

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*}

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

²Instituto 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

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

27

28 [H1] **Introduction**

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

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

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

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

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

75 76 **[H1] Direct modification of native proteins**

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

92
93 **[H2] *N-/C-terminus Modification***

94 Terminus modifications have the potential to be more generally applicable —most protein
95 termini are accessible and have chemical environments distinct from the remainder of the protein.
96 Several recent discoveries have been made that improve method versatility and terminal AA
97 tolerance while addressing past problems.¹⁷ For C-terminal modifications, targeting relies on the
98 difference in oxidation potential between terminal carboxylic acids and in-chain glutamic acid
99 (Glu)/Asp residues. MacMillan and coworkers have reported a method based on visible-light-
100 mediated single-electron transfer (SET) to perform decarboxylative alkylation at C-terminal
101 residues (3 equiv. photocatalyst, 10 equiv. Michael acceptor, 8 h, r.t., pH 3.5, 41–49% conversion,
102 Figure 2i).¹⁶ Unlike prior amide coupling and esterification strategies, this SET reaction favors the
103 C-terminus over carboxylic acid moieties in Glu and Asp residues.^{20–22} Such selectivity originates
104 from increased stability of the C-terminal carbon-centered radical.¹⁶ The aforementioned
105 conditions allow for the selective alkylation of human insulin at the A chain C-terminus. However,
106 using alternate decarboxylative conditions, originally studied for targeting tryptophan residues at
107 the peptide level, selective modification of the B chain C-terminus was observed (0.5-1 equiv.
108 photocatalyst, 10-30 equiv. Michael acceptor, 10 equiv. K₂HPO₄, 2.5-3 h, DMF, Merck
109 Photoreactor (450 nm), 15-38% conversion).²³ Though the precise reasons for this selectivity
110 remain as yet undetermined, the divergence in these C-terminal modification products showcases
111 the number of factors at play when performing protein conjugation.

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

117 residues are lower in abundance and may not be as accessible as the N-terminus. Thus, Cys-
118 modification can be avoided by using protection steps or by careful, amine-reactive reagent
119 selection.^{25,26} Therefore, targeting the N-terminus in a single chain protein can lead to site-specific
120 modification. Even with such promise, the earliest reported methods targeting the N-terminus
121 required high concentrations of modifying reagents and long reaction times.^{27,28} Numerous
122 methods that address this, involving one step, low reagent concentrations, and reasonable reaction
123 times have been reported: oxidative coupling with *o*-aminophenols (5 equiv. *o*-aminophenol, 250
124 equiv. oxidant, 30 min, room temperature (r.t.), pH 7.5, Figure 2ii), addition of 2-
125 pyridinecarboxaldehydes (2-PCA, 400 equiv., 16 h, 37 °C, pH 7.5, 33-95% conversion, Figure
126 2iii), and reductive alkylation with aldehydes (2 equiv. aldehyde, 5 equiv. reductant, 6-48 h, r.t.,
127 pH 6.1, 30-70% conversion, Figure 2iv).²⁴⁻²⁶ While some N-terminal residue types show higher
128 conversion values than others, only a few are found to be incompatible (for example, an N-terminal
129 Cys leads to reductive alkylation).^{24,25} Such versatility can be even further extended by enzyme-
130 mediated modification at the N-terminus. Butelase 1 has been reported as an asparagine/aspartic
131 acid (Asn/Asp) specific ligase to efficiently conjugate thiopeptide substrates to N-terminal
132 residues (0.001 equiv. butelase 1, 5 equiv. substrate, 2.5 h, 42 °C, pH 6.5, 70-82% conversion,
133 Figure 2v). Butelase 1 has a high substrate tolerance, and can successfully perform conjugation
134 with any N-terminal AA other than proline. Reaction tends to be favoured by the presence of a
135 hydrophobic residue in the second position.²⁹ Having several conjugation strategies available
136 allows navigation of the complex combinations of disadvantages and advantages in order to choose
137 the optimal method for a diverse range of proteins and applications.

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

140 for every terminal residue type. Further advantages and limitations stem from specific reagents or
141 conditions used. With *o*-aminophenol oxidative coupling, double modification can be seen at basic
142 pH either due to Lys modification or dimerization.²⁶ In the case of 2-PCA, after 12 h at 37 °C, 20-
143 30% of the modified protein is lost.²⁵ For the method relying on butelase 1, the short half-life of
144 the thiopeptide reagents must be acknowledged and reaction conditions adjusted.²⁹ These
145 limitations highlight why method diversity is needed: to allow for the choice of a modification
146 method with limitations that will not affect the conjugate application. Such diversity also applies
147 to the various advantages of methods: the use of reductive alkylation maintains the charge on the
148 N-terminal amine which may be necessary for protein function, and the use of 2-PCA or *o*-
149 aminophenol can facilitate multiple, site-specific modifications.²⁴⁻²⁶ Optimization of older
150 methods has also led to the progression of N-/C-terminal modification strategies. As an example,
151 Rapoport's Salt (RS) for oxidation at the N-terminus shows an expanded tolerance for different N-
152 terminal residues in higher pH environments (~ pH 8.5).³⁰ The important advantages to incorporate
153 and limitations to avoid are depend on the application of interest. However, the structure and
154 makeup of the protein, the presence of post translational modifications (PTMs), or the need for the
155 termini to be available for protein function, mean that other protein modification strategies (those
156 that target in-chain sites) are also in high demand.^{17,31}

157

158 [H2] *In-chain Residue Modification*

159

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

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

171 Due to their innate nucleophilicity, Lys residues represent one of the more abundant AAs
172 (~6%) commonly targeted for protein modification.¹⁸ While Lys abundance can lead to extreme
173 product heterogeneity with conventional modification methods (for example amidation with
174 succinimide-based esters), the tertiary structure of proteins can enable higher levels of specificity
175 as a result of unique, microenvironment-driven pK_a changes.^{1,18,19} A recent study by Cravatt and
176 coworkers observed and quantified the reactivities of 9,000 Lys residues in the human proteome,
177 finding several hundred hyperreactive cases.¹⁸ Therefore, specific reagents or conditions have been
178 developed to modify the most reactive Lys residues in different proteins. By adjusting the
179 electrophilicity of Michael acceptors, our group has targeted the most reactive Lys residues in
180 proteins while leaving other Lys and Cys residues untouched (1 equiv., 1-2 h, 25-37 °C, pH 8.0,
181 >95% conversion, Figure 3ii). The design of sulfonyl acrylate reagents was assisted by
182 computational calculations that determined transition states facilitating the desired bond
183 formation.¹² Similarly, a multicomponent reaction involving formaldehyde, acetylene, and a Cu-
184 ligand complex as a catalyst was found to modify a single Lys residue in a handful of different
185 proteins (100 equiv. aldehyde, alkyne, and CuI, 400 equiv. ligand, 72 h, r.t., pH 7.8, 40-99%
186 conversion, Figure 3iii). Conveniently, the N-terminus is reversibly protected by the

