in an affibody, the insertion into the full length antibody was accomplished with a sortase A mediated method. [88] Even so, this method shows promise for direct modification of the oxidized Sec in full length antibodies and larger proteins. All of the methods in this section, while creating novel solutions and tools to unlock unexplored directions, represent an ongoing battle against lower reactivity, lower conversion, and lower selectivity when attempting to target inserted, low-abundance AAs (with the exception of Cys).

[H2] *Noncanonical Amino Acid Insertion*

One of the most reliable methods to achieve site-specificity for protein modification involves the genetic insertion of ncAAs, containing either unique abiotic or biotic functionalities

for subsequent bioorthogonal reactions. In the 1960's, translation of the genetic code was thought to be rigid, but the discovery of selenocysteine (Sec) and pyrrolysine (Pyl) indicated an inherent flexibility.[89] Schultz and coworkers spearheaded the movement toward the broadly applicable use of mutually orthogonal tRNA/aminoacyl-tRNA synthetase (RS) pairs (i.e. no native RS aminoacylates the orthogonal tRNA and no naturally occurring tRNA is modified by the orthogonal RS) for recognizing and inserting ncAAs at the site of the amber nonsense codon in *E. coli*.[90–93] The amber nonsense codon (i.e. UAG) was chosen as a model method due to efficient incorporation with low levels of error.[89] With such promising foundational methods established, protein modification via genetically inserted ncAAs has progressed both in insertion methodology and the inclusion of novel reactive groups (Figure 6).

Methods for the insertion of ncAAs aside from orthogonal aminoacyl-RS/tRNA pairs for amber suppression have been developed over the past few years. These incorporation strategies have focused on the reassignment of the rare arginine-tRNA sense codon (AGG); DNA hybridization chromatography for depletion and replacement of certain tRNAs; a cell-free translation system with *in vitro* transcripts of tRNA; and engineered orthogonal ribosomes.[89,94–96] However, to be incorporated into a protein, the ncAA must be synthesized. A review has recently been published on strategies being developed to make this synthetic process more effective.[97] While the method of insertion is crucial for expression efficiency and performing multiple insertions, the abiotic, bioorthogonal chemical handles incorporated by ncAAs represent a critical method for the furthering of site-selective protein modification. Because of the insertion of unique chemistries that respond to a specific reactive partner and are inert to native entities in biological environments, clean and efficient reactions in cells, direct functionalization of the ncAA with click chemistry, and the precise placement of PTM mimics have been performed. The breadth of

applications and functionalities based on the genetic incorporation of ncAAs is such that we have only discussed here examples that we consider to be the most representative of the accomplishments of this strategy. accomplishments

Recently, chemistries have been explored to install accurate PTMs or chemical handles for subsequent crosslinking.[98,99] Site-selectivity is vital for the evaluation of specific PTMs, as these epigenetic modifications have different consequences when translated by corresponding "reader" proteins in varying contexts.[100] Understanding the role of certain PTMs, especially in histone proteins, can be accomplished through ncAA insertion. Such insertions help decode protein–protein interactions (PPIs), including those necessary for gene regulation and apoptosis.[92] A novel ncAA, ε-*N*-2-hydroxyisobutyryl-Lys (HibK), representing a PTM observed in histones, was recently incorporated in histone proteins using an orthogonal amber suppressor pyrrolysl-RS pair (3.6-11.9 mg/L expression yield). By altering the charge of the Lys residue and adding steric bulk, site-specific insertion of HibK will help determine how the PTM affects chromatin structure.[101] Additionally, PTMs can be inserted in protected forms if the native form is unstable or too reactive. Phosphotyrosine (pTyr) was inserted as a neutral analog that is both cell permeable and stable in cells (1.0-1.75 mg/L expression yield). Acidic conditions were used to reveal the native phosphotyrosine (16-48 h, 4 ºC, pH 1.0-2.0).[102] Similarly, a protected allysine residue in the form of ε-*N*-(4-azidobenzoxycarbonyl)-δ,ε-dehydrolysine (AcdK) was inserted into histone proteins and epigenetic enzymes (7.0 mg/L expression yield). The AcdK undergoes reduction by phosphines to reveal allysine (TCEP, 2 h, r.t., pH 7.0), which hydrolyzes and is modified by reductive amination ($NaCNBH_3$, 8 h, r.t., pH 7.0) to insert either monomethyllysine or dimethyllysine.[103] Both lysine-methylation and phosphorylation PTMs have significant impact on

cell cycle progression and development, and ncAA insertion has provided methods to facilitate their comprehension.

Unlike the aforementioned examples, novel photo-lysine ncAAs can insert photo-derivatives (4-40% incorporation), containing diazirine rings, of lysine-based PTMs. These derivatives both approximate and identify possible native PPIs. The method has yet to be demonstrated in a site-specific manner, but a residue specific manner based on growing cells in media containing photo-lysine has confirmed its potential for selective Lys replacement in native proteins. Noncovalent, transient PTM interactions can then be captured by covalent bonds formed after the photo-cross-linking of the inserted photo-lysine and the protein of interest (1 equiv., 30-60 min, 25-37 ºC, pH 7.5, 365 nm irradiation).[13] Such cross-linking methods also represent an alternative application for ncAAs: the insertion of chemical handles for subsequent functionalization.

The ability of ncAA insertion to install reactive groups that enable bioorthogonal functionalization has been invaluable to site-selective protein modification research. Chemical handles recently inserted into proteins include: fluorine-activated aryl carbamates (FPheK, 3-8 mg/L expression yield), aryl isothiocyanates (pNCSF, 8-16 mg/L expression yield), and thioester derivatives of Asp acid (ThioD, 8 mg/L expression yield).[104–106] FPheK, once inserted into a protein, reacts with amines, thiols, and phenols to produce intra- or inter-molecular cross-linking (2 equiv. nucleophile, 8 h, 37 ºC, pH 8.5).[104] Similarly, pNCSF forms urea-type, cross-linking bridges between proteins or between proteins and small molecules (3-100 equiv. amine-containing nucleophile, 3-24 h, 37 ºC, pH 7.4-8.5).[105] Meanwhile, ThioD can be modified by amine-based nucleophiles (100 equiv. nucleophile, 20 h, 37 ºC, pH 7.4).[106] While these insertions broaden the scope of reactive groups available for modification in proteins, the sequential steps and long

536 modification reaction times are suboptimal. Photo-mediated methods can circumnavigate these
537 limitations by requiring less time and avoiding the use of excess reagents.[98]

538 The most recent ncAAs developed for photo-mediation post insertion are 2-aryl-5-
539 carboxytetrazole-lysine derivatives (ACTKs, 0.8 mg/L expression yield, 15 min, UV radiation 302
540 nm), benzyloxycarbonyl-lysine derivatives (AmAzZLys with an amine and azide functionality, 53
541 mg/L expression yield, 15 min, UV radiation 365 nm), and photoswitchable click AAs (PSCaas
542 equipped with azobenzenes modified with an alkene, ketone, or chloride; 1.2-1.8 mg/mL
543 expression yield; 2 min; 365 nm).[107–109] While AmAzZLys and PSCaas both need further chemical
544 modification in addition to photo-cross-linking or conformation change via photoswitching, these
545 strategies still represent a progression toward photo-based reactions in this research field.[107,108] Of
546 all the strategies discussed here, the insertion of photo-lysine represents most clearly the overall
547 direction ncAA insertion strategies are moving. As an elegant insertion of multifaceted
548 functionality capable of providing information both on PTMs but also for photo-cross-linking and
549 modification, the insertion of photo-lysine has the potential for many applications, especially once
550 site-specific insertion of these ncAAs has been achieved.[13]

551 [H2] *Motif Insertion and Enzymatic Methods*

552 To circumvent the synthesis and expression of ncAAs while retaining high levels of
553 specificity, canonical AA motifs have been designed for insertion into protein sequences to modify
554 specific residues. The target residue in the motif is activated by microenvironment manipulation
555 based on the identity of the surrounding AAs or recognition of the motif by a specific
556 enzyme.[3,11,110] Several reviews on enzymatic methods for site selective protein modification have
557 been published recently.[111,112] Thus, the limitations inherent in targeting ncAAs or native residues
558 based on chemical functional group alone can be overcome. However, the size and position of the

559 inserted motif can cause challenges. In some cases, motifs can only be added at the extremities of
560 proteins due to need for increased accessibility or are large enough that insertion compromises
561 protein activity.[113,114] While the inherent specificity of enzymatic modification is a substantial
562 advantage, the enzyme must be easily obtained and achieve high conversions to be industrially
563 useful.[3]

564 Based on the advantages of motif insertion, studies exploring novel fusion proteins as well
565 as enzymatic tags have been reported recently. Even though fluorescent proteins, such as
566 SNAP-tag, HaloTag, and CLIP-tag, have been proven valuable for fusion to termini of target
567 proteins, the attachment of a whole protein may disrupt activity.[115] Therefore, shorter fusion tags,
568 such as fluorophore-binding peptides (i.e. "fluorettes"), have taken precedence. A method to install
569 TexasRed covalently to a target protein was recently reported using a TexasRed fluorette, TR512
570 (1.5-4 equiv. probe, 10-40 equiv. TCEP, 30-60 min, 37 ºC, pH 7.9, 76% conversion, 34-AA tag,
571 Figure 7i). The fluorette was added to the N-termini of target proteins via linkers of two Cys
572 residues to ensure covalent binding to TexasRed promoted by proximity.[116] While this strategy
573 relies on exclusive fluorette specificity to certain fluorophores, other inserted motifs have more
574 general applications.

575 A broader substrate range for more general modification is possible with enzyme-
576 mediation (e.g. tubulin Tyr ligase – TTL – and trypsiligase).[114,117,118] TTL attaches Tyr derivatives
577 to the C-terminal residue in an inserted Tub-tag at the C-terminus of the target protein (200 equiv.
578 substrate, 0.02-0.2 equiv. TTL, 1-3 h, 37 ºC, pH 7, 99% conversion,14-AA tag, Figure 7ii).[114]
579 Meanwhile, due to reversibility often observed in enzymatic reactions, the proteinase trypsiligase
580 can also be used for ligation. While competing hydrolysis reactions generally limit this activity,
581 an activation domain on trypsiligase that only allows proteinase activity when interacting with

582 specific substrates (i.e. “substrate-activated catalysis”) allows for promotion of the ligase activity.
583 Therefore, the YRH recognition tag for trypsiligase can be inserted at the N-terminus of the target
584 protein, the Y-R bond broken, and the guanidinophenyl ester derivative (OGp) added (first step:
585 0.05 equiv. trypsiligase, 0.5 equiv. Zn(II) additive, 1-18 h, second step: 3-5 equiv. OGp, 30-60
586 min, 4-20 ºC, pH 7.8, >95% conversion, 3-AA tag, Figure 7iii).[118] While still limited to
587 modification at the protein extremities, a diverse population of substrates can be used, and the
588 specificity of the enzyme can lead to higher likelihood of orthogonality with other modification
589 methods.

590 Due to the recent discovery of short sequences that each feature a particularly reactive
591 amino acid (i.e. “clever” peptides), enzymes are sometimes unnecessary for site-specific
592 modification directed by motif insertion. Most recently, the activation of specific Cys residues for
593 modification by aza-dibenzocyclooctyne (DBCO), 2-cyanobenzothiazole (CBT), and
594 perfluoroaromatic reagents have been reported.[14,119,120] The DBCO-tag facilitates modification of
595 a Cys residue by DBCO derivatives at either terminus of a target protein. Thiol–yne reactions have
596 gained interest recently as an underdeveloped click reaction but have struggled with site-
597 specificity, the use of a tag enables a more selective and rapid reaction to occur (20 equiv. DBCO,
598 20-100 equiv. DTT, 4-16 h, 37 ºC, pH 8.0, 80-90% conversion, 7-AA tag, Figure 7iv).[119] Similarly,
599 the fusion tag targeting Cys-CBT reactivity, installed at the N-terminus, avoids protection or
600 proteinase steps prior to modification (100 equiv. CBT, 200 equiv. TCEP, 1 h, 37 ºC, pH 7.4-8.5,
601 12-71% conversion,11-AA tag, Figure 7v).[120] Meanwhile, the π-clamp for the targeting of Cys-
602 perfluoroaromatic reactivity achieved the goal of site-specificity with only 4 AAs in the motif (20-
603 26 equiv. perfluoroaromatic, 400 equiv. TCEP, 2-6 h, 37 ºC, pH 8.0, >95% conversion, 4-AA tag,
604 Figure 7vi). Based on computationally calculated peptide conformations and energy pathways, the

perfluoroaromatic reagents are hypothesized to be recognized by the phenylalanine residues, bringing the reagent into the vicinity of the activated Cys.[14] Recently, another Cys activation-based tag (Dis-tag) was reported that differentiates between free Cys and disulfide bond reactivity, allowing for the free Cys to first be modified with a maleimide reagent followed by the reduction and rebridging of the disulfide bond (6-AA tag, free Cys modification: 2 equiv. maleimide reagent, overnight, 15 ºC, pH 7.4, 75% conversion; disulfide rebridging: 2 equiv. allyl sulfone reagent, 2 equiv. TCEP, 24 h, 15 ºC, pH 7.8, 55% conversion). The Dis-tag allows more facile access to dual modifications by incorporating both sites for modification within a 6-residue minimal distance.[121] With no need for enzymatic mediation, the potential for insertion and modification of these Cys residues at in-chain positions is higher based on the easier access of smaller molecules to sterically hindered sites.