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

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

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

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

233 the size and bulkiness of rebridging reagents must be limited to avoid disruption of the structure
234 of the target protein. Recently, commercially available oxetane reagents have demonstrated the
235 ideal distance for disulfide rebridging while also improving the stability and activity of therapeutic
236 proteins and antibodies (20-60 equiv. oxetane, 6-12 equiv. tris(2-carboxyethyl)phosphine (TCEP),
237 24-48 h, 25-37 °C, pH 8.5, >95% conversion, Figure 3viii).⁴³ Alternatively, unlike their bis-sulfone
238 counterparts⁴⁴, allyl sulfones have been proposed as viable disulfide rebridging reagents with high
239 aqueous solubility and reactivity (2 equiv. allyl sulfone, ~1 equiv. TCEP, 24 h, r.t., pH 7.8, 19%
240 and 28% isolated yield with insulin and lysozyme as model proteins, Figure 3ix).⁴⁵ Meanwhile,
241 divinylpyrimidine (DVP) was just reported as a stable rebridging agent for antibody and protein
242 conjugates (10-15 equiv. DVP, 5 equiv. TCEP, 1-2 h, 37 °C, pH 8.0, >95% conversion).⁴⁶ A type
243 of disulfide rebridging reagent for obtaining highly homogenous antibody drug conjugates (ADCs)
244 with drug-to-antibody ratios (DARs) of 2 was reported: dibromopyridazinedione (dibromo-PBD)
245 derivatives (16 equiv., 80 equiv. TCEP, 16 h, 4 °C, pH 8.0, Figure 3x). To achieve this, one
246 compound containing two dibromo-PBD derivatives connected by a linker crosslinks two disulfide
247 bridges. Since there are four interchain disulfide bridges available in IgG antibodies, two such
248 dibromo-PBD-linker compounds are used, each with one payload incorporated.⁴⁷ Further reports
249 of dibromo-PBD rebridging reagents for a “plug and play approach” for the production of ADCs
250 and for the attachment of antibody fragments in a specific, favorable orientation to nanoparticles
251 have been published recently.^{48,49} Even photomediated disulfide bridging has been explored using
252 a one-pot thiol-yne conjugation strategy on a reduced antibody fragment (0.8 equiv. initiator, 1
253 equiv. alkyne reagent, 4 h, 0 °C, UV radiation (365 nm), 40% conversion, Figure 3xi). The method
254 has proven applicable for a variety of alkyne reagents and has exhibited the potential of
255 photomediated disulfide rebridging for therapeutic applications.⁵⁰

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

265 266 **[H1] Modification via Genetic Manipulation**

267 The genetic engineering of proteins has proven to be an indispensable methodology for
268 site-selective protein modification and the production of homogenous bioconjugates. Rather than
269 working only with a scaffold of endogenous AAs, genetic engineering allows for the controlled
270 introduction of a variety of abiotic and biotic chemical handles at designated sites.¹ The potential
271 versatility and generality inherent in the genetic engineering platform make these strategies
272 invaluable for the exploration and modulation of protein function. The discussion of the various,
273 genetically-based strategies that follows, focuses on the diversity and complementarity of the
274 different method components and potential applications. Such diversity can be evaluated based on
275 certain attributes: multiple modifications; biologically compatible conditions; reaction selectivity;
276 directly added functionality vs. added bioorthogonal handle for subsequent modification; reaction
277 efficiency and stability; method applicability; and modification reversibility. Additionally, the
278 strategies presented here have been categorized into three broader insertion groups: canonical AAs,
279 non-canonical (ncAAs), and motifs or enzymatic tags.
280

281 [H2] *Canonical Amino Acid Insertion*

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

292 [H3] *Cysteine insertion*

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

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

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

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

327 a high potential for rapid reactions and forms highly-stable products (12 equiv. CBTF, 1.1-2.2
328 equiv. TCEP, 12 h, r.t., pH 6.8, Figure 4iii).^{61,62} Alternatively, by decorating the aromatic ring with
329 a variety of electron withdrawing groups (EWGs) and fluorine substituents, fluorobenzene type
330 reagents have also been explored for the stable and selective S-arylation of Cys residues (2-67
331 equiv., 16 h, 37 °C, pH 8.3, Figure 4iv).⁶³ Other reagents that form stable, irreversible linkages
332 while also introducing desired functionality are: carbonylacrylic, allenamide, and cyclopropenyl
333 ketone derivatives.^{15,64,65} Carbonylacrylic reagents succeed in performing the desired modification
334 via a rapid Michael addition in a stoichiometric manner (1-50 equiv., 1-2 h, r.t.-37 °C, pH 8.0,
335 >95% conversion, Figure 4v).¹⁵ Allenamide reagents, being less electron deficient than ketones
336 and esters, react specifically with Cys and avoid modification by common biological nucleophiles
337 (10-25 equiv. allenamide, 10 equiv. dithiothreitol (DTT), 30 min, 37 °C, pH 8.0, Figure 4vi).⁶⁴
338 Unlike the carbonylacrylic and allenamide reagents, the cyclopropenyl ketone derivatives,
339 functionalized via amide connections formed by NHS-ester promoted reactions, rely on ring strain
340 to promote irreversible, rapid formation of stable products (50 equiv. cyclopropenyl ketone, 5
341 equiv. tris(hydroxypropyl)phosphine, 10 min, r.t., pH 6.0, Figure 4vii).⁶⁵ Furthermore,
342 dichlorotetrazine (10 equiv., 1 h, r.t., pH 5.2, Figure 4viii) and 2-azidoacrylate reagents (10 equiv.,
343 12 h, r.t., pH 7.4, >85% conversion, Figure 4ix) have been proposed as both being able to directly
344 add functionality in one step in addition to introducing bioorthogonal sites for further
345 functionalization.^{66,67} These dual functionalization strategies circumvent the need for two ncAA
346 insertions and are very promising for improved theranostic applications.⁶⁷

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

350 difficulties with complex biomolecules: side reactions with endogenous protic, basic, or thiol-
351 containing moieties; heterogeneous product mixtures; catalyst deactivation; and incompatible
352 protein modification conditions.^{68,69} Only recently has the specificity, efficiency, and versatility of
353 transition-metal based chemistry been harnessed thanks to judicious metal and ligand choices.⁷⁰
354 Au(III) (15-20 equiv., 30 min, r.t., pH 8.0, Figure 4x) and Pd(II) complexes (1.1-10 equiv., 24 h,
355 37 °C, pH 8.5, >94% conversion, Figure 4xi) have been used for the production of stable S-arylated
356 products by targeting Cys residues. The Au(III) and Pd(II) catalyzed systems are based on two-
357 electron strategies. The S-arylation products are stable and the conjugation irreversible, making
358 these methods useful for many potential therapeutic applications.^{68,70-72}

359 Reversible conjugation processes can, in the right context, be useful. These methods
360 include the efficient Michael addition of either 5-methylene pyrrolones (5MPs, 200-500 equiv., 2
361 h, 37 °C, pH 7.5, Figure 4xii) or 4-acetoxy cyclopentenones (50 equiv., 1-2 h, r.t., pH 7.0-7.4, 22-
362 95% conversion, Figure 4xiii) with controlled release by increasing pH/thiol exchange or Michael
363 donor addition respectively; the formation of a thiazolidino boronate (TzB) product at N-terminal
364 Cys residues by 2-formyl phenylboronic acid (2-FPBA) that dissociates in slightly acidic
365 environments (this particular case shows modification of native N-terminal Cys, 1 equiv., 30 min,
366 r.t., pH 7.0, >95%, Figure 4xiv); and the formation of a thioether bond by a fast reaction driven by
367 irradiation at 350 nm of 3-(hydroxymethyl)-2-naphthols (naphthoquinone methide precursors,
368 NQMPs) that can be reversed by irradiation of a dilute solution of labeled conjugate or when mixed
369 with vinyl ether (8-9 equiv., 2-6 min, r.t., pH 7.4, 350 nm irradiation, Figure 4xv).⁷³⁻⁷⁷ Such
370 variable processes for controlled release of a Cys residue can lead to information on critical
371 epigenetic roles or reversible modulation of protein function.^{73,76} Additionally, the acidic
372 environment-driven release of 2-FPBA could be used for endosomal release and delivery of

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

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

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

388
389 To supplement engineered Cys residue modification, methods targeting alternative, low-
390 abundance AAs have been explored. Genetically inserted residues that have captured particular
391 interest are: Trp, Tyr, Met and His. Apart from all four being of relative low-abundance to other
392 canonical AAs, endogenous Met and Trp residues are also most commonly found in the
393 hydrophobic interior of the protein.^{78,79} Both low abundance and positioning can increase the
394 likelihood that a genetically inserted copy is the only instance of that residue available for
395 modification. In other words, genetic manipulation can allow for the formation of a unique
396 chemical handle. However, due to the lower reactivity of these four sidechains in comparison to

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

402 Two methods reported in the last five years, one metal-mediated and the other
403 organoradical-based, target Trp residues.^{79,83} The metal-mediated method for Trp uses
404 1-[(triisopropylsilyl)ethynyl]-1,2-benziodoxol-3(1H)-one (TIPS-EBX). Catalyzed by
405 [AuCl(SMe₂)], the hypervalent iodide TIPS-EBX adds regioselectively to the Trp indole (10 equiv.
406 TIPS-EBX, 5 equiv. catalyst, overnight, r.t., ~90% conversion, Figure 5i), rendering this the first
407 Trp modification method both selective for Trp and able to modify a specific site on Trp residues
408 (i.e. modification occurs at Trp C2).⁸³ Meanwhile, the reagents used for the organoradical,
409 selective modification of Trp are derivatives of 9-azabicyclo[3.3.1]nonane-3-one-N-oxyl (keto-
410 ABNO). Stabilized on the oxygen bonded to the nitrogen, the radical adds to the indole ring of Trp
411 to achieve a highly homogenous product (1-5 equiv. keto-ABNO, 0.6-3 equiv. NaNO₂, 30 min,
412 r.t., pH 7.4, 11-64% conversion, Figure 5ii).⁷⁹ Based on some drawbacks to these methods,
413 including the acidic conditions required for high conversion with keto-ABNO conjugation, metal-
414 mediated strategies act as valuable alternatives for proteins and applications requiring alternate
415 conditions.⁷⁹

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

420 addition of *N*-methylated luminol derivatives to the *ortho* Tyr position in the presence of H₂O₂ (1
421 equiv. hemin, 100 equiv. peroxide, 100 equiv. luminol derivative, 1 h, r.t., pH 7.4, Figure 5iii).
422 The production of a reactive cyclic diazodicarboxamide intermediate *in situ* drives the reaction
423 forward.⁸⁴ A similar intermediate has been used as the starting reagent for Tyr modification, but
424 the hemin-mediated method avoids the need to store unstable reagents.^{84,85} Whether through
425 coordination or covalent binding, some methods require direct metalation of the protein rather than
426 metal mediation of the modification. A method for Tyr modification has been established recently
427 that uses rhodium(III) chloride and boronic acid to link arene complexes to the *ortho* position in
428 Tyr (50 equiv. rhodium(III) chloride and boronic acid, overnight, r.t., pH 9.4, Figure 5iv). The
429 Rh(III) Tyr complexes maintain both a metastability and controlled reversibility (via DTT or H₂O₂)
430 due to the inorganic linkage.⁸⁶ Although still not as highly regarded as metal-free methods,
431 advantages of metal-based reactions have become more apparent and have led to further
432 exploration of novel, specific interactions.

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

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

451 Outside of the proteinogenic, canonical AAs, selenocysteine (Sec) has also made an
452 appearance quite recently in reports of site-specific protein modification (oxidation of Sec: 20
453 equiv. 2,2'-dithiobis(5-nitropyridine), 15 min, r.t., acidic conditions, not isolated; arene addition:
454 10 equiv. arene, 5 h, 37 °C, pH 8.0, 23% isolated yield).⁸⁸ This report acknowledges the ability for
455 electron-rich arenes (e.g. vancomycin) to attach at the site of an oxidized Sec residue, containing
456 an electrophilic Se-S bond. While modification of the oxidized Sec was shown directly in an
457 affibody, the insertion into the full length antibody was accomplished with a sortase A mediated
458 method.⁸⁸ Even so, this method shows promise for direct modification of the oxidized Sec in full
459 length antibodies and larger proteins. All of the methods in this section, while creating novel
460 solutions and tools to unlock unexplored directions, represent an ongoing battle against lower
461 reactivity, lower conversion, and lower selectivity when attempting to target inserted, low-
462 abundance AAs (with the exception of Cys).

463
464 [H2] *Noncanonical Amino Acid Insertion*
465

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

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

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

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

494 Recently, chemistries have been explored to install accurate PTMs or chemical handles for
495 subsequent crosslinking.^{98,99} Site-selectivity is vital for the evaluation of specific PTMs, as these
496 epigenetic modifications have different consequences when translated by corresponding “reader”
497 proteins in varying contexts.¹⁰⁰ Understanding the role of certain PTMs, especially in histone
498 proteins, can be accomplished through ncAA insertion. Such insertions help decode protein–
499 protein interactions (PPIs), including those necessary for gene regulation and apoptosis.⁹² A novel
500 ncAA, ϵ -*N*-2-hydroxyisobutyryl-Lys (HibK), representing a PTM observed in histones, was
501 recently incorporated in histone proteins using an orthogonal amber suppressor pyrrolysl-RS pair
502 (3.6-11.9 mg/L expression yield). By altering the charge of the Lys residue and adding steric bulk,
503 site-specific insertion of HibK will help determine how the PTM affects chromatin structure.¹⁰¹
504 Additionally, PTMs can be inserted in protected forms if the native form is unstable or too reactive.
505 Phosphotyrosine (pTyr) was inserted as a neutral analog that is both cell permeable and stable in
506 cells (1.0-1.75 mg/L expression yield). Acidic conditions were used to reveal the native
507 phosphotyrosine (16-48 h, 4 °C, pH 1.0-2.0).¹⁰² Similarly, a protected allysine residue in the form
508 of ϵ -*N*-(4-azidobenzoxycarbonyl)- δ,ϵ -dehydrolysine (AcdK) was inserted into histone proteins
509 and epigenetic enzymes (7.0 mg/L expression yield). The AcdK undergoes reduction by
510 phosphines to reveal allysine (TCEP, 2 h, r.t., pH 7.0), which hydrolyzes and is modified by
511 reductive amination (NaCNBH₃, 8 h, r.t., pH 7.0) to insert either monomethyllysine or
512 dimethyllysine.¹⁰³ Both lysine-methylation and phosphorylation PTMs have significant impact on

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

515 Unlike the aforementioned examples, novel photo-lysine ncAAs can insert photo-
516 derivatives (4-40% incorporation), containing diazirine rings, of lysine-based PTMs. These
517 derivatives both approximate and identify possible native PPIs. The method has yet to be
518 demonstrated in a site-specific manner, but a residue specific manner based on growing cells in
519 media containing photo-lysine has confirmed its potential for selective Lys replacement in native
520 proteins. Noncovalent, transient PTM interactions can then be captured by covalent bonds formed
521 after the photo-cross-linking of the inserted photo-lysine and the protein of interest (1 equiv., 30-
522 60 min, 25-37 °C, pH 7.5, 365 nm irradiation).¹³ Such cross-linking methods also represent an
523 alternative application for ncAAs: the insertion of chemical handles for subsequent
524 functionalization.