To complement Cys-targeted methods, motifs that target Lys residues have also been developed. Cyclohexene sulfonamide reagents are known to modify a specific Lys residue (Lys64) over all other endogenous Lys residues in human serum albumin (HSA). Therefore, the specific domain, HSAdI, has been fused to protein termini as a reactive platform for Lys site-specific modification (1-50 equiv., 2-24 h, 37 ºC, pH 7.4, 197-AA HSAdI, Figure 7vii).[115]

In a similar fashion, the unusual activity of the enzyme sortase A has been used to mediate the formation of an isopeptide bond between the ε-amino group of a Lys residue in an inserted pilin domain with the threonine carboxyl group from an LPXTG (where X = any AA) tag-containing substrate. (10-100 equiv. LPXTG-containing substrate, 2 equiv. enzyme, overnight, 32 ºC, 75% conversion, 11-AA domain, Figure 7viii).[122]

Beyond the targeting of Cys and Lys residues, a recently reported method targets a His-Gly-His (His_2-tag, 1-5 equiv. PEGylated bis-sulfone, 16 h, 20 ºC, 28-39% conversion, 6-AA tag).[123] With

no need for metal chelation, bis-sulfone modified PEG chains were shown to selectively modify the two inserted His residues in the tag over other His residues in the protein due to their close proximity. Performing the reaction at pH = 5 avoids modification of other residues and, combined with the low abundance of His, leads to high levels of selectivity. [123]

Even with the discovery of novel activities of known enzymes, enzymatic methods learned previously are most commonly optimized by the positioning and length of the recognition tag. In several cases, enzymatic recognition tags have been shortened or adjusted to create more reactive microenvironments.[124,125] In others, methods have been effective in various positions within the protein structure, allowing for multiple tag insertions without affecting protein activity.[113,125] Even the choice of slightly different reagents and methodology can allow for higher control over product identity as well as the option for a reversible conjugation.[126] Furthermore, the kinetics of enzymatic recognition and transformation, such as for sortase A, can be manipulated and improved through engineering an intramolecular reaction.[127,128] Successful attempts at one-pot processes have also been performed using several enzymes either in a tandem reaction to fine-tune the resulting modification or in a simultaneous, dual modification based on orthogonal recognition tags.[129,130] Even with these improvements, the discovery of the π-clamp represents an influential benchmark in this field. This elegant, short, and computationally designed motif accomplishes efficient modification at protein termini and shows promise for in-chain position insertion.[14] Such attributes indicate future studies will most likely include: heavier emphasis on computational methods and creative manipulations of microenvironments rather than randomized peptide assays for the discovery of novel insertion motifs.

[H1] Downstream Functionalization

When direct modification of protein sequence is not possible, functionalization can be achieved by inserting or attaching areactive functional group ("a chemical handle"). An entire class of selective bioorthogonal reactions has been developed for this purpose and has been promoted by the introduction of ncAAs. This diverse class of reactive pairs has been critical for installing unique functionalities in proteins. If insertion of a specific ncAA proves challenging due to limited cellular machinery, chemical handles might also be installed by enzymatic or chemical modification of a residue – though this is less attractive purely by virtue of requiring two synthetic steps. Therefore, the toolkit for efficient protein modification is augmented by both new methods for ncAA insertion and also the discovery of new bioorthogonal reactions.

However, with each major development, limitations and problems have arisen. The use of large protein tags (e.g. GFP) risks affecting protein activity post conjugation, and the introduction of copper-catalysed azide–alkyne cycloaddition (CuAAC) allowed the attachment of small molecule tags. The CuAAC is one example of "click chemistry" — chemistry defined by high reaction and conversion rate, green solvent systems, low levels of byproducts, and broad functional group applicability.[131–133] The potential toxicity of Cu(I), led to the development of strain-promoted (and copper-free) azide–alkyne cycloaddition (SPAAC).[133,134] In attempts to further improve the reaction kinetics, inverse-electron-demand Diels–Alder (IEDDA) reactions were developed.[134] Reaction rate is a highly important criterion in the development of bioorthogonal reactions (rates span from 10^{-5} M^{-1} s^{-1} to 10^{5} M^{-1} s^{-1}) — molar substrate concentrations can be naturally limited when dealing with large molecular weight biomolecules and in the case of radiolabelling experiments reactions must be completed before decay is complete. . However, more reactive agents used to achieve faster reactions were also observed to be less stable (e.g. *trans*-cyclooctyne – TCO – and tetrazine derivatives).[135–137] Increased reaction rate and reactivity

675 also generally led to difficulties with complementary bioorthogonal reactions for multi-site
676 modification.[137] Therefore, bioorthogonal reactions must be designed keeping in mind competing
677 needs for high reaction rates and for reagents to stable in biological environments, which differ
678 from those simulated with *in vitro* testing.[131]

679 To expand the toolkit of bioorthogonal reactions available, novel reaction partners have
680 been reported with comparable reaction rates and increased stability relative to already discovered
681 methods. Most of these recent methods take advantage of alkyne or aldehyde reactivity. Based on
682 SPAAC, a strain-promoted oxidation-controlled cyclooctyne-1,2-quinone cycloaddition (SPOCQ)
683 was established to add temporal control to the reaction. Using periodate oxidation, 1,2-catechols
684 are oxidized to 1,2-quinones that perform SPOCQ cycloaddition with the strained alkyne
685 bicyclo[6.1.0]nonyne (BCN) as an inserted ncAA (4 equiv. quinone, 1 h, r.t., pH 7.4, 90%
686 conversion, $k = 496 \pm 70\ M^{-1}\ s^{-1}$, Figure 8i).[138] Also using an inserted BCN ncAA, phenyl sydnones
687 with a 1,3-dipole were shown to undergo a [3+2] cycloaddition to produce a stable pyrazole
688 functionality (50 equiv. sydnone, 6 h, 37 ºC, pH 8.0, >95% conversion, $k = 0.054 \pm 0.00067\ M^{-1}$
689 s^{-1}, Figure 8ii). Although slower than SPOCQ, the sydnone-BCN reaction does display comparable
690 rates to SPAAC and cross-metathesis reactions.[135] An alkyne functionalized ncAA can be coupled
691 with a second, alkyne functionalized reactant using a Glaser–Hay couplingproducing a linear,
692 stable diyne product (10 equiv. alkyne, >50 equiv. CuI/tetramethylethylenediamine, 4-6 h, 4 ºC,
693 71-93% conversion, Figure 8iii). Glaser–Hay couplings have recently been optimized for use in
694 an aqueous environment, with the installed diyne product amenable to further
695 modification.[139]Using a similar ncAA handle, ruthenium catalysed alkyne hydrosilylation has
696 recent been used to form a C–Si bond (10-300 equiv. hydrosilane, 0.05-4.5 equiv. catalyst, 2-24 h,
697 37 ºC, pH 7.4-8.0, 45-50% conversion, $k = 1.0\ M^{-1}\ s^{-1}$, Figure 8iv). The *gem*-disubstituted

698 vinylsilane product lends itself to additional modification, and the overall reaction is
699 complementary to hydrazone formation, allowing for multiple modifications.[140]

700 Aldehyde and ketone reactive handles offer selective reactions and are widely tolerated
701 electrophiles, and on this basis interest in the use of hydrazone and oxime linkages has grown.[141]
702 Many biological applications require highly stable products and thus C–C bond forming reactions
703 are desirable.[142] Based on this reasoning, two Knoevenagel-type condensations were reported:
704 trapped (8 equiv. pyrazolone reagent, 16 h, 37 ºC, pH 5.5, $k = 0.20\ M^{-1}\ s^{-1}$, Figure 8v) and tandem
705 (8 equiv. pyrazolone, 16 h, 37 ºC, pH 7.2, Figure 8vi). Both target a formylglycine (fGly) residue,
706 inserted by formylglycine generating enzyme (FGE).[143,144] Also targeting C–C bond formation, a
707 direct aldol reaction of 2,4-thiazolidinediones with an N-terminal aldehyde (itself produced by a
708 sodium periodate oxidation of a 1,2-aminothiol moiety) was also successful (1000 equiv., 3 h, 37
709 ºC, pH 6.5, 83% conversion, $k = 0.0078\ M^{-1}\ s^{-1}$, Figure 8vii).[145] An aldol ligation reaction catalysed
710 by an L-proline derivative was reported (2-20 equiv. aldehyde, 100-500 equiv. catalyst, 1-6 h, 37
711 ºC, pH 7.5, >95% conversion, $k = 24\ M^{-1}\ s^{-1}$, Figure 8viii). The aldehyde functionality was inserted
712 as a protected analogue via a thiazolidine-Lys (ThzK) ncAA, allowing for both in-chain and
713 extremity modification. An additional aldehyde functionality was then needed to complete the
714 organocatalyst-mediated protein aldol ligation (OPAL) reaction. OPAL products can then be
715 further modified through oxime ligation.[142]

716 Methods outside alkyne- and aldehyde-based reactions have also been reported. Generally,
717 the incorporation of a new ncAA allows for these new protocols to be developed. One such ncAA
718 is *N*-acryloyl-Lys (AcrK). Alkyl phosphine reagents were used to modify AcrK through a
719 phospha-Michael addition, which occurs at a faster rate than thiol addition as phosphine reagents
720 have been used to activate electrophiles in thiol-ene reactions (30-40 equiv., 1-5 h, 25-37 ºC, pH

6.8-8.8, 80-90% conversion, $k = 0.06 \pm 0.01$ M^{-1} s^{-1}, Figure 8ix).[146] A quadricyclane (QC)-containing ncAA has also been successfully inserted into proteins. The strained, hydrocarbon reagent was functionalized with nickel bis(dithiolene), leading to a QC ligation cleavable by UV irradiation and orthogonal to common reactions with aldehyde/aminooxy and alkyne/azide pairs (1-2 h, r.t., pH 7.4, Figure 8x).[147] Lastly, the reaction between cyclopropenones, inserted as ncAAs, and triarylphosphines was shown to produce α,β-unsaturated amides (20 equiv., 1-4 h, 37 ºC, pH 7.0, >95% conversion, $k \geq 20$ M^{-1} s^{-1}, Figure 8xi).[148,149] The introduction of reversibility and orthogonality into bioorthogonal reactions paves the way for traceless or multi-site conjugation applications. In particular, the increase in bioorthogonal reactions discovered raises the likelihood for complementary/orthogonal reactions to achieve multiple, distinct modifications.

In addition to "click-type" reactions, accurate approximations of PTMs can be installed using methods established for the modification of Dha, which is itself commonly installed by the reduction of an inserted Cys residue.[6,150] PTMs have also been inserted based on synthetic and ncAA methods. The synthetic approach is inconvenient as the PTM can only be incorporated easily into shorter peptides, and the insertion of ncAAs involves being able to obtain the appropriate genetic machinery.[151] Thus, the formation and modification of Dha is a much more attractive option. As Dha insertion marks the addition of an electrophilic moiety, a type of functional group not endogenous to proteins, novel methods to modify Dha sidechains at both terminal and in-chain positions have been ardently pursued.[152] Most recently, aza-Michael additions have been demonstrated with amine-based nucleophiles to produce secondary and tertiary amine products (>300 equiv., 1-4 h, 25-37 ºC, pH 8.0-9.0, 40-95% conversion, $k = 6.1 \times 10^{-5}$ M^{-1} s^{-1}, Figure 8xii).[153,154] These reactions avoid the use of thiol-based nucleophiles to avoid risking the disruption of surface disulfide bonds. While these C–N bonds are representative of a common trope in nature

744 and stable between pH 2.8-12.8, more accurate representations of PTMs need to be attached by C–
745 C bonds at the site of Dha modification.[153] Two recently proposed radical-based mechanisms were
746 offered as solutions. The first used an alkyl-halide (either iodide- or bromide) for the initiation of
747 radical species combined with $NaBH_4$ for the prevention of unwanted oxidation and disubstitution
748 (100-2000 equiv. alkyl-halide, 30 min, 4 ºC, pH 4.0-8.0, Figure 8xiii). Sidechains from nonpolar
749 to polar and even charged PTMs were installed using this method.[151] The second took advantage
750 of O-phosphoserine (Sep) insertion followed by dephosphorylation to obtain Dha. Alkyl iodides
751 were then used to modify the Dha using the transmetalation from zinc to copper to form
752 organocopper reagents and produce a radical alkyl species (300 equiv. alkyl iodide, 300 equiv.
753 zinc powder, 100 equiv. organocopper, 30 min, r.t., pH 4.5, >80% conversion, Figure 8xiii). Using
754 this method, all methylated forms of Lys PTMs were successfully formed.[155] Beyond the insertion
755 of PTMs into proteins, quite recently, Dha residues were harnessed as site-specific handles to
756 access isotopic replacement techniques in proteins by performing a hydrogen–deuterium exchange
757 at the α-carbon of the Dha residue, a nonexchangeable site in the protein backbone (first step after
758 Dha formation in deuterated buffer: 15 equiv. Na_3SPO_3, 1 h, 37 ºC, pH 8.6; second step: PP1
759 phosphatase, 1 h, 30 ºC, pH 8). Aside from some limitations that result from the formation of
760 epimers at the deuterated site, the ability to isotopically label a protein site-specifically while
761 avoiding complex biosynthetic methods allows for great potential in the monitoring and probing
762 of modification mechanisms.[156] Even with a lack of stereocontrol, the open-ended diversity for
763 insertion at Dha sites has the potential to both unlock unknown protein functionality and redesign
764 others.[151]

765 While the toolkit of promising novel bioorthogonal reactions continues to expand,
766 improvements push older bioorthogonal reactions closer to achieving a balance of reactivity and

767 stability for broader applicability. These improvements include: the discovery of a supramolecular-
768 mediated azide-alkyne reaction using cucurbit[6]uril that increases the solubility of reagents and
769 facilitates reactions; the implementation of smaller, more stable 1,2,4-triazines in the place of
770 tetrazines for IEDDA; the fusing of dioxolane to trans-cyclooctene (d-TCO) for increased stability
771 and solubility while conserving high reaction rate (on the order of 10^5 M^{-1} s^{-1}); an increase in
772 selectivity for a diazo-coupling reaction by using 5-hydroxytryptophan instead of Tyr for a reaction
773 with aromatic diazonium ions; and a method that takes advantage of prior knowledge on boronic
774 acid tag capabilities by using the dynamic covalent character of boronic acid interactions with diols
775 as a purification system before reacting the boronic acid with functionalized salicylhydroxamates
776 to form a more stable product.[133,136,137,157,158] Even with these improvements, strategies must
777 continuously take into account the additional complications added when applications are meant
778 for *in vivo* use (e.g. sodium periodate could not be used for oxidation *in vivo*, but enzymatic
779 methods for oxidation may be able to replace it).[145] An account highlighting the process of
780 developing new bioorthogonal reagents and what is still lacking in the current toolbox of reactions
781 and reactive pairs was recently published.[159] At some point, the motivation for bioorthogonal
782 research will have to change from attempts to find novel reactive pairs to optimizing those already
783 discovered for efficient use.[147]
784
785 **[H1] Therapeutic and Diagnostic Applications**
786 There have been many reports of applications for protein conjugation, but we have limited
787 discussion here to only the most recent reports that pertain to diagnostics and therapeutic
788 intervention. Recently, there have been many discoveries of novel, controlled and targeted systems
789 for radioimaging and the delivery of therapeutic protein conjugates.[160–163] These systems rely on a

variety of targeting mechanisms: antibodies, nanobodies, or cyclic peptides, to gain this increased specificity and control. Both the targeting method and type of payload can be independently varied. In many cases, the stability and efficacy of the conjugate depends on the payload linker connecting the components.[160,164,165] By combining creative targeting and payload choices, site-selective protein modification will continue to open doors to impactful and novel biological applications.

In the field of therapeutic conjugates, ADCs capture most of the attention.[160] However, novel methods have been reported recently to either improve the ADC mechanism of action or to apply the idea of targeted delivery to alternate payloads (e.g. radioligands). All of these improvements require the assistance of selective protein conjugation methods.[164,166–178] Strategies ranging from conventional to site-specific conjugation are still in use to drive these adjustments and optimizations. Conventional, selective Lys amidation has very recently reported as a method for conjugating thiol-based histone deacetylase (HDAC) inhibitors to cetuximab antibodies targeting EGFR (Table 2i). Less than 1% of the injected dosage of ADCs are expected to reach and be internalized by a target tumor, and as a result payloads for ADCs have been thought to require sub-nanomolar IC_{50} values. Meanwhile, HDAC inhibitors only have about a 0.07 μM IC_{50} value. As the ADCs incorporating HDAC exhibited anti-tumor effects, this results suggests such highly toxic drugs are unnecessary and that off-site toxicity may be avoided by using lower potency drugs.[166] Other payloads outside of the class of highly cytotoxic, anticancer drugs have also been successfully conjugated to antibodies (see Table 2 entries i-vi).[168–172] Amazingly, the wide range of applications represented by these antibody-payload constructs is made possible by altering only the payload, conjugation chemistry, and linker, exemplifying how antibody conjugates can be viewed as a modular concept.

Similarly, system variability is possible by moving away from the use of full-length antibodies and toward smaller antibody formats or even small molecule ligands as targeting mechanisms. Additionally, while some of these strategies were successful with conventional conjugation methods, a higher level of specificity is observed when more homogenous conjugates have been used with improvements in efficacy, pharmacokinetic properties, and diminished off-site delivery.[160] The instability and tissue penetration issues associated with the use of full-length antibodies for targeting has led to the use of alternative biomolecules with affinity-based targeting abilities (Table 2 entries vii-x).[173–176] Rather than determining new functionality based on altering the payload identity, emphasizing the use of smaller targeting mechanisms has led to an increase in antibody-payload constructs efficacy. More specifically, such optimizations have uncovered methods for safer payload delivery, more effective payload distribution, and improved accessibility to medically relevant areas of the body that have not been explored by systems incorporatting full-length antibodies.[173-176]

In addition to methods targeting particular disease treatment, several recent studies have been aimed at illuminating the unknown mechanistic aspects of ADC approaches to drug delivery or to overcome foreseeable future issues related to antibody-targeting applications.[164,167] As knowledge of the internalization and subsequent intracellular trafficking of ADCs remains quite limited, clever strategies are needed to enhance efficacy. One such strategy uses fluorescence resonance energy transfer (FRET) pairing to gain insight (Table 2xi). By incorporating a cleavable linker with one FRET fluorophore on the antibody side, attached by a maleimide-engineered Cys linkage, and one on the warhead side of the linker, when the linker is cleaved, both the antibody and payload can still be visualized and monitored. This method revealed the critical role that the cellular background has in internalization of the antibody.[167] Beyond learning more information is

835 the anticipation of future problems in ADC performance, including the evolution of increased drug
836 resistance. Therefore, a recently published report establishes THIO-SELENOMABs through the
837 site-specific insertion of Cys and Sec to enable dual modification of the antibody (Table 2xii).
838 Such dual modification would allow two different drugs with two different mechanisms of action
839 to be delivered to the target cells, and, thus, potentially hinder the onslaught of resistance.[164]
840 Similarly, a multidomain protein therapeutic has been designed by biotinylation of somatostatin
841 (SST) and a Rho inhibitor (C3, Table 2xiii). The construct (SST3-Avi-C3) is made by the binding
842 of three SSTs to avidin with one binding site left for the binding of C3. The C3 toxin can work in
843 concert with doxorubicin to increase antitumor activity through the synergy of the two different
844 mechanisms of attack.[177]

845 Rather than the targeted delivery of a payload, a recent perspective article discusses the
846 advantages of using the antibody-antigen specific relationship for the creation of synthetic
847 vaccines. Synthetic vaccines are generally composed of antigens conjugated to proteins (using a
848 variety of techniques) which if proven viable would have a higher safety profile in comparison to
849 whole organism-based vaccines. While synthetic vaccines still take advantage of the antibody–
850 antigen specific relationship, therapeutic applications involving protein conjugates also exist
851 outside of antibody-related targeting.[178] One such method involves the novel, N-terminal selective
852 modification of cowpea chlorotic mottle virus (CCMV) capsid, a virus-like particle, using sortase
853 A, allowing for higher encapsulation efficiencies of therapeutics for subsequent delivery.[179]
854 Alternatively, protein–polymer conjugates, PEGylated and beyond, have been reviewed recently
855 due to their therapeutic relevance.[165] Aside from therapeutic and diagnostic applications, methods
856 for the profiling and modulation of protein function are reliant upon the production of protein
857 conjugates. For example, methods to explore histone PTMs using protein modification have

858 recently been reviewed. [100] While we have focused on therapeutic conjugates here, (with a
859 particular emphasis on antibody-based targeting strategies), the conjugation methods presented in
860 this review continue to be applied over many different fields of research, and the methods need to
861 adapt and expand to meet ever-changing demands and needs.

862 Of note, the reaction conditions reported throughout this review reflect the information
863 reported and available. For example, if certain methods do not have pH or conversion values in
864 the list of conditions, it is due to the information not being clearly stated or reported.

865
866
867 **[H1] Summary and outlook**

868 While the methods in this review exemplify the major progress made in site-selective
869 protein modification over the last five years, scientists have also been determining which direction
870 the field needs to take to move forward. Whether incorporating new reactivities or refurbishing
871 established chemistries, modification requirements are determined primarily by the complexity of
872 the targeted biological system. For the modification of endogenous AA sidechains, N-/C-terminus
873 and in-chain residue targeted methods need to be either tolerant of varying terminal AA types or
874 use the unique tertiary structure of the target protein to improve specificity. These methods have
875 the potential for high-yielding, one-step direct modification that avoids genetic engineering
876 complexities. Both the identity of the target protein and the importance of product homogeneity to
877 the application determine whether these methods may be used. On the other hand, genetic
878 manipulation of the protein before modification allows for exquisite selectivity and versatility.
879 While installed functionalities are limited based on natural translational tools and expression
880 yields, high selectivity makes genetic manipulation the most desirable method to achieve
881 homogeneity and has promoted the growth of bioorthogonal reaction types. Both endogenous AA

882 sidechain modification and genetic manipulation have played prominent roles in biological
883 applications, especially in therapeutic and diagnostic areas. However, such prominence has also
884 revealed the many ways in which the field can still be improved or expanded.

885 In general, an accurate prediction of future techniques can be determined by looking to
886 new methods for peptide modification. Many protein modification methods are first proven using
887 small molecules to exhibit functional group reactivity and before moving on to peptides and
888 ultimately whole proteins. Peptide studies allow comparisons of varying AA sidechains before
889 determining selectivity in longer AA sequences with complex tertiary structure. As such, methods
890 that have been proven to show selectivity on peptides have strong potential for implementation in
891 proteins. However, due to protein tertiary structure and large size, approximate reactivity with
892 peptides does not necessarily reflect the protein interactions that will occur. A recently reported
893 method addresses this by installing the reactive groups on well-known protein interaction faces.[180]
894 Strategies similar to this, including computational design, calculations, and modeling, improve the
895 efficiency and testing for new protein modification chemistries.[181–183] Alternatively, it is important
896 to acknowledge that older methods continue to develop and improve. This includes adjustments to
897 allow for the installation of several similar or varying functionalities and for use in biological
898 applications even outside of protein conjugation.[184–186] While refining methodology can facilitate
899 the discovery of new modification chemistries, the requirements that these new chemistries must
900 fill are reliant on the purpose and demand of the application.