525 The ability of ncAA insertion to install reactive groups that enable bioorthogonal
526 functionalization has been invaluable to site-selective protein modification research. Chemical
527 handles recently inserted into proteins include: fluorine-activated aryl carbamates (FPheK, 3-8
528 mg/L expression yield), aryl isothiocyanates (pNCSF, 8-16 mg/L expression yield), and thioester
529 derivatives of Asp acid (ThioD, 8 mg/L expression yield).¹⁰⁴⁻¹⁰⁶ FPheK, once inserted into a
530 protein, reacts with amines, thiols, and phenols to produce intra- or inter-molecular cross-linking
531 (2 equiv. nucleophile, 8 h, 37 °C, pH 8.5).¹⁰⁴ Similarly, pNCSF forms urea-type, cross-linking
532 bridges between proteins or between proteins and small molecules (3-100 equiv. amine-containing
533 nucleophile, 3-24 h, 37 °C, pH 7.4-8.5).¹⁰⁵ Meanwhile, ThioD can be modified by amine-based
534 nucleophiles (100 equiv. nucleophile, 20 h, 37 °C, pH 7.4).¹⁰⁶ While these insertions broaden the
535 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 °C, 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 °C, 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 °C, 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 °C, 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 °C, 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 °C, pH 8.0, >95% conversion, 4-AA tag,
604 Figure 7vi). Based on computationally calculated peptide conformations and energy pathways, the

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

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

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

626 Beyond the targeting of Cys and Lys residues, a recently reported method targets a His-Gly-His
627 (His₂-tag, 1-5 equiv. PEGylated bis-sulfone, 16 h, 20 °C, 28-39% conversion, 6-AA tag).¹²³ With

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

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

649
650 **[H1] Downstream Functionalization**
651

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

661 However, with each major development, limitations and problems have arisen. The use of
662 large protein tags (e.g. GFP) risks affecting protein activity post conjugation, and the introduction
663 of copper-catalysed azide–alkyne cycloaddition (CuAAC) allowed the attachment of small
664 molecule tags. The CuAAC is one example of “click chemistry” — chemistry defined by high
665 reaction and conversion rate, green solvent systems, low levels of byproducts, and broad functional
666 group applicability.^{131–133} The potential toxicity of Cu(I), led to the development of strain-
667 promoted (and copper-free) azide–alkyne cycloaddition (SPAAC).^{133,134} In attempts to further
668 improve the reaction kinetics, inverse-electron-demand Diels–Alder (IEDDA) reactions were
669 developed.¹³⁴ Reaction rate is a highly important criterion in the development of bioorthogonal
670 reactions (rates span from $10^{-5} \text{ M}^{-1} \text{ s}^{-1}$ to $10^5 \text{ M}^{-1} \text{ s}^{-1}$) — molar substrate concentrations can be
671 naturally limited when dealing with large molecular weight biomolecules and in the case of
672 radiolabelling experiments reactions must be completed before decay is complete. . However,
673 more reactive agents used to achieve faster reactions were also observed to be less stable (e.g.
674 *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, $k = 496 \pm 70 \text{ M}^{-1} \text{ s}^{-1}$, 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 °C, pH 8.0, >95% conversion, $k = 0.054 \pm 0.00067 \text{ M}^{-1}$
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 coupling producing a linear,
692 stable diyne product (10 equiv. alkyne, >50 equiv. CuI/tetramethylethylenediamine, 4-6 h, 4 °C,
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 °C, pH 7.4-8.0, 45-50% conversion, $k = 1.0 \text{ M}^{-1} \text{ s}^{-1}$, 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 °C, pH 5.5, $k = 0.20 \text{ M}^{-1} \text{ s}^{-1}$, Figure 8v) and tandem
705 (8 equiv. pyrazolone, 16 h, 37 °C, 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 °C, pH 6.5, 83% conversion, $k = 0.0078 \text{ M}^{-1} \text{ s}^{-1}$, 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 °C, pH 7.5, >95% conversion, $k = 24 \text{ M}^{-1} \text{ s}^{-1}$, 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 phospho-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 °C, pH

721 6.8-8.8, 80-90% conversion, $k = 0.06 \pm 0.01 \text{ M}^{-1} \text{ s}^{-1}$, Figure 8ix).¹⁴⁶ A quadricyclane (QC)-
722 containing ncAA has also been successfully inserted into proteins. The strained, hydrocarbon
723 reagent was functionalized with nickel bis(dithiolene), leading to a QC ligation cleavable by UV
724 irradiation and orthogonal to common reactions with aldehyde/aminooxy and alkyne/azide pairs
725 (1-2 h, r.t., pH 7.4, Figure 8x).¹⁴⁷ Lastly, the reaction between cyclopropenones, inserted as ncAAs,
726 and triarylphosphines was shown to produce α,β -unsaturated amides (20 equiv., 1-4 h, 37 °C, pH
727 7.0, >95% conversion, $k \geq 20 \text{ M}^{-1} \text{ s}^{-1}$, Figure 8xi).^{148,149} The introduction of reversibility and
728 orthogonality into bioorthogonal reactions paves the way for traceless or multi-site conjugation
729 applications. In particular, the increase in bioorthogonal reactions discovered raises the likelihood
730 for complementary/orthogonal reactions to achieve multiple, distinct modifications.

731 In addition to “click-type” reactions, accurate approximations of PTMs can be installed
732 using methods established for the modification of Dha, which is itself commonly installed by the
733 reduction of an inserted Cys residue.^{6,150} PTMs have also been inserted based on synthetic and
734 ncAA methods. The synthetic approach is inconvenient as the PTM can only be incorporated easily
735 into shorter peptides, and the insertion of ncAAs involves being able to obtain the appropriate
736 genetic machinery.¹⁵¹ Thus, the formation and modification of Dha is a much more attractive
737 option. As Dha insertion marks the addition of an electrophilic moiety, a type of functional group
738 not endogenous to proteins, novel methods to modify Dha sidechains at both terminal and in-chain
739 positions have been ardently pursued.¹⁵² Most recently, aza-Michael additions have been
740 demonstrated with amine-based nucleophiles to produce secondary and tertiary amine products
741 (>300 equiv., 1-4 h, 25-37 °C, pH 8.0-9.0, 40-95% conversion, $k = 6.1 \times 10^{-5} \text{ M}^{-1} \text{ s}^{-1}$, Figure
742 8xii).^{153,154} These reactions avoid the use of thiol-based nucleophiles to avoid risking the disruption
743 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 NaBH₄ for the prevention of unwanted oxidation and disubstitution
748 (100-2000 equiv. alkyl-halide, 30 min, 4 °C, 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. Na₃SPO₃, 1 h, 37 °C, pH 8.6; second step: PP1
759 phosphatase, 1 h, 30 °C, 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 $10^5 \text{ M}^{-1} \text{ s}^{-1}$); 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

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

795 In the field of therapeutic conjugates, ADCs capture most of the attention.¹⁶⁰ However,
796 novel methods have been reported recently to either improve the ADC mechanism of action or to
797 apply the idea of targeted delivery to alternate payloads (e.g. radioligands). All of these
798 improvements require the assistance of selective protein conjugation methods.^{164,166–178} Strategies
799 ranging from conventional to site-specific conjugation are still in use to drive these adjustments
800 and optimizations. Conventional, selective Lys amidation has very recently reported as a method
801 for conjugating thiol-based histone deacetylase (HDAC) inhibitors to cetuximab antibodies
802 targeting EGFR (Table 2i). Less than 1% of the injected dosage of ADCs are expected to reach
803 and be internalized by a target tumor, and as a result payloads for ADCs have been thought to
804 require sub-nanomolar IC₅₀ values. Meanwhile, HDAC inhibitors only have about a 0.07 μM IC₅₀
805 value. As the ADCs incorporating HDAC exhibited anti-tumor effects, this results suggests such
806 highly toxic drugs are unnecessary and that off-site toxicity may be avoided by using lower
807 potency drugs.¹⁶⁶ Other payloads outside of the class of highly cytotoxic, anticancer drugs have
808 also been successfully conjugated to antibodies (see Table 2 entries i-vi).^{168–172} Amazingly, the
809 wide range of applications represented by these antibody-payload constructs is made possible by
810 altering only the payload, conjugation chemistry, and linker, exemplifying how antibody
811 conjugates can be viewed as a modular concept.