901 Overall, these motivations and future directions allow insight into how best to assess and
902 select modification methods and conditions to produce protein conjugates. By keeping in mind the
903 intended application, whether targeted delivery or probing of a biological system, appropriate
904 proteins should first be identified. Subsequently, the protein then determines if direct, native

modification techniques are possible or if a chemical handle should be installed genetically to promote a bioorthogonal reaction. Based on this determination, a more specific method (e.g. targeting N-/C-terminus, motif insertion, etc.) can be chosen based on how best to retain protein activity. If a bioorthogonal reaction is necessary, the application determines the stability, kinetics, and reactivity needed. By highlighting specific conditions, this review aims to guide scientists to helpful methods based on the specific limitations of their circumstances. Between modification methods already available and promising discoveries on the horizon, site-selective protein modification will lead to versatile biological applications more capable of providing critical information not only for therapeutic and diagnostic purposes, but also for profiling and modulating protein function to probe and manipulate novel complex systems.

[H1] References

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

2. de Graaf, A. J., Kooijman, M., Hennink, W. E. & Mastrobattista, E. Nonnatural Amino Acids for Site-Specific Protein Conjugation. *Bioconjugate Chem.* **20,** 1281–1295 (2009).

3. Milczek, E. M. Commercial Applications for Enzyme-Mediated Protein Conjugation: New Developments in Enzymatic Processes to Deliver Functionalized Proteins on the Commercial Scale. *Chem. Rev.* **118,** 119–141 (2018).

4. Takaoka, Y., Ojida, A. & Hamachi, I. Protein Organic Chemistry and Applications for Labeling and Engineering in Live-Cell Systems. *Angew. Chem., Int. Ed.* **52,** 4088–4106 (2013).

5. Boutureira, O. & Bernardes, G. J. L. Advances in Chemical Protein Modification. *Chem. Rev.* **115,** 2174–2195 (2015).

6. Gunnoo, S. B. & Madder, A. Chemical Protein Modification through Cysteine. *ChemBioChem* **17,** 529–553 (2016).

7. Koniev, O. & Wagner, A. Developments and recent advancements in the field of

endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc. Rev.* **44,** 5495–5551 (2015).

8. Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A 'Tag-and-Modify' Approach to Site-Selective Protein Modification. *Acc. Chem. Res.* **44,** 730–741 (2011).

9. Schmidt, M. J. & Summerer, D. A Need for Speed: Genetic Encoding of Rapid Cycloaddition Chemistries for Protein Labelling in Living Cells. *ChemBioChem* **13,** 1553–1557 (2012).

10. Boyce, M. & Bertozzi, C. R. Bringing chemistry to life. *Nat. Methods* **8,** 638–642 (2011).

11. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea of Functionality. *Angew. Chem., Int. Ed.* **48,** 6974–6998 (2009).

12. Matos, M. J. *et al.* Chemo- and Regioselective Lysine Modification on Native Proteins. *J. Am. Chem. Soc.* **140,** 4004–4017 (2018).

13. Yang, T., Li, X.-M., Bao, X., Fung, Y. M. E. & Li, X. D. Photo-lysine captures proteins that bind lysine post-translational modifications. *Nat. Chem. Biol.* **12,** 70 (2015).

14. Zhang, C. *et al.* π-Clamp-mediated cysteine conjugation. *Nat. Chem.* **8,** 120–128 (2016).

15. Bernardim, B. *et al.* Stoichiometric and irreversible cysteine-selective protein modification using carbonylacrylic reagents. *Nat. Commun.* **7,** 13128 (2016).

16. Bloom, S. *et al.* Decarboxylative alkylation for site-selective bioconjugation of native proteins via oxidation potentials. *Nat. Chem.* **10,** 205–211 (2018).

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

18. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human proteome. *Nat. Chem.* **9,** 1181–1190 (2017).

19. Huang, F., Nie, Y., Ye, F., Zhang, M. & Xia, J. Site Selective Azo Coupling for Peptide Cyclization and Affinity Labeling of an SH3 Protein. *Bioconjugate Chem.* **26,** 1613–1622 (2015).

20. Zuo, Z. & MacMillan, D. W. C. Decarboxylative Arylation of α-Amino Acids via Photoredox Catalysis: A One-Step Conversion of Biomass to Drug Pharmacophore. *J. Am. Chem. Soc.* **136,** 5257–5260 (2014).

21. McGrath, N. A., Andersen, K. A., Davis, A. K. F., Lomax, J. E. & Raines, R. T. Diazo compounds for the bioreversible esterification of proteins. *Chem. Sci.* **6,** 752–755 (2015).

22. Totaro, K. A. *et al.* Systematic Investigation of EDC/sNHS-Mediated Bioconjugation Reactions for Carboxylated Peptide Substrates. *Bioconjugate Chem.* **27,** 994–1004 (2016).

23. Yu, Y. *et al.* Chemoselective Peptide Modification via Photocatalytic Tryptophan β-Position Conjugation. *J. Am. Chem. Soc.* **140,** 6797–6800 (2018).

24. Chen, D., Disotuar, M. M., Xiong, X., Wang, Y. & Chou, D. H.-C. Selective N-terminal functionalization of native peptides and proteins. *Chem. Sci.* **8,** 2717–2722 (2017).

25. MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11,** 326–331 (2015).

26. Obermeyer, A. C., Jarman, J. B. & Francis, M. B. N-Terminal Modification of Proteins with o-Aminophenols. *J. Am. Chem. Soc.* **136,** 9572–9579 (2014).

27. 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.* **134,** 5887–5892 (2012).

28. Antos, J. M. *et al.* Site-Specific N- and C-Terminal Labeling of a Single Polypeptide Using Sortases of Different Specificity. *J. Am. Chem. Soc.* **131,** 10800–10801 (2009).

29. Nguyen, G. K. T., Cao, Y., Wang, W., Liu, C. F. & Tam, J. P. Site-Specific N-Terminal Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide. *Angew. Chem., Int. Ed.* **54,** 15694–15698 (2015).

30. Palla, K. S., Witus, L. S., Mackenzie, K. J., Netirojjanakul, C. & Francis, M. B. Optimization and Expansion of a Site-Selective N-Methylpyridinium-4-carboxaldehyde-Mediated Transamination for Bacterially Expressed Proteins. *J. Am. Chem. Soc.* **137,** 1123–1129 (2015).

31. Murza, A. *et al.* C-Terminal Modifications of Apelin-13 Significantly Change Ligand Binding, Receptor Signaling, and Hypotensive Action. *J. Med. Chem.* **58,** 2431–2440 (2015).

32. Willwacher, J., Raj, R., Mohammed, S. & Davis, B. G. Selective Metal-Site-Guided Arylation of Proteins. *J. Am. Chem. Soc.* **138,** 8678–8681 (2016).

33. Imiolek, M. *et al.* Selective Radical Trifluoromethylation of Native Residues in Proteins. *J. Am. Chem. Soc.* **140,** 1568–1571 (2018).

34. Chilamari, M., Purushottam, L. & Rai, V. Site-Selective Labeling of Native Proteins by a Multicomponent Approach. *Chem. - Eur. J.* **23,** 3819–3823 (2017).

35. Ball, Z. T. Molecular recognition in protein modification with rhodium metallopeptides. *Curr. Opin. Chem. Biol.* **25,** 98–102 (2015).

36. Pentelute, B., Zhang, C., Vinogradova, E., Spokoyny, A., & Buchwald, S., Arylation Chemistry for Bioconjugation. *Angew. Chem., Int. Ed.* (2018) Accepted Author Manuscript.

37. Lohse, J. *et al.* Targeted Diazotransfer Reagents Enable Selective Modification of Proteins with Azides. *Bioconjugate Chem.* **28,** 913–917 (2017).

38. Vohidov, F., Coughlin, J. M. & Ball, Z. T. Rhodium(II) Metallopeptide Catalyst Design Enables Fine Control in Selective Functionalization of Natural SH3 Domains. *Angew. Chem., Int. Ed.* **54,** 4587–4591 (2015).

39. Ohata, J. & Ball, Z. T. A Hexa-rhodium Metallopeptide Catalyst for Site-Specific Functionalization of Natural Antibodies. *J. Am. Chem. Soc.* **139,** 12617–12622 (2017).

40: Yu, C. *et al*. Proximity-Induced Site-Specific Antibody Conjugation. *Bioconjugate Chem.* **29**, 3522-3526 (2018).

41. Adusumalli, S. R. *et al*. Single-Site Labeling of Native Proteins Enabled by a Chemoselective and Site-Selective Chemical Technology. *J. Am. Chem. Soc.* **140**, 15114-15123 (2018).

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

43. Martinez-Saez, N. *et al.* Oxetane Grafts Installed Site-Selectively on Native Disulfides to Enhance Protein Stability and Activity In Vivo. *Angew. Chem., Int. Ed.* **56,** 14963–14967 (2017).

44. Badescu, G. *et al.* Bridging Disulfides for Stable and Defined Antibody Drug Conjugates. *Bioconjugate Chem.* **25,** 1124–1136 (2014).

45. Wang, T. *et al.* Water-soluble allyl sulfones for dual site-specific labelling of proteins and cyclic peptides. *Chem. Sci.* **7,** 3234–3239 (2016).

46. Walsh, S. J. *et al*. A general approach for the site-selective modification of native proteins, enabling the generation of stable and functional antibody–drug conjugates. *Chem. Sci.* (2019).

47. Lee, M. T. W. *et al.* Enabling the controlled assembly of antibody conjugates with a loading of two modules without antibody engineering. *Chem. Sci.* **8,** 2056–2060 (2017).

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

49. Greene, M. K. *et al*. Forming next-generation antibody–nanoparticle conjugates through the oriented installation of non-engineered antibody fragments. *Chem. Sci.* **9**, 79-87 (2018).

50. Griebenow, N., Dilmaç, A. M., Greven, S. & Bräse, S. Site-Specific Conjugation of Peptides and Proteins via Rebridging of Disulfide Bonds Using the Thiol–Yne Coupling Reaction. *Bioconjugate Chem.* **27,** 911–917 (2016).

51. Cal, P. M. S. D., Bernardes, G. J. L. G. J. L. & Gois, P. M. P. Cysteine-Selective Reactions for Antibody Conjugation. *Angew. Chem., Int. Ed.* **53,** 10585–10587 (2014).

52. Lyon, R. P. *et al.* Self-hydrolyzing maleimides improve the stability and pharmacological properties of antibody-drug conjugates. *Nat. Biotechnol.* **32,** 1059–1062 (2014).

53. Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W. & Santi, D. V. Long-Term Stabilization of Maleimide-Thiol Conjugates. *Bioconjugate Chem.* **26,** 145–152 (2015).

54. Christie, R. *et al.* Stabilization of cysteine-linked antibody drug conjugates with N-aryl maleimides. *J. Control. Release* **220,** 660–670 (2015).

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

56. Tumey, L. N. *et al.* Mild Method for Succinimide Hydrolysis on ADCs: Impact on ADC Potency, Stability, Exposure, and Efficacy. *Bioconjugate Chem.* **25,** 1871–1880 (2014).

57. Morais, M. *et al.* Optimisation of the dibromomaleimide (DBM) platform for native antibody conjugation by accelerated post-conjugation hydrolysis. *Org. Biomol. Chem.* **15,** 2947–2952 (2017).

58. Kalia, D., Pawar, S. P. & Thopate, J. S. Stable and Rapid Thiol Bioconjugation by Light-Triggered Thiomaleimide Ring Hydrolysis. *Angew. Chem., Int. Ed.* **56,** 1885–1889 (2017).

59. Boutureira, O. *et al.* Site-Selective Modification of Proteins with Oxetanes. *Chem. - Eur. J.* **23,** 6483–6489 (2017).

60. Lee, B., Sun, S., Jiménez-moreno, E., Neves, A. A. & Bernardes, G. J. L. Site-selective installation of an electrophilic handle on proteins for bioconjugation. *Bioorg. Med. Chem.* **26,** 3060–3064 (2018).

61. Kolodych, S. *et al.* CBTF: New Amine-to-Thiol Coupling Reagent for Preparation of Antibody Conjugates with Increased Plasma Stability. *Bioconjugate Chem.* **26,** 197–200 (2015).

62. Koniev, O. *et al.* Selective Irreversible Chemical Tagging of Cysteine with 3-Arylpropiolonitriles. *Bioconjugate Chem.* **25,** 202–206 (2014).

63. Embaby, A. M., Schoffelen, S., Kofoed, C., Meldal, M. & Diness, F. Rational Tuning of Fluorobenzene Probes for Cysteine-Selective Protein Modification. *Angew. Chem., Int. Ed.* **57,** 8022–8026 (2018).

64. Abbas, A., Xing, B. & Loh, T.-P. Allenamides as Orthogonal Handles for Selective Modification of Cysteine in Peptides and Proteins. *Angew. Chem., Int. Ed.* **53,** 7491–7494 (2014).

65. Smith, N. J. *et al.* Fast, irreversible modification of cysteines through strain releasing conjugate additions of cyclopropenyl ketones. *Org. Biomol. Chem.* **16,** 2164–2169 (2018).

66. Canovas, C. *et al.* Site-Specific Dual Labeling of Proteins on Cysteine Residues with Chlorotetrazines. *Angew. Chem., Int. Ed.* **57,** 10646–10650 (2018).

67. Ariyasu, S., Hayashi, H., Xing, B. & Chiba, S. Site-Specific Dual Functionalization of Cysteine Residue in Peptides and Proteins with 2-Azidoacrylates. *Bioconjugate Chem.* **28,** 897–902 (2017).

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

69. Vara, B. A. *et al.* Scalable thioarylation of unprotected peptides and biomolecules under Ni/photoredox catalysis. *Chem. Sci.* **9,** 336–344 (2018).

70. Messina, M. S. *et al.* Organometallic Gold(III) Reagents for Cysteine Arylation. *J. Am. Chem. Soc.* **140,** 7065–7069 (2018).

71. Rojas, A. J., Pentelute, B. L. & Buchwald, S. L. Water-Soluble Palladium Reagents for Cysteine S-Arylation under Ambient Aqueous Conditions. *Org. Lett.* **19,** 4263–4266 (2017).

72. Kubota, K., Dai, P., Pentelute, B. L. & Buchwald, S. L. Palladium Oxidative Addition Complexes for Peptide and Protein Cross-linking. *J. Am. Chem. Soc.* **140,** 3128–3133 (2018).

73. Zhang, Y. *et al.* Thiol Specific and Tracelessly Removable Bioconjugation via Michael Addition to 5-Methylene Pyrrolones. *J. Am. Chem. Soc.* **139,** 6146–6151 (2017).

74. Bandyopadhyay, A. *et al.* Fast and selective labeling of N-terminal cysteines at neutral pH via thiazolidino boronate formation. *Chem. Sci.* **7,** 4589–4593 (2016).

75. Arumugam, S. *et al.* Selective and reversible photochemical derivatization of cysteine residues in peptides and proteins. *Chem. Sci.* **5,** 1591–1598 (2014).

76. Yu, J., Yang, X., Sun, Y. & Yin, Z. Highly Reactive and Tracelessly Cleavable Cysteine-Specific Modification of Proteins via 4-Substituted Cyclopentenone. *Angew. Chem., Int. Ed.* **57,** 11598–11602 (2018).

77. 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.* **7,** 5052–5058 (2016).

78. Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation. *Science* **355,** 597–602 (2017).

79. Seki, Y. *et al.* Transition Metal-Free Tryptophan-Selective Bioconjugation of Proteins. *J. Am. Chem. Soc.* **138,** 10798–10801 (2016).

80. Wang, L., Gruzdys, V., Pang, N., Meng, F. & Sun, X.-L. Primary arylamine-based tyrosine-targeted protein modification. *RSC Adv.* **4,** 39446–39452 (2014).

81. Wong, C.-Y. *et al.* A Ruthenium(II) Complex Supported by Trithiacyclononane and Aromatic Diimine Ligand as Luminescent Switch-On Probe for Biomolecule Detection and Protein Staining. *Sci. Rep.* **4,** 7136 (2014).

82. Solomatina, A. I. *et al.* Coordination to Imidazole Ring Switches on Phosphorescence of Platinum Cyclometalated Complexes: The Route to Selective Labeling of Peptides and Proteins via Histidine Residues. *Bioconjugate Chem.* **28,** 426–437 (2017).

83. Hansen, M. B., Hubálek, F., Skrydstrup, T. & Hoeg-Jensen, T. Chemo- and Regioselective Ethynylation of Tryptophan-Containing Peptides and Proteins. *Chem. - Eur. J.* **22,** 1572–1576 (2015).

84. Sato, S., Nakamura, K. & Nakamura, H. Tyrosine-Specific Chemical Modification with in Situ Hemin-Activated Luminol Derivatives. *ACS Chem. Biol.* **10,** 2633–2640 (2015).

85. Ban, H., Gavrilyuk, J. & Barbas, C. F. Tyrosine Bioconjugation through Aqueous Ene-Type Reactions: A Click-Like Reaction for Tyrosine. *J. Am. Chem. Soc.* **132,** 1523–1525 (2010).

80. Ohata, J., Miller, M. K., Mountain, C. M., Vohidov, F. & Ball, Z. T. A Three-Component Organometallic Tyrosine Bioconjugation. *Angew. Chem., Int. Ed.* **57,** 2827–2830 (2018).

87. Taylor, M. T., Nelson, J. E., Suero, M. G. & Gaunt, M. J. A protein functionalization platformbased on selective reactions at methionine residues. *Nature* **562**, 563-568 (2018).

88. Cohen, D. T. *et al.* A chemoselective strategy for late-stage functionalization of complex small molecules with polypeptides and proteins. *Nat. Chem.* (2018).

89. Cui, Z. *et al.* Combining Sense and Nonsense Codon Reassignment for Site-Selective Protein Modification with Unnatural Amino Acids. *ACS Synth. Biol.* **6,** 535–544 (2017).

90. Xie, J. & Schultz, P. G. A chemical toolkit for proteins — an expanded genetic code. *Nat. Rev. Mol. Cell Biol.* **7,** 775 (2006).

91. Haney, C. M., Wissner, R. F. & Petersson, E. J. Multiply labeling proteins for studies of folding and stability. *Curr. Opin. Chem. Biol.* **28,** 123–130 (2015).

92. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of Escherichia coli. *Science* **292,** 498 (2001).

93. Lang, K. & Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and Bioorthogonal Labeling of Proteins. *Chem. Rev.* **114,** 4764–4806 (2014).

94. Lee, B. S. *et al.* Incorporation of Unnatural Amino Acids in Response to the AGG Codon. *ACS Chem. Biol.* **10,** 1648–1653 (2015).

95. Iwane, Y. *et al.* Expanding the amino acid repertoire of ribosomal polypeptide synthesis via the artificial division of codon boxes. *Nat. Chem.* **8,** 317 (2016).

96. Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. *Nat. Chem.* **6,** 393 (2014).

97. Almhjell, P. J., Boville, C. E. & Arnold, F. H. Engineering enzymes for noncanonical amino acid synthesis. *Chem. Soc. Rev.* (2018).

98. Schmidt, M. J. & Summerer, D. Genetic code expansion as a tool to study regulatory processes of transcription. *Front. Chem.* **2,** 7 (2014).

99. Ovaa, H. & wals, kim. Unnatural amino acid incorporation in E. coli: current and future applications in the design of therapeutic proteins. *Front. Chem.* **2,** 15 (2014).

100. Nadal, S., Raj, R., Mohammed, S. & Davis, B. G. Synthetic post-translational modification of histones. *Curr. Opin. Chem. Biol.* **45,** 35–47 (2018).

101. Xiao, H., Xuan, W., Shao, S., Liu, T. & Schultz, P. G. Genetic incorporation of ε-N-2-hydroxyisobutyryl-lysine into recombinant histones. *ACS Chem. Biol.* **10,** 1599–1603 (2015).

102. Hoppmann, C. *et al.* Site-specific incorporation of phosphotyrosine using an expanded genetic code. *Nat. Chem. Biol.* **13,** 842 (2017).

103. Wang, Z. A. *et al.* A Genetically Encoded Allysine for the Synthesis of Proteins with Site-Specific Lysine Dimethylation. *Angew. Chem., Int. Ed.* **56,** 212–216 (2016).

104. Xuan, W., Shao, S. & Schultz, P. G. Protein Crosslinking by Genetically Encoded Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angew. Chem., Int. Ed.* **56,** 5096–5100 (2017).

105. Xuan, W., Li, J., Luo, X. & Schultz, P. G. Genetic Incorporation of a Reactive Isothiocyanate Group into Proteins. *Angew. Chem., Int. Ed.* **55,** 10065–10068 (2016).

106. Xuan, W. *et al.* Site-Specific Incorporation of a Thioester Containing Amino Acid into Proteins. *ACS Chem. Biol.* **13,** 578–581 (2018).

107. Tian, Y. *et al.* Genetically Encoded 2-Aryl-5-carboxytetrazoles for Site-Selective Protein Photo-Cross-Linking. *J. Am. Chem. Soc.* **139,** 6078–6081 (2017).

108. Yamaguchi, A. *et al.* Incorporation of a Doubly Functionalized Synthetic Amino Acid into Proteins for Creating Chemical and Light-Induced Conjugates. *Bioconjugate Chem.* **27,** 198–206 (2016).

109. Hoppmann, C. *et al.* Genetically Encoding Photoswitchable Click Amino Acids in Escherichia coli and Mammalian Cells. *Angew. Chem., Int. Ed.* **53,** 3932–3936 (2014).

110. Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic Labeling of Proteins: Techniques and Approaches. *Bioconjugate Chem.* **24,** 1277–1294 (2013).

111. Zhang, Y., Park, K.-Y., Suazo, K. F. & Distefano, M. D. Recent progress in enzymatic protein labelling techniques and their applications. *Chem. Soc. Rev.* (2018) Advance Article.

112. Li, C. & Wang, L.-X. Chemoenzymatic Methods for the Synthesis of Glycoproteins. *Chem. Rev.* **118**, 8359-8413 (2018).

113. Plaks, J. G., Falatach, R., Kastantin, M., Berberich, J. A. & Kaar, J. L. Multisite Clickable Modification of Proteins Using Lipoic Acid Ligase. *Bioconjugate Chem.* **26,** 1104–1112 (2015).

114. Schumacher, D. *et al.* Versatile and Efficient Site-Specific Protein Functionalization by Tubulin Tyrosine Ligase. *Angew. Chem., Int. Ed.* **54,** 13787–13791 (2015).

115. Patterson, J. T. *et al.* Human Serum Albumin Domain I Fusion Protein for Antibody Conjugation. *Bioconjugate Chem.* **27,** 2271–2275 (2016).

116. Sunbul, M., Nacheva, L. & Jäschke, A. Proximity-Induced Covalent Labeling of Proteins with a Reactive Fluorophore-Binding Peptide Tag. *Bioconjugate Chem.* **26,** 1466–1469 (2015).

117. Meyer, C., Liebscher, S. & Bordusa, F. Selective Coupling of Click Anchors to Proteins via Trypsiligase. *Bioconjugate Chem.* **27,** 47–53 (2016).

118. Liebscher, S. *et al.* N-Terminal Protein Modification by Substrate-Activated Reverse Proteolysis. *Angew. Chem., Int. Ed.* **53,** 3024–3028 (2014).

119. Zhang, C., Dai, P., Vinogradov, A. A., Gates, Z. P. & Pentelute, B. L. Site-Selective Cysteine – Cyclooctyne Conjugation. *Angew. Chem., Int. Ed.* **57,** 6459–6463 (2018).

120. Ramil, C. P., An, P., Yu, Z. & Lin, Q. Sequence-Specific 2-Cyanobenzothiazole Ligation. *J. Am. Chem. Soc.* **138,** 5499–5502 (2016).

121. Agrawalla, B. K. *et al*. Chemoselective Dual Labeling of Native and Recombinant Proteins. *Bioconjugate Chem.* **29**, 29-34 (2018).

122. Bellucci, J. J., Bhattacharyya, J. & Chilkoti, A. A Noncanonical Function of Sortase Enables Site-Specific Conjugation of Small Molecules to Lysine Residues in Proteins. *Angew. Chem., Int. Ed.* **54,** 441–445 (2015).

123. Peciak, K., Laurine, E., Tommasi, R., Choi, J.-w. & Brocchini, S. Site-selective protein conjugation at histidine. *Chem. Sci.* (2019) Accepted Author Manuscript.