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

825 In addition to methods targeting particular disease treatment, several recent studies have
826 been aimed at illuminating the unknown mechanistic aspects of ADC approaches to drug delivery
827 or to overcome foreseeable future issues related to antibody-targeting applications.^{164,167} As
828 knowledge of the internalization and subsequent intracellular trafficking of ADCs remains quite
829 limited, clever strategies are needed to enhance efficacy. One such strategy uses fluorescence
830 resonance energy transfer (FRET) pairing to gain insight (Table 2xi). By incorporating a cleavable
831 linker with one FRET fluorophore on the antibody side, attached by a maleimide-engineered Cys
832 linkage, and one on the warhead side of the linker, when the linker is cleaved, both the antibody
833 and payload can still be visualized and monitored. This method revealed the critical role that the
834 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

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

915

916

917 [H1] References

918 1. Krall, N., da Cruz, F. P., Boutureira, O. & Bernardes, G. J. L. Site-selective protein-
919 modification chemistry for basic biology and drug development. *Nat. Chem.* **8**, 103–113
920 (2016).

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

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

926 4. Takaoka, Y., Ojida, A. & Hamachi, I. Protein Organic Chemistry and Applications for
927 Labeling and Engineering in Live-Cell Systems. *Angew. Chem., Int. Ed.* **52**, 4088–4106
928 (2013).

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

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

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

- 934 endogenous amino acid selective bond forming reactions for bioconjugation. *Chem. Soc.*
935 *Rev.* **44**, 5495–5551 (2015).
- 936 8. Chalker, J. M., Bernardes, G. J. L. & Davis, B. G. A ‘Tag-and-Modify’ Approach to Site-
937 Selective Protein Modification. *Acc. Chem. Res.* **44**, 730–741 (2011).
- 938 9. Schmidt, M. J. & Summerer, D. A Need for Speed: Genetic Encoding of Rapid
939 Cycloaddition Chemistries for Protein Labelling in Living Cells. *ChemBioChem* **13**,
940 1553–1557 (2012).
- 941 10. Boyce, M. & Bertozzi, C. R. Bringing chemistry to life. *Nat. Methods* **8**, 638–642 (2011).
- 942 11. Sletten, E. M. & Bertozzi, C. R. Bioorthogonal Chemistry: Fishing for Selectivity in a Sea
943 of Functionality. *Angew. Chem., Int. Ed.* **48**, 6974–6998 (2009).
- 944 12. Matos, M. J. *et al.* Chemo- and Regioselective Lysine Modification on Native Proteins. *J.*
945 *Am. Chem. Soc.* **140**, 4004–4017 (2018).
- 946 13. Yang, T., Li, X.-M., Bao, X., Fung, Y. M. E. & Li, X. D. Photo-lysine captures proteins
947 that bind lysine post-translational modifications. *Nat. Chem. Biol.* **12**, 70 (2015).
- 948 14. Zhang, C. *et al.* π -Clamp-mediated cysteine conjugation. *Nat. Chem.* **8**, 120–128 (2016).
- 949 15. Bernardim, B. *et al.* Stoichiometric and irreversible cysteine-selective protein
950 modification using carbonylacrylic reagents. *Nat. Commun.* **7**, 13128 (2016).
- 951 16. Bloom, S. *et al.* Decarboxylative alkylation for site-selective bioconjugation of native
952 proteins via oxidation potentials. *Nat. Chem.* **10**, 205–211 (2018).
- 953 17. Rosen, C. B. & Francis, M. B. Targeting the N terminus for site-selective protein
954 modification. *Nat. Chem. Biol.* **13**, 697–705 (2017).
- 955 18. Hacker, S. M. *et al.* Global profiling of lysine reactivity and ligandability in the human
956 proteome. *Nat. Chem.* **9**, 1181–1190 (2017).
- 957 19. Huang, F., Nie, Y., Ye, F., Zhang, M. & Xia, J. Site Selective Azo Coupling for Peptide
958 Cyclization and Affinity Labeling of an SH3 Protein. *Bioconjugate Chem.* **26**, 1613–1622
959 (2015).
- 960 20. Zuo, Z. & MacMillan, D. W. C. Decarboxylative Arylation of α -Amino Acids via
961 Photoredox Catalysis: A One-Step Conversion of Biomass to Drug Pharmacophore. *J. Am.*
962 *Chem. Soc.* **136**, 5257–5260 (2014).
- 963 21. McGrath, N. A., Andersen, K. A., Davis, A. K. F., Lomax, J. E. & Raines, R. T. Diazo
964 compounds for the bioreversible esterification of proteins. *Chem. Sci.* **6**, 752–755 (2015).
- 965 22. Totaro, K. A. *et al.* Systematic Investigation of EDC/sNHS-Mediated Bioconjugation
966 Reactions for Carboxylated Peptide Substrates. *Bioconjugate Chem.* **27**, 994–1004 (2016).

- 967 23. Yu, Y. *et al.* Chemoselective Peptide Modification via Photocatalytic Tryptophan β -
968 Position Conjugation. *J. Am. Chem. Soc.* **140**, 6797–6800 (2018).
- 969 24. Chen, D., Disotuar, M. M., Xiong, X., Wang, Y. & Chou, D. H.-C. Selective N-terminal
970 functionalization of native peptides and proteins. *Chem. Sci.* **8**, 2717–2722 (2017).
- 971 25. MacDonald, J. I., Munch, H. K., Moore, T. & Francis, M. B. One-step site-specific
972 modification of native proteins with 2-pyridinecarboxyaldehydes. *Nat. Chem. Biol.* **11**,
973 326–331 (2015).
- 974 26. Obermeyer, A. C., Jarman, J. B. & Francis, M. B. N-Terminal Modification of Proteins
975 with o-Aminophenols. *J. Am. Chem. Soc.* **136**, 9572–9579 (2014).
- 976 27. Casi, G., Huguenin-Dezot, N., Zuberbühler, K., Scheuermann, J. & Neri, D. Site-Specific
977 Traceless Coupling of Potent Cytotoxic Drugs to Recombinant Antibodies for
978 Pharmacodelivery. *J. Am. Chem. Soc.* **134**, 5887–5892 (2012).
- 979 28. Antos, J. M. *et al.* Site-Specific N- and C-Terminal Labeling of a Single Polypeptide
980 Using Sortases of Different Specificity. *J. Am. Chem. Soc.* **131**, 10800–10801 (2009).
- 981 29. Nguyen, G. K. T., Cao, Y., Wang, W., Liu, C. F. & Tam, J. P. Site-Specific N-Terminal
982 Labeling of Peptides and Proteins using Butelase 1 and Thiodepsipeptide. *Angew. Chem.,*
983 *Int. Ed.* **54**, 15694–15698 (2015).
- 984 30. Palla, K. S., Witus, L. S., Mackenzie, K. J., Netirojjanakul, C. & Francis, M. B.
985 Optimization and Expansion of a Site-Selective N-Methylpyridinium-4-carboxaldehyde-
986 Mediated Transamination for Bacterially Expressed Proteins. *J. Am. Chem. Soc.* **137**,
987 1123–1129 (2015).
- 988 31. Murza, A. *et al.* C-Terminal Modifications of Apelin-13 Significantly Change Ligand
989 Binding, Receptor Signaling, and Hypotensive Action. *J. Med. Chem.* **58**, 2431–2440
990 (2015).
- 991 32. Willwacher, J., Raj, R., Mohammed, S. & Davis, B. G. Selective Metal-Site-Guided
992 Arylation of Proteins. *J. Am. Chem. Soc.* **138**, 8678–8681 (2016).
- 993 33. Imiolek, M. *et al.* Selective Radical Trifluoromethylation of Native Residues in Proteins.
994 *J. Am. Chem. Soc.* **140**, 1568–1571 (2018).
- 995 34. Chilamari, M., Purushottam, L. & Rai, V. Site-Selective Labeling of Native Proteins by a
996 Multicomponent Approach. *Chem. - Eur. J.* **23**, 3819–3823 (2017).
- 997 35. Ball, Z. T. Molecular recognition in protein modification with rhodium metallopeptides.
998 *Curr. Opin. Chem. Biol.* **25**, 98–102 (2015).
- 999 36. Pentelute, B., Zhang, C., Vinogradova, E., Spokoyny, A., & Buchwald, S., Arylation
1000 Chemistry for Bioconjugation. *Angew. Chem., Int. Ed.* (2018) Accepted Author
1001 Manuscript.