124. Grünewald, J. *et al.* Optimization of an Enzymatic Antibody–Drug Conjugation Approach Based on Coenzyme A Analogs. *Bioconjugate Chem.* **28,** 1906–1915 (2017).

125. Grünewald, J. *et al.* Efficient Preparation of Site-Specific Antibody–Drug Conjugates Using Phosphopantetheinyl Transferases. *Bioconjugate Chem.* **26,** 2554–2562 (2015).

126. Montanari, E. *et al.* Tyrosinase-Mediated Bioconjugation. A Versatile Approach to Chimeric Macromolecules. *Bioconjugate Chem.* **29,** 2550–2560 (2018).

127. Siegmund, V. *et al.* Locked by Design: A Conformationally Constrained Transglutaminase Tag Enables Efficient Site-Specific Conjugation. *Angew. Chem., Int. Ed.* **54,** 13420–13424 (2015).

128. Wang, H. H., Altun, B., Nwe, K. & Tsourkas, A. Proximity-Based Sortase-Mediated Ligation. *Angew. Chem., Int. Ed.* **56,** 5349–5352 (2017).

129. Struck, A.-W. *et al.* An Enzyme Cascade for Selective Modification of Tyrosine Residues in Structurally Diverse Peptides and Proteins. *J. Am. Chem. Soc.* **138,** 3038–3045 (2016).

130. Zhang, Y. *et al.* Simultaneous Site-Specific Dual Protein Labeling Using Protein Prenyltransferases. *Bioconjugate Chem.* **26,** 2542–2553 (2015).

131. King, M. & Wagner, A. Developments in the Field of Bioorthogonal Bond Forming Reactions—Past and Present Trends. *Bioconjugate Chem.* **25,** 825–839 (2014).

132. Haldon, E., Nicasio, M. C. & Perez, P. J. Copper-catalysed azide-alkyne cycloadditions (CuAAC): an update. *Org. Biomol. Chem.* **13,** 9528–9550 (2015).

133. Finbloom, J. A., Han, K., Slack, C. C., Furst, A. L. & Francis, M. B. Cucurbit[6]uril-Promoted Click Chemistry for Protein Modification. *J. Am. Chem. Soc.* **139,** 9691–9697 (2017).

134. Yang, M., Li, J. & Chen, P. R. Transition metal-mediated bioorthogonal protein chemistry in living cells. *Chem. Soc. Rev.* **43,** 6511–6526 (2014).

135. Wallace, S. & Chin, J. W. Strain-promoted sydnone bicyclo[6.1.0]nonyne cycloaddition. *Chem. Sci.* **5,** 1742–1744 (2014).

136. Darko, A. *et al.* Conformationally strained trans-cyclooctene with improved stability and excellent reactivity in tetrazine ligation. *Chem. Sci.* **5,** 3770–3776 (2014).

137. Kamber, D. N. *et al.* 1,2,4-Triazines Are Versatile Bioorthogonal Reagents. *J. Am. Chem. Soc.* **137,** 8388–8391 (2015).

138. Borrmann, A. *et al.* Strain-Promoted Oxidation-Controlled Cyclooctyne–1,2-Quinone Cycloaddition (SPOCQ) for Fast and Activatable Protein Conjugation. *Bioconjugate Chem.* **26,** 257–261 (2015).

139. Lampkowski, J. S., Villa, J. K., Young, T. S. & Young, D. D. Development and Optimization of Glaser–Hay Bioconjugations. *Angew. Chem., Int. Ed.* **54,** 9343–9346 (2015).

140. Kwan, T. T. L. *et al.* Protein modification via alkyne hydrosilylation using a substoichiometric amount of ruthenium(II) catalyst. *Chem. Sci.* **8,** 3871–3878 (2017).

141. Kölmel, D. K. & Kool, E. T. Oximes and Hydrazones in Bioconjugation: Mechanism and Catalysis. *Chem. Rev.* **117,** 10358–10376 (2017).

142. Spears, R. J. *et al.* Site-selective C–C modification of proteins at neutral pH using organocatalyst-mediated cross aldol ligations. *Chem. Sci.* **9,** 5585–5593 (2018).

143. Kudirka, R. *et al.* Generating site-specifically modified proteins via a versatile and stable nucleophilic carbon ligation. *Chem. Biol.* **22,** 293–298 (2015).

144. Kudirka, R. A. *et al.* Site-Specific Tandem Knoevenagel Condensation-Michael Addition to Generate Antibody-Drug Conjugates. *ACS Med. Chem. Lett.* **7,** 994–998 (2016).

145. Wang, P. *et al.* Site-Specific Chemical Modification of Peptide and Protein by Thiazolidinediones. *Org. Lett.* **17,** 1361–1364 (2015).

146. Lee, Y.-J., Kurra, Y. & Liu, W. R. Phospha-Michael Addition as a New Click Reaction for Protein Functionalization. *ChemBioChem* **17,** 456–461 (2016).

147. Tomlin, F. M. *et al.* Site-specific incorporation of quadricyclane into a protein and photocleavage of the quadricyclane ligation adduct. *Bioorg. Med. Chem.* (2018). doi:https://doi.org/10.1016/j.bmc.2018.04.009

148. 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.* **139,** 7370–7375 (2017).

149. Shih, H.-W. & Prescher, J. A. A Bioorthogonal Ligation of Cyclopropenones Mediated by Triarylphosphines. *J. Am. Chem. Soc.* **137,** 10036–10039 (2015).

150. Bernardes, G. J. L., Chalker, J. M., Errey, J. C. & Davis, B. G. Facile Conversion of Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **130,** 5052–5053 (2008).

151. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring protein side-chain diversity. *Science* **354,** 597 (2016).

152. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5,** 4740 (2014).

153. Freedy, A. M. *et al.* Chemoselective Installation of Amine Bonds on Proteins through Aza-Michael Ligation. *J. Am. Chem. Soc.* **139,** 18365–18375 (2017).

154. Dadova, J. *et al.* Precise probing of residue roles by post-translational β,γ-C,N aza-Michael mutagenesis in enzyme active sites. *ACS Cent. Sci.* **3,** 1168–1173 (2017).

155. Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications. *Science* **354,** 623–626 (2016).

156. Galan, S. R. G. *et al*. Post-translational site-selective protein backbone α-deuteration. *Nat. Chem. Biol.* **14**, 955-963 (2018).

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

158. Zegota, M. M., Wang, T., Seidler, C., Wah Ng, D. Y., Kuan, S. L. & Weil, T. "Tag and Modify" Protein Conjugation with Dynamic Covalent Chemistry. *Bioconjugate Chem.* **29**, 2665-2670 (2018).

159. Row, R. D. & Prescher, J. A. Constructing New Bioorthogonal Reagents and Reactions. *Acc. Chem. Res.* **51**, 1073-1081 (2018).

160. Beck, A., Goetsch, L., Dumontet, C. & Corvaia, N. Strategies and challenges for the next generation of antibody-drug conjugates. *Nat. Rev. Drug Discov.* **16,** 315–337 (2017).

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

162. Akkapeddi, P. *et al.* Construction of homogeneous antibody-drug conjugates using site-selective protein chemistry. *Chem. Sci.* **7,** 2954–2963 (2016).

163. Kelkar, S. S. & Reineke, T. M. Theranostics: Combining Imaging and Therapy. *Bioconjugate Chem.* **22,** 1879–1903 (2011).

164. Li, X. *et al.* Site-Specific Dual Antibody Conjugation via Engineered Cysteine and Selenocysteine Residues. *Bioconjugate Chem.* **26,** 2243–2248 (2015).

165. Pelegri-O'Day, E. M., Lin, E.-W. & Maynard, H. D. Therapeutic Protein–Polymer Conjugates: Advancing Beyond PEGylation. *J. Am. Chem. Soc.* **136,** 14323–14332 (2014).

166. Cini, E. *et al.* Antibody drug conjugates (ADCs) charged with HDAC inhibitor for targeted epigenetic modulation. *Chem. Sci.* **9,** 6490–6496 (2018).

167. Lee, B.-C. *et al.* FRET Reagent Reveals the Intracellular Processing of Peptide-Linked Antibody–Drug Conjugates. *Bioconjugate Chem.* **29,** 2468–2477 (2018).

168. Wang, R. E. *et al.* An Immunosuppressive Antibody-dasatinib conjugate and its Lck inhibitory activity. *J. Am. Chem. Soc.* **137,** 3229–3232 (2015).

169. Lehar, S. M. *et al.* Novel antibody-antibiotic conjugate eliminates intracellular S. aureus. *Nature* **527,** 323–328 (2015).

170. Balintová, J., Welter, M. & Marx, A. Antibody–nucleotide conjugate as a substrate for DNA polymerases. *Chem. Sci.* **9,** 7122–7125 (2018).

171. Adumeau, P., Davydova, M. & Zeglis, B. M. Thiol-Reactive Bifunctional Chelators for the Creation of Site-Selectively Modified Radioimmunoconjugates with Improved Stability. *Bioconjugate Chem.* **29,** 1364–1372 (2018).

172. Lerchen, H.-G. *et al.* Antibody–Drug Conjugates with Pyrrole-Based KSP Inhibitors as the Payload Class. *Angew. Chem., Int. Ed.* **57**, 15243-15247 (2018).

173. Kelemen, R. E. *et al.* A Precise Chemical Strategy To Alter the Receptor Specificity of the Adeno-Associated Virus. *Angew. Chem., Int. Ed.* **55,** 10645–10649 (2016).

174. Greineder, C. F. *et al.* Site-Specific Modification of Single-Chain Antibody Fragments for Bioconjugation and Vascular Immunotargeting. *Bioconjugate Chem.* **29,** 56–66 (2018).

175. Wu, T. *et al.* Transglutaminase mediated PEGylation of nanobodies for targeted nano-drug delivery. *J. Mater. Chem. B* **6,** 1011–1017 (2018).

176. Brasino, M. *et al*. Anti-EGFR Affibodies with Site-Specific Photo-Cross-Linker Incorporation Show Both Directed Target-Specific Photoconjugation and Increased Retention in Tumors. *J. Am. Chem. Soc.* **140**, 11820-11828 (2018).

177. Kuan, S. L. *et al*. Boosting Antitumor Drug Efficacy with Chemically Engineered Multidomain Proteins. *Adv. Sci.* **5**, 1701036-1701036 (2018).

178. Jones, L. H. Recent advances in the molecular design of synthetic vaccines. *Nat. Chem.* **7,** 952 (2015).

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

180. Kobayashi, T., Hoppmann, C., Yang, B. & Wang, L. Using Protein-Confined Proximity To Determine Chemical Reactivity. *J. Am. Chem. Soc.* **138,** 14832–14835 (2016).

181. Strumillo, M. & Beltrao, P. Towards the computational design of protein post-translational regulation. *Bioorg. Med. Chem.* **23,** 2877–2882 (2015).

182. Wagner, J. A., Mercadante, D., Nikić, I., Lemke, E. A. & Gräter, F. Origin of Orthogonality of Strain-Promoted Click Reactions. *Chem. - Eur. J.* **21,** 12431–12435 (2015).

183. Narayanam, M. K., Liang, Y., Houk, K. N. & Murphy, J. M. Discovery of new mutually orthogonal bioorthogonal cycloaddition pairs through computational screening. *Chem. Sci.*

7, 1257–1261 (2016).

184. Knall, A.-C., Hollauf, M., Saf, R. & Slugovc, C. A trifunctional linker suitable for conducting three orthogonal click chemistries in one pot. *Org. Biomol. Chem.* **14,** 10576–10580 (2016).

185. Bryden, F. *et al.* Assembly of High-Potency Photosensitizer–Antibody Conjugates through Application of Dendron Multiplier Technology. *Bioconjugate Chem.* **29,** 176–181 (2018).

186. Cazzamalli, S., Dal Corso, A., Widmayer, F. & Neri, D. Chemically Defined Antibody– and Small Molecule–Drug Conjugates for in Vivo Tumor Targeting Applications: A Comparative Analysis. *J. Am. Chem. Soc.* **140,** 1617–1621 (2018).

Acknowledgements

We thank FCT Portugal (FCT Investigator to G.J.L.B., IF/00624/2015 and Postdoctoral Fellowship to P.M.S.D.C.), the Herchel Smith Fund (PhD Studentship to E.A.H.) and the EU Horizon 2020 programme (Marie Skłodowska-Curie ITN grant agreement No. 675007 to G.J.L.B. and Marie Skłodowska-Curie IEF grant agreement No. 702574 to B.L.O.) for funding. G.J.L.B. is a Royal Society University Research Fellow (UF110046 and URF\R\180019) and the recipient of a European Research Council Starting Grant (TagIt, grant agreement No. 676832).