- 1002 37. Lohse, J. *et al.* Targeted Diazotransfer Reagents Enable Selective Modification of Proteins
1003 with Azides. *Bioconjugate Chem.* **28**, 913–917 (2017).
- 1004 38. Vohidov, F., Coughlin, J. M. & Ball, Z. T. Rhodium(II) Metallopeptide Catalyst Design
1005 Enables Fine Control in Selective Functionalization of Natural SH3 Domains. *Angew.*
1006 *Chem., Int. Ed.* **54**, 4587–4591 (2015).
- 1007 39. Ohata, J. & Ball, Z. T. A Hexa-rhodium Metallopeptide Catalyst for Site-Specific
1008 Functionalization of Natural Antibodies. *J. Am. Chem. Soc.* **139**, 12617–12622 (2017).
- 1009 40: Yu, C. *et al.* Proximity-Induced Site-Specific Antibody Conjugation. *Bioconjugate Chem.*
1010 **29**, 3522–3526 (2018).
- 1011 41. Adusumalli, S. R. *et al.* Single-Site Labeling of Native Proteins Enabled by a
1012 Chemoselective and Site-Selective Chemical Technology. *J. Am. Chem. Soc.* **140**, 15114-
1013 15123 (2018).
- 1014 42. Kuan, S. L., Wang, T. & Weil, T. Site-Selective Disulfide Modification of Proteins:
1015 Expanding Diversity beyond the Proteome. *Chem. - Eur. J.* **22**, 17112–17129 (2016).
- 1016 43. Martinez-Saez, N. *et al.* Oxetane Grafts Installed Site-Selectively on Native Disulfides to
1017 Enhance Protein Stability and Activity In Vivo. *Angew. Chem., Int. Ed.* **56**, 14963–14967
1018 (2017).
- 1019 44. Badescu, G. *et al.* Bridging Disulfides for Stable and Defined Antibody Drug Conjugates.
1020 *Bioconjugate Chem.* **25**, 1124–1136 (2014).
- 1021 45. Wang, T. *et al.* Water-soluble allyl sulfones for dual site-specific labelling of proteins and
1022 cyclic peptides. *Chem. Sci.* **7**, 3234–3239 (2016).
- 1023 46. Walsh, S. J. *et al.* A general approach for the site-selective modification of native proteins,
1024 enabling the generation of stable and functional antibody–drug conjugates. *Chem. Sci.*
1025 (2019).
- 1026 47. Lee, M. T. W. *et al.* Enabling the controlled assembly of antibody conjugates with a
1027 loading of two modules without antibody engineering. *Chem. Sci.* **8**, 2056–2060 (2017).
- 1028 48. Maruani, A., Smith, M. E. B., Miranda, E., Chester, K. A., Chudasama, V. & Caddick, S.
1029 A plug-and-play approach to antibody-based therapeutics via a chemoselective dual click
1030 strategy. *Nat. Commun.* **6**, 6645 (2015).
- 1031 49. Greene, M. K. *et al.* Forming next-generation antibody–nanoparticle conjugates through
1032 the oriented installation of non-engineered antibody fragments. *Chem. Sci.* **9**, 79–87
1033 (2018).
- 1034 50. Griebenow, N., Dilmaç, A. M., Greven, S. & Bräse, S. Site-Specific Conjugation of
1035 Peptides and Proteins via Rebridging of Disulfide Bonds Using the Thiol–Yne Coupling
1036 Reaction. *Bioconjugate Chem.* **27**, 911–917 (2016).

- 1037 51. Cal, P. M. S. D., Bernardes, G. J. L. G. J. L. & Gois, P. M. P. Cysteine-Selective
1038 Reactions for Antibody Conjugation. *Angew. Chem., Int. Ed.* **53**, 10585–10587 (2014).
- 1039 52. Lyon, R. P. *et al.* Self-hydrolyzing maleimides improve the stability and pharmacological
1040 properties of antibody-drug conjugates. *Nat. Biotechnol.* **32**, 1059–1062 (2014).
- 1041 53. Fontaine, S. D., Reid, R., Robinson, L., Ashley, G. W. & Santi, D. V. Long-Term
1042 Stabilization of Maleimide-Thiol Conjugates. *Bioconjugate Chem.* **26**, 145–152 (2015).
- 1043 54. Christie, R. *et al.* Stabilization of cysteine-linked antibody drug conjugates with N-aryl
1044 maleimides. *J. Control. Release* **220**, 660–670 (2015).
- 1045 55. Kalia, D., Malekar, P. V & Parthasarathy, M. Exocyclic Olefinic Maleimides: Synthesis
1046 and Application for Stable and Thiol-Selective Bioconjugation. *Angew. Chem., Int. Ed.*
1047 **55**, 1432–1435 (2016).
- 1048 56. Tumey, L. N. *et al.* Mild Method for Succinimide Hydrolysis on ADCs: Impact on ADC
1049 Potency, Stability, Exposure, and Efficacy. *Bioconjugate Chem.* **25**, 1871–1880 (2014).
- 1050 57. Morais, M. *et al.* Optimisation of the dibromomaleimide (DBM) platform for native
1051 antibody conjugation by accelerated post-conjugation hydrolysis. *Org. Biomol. Chem.* **15**,
1052 2947–2952 (2017).
- 1053 58. Kalia, D., Pawar, S. P. & Thopate, J. S. Stable and Rapid Thiol Bioconjugation by Light-
1054 Triggered Thiomaleimide Ring Hydrolysis. *Angew. Chem., Int. Ed.* **56**, 1885–1889 (2017).
- 1055 59. Boutureira, O. *et al.* Site-Selective Modification of Proteins with Oxetanes. *Chem. - Eur.*
1056 *J.* **23**, 6483–6489 (2017).
- 1057 60. Lee, B., Sun, S., Jiménez-moreno, E., Neves, A. A. & Bernardes, G. J. L. Site-selective
1058 installation of an electrophilic handle on proteins for bioconjugation. *Bioorg. Med. Chem.*
1059 **26**, 3060–3064 (2018).
- 1060 61. Kolodych, S. *et al.* CBTF: New Amine-to-Thiol Coupling Reagent for Preparation of
1061 Antibody Conjugates with Increased Plasma Stability. *Bioconjugate Chem.* **26**, 197–200
1062 (2015).
- 1063 62. Koniev, O. *et al.* Selective Irreversible Chemical Tagging of Cysteine with 3-
1064 Arylpropionitriles. *Bioconjugate Chem.* **25**, 202–206 (2014).
- 1065 63. Embaby, A. M., Schoffelen, S., Kofoed, C., Meldal, M. & Diness, F. Rational Tuning of
1066 Fluorobenzene Probes for Cysteine-Selective Protein Modification. *Angew. Chem., Int.*
1067 *Ed.* **57**, 8022–8026 (2018).
- 1068 64. Abbas, A., Xing, B. & Loh, T.-P. Allenamides as Orthogonal Handles for Selective
1069 Modification of Cysteine in Peptides and Proteins. *Angew. Chem., Int. Ed.* **53**, 7491–7494
1070 (2014).