Author contributions

All authors made substantial contributions to the discussion and organization of the content as well as reviewed and/or edited the manuscript before submission. Additionally, E.A.H. conducted the research, wrote the main body of the paper, edited the figures, and put the complete manuscript together; P.M.S.D.C, played an integral role in writing the introduction, researching, and providing content guidance at all stages of manuscript preparation; and B.L.O. designed and created the figures. G.J.L.B. coordinated the research and writing.

Competing interests

The authors declare no competing interests.

Publisher's note

Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.

Figure captions

Figure 1 | Juxtaposition of classical and modern protein modification methods. On the left: classical methods focused primarily on the modification of Cys and Lys sidechains. The reactions most commonly focused on and depicted here include: thiol-exchange (i), alkylation of α-halocarbonyl electrophiles (ii), maleimide Michael addition (iii), NHS-ester amidation (iv), isothiocyanate or isocyanate addition (v), and reductive amination (vi).[6,7] On the right: modern methods focused on the improvement of selectivity, reaction efficiency, and generality of application. The general modification classifications discussed in this review, as shown by specific examples, are represented: native protein/endogenous AA sidechain modification, engineered canonical AA insertion, engineered ncAA insertion, and motif/tag insertion.[13–16]

Figure 2 | N-/C-terminal selective protein modification. Above: C-terminal modification based on a decarboxylative strategy facilitated by a photocatalyst (i).[16] Below: N-terminal modification techniques based on oxidative (ii), direct (iii), reductive (iv), or enzymatic type reactions (v).[24–26,29] To mention a few more specific points on each reaction: (i) While not tested with glycine, phenylalanine, or proline as C-terminal residues, desired products were observed for all other AAs with lower yields for histidine (His), tyrosine (Tyr), and lysine (Lys) terminal AAs.[16] (ii) Best with proline in the terminal position (only residue allowing for high yields at the protein level) and not performed with cysteine in the terminal position unless only cysteine in protein sequence as method modifies cysteine residues regardless of terminal position.[26] (iii) This method has been exhibited with all AAs in terminal positions but cannot be performed on proteins with proline in the second position.[25] (iv) With the exception of cysteine as the terminal residue due to thiazolidine side products, all types of terminal AAs achieve high yields.[24] (v) Butelase 1 can interact with a variety of terminal AA types aside from proline and prefers the second position to be either, isoleucine, valine, leucine, or cysteine. Most of the terminal AA type compatibility types were performed on peptide platforms.[29]

Figure 3 | In-chain endogenous sidechain modification. Methods for endogenous sidechain modification of in-chain residues. The recent discoveries and reports of successful strategies have followed three different trends: modification based on selection of conditions and reagents to target the most reactive instance of a repeated sidechain, site-selective modification based on the direction of metals and ligands to native binding sites, and modification via disulfide rebridging. Each trend has several valuable examples that have been established in the last five years and are represented. Reagent- or condition-based targeting of the most reactive instance of an AA: selective trifluoromethylation of tryptophan residues (i), sulfonyl acrylate modification of most reactive lysine (ii), a three-component reaction for the modification of a single lysine (iii).[12,33,34] Ligand- or metal-binding site directed methods: Cys arylation based on proximity to Asp-regulated binding site (iv), selective modification of Tyr residues proximal to the SH3 binding domain (v), diazotransfer to lysines proximal to ligand site (vi), metallopeptide targeted Asp modification in antibodies using Fc-binding peptides (vii).[19,32,37,39] Disulfide rebridging: by oxetanes (viii), by water-soluble allyl sulfones (ix), by dibromide-based moieties (x), and by thiol-yne coupling (xi).[43,45–50] The colored highlights represent different reactive handles or functionalities added to the protein by the conjugation reactions.

Figure 4 | Modification methods for engineered Cys residues. Out of all of the canonical AAs for the installation of chemical handles in proteins, Cys residues have the broadest reactivity profile as represented by these examples. The methods have been separated into three classifying groups: metal-free, metal-assisted, and reversible type reactions. As the eventual application for the bioconjugate determines the type of chemistry linking the added functionality and the protein, these three categories all represent different mechanistic strategies that would allow these methods to be useful under various circumstances (e.g. cleavable linkers for ADCs to prevent reliance on release of the attached drug by native cell processes). Metal-free methods: addition of alkyl bromide electrophilic handles via an isobutylene and oxetane type chemical handle (i and ii), amine functionalization followed by addition of CBTF (iii), S-arylation by fluorobenzene derivatives (iv), addition of carbonylacrylic reagents (v), allenamide addition (vi), cyclopropenyl ketone addition (vii), chlorotetrazine addition (viii), and 2-azidoacrylate addition (ix).[15,59–61,63–67] Metal-assisted methods: S-arylation by way of Au(III) (x) or Pd(II) catalyst (xi).[70,72] Reversible methods: addition of 5MP derivatives (xii), addition of 4-acetoxy cyclopentenone (xiii), 2-FPBA addition at the N-terminus (xiv), and the addition of NQMPs (xv).[73–76]

Figure 5 | Insertion of canonical amino acids aside from Cys. Site-selective protein modification based on the genetic insertion of canonical AAs aside from Cys. Forming unique reactive handles when inserted due to low abundance or being unlikely to be expressed endogenously in a solvent accessible position, these inserted AAs (Trp, Tyr, His, and Met) allow modifications that will lead to products with higher levels of homogeneity. Trp modification: a metal-assisted reaction with TIPS-EBX to install a protected alkyne reactive handle regioselectively at the C2 position on the indole rings of installed Trp residues (i);[83] using an organoradical-driven mechanism, keto-ABNO derivatives add to the indole ring of Trp residues with elevated conversion in acidic environments (ii);[79] in the presence of H_2O_2 and hemin, luminol derivatives are added to Tyr *ortho* positions (iii);[84] Rh(III)-mediated reaction for the addition of arene complexes to Tyr *ortho* positions (iv).[86] Pt(II)- and Ru(II)-driven selective complexation with His residues (v and vi);[81,82] and a metal-free, redox-based reaction targeting Met residues with oxaziridine derivatives (vii).[78]

Figure 6 | Insertion of ncAAs. The most common method and novel chemistries made available by ncAA insertion over the last five years: a) Depiction of the groundbreaking use of orthogonal aminoacyl tRNA synthetase/tRNA pairs for the insertion of ncAAs.[90–93] 1 - Binding of tRNA and ncAA to aminoacyl tRNA synthetase (aaRS). 2 - Attachment of ncAA to tRNA by aaRS. 3 – Recognition of the amber codon by the ncAA-equipped tRNA. 4 – Incorporation of ncAA into the protein sequence by a native ribosome. b) examples of ncAAs that have been synthesized and inserted into proteins for the first time over the last five years. The colored highlights on the ncAA examples represent different chemical handles or functionalities added to the protein when the respective ncAAs are inserted. The red-highlight for HibK is to signify the PTM that this ncAA directly inserts.[101] Orange highlights represent protection groups that must be removed to reveal a phosphorylation PTM or an allysine residue that can be hydrolyzed to an aldehyde functionality (pTyr and AcdK respectively).[102,103] Blue highlights shows the photo-reactive moieties that allow for further functionalization of the protein or crosslinking.[13,107–109] While the green highlights different electrophilic handles incorporated by these ncAAs, the purple highlight signifies a nucleophilic site for subsequent reactions.[104–106,108,109]

Figure 7 | Motif and enzymatic tag insertion. Rather than the insertion of single residues and reliance only on the chemistry of the added functional group to drive selectivity, the insertion of tags allows for manipulation of the microenvironment around specific AAs. Such manipulation can lead to elevated reactivity of the targeted AA or to the enzymatic recognition of the inserted tag. In either case, the site-specific modification of the targeted residue within the motif or tag occurs. Ideal motif insertion methods allow for site-specific modification at either in-chain or terminal sites as well as cause minimal disruption of the protein structure, even with multiple instances of the tag inserted. Methods discussed here: fluorette fusion for fluorophore functionalization (i), modification mediated by TTL (ii), modification mediated by trypsiligase (iii), DBCO tag for Cys modification (iv), tag for CBT modification of Cys (v), π-clamp for Cys modification (vi), Lys activation by HSAdI (vii), noncanonical function of sortase A allows for Lys-specific modification of inserted pilin domain (viii).[14,114–116,118–122]

Figure 8 | Downstream functionalization methods. Clean and efficient modification of unique, bioorthogonal chemical handles. A general scheme of the concept of "click chemistry" is included at the top of the diagram.[131–133] The methods highlighted here fall under four classifications: alkyne-based reactions, aldehyde-based reactions, methods aside from alkyne- and aldehyde-based reactions, and Dha functionalization. Alkyne-based reactions: SPOCQ cycloaddition (i), phenyl sydnones [3+2] cycloaddition with BCN (ii), Glaser-Hay coupling (iii), Ru(II)-catalyzed alkyne hydrosilylation (iv).[135,138–140] Aldehyde-based reactions: trapped Knoevenagel-type condensation (v), tandem Knoevenagel-type condensation (vi), aldol reaction with 2,4-thiazolidinediones and an N-terminal aldehyde (vii), OPAL with an inserted aldehyde functionality (viii).[142–145] Methods aside from alkyne- and aldehyde-based strategies: phospha-Michael addition (ix), QC ligation (x), triarylphosphine-mediated addition to cyclopropenone (xi).[146–148] Methods for Dha modification: aza-Michael addition to Dha (xii), radical-based reactions for the formation of C–C bonds with Dha (xiii).[151,153–155]

1498 **Table 1. Overview of modification method components and key features**
1499

Approach		Possible reagents for modification or insertion	Key features
Direct modification of native proteins	N-/C-terminus modification	Michael acceptors (visible-light-mediated SET method); *o*-aminophenols; 2-PCA; aldehydes; thiodepsipeptides (mediated by butelase 1)	Advantages: -General method for site-specific modification of single chain native proteins due to distinct termini microenvironments -Termini tend to be solvent accessible -No genetic engineering needed Limitations: -Termini need to be available for modification (i.e. not vital for protein function and no PTMs) -Sometimes dependent on identity of terminal AA residue
	In-chain residue modification	Reactive-residue targeting: Sodium trifluoromethanesulfinate (Trp modification); sulfonyl acrylates (Lys modification); multicomponent reaction with aldehydes, alkynes, and copper(I) iodide (Lys modification) Proximity-induced: aryl halides for Cys arylation based on Pd(II) binding site; aryldiazonium addition to Tyr guided by SH3 domain-binding peptides; Cu(II) catalyzed diazotransfer to proximal Lys at binding sites; metallopeptide addition based on SH3 and Fc	Advantages: -With careful reaction/reagent selection, can rely on distinct microenvironment for targeting a single residue -Possible to use substoichiometric amount of targeting component when using binding site for proximity-induced modifications -Disulfide rebridging allows control over modification site -No genetic engineering needed Limitations: -Necessary to either have a distinct microenvironment to enhance reactivity of a specific residue or have a native metal- or ligand-

		domain-binding peptides; antibody heavy chain junction-binding protein targeted modification of lysine; LDM method for modification of His Disulfide rebridging: oxetanes; allyl sulfones; DVP; dibromo-PBD derivatives; alkynes for photomediated thiol-yne reactions	binding site as well as a proximal reactive residue -Smaller size needed for disulfide rebridging reagents -Little control over choice of modification site
Protein modification via genetic manipulation	Canonical AA insertion: cysteine	Isobutylene and oxetane bromo-derivatives; CBTF; fluorobenzenes; carbonylacrylic derivatives; allenamides; cyclopropenyl ketones; dichlorotetrazines; 2-azidoacrylate reagents; Au(III) and Pd(II) complexes for S-arylation; 5MPs; 4-acetoxy cyclopentenones; 2-FPBA; NQMPs	Advantages: -Broad reactivity profile of cysteine -Numerous previously determined methods for modification -Low abundance of Cys allowing for higher modification site selectivity -Easier to express mutations for canonical AAs than ncAAs Limitations: -Possible disulfide disruption or shuffling with reduction step needed to free Cys thiol for modification -Cannot be used in proteins where reactive Cys residue plays a critical role in protein activity
	Canonical AA insertion: other low-abundance canonical AAs	TIPS-EBX (metal-mediated Trp modification); keto-ABNO (organoradical Trp modification); N-methylated luminol derivatives (hemin-catalyzed method for Tyr modification); arene complexes (modification method for Tyr using	Advantages: -Residues focused on here (Trp, Tyr, Met, His, and Sec) have higher chances to form unique chemical handles based on low abundance and expected positioning within the protein structure -Easier to express mutations for canonical AAs than ncAAs