- 1071 65. Smith, N. J. *et al.* Fast, irreversible modification of cysteines through strain releasing
1072 conjugate additions of cyclopropenyl ketones. *Org. Biomol. Chem.* **16**, 2164–2169 (2018).
- 1073 66. Canovas, C. *et al.* Site-Specific Dual Labeling of Proteins on Cysteine Residues with
1074 Chlorotetrazines. *Angew. Chem., Int. Ed.* **57**, 10646–10650 (2018).
- 1075 67. Ariyasu, S., Hayashi, H., Xing, B. & Chiba, S. Site-Specific Dual Functionalization of
1076 Cysteine Residue in Peptides and Proteins with 2-Azidoacrylates. *Bioconjugate Chem.* **28**,
1077 897–902 (2017).
- 1078 68. Vinogradova, E. V, Zhang, C., Spokoyny, A. M., Pentelute, B. L. & Buchwald, S. L.
1079 Organometallic palladium reagents for cysteine bioconjugation. *Nature* **526**, 687–691
1080 (2015).
- 1081 69. Vara, B. A. *et al.* Scalable thioarylation of unprotected peptides and biomolecules under
1082 Ni/photoredox catalysis. *Chem. Sci.* **9**, 336–344 (2018).
- 1083 70. Messina, M. S. *et al.* Organometallic Gold(III) Reagents for Cysteine Arylation. *J. Am.*
1084 *Chem. Soc.* **140**, 7065–7069 (2018).
- 1085 71. Rojas, A. J., Pentelute, B. L. & Buchwald, S. L. Water-Soluble Palladium Reagents for
1086 Cysteine S-Arylation under Ambient Aqueous Conditions. *Org. Lett.* **19**, 4263–4266
1087 (2017).
- 1088 72. Kubota, K., Dai, P., Pentelute, B. L. & Buchwald, S. L. Palladium Oxidative Addition
1089 Complexes for Peptide and Protein Cross-linking. *J. Am. Chem. Soc.* **140**, 3128–3133
1090 (2018).
- 1091 73. Zhang, Y. *et al.* Thiol Specific and Tracelessly Removable Bioconjugation via Michael
1092 Addition to 5-Methylene Pyrrolones. *J. Am. Chem. Soc.* **139**, 6146–6151 (2017).
- 1093 74. Bandyopadhyay, A. *et al.* Fast and selective labeling of N-terminal cysteines at neutral pH
1094 via thiazolidino boronate formation. *Chem. Sci.* **7**, 4589–4593 (2016).
- 1095 75. Arumugam, S. *et al.* Selective and reversible photochemical derivatization of cysteine
1096 residues in peptides and proteins. *Chem. Sci.* **5**, 1591–1598 (2014).
- 1097 76. Yu, J., Yang, X., Sun, Y. & Yin, Z. Highly Reactive and Tracelessly Cleavable Cysteine-
1098 Specific Modification of Proteins via 4-Substituted Cyclopentenone. *Angew. Chem., Int.*
1099 *Ed.* **57**, 11598–11602 (2018).
- 1100 77. Faustino, H., Silva, M. J. S. A., Veiros, L. F., Bernardes, G. J. L. & Gois, P. M. P.
1101 Iminoboronates are efficient intermediates for selective, rapid and reversible N-terminal
1102 cysteine functionalisation. *Chem. Sci.* **7**, 5052–5058 (2016).
- 1103 78. Lin, S. *et al.* Redox-based reagents for chemoselective methionine bioconjugation.
1104 *Science* **355**, 597–602 (2017).

- 1105 79. Seki, Y. *et al.* Transition Metal-Free Tryptophan-Selective Bioconjugation of Proteins. *J.*
1106 *Am. Chem. Soc.* **138**, 10798–10801 (2016).
- 1107 80. Wang, L., Gruzdy, V., Pang, N., Meng, F. & Sun, X.-L. Primary arylamine-based
1108 tyrosine-targeted protein modification. *RSC Adv.* **4**, 39446–39452 (2014).
- 1109 81. Wong, C.-Y. *et al.* A Ruthenium(II) Complex Supported by Trithiacyclononane and
1110 Aromatic Diimine Ligand as Luminescent Switch-On Probe for Biomolecule Detection
1111 and Protein Staining. *Sci. Rep.* **4**, 7136 (2014).
- 1112 82. Solomatina, A. I. *et al.* Coordination to Imidazole Ring Switches on Phosphorescence of
1113 Platinum Cyclometalated Complexes: The Route to Selective Labeling of Peptides and
1114 Proteins via Histidine Residues. *Bioconjugate Chem.* **28**, 426–437 (2017).
- 1115 83. Hansen, M. B., Hubálek, F., Skrydstrup, T. & Hoeg-Jensen, T. Chemo- and
1116 Regioselective Ethynylation of Tryptophan-Containing Peptides and Proteins. *Chem. -*
1117 *Eur. J.* **22**, 1572–1576 (2015).
- 1118 84. Sato, S., Nakamura, K. & Nakamura, H. Tyrosine-Specific Chemical Modification with in
1119 Situ Hemin-Activated Luminol Derivatives. *ACS Chem. Biol.* **10**, 2633–2640 (2015).
- 1120 85. Ban, H., Gavrilyuk, J. & Barbas, C. F. Tyrosine Bioconjugation through Aqueous Ene-
1121 Type Reactions: A Click-Like Reaction for Tyrosine. *J. Am. Chem. Soc.* **132**, 1523–1525
1122 (2010).
- 1123 80. Ohata, J., Miller, M. K., Mountain, C. M., Vohidov, F. & Ball, Z. T. A Three-Component
1124 Organometallic Tyrosine Bioconjugation. *Angew. Chem., Int. Ed.* **57**, 2827–2830 (2018).
- 1125 87. Taylor, M. T., Nelson, J. E., Suero, M. G. & Gaunt, M. J. A protein functionalization
1126 platform based on selective reactions at methionine residues. *Nature* **562**, 563–568 (2018).
- 1127 88. Cohen, D. T. *et al.* A chemoselective strategy for late-stage functionalization of complex
1128 small molecules with polypeptides and proteins. *Nat. Chem.* (2018).
- 1129 89. Cui, Z. *et al.* Combining Sense and Nonsense Codon Reassignment for Site-Selective
1130 Protein Modification with Unnatural Amino Acids. *ACS Synth. Biol.* **6**, 535–544 (2017).
- 1131 90. Xie, J. & Schultz, P. G. A chemical toolkit for proteins — an expanded genetic code. *Nat.*
1132 *Rev. Mol. Cell Biol.* **7**, 775 (2006).
- 1133 91. Haney, C. M., Wissner, R. F. & Petersson, E. J. Multiply labeling proteins for studies of
1134 folding and stability. *Curr. Opin. Chem. Biol.* **28**, 123–130 (2015).
- 1135 92. Wang, L., Brock, A., Herberich, B. & Schultz, P. G. Expanding the Genetic Code of
1136 *Escherichia coli*. *Science* **292**, 498 (2001).
- 1137 93. Lang, K. & Chin, J. W. Cellular Incorporation of Unnatural Amino Acids and
1138 Bioorthogonal Labeling of Proteins. *Chem. Rev.* **114**, 4764–4806 (2014).