		Rh(III)); Pt(II) and Ru(II) complexes (His modification); oxaziridine reagents (Met modification); hypervalent iodonium salts (Met modification); electron-rich arenes (Sec modification)	Limitations: -More difficult to use biologically compatible conditions to modify these less reactive side chains and often observe lower conversions -Difficult to achieve high selectivity due to the need for more reactive modifying reagents or conditions
	ncAA insertion	HibK; pTyr; AcdK; photo-lysine; FPheK; pNCSF; ThioD; ACTKs; AmAzLys; PSCaas	Advantages: -The inserted ncAA chemical handle is unique and can be matched with a reactive pair for site-specific modification -Different methods for incorporating ncAAs have been discovered -Allows clean and efficient reactions that can be done in cells -Can mimic precise placement of PTMs Limitations: -Multiple step syntheses for the production of ncAAs -Complexities inherent in expression technologies and capabilities often leading to low levels of expression of the mutated protein and limited insertion of which functional groups able to be inserted
	Motif insertion and enzymatic methods	TexasRed fluorette for the covalent attachment of TexasRed to Cys; TTL for attachment of Tyr derivatives to Tub-tag; trypsiligase for the attachment of OGp to YRH tag; DBCO-tag, CBT-tag, π-clamp (for perfluoroaromatic	Advantages: -Expression of canonical AA mutations more successful than ncAA -Target residue activated by microenvironment manipulation or enzymatic recognition leading to site-specific reactions

		reagents), and Dis-tag for Cys activation; cyclohexene sulfonamide and LPXTG substrates for the modification of Lys mediated by inserted HSAdI and pilin domains respectively; His_2-tag for modification of two His residues proximal to each other	Limitations: -Tag size cannot be too large and positions for insertion are limited based on retaining protein function -For enzymatic modification, enzymes need to be readily available and affordable

Table 2. Therapeutic and Diagnostic Applications of Protein Conjugates

Targeting component	Added functionality	Conjugation method	Specified conditions	Relevance in research field
(i) cetuximab (targeting EGFR antigen)[166]	HDAC inhibitors (IC_{50} = 0.07 μM)	conventional (Lys selective amidation)	Step 1: 1 equiv. payload, 1.75 equiv. N,N'-dicyclohexylcarbodiimide, 1.5 equiv. N-hydroxysuccinimide, 16 h, r.t., DMF Step 2: 20 equiv. activated payload, 1 h, r.t., pH 7.4	Therapeutic delivery of medium-cytotoxic drugs for the treatment of neurological disorders, inflammation, viral and protozoal infections, cardiovascular disorders, and cancer
(ii) anti-CXCR4 (targeting T-lymphocyte antigen)[168]	dasatinib (Lck inhibitor, IC_{50} < 1 nM)	Step 1: conventional (Lys selective amidation) Step 2: aldehyde-based click reaction (oxime formation)	Step 1: 30 equiv. N-succinimidyl-4-formylbenzamine, 3 h, r.t., pH 7.4 Step 2: 30 equiv. dasatinib, 24 h, 37 ºC, pH 5-7, >95% conversion	Suppression of T-cell activation and cytokine expression for the treatment of T-cell mediated immune disorders
(iii) anti-WTA (targeting wall-teichoic acids of *S. aureus*)[169]	rifalogue antibiotic (kills both replicating and non-replicating intracellular bacteria)	canonical AA insertion (Cys selective maleimide-based conjugation)	3 equiv. payload, 1 h	Elimination of intracellular *S. aureus* (a major contributor to invasive infections and is resistant to regular antibiotic treatments), and the method holds potential for the treatment of other intracellular pathogens
(iv) mouse IgG1-Apoliprotein A1[a] (targeted by a secondary antibody - goat anti-mouse IgG H+L)[170]	nucleotide (azide-functionalized)	Step 1: conventional (Lys selective amidation) Step 2: alkyne-based click chemistry (SPAAC)	Step 1: 5 equiv. DBCO reagent, 2 h, r.t., pH 7.4 Step 2: 10 equiv. nucleotide, 16 h, 4 ºC, pH 7.4	A colorimetric read-out of nucleotide incorporation by using enzyme-linked immunosorbent assays based on the targeting of an antibody conjugate incorporated into the DNA
(v) trastuzumab (HER2 targeting antibody)[171]	thiol-reactive bifunctional chelators to allow labeling by ^{89}Zr and ^{177}Lu	canonical AA insertion (Cys selective reaction with phenyloxadiazolyl methylsulfone (PODS) reagents)	10 equiv. PODS reagent, 10 equiv. TCEP, 2 h, r.t., pH 7.4	Improved stability of radiolabeling of bioconjugates for PET with lower background signals

(vi) trastuzumab (HER2 targeting antibody)[172]	KSPis (pyrrole subclass)	Step 1: conventional (Cys modification by partial reduction of disulfide bonds with maleimide reagent) Step 2: promotion of thiosuccinimde ring hydrolysis and stabilization of ADC product	Step 1: 1 h, r.t. Step 2: overnight, r.t., pH 8	Introduces new antitumor payload for the creation of ADCs. KSPis follow an alternative mechanism to those of usual payload classes (DNA intercalators and tubulin inhibitors). KSPis prevent centrosome separation during the cell cycle.
(vii) cyclic-RGD peptides (targeting $\alpha_v\beta_3$ integrin receptors in tumor vasculature)[173]	adeno-associated virus capsid (functionalized via azido-Lys ncAA)	Step 1: ncAA insertion (azido-Lys) Step 2: alkyne-based click chemistry (SPAAC)	2 h, r.t.	Creates targeted delivery for safer gene therapy for anticancer treatment by redirecting binding target of the adeno-associated virus capsid
(viii) single-chain antibody fragments (three different endothelial-targeting fragments)[174]	azide-containing peptide and antioxidant enzyme	Step 1: enzymatic tag insertion (sortase A mediated conjugation) Step 2: alkyne-based click chemistry (SPAAC)	Step 1: 1 equiv. scFv, 1 equiv. sortase A, 5 equiv. azide-containing peptide, 16 h, r.t., pH 7.5, conversion >95% Step 2: 4 equiv. scFv, 1 equiv. DBCO-functionalized catalase, overnight, r.t.	Targeted-therapeutic for vascular endothelial cells as they act as sites of interest in thrombotic, ischemic, and inflammatory conditions and could furthermore modulate passage of macromolecules or drug carriers from vasculature areas to target organs
(ix) nanobody (anti-EGFR)[175]	upconversion nanoparticles loaded with the drug doxorubicin	enzymatic tag insertion (C-terminal conjugation with microbial transglutaminase)	1 h, r.t.	Targeted delivery of anticancer drugs demonstrated using PEGylated nanobodies tethered to human serum albumin coated upconversion nanoparticles loaded with doxorubicin
(x) affibodies (anti-EGFR)[176]	Benzophenone for photo-cross-linking to EGFR receptor	Step 1: canonical AA insertion (engineered Cys modification by maleimide moiety) Step 2:	Step 1: 20 equiv. 4N-maleimido-benzophenone, overnight, r.t., pH 7.4 Step 2: 127 equiv. EGFR extracellular domain, near UV (365 nm), 1 h	Smaller size and lesser affinity for EGFR allows for an affibody-targeted system to have deeper penetration into a solid tumor environment. Crosslinking the

		once modified affibodies throughout tumor environment, upconversion nanoparticles used in tandem to deliver local irradiation for crosslinking		affibody to the EGFR receptor allows for the distribution of the affibody construct to be retained for longer in the tumor environment.
(xi) anti-HER2 and anti-TenB2 (HER2 and tomoregulin targeting antibodies)[167]	maytansinoid DM1 cytotoxic drug linked to antibodies via a linker with two fluorophores present on either side of a cleavage site	canonical AA insertion (engineered Cys modification by maleimide moiety)	no specifics reported	Due to having a cleavable linker with FRET fluorophores on either side, the antibody and payload can be tracked even after the linker has been cleaved and provide more information on internalization and intracellular trafficking.
(xii) trastuzumab (HER2 targeting antibody)[164]	biotin and fluorophore functionalities	canonical AA insertion (engineered Cys selective reaction and engineered Sec selective reaction)	Step 1 (Sec-modification): derivatives (iodoacetamide or methylsulfone based), 25 equiv. DTT, 30-60 min, r.t., pH 5.2, 2.5 equiv. biotin Step 2 (Cys modification): 5 equiv. methylsulfone-functionalized fluorophore, 1 h, r.t., pH 7.4	Possible strategy to overcome the developing resistance seen to current ADCs, site-specific conjugation of two different drugs
(xiii) SST (targeting SST-2 receptors on cancer cells that have been biotinylated)[177]	C3 (a Rho inhibitor that has been biotinylated)	For SST: in-chain residue modification (disulfide rebridging) For C3: canonical AA insertion (engineered Cys to be modified by maleimide)	For SST: pH 7.8 For C3: 30 equiv. biotinylated reagent, 3 h, r.t., pH 7.4	An alternative strategy for overcoming cancer cell resistance to mechanisms of certain drugs.

[a]In this case, the antibody involved in the nucleotide conjugate is the target of a secondary antibody that allows the colorimetric assay to visualize the nucleotide-antibody conjugate incorporation into DNA.

Glossary:

Site-selective:
Modification methods that target a certain residue over other types of amino acids.

Site-specific:
Modification methods that target a single occurrence of a particular type of amino acid.

Endogenous residues:
Amino acid residues that are present in the native sequence of a protein based on the unaltered genes of the host organism.

Canonical amino acids:
The standard 20 amino acid types encoded and inserted naturally by the genetic code and by native protein biosynthesis systems.

Protein microenvironment:
The manipulation of amino acid sidechain properties (e.g. steric or electric characteristics) and reactivity based upon the identity of surrounding amino acids in the protein sequence.

Noncanonical amino acids:
Amino acids that are most often synthesized and non-proteinogenic (with the exception of selenocysteine and pyrrolysine) and can either be inserted residue- or site-specifically into protein sequences.

Heterogenous products:
Protein conjugates that have different constitutions based on the conjugation method binding differing amino acid types or various occurrences of the same amino acid type.

Human insulin:
A protein that is made up of two separate chains of amino acids labeled as A and B, bound together by two disulfide bridges.

Post translational modifications:
Post-translational, covalent protein modifications that have critical roles in cell signaling and control of protein activation or function.

SH3 domain proteins:
Proteins that contain SH3 domains for the regulation of cytoplasmic signaling pathways.

Bioorthogonal reactions:
Chemical reactions that can be executed in the complex environment of living systems (i.e. in the presence of many nucleophiles, reductants, etc.) without altering or affecting native processes.

Antibodies:
Proteins that are composed of two main regions: Fc regions (constant regions for the support and stability of the antibody) and Fab regions (variable regions of the antibody that must be preserved in order to retain affinity and specificity for a corresponding antigen).

Click chemistry:
Chemical reactions that can be defined based on high reaction and conversion rate, green solvent systems, low byproduct levels, and broad functional group applicability.

Conjugate payload:
The chemical linker and added functionality (e.g. fluorophore, cytotoxic drug, etc.) in a protein conjugate.

Disulfide rebridging:
Process by which two cysteine residues, revealed by disulfide reduction, reform the disrupted disulfide either through the construction of a mixed disulfide or through the introduction of a synthetic stapling molecule to connect the two residues

Orthogonal tRNA/RS pairs:
These orthogonal pairs can use native protein biosynthesis machinery for the site-specific insertion of noncanonical amino acids and require that no native RS be able to aminoacylate the incorporated tRNA and no native tRNA be modified by the incorporated RS.

Fusion protein:
Proteins that are produced by combining parts from different proteins or proteins with smaller amino acid sequences/tags to create one expressed entity.

Upconversion nanoparticles:
Nanoscale particles that allow for photon upconversion (the absorption of two lower-energy photons to create one higher-energy, emitted photon) for imaging and sensors in deep tissue environments.