- 1139 94. Lee, B. S. *et al.* Incorporation of Unnatural Amino Acids in Response to the AGG Codon.
1140 *ACS Chem. Biol.* **10**, 1648–1653 (2015).
- 1141 95. Iwane, Y. *et al.* Expanding the amino acid repertoire of ribosomal polypeptide synthesis
1142 via the artificial division of codon boxes. *Nat. Chem.* **8**, 317 (2016).
- 1143 96. Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables
1144 spontaneous protein double-labelling and FRET. *Nat. Chem.* **6**, 393 (2014).
- 1145 97. Almhjell, P. J., Boville, C. E. & Arnold, F. H. Engineering enzymes for noncanonical
1146 amino acid synthesis. *Chem. Soc. Rev.* (2018).
- 1147 98. Schmidt, M. J. & Summerer, D. Genetic code expansion as a tool to study regulatory
1148 processes of transcription. *Front. Chem.* **2**, 7 (2014).
- 1149 99. Ovaa, H. & wals, kim. Unnatural amino acid incorporation in E. coli: current and future
1150 applications in the design of therapeutic proteins. *Front. Chem.* **2**, 15 (2014).
- 1151 100. Nadal, S., Raj, R., Mohammed, S. & Davis, B. G. Synthetic post-translational
1152 modification of histones. *Curr. Opin. Chem. Biol.* **45**, 35–47 (2018).
- 1153 101. Xiao, H., Xuan, W., Shao, S., Liu, T. & Schultz, P. G. Genetic incorporation of ϵ -N-2-
1154 hydroxyisobutyryl-lysine into recombinant histones. *ACS Chem. Biol.* **10**, 1599–1603
1155 (2015).
- 1156 102. Hoppmann, C. *et al.* Site-specific incorporation of phosphotyrosine using an expanded
1157 genetic code. *Nat. Chem. Biol.* **13**, 842 (2017).
- 1158 103. Wang, Z. A. *et al.* A Genetically Encoded Allysine for the Synthesis of Proteins with Site-
1159 Specific Lysine Dimethylation. *Angew. Chem., Int. Ed.* **56**, 212–216 (2016).
- 1160 104. Xuan, W., Shao, S. & Schultz, P. G. Protein Crosslinking by Genetically Encoded
1161 Noncanonical Amino Acids with Reactive Aryl Carbamate Side Chains. *Angew. Chem.,
1162 Int. Ed.* **56**, 5096–5100 (2017).
- 1163 105. Xuan, W., Li, J., Luo, X. & Schultz, P. G. Genetic Incorporation of a Reactive
1164 Isothiocyanate Group into Proteins. *Angew. Chem., Int. Ed.* **55**, 10065–10068 (2016).
- 1165 106. Xuan, W. *et al.* Site-Specific Incorporation of a Thioester Containing Amino Acid into
1166 Proteins. *ACS Chem. Biol.* **13**, 578–581 (2018).
- 1167 107. Tian, Y. *et al.* Genetically Encoded 2-Aryl-5-carboxytetrazoles for Site-Selective Protein
1168 Photo-Cross-Linking. *J. Am. Chem. Soc.* **139**, 6078–6081 (2017).
- 1169 108. Yamaguchi, A. *et al.* Incorporation of a Doubly Functionalized Synthetic Amino Acid into
1170 Proteins for Creating Chemical and Light-Induced Conjugates. *Bioconjugate Chem.* **27**,
1171 198–206 (2016).

- 1172 109. Hoppmann, C. *et al.* Genetically Encoding Photoswitchable Click Amino Acids in
1173 Escherichia coli and Mammalian Cells. *Angew. Chem., Int. Ed.* **53**, 3932–3936 (2014).
- 1174 110. Rashidian, M., Dozier, J. K. & Distefano, M. D. Enzymatic Labeling of Proteins:
1175 Techniques and Approaches. *Bioconjugate Chem.* **24**, 1277–1294 (2013).
- 1176 111. Zhang, Y., Park, K.-Y., Suazo, K. F. & Distefano, M. D. Recent progress in enzymatic
1177 protein labelling techniques and their applications. *Chem. Soc. Rev.* (2018) Advance
1178 Article.
- 1179 112. Li, C. & Wang, L.-X. Chemoenzymatic Methods for the Synthesis of Glycoproteins.
1180 *Chem. Rev.* **118**, 8359-8413 (2018).
- 1181 113. Plaks, J. G., Falatach, R., Kastantin, M., Berberich, J. A. & Kaar, J. L. Multisite Clickable
1182 Modification of Proteins Using Lipoic Acid Ligase. *Bioconjugate Chem.* **26**, 1104–1112
1183 (2015).
- 1184 114. Schumacher, D. *et al.* Versatile and Efficient Site-Specific Protein Functionalization by
1185 Tubulin Tyrosine Ligase. *Angew. Chem., Int. Ed.* **54**, 13787–13791 (2015).
- 1186 115. Patterson, J. T. *et al.* Human Serum Albumin Domain I Fusion Protein for Antibody
1187 Conjugation. *Bioconjugate Chem.* **27**, 2271–2275 (2016).
- 1188 116. Sunbul, M., Nacheva, L. & Jäschke, A. Proximity-Induced Covalent Labeling of Proteins
1189 with a Reactive Fluorophore-Binding Peptide Tag. *Bioconjugate Chem.* **26**, 1466–1469
1190 (2015).
- 1191 117. Meyer, C., Liebscher, S. & Bordusa, F. Selective Coupling of Click Anchors to Proteins
1192 via Trypsin Ligase. *Bioconjugate Chem.* **27**, 47–53 (2016).
- 1193 118. Liebscher, S. *et al.* N-Terminal Protein Modification by Substrate-Activated Reverse
1194 Proteolysis. *Angew. Chem., Int. Ed.* **53**, 3024–3028 (2014).
- 1195 119. Zhang, C., Dai, P., Vinogradov, A. A., Gates, Z. P. & Pentelute, B. L. Site-Selective
1196 Cysteine – Cyclooctyne Conjugation. *Angew. Chem., Int. Ed.* **57**, 6459–6463 (2018).
- 1197 120. Ramil, C. P., An, P., Yu, Z. & Lin, Q. Sequence-Specific 2-Cyanobenzothiazole Ligation.
1198 *J. Am. Chem. Soc.* **138**, 5499–5502 (2016).
- 1199 121. Agrawalla, B. K. *et al.* Chemoselective Dual Labeling of Native and Recombinant
1200 Proteins. *Bioconjugate Chem.* **29**, 29-34 (2018).
- 1201 122. Bellucci, J. J., Bhattacharyya, J. & Chilkoti, A. A Noncanonical Function of Sortase
1202 Enables Site-Specific Conjugation of Small Molecules to Lysine Residues in Proteins.
1203 *Angew. Chem., Int. Ed.* **54**, 441–445 (2015).
- 1204 123. Peciak, K., Laurine, E., Tommasi, R., Choi, J.-w. & Brocchini, S. Site-selective protein
1205 conjugation at histidine. *Chem. Sci.* (2019) Accepted Author Manuscript.

- 1206 124. Grünewald, J. *et al.* Optimization of an Enzymatic Antibody–Drug Conjugation Approach
1207 Based on Coenzyme A Analogs. *Bioconjugate Chem.* **28**, 1906–1915 (2017).
- 1208 125. Grünewald, J. *et al.* Efficient Preparation of Site-Specific Antibody–Drug Conjugates
1209 Using Phosphopantetheinyl Transferases. *Bioconjugate Chem.* **26**, 2554–2562 (2015).
- 1210 126. Montanari, E. *et al.* Tyrosinase-Mediated Bioconjugation. A Versatile Approach to
1211 Chimeric Macromolecules. *Bioconjugate Chem.* **29**, 2550–2560 (2018).
- 1212 127. Siegmund, V. *et al.* Locked by Design: A Conformationally Constrained
1213 Transglutaminase Tag Enables Efficient Site-Specific Conjugation. *Angew. Chem., Int.*
1214 *Ed.* **54**, 13420–13424 (2015).
- 1215 128. Wang, H. H., Altun, B., Nwe, K. & Tsourkas, A. Proximity-Based Sortase-Mediated
1216 Ligation. *Angew. Chem., Int. Ed.* **56**, 5349–5352 (2017).
- 1217 129. Struck, A.-W. *et al.* An Enzyme Cascade for Selective Modification of Tyrosine Residues
1218 in Structurally Diverse Peptides and Proteins. *J. Am. Chem. Soc.* **138**, 3038–3045 (2016).
- 1219 130. Zhang, Y. *et al.* Simultaneous Site-Specific Dual Protein Labeling Using Protein
1220 Prenyltransferases. *Bioconjugate Chem.* **26**, 2542–2553 (2015).
- 1221 131. King, M. & Wagner, A. Developments in the Field of Bioorthogonal Bond Forming
1222 Reactions—Past and Present Trends. *Bioconjugate Chem.* **25**, 825–839 (2014).
- 1223 132. Haldon, E., Nicasio, M. C. & Perez, P. J. Copper-catalysed azide-alkyne cycloadditions
1224 (CuAAC): an update. *Org. Biomol. Chem.* **13**, 9528–9550 (2015).
- 1225 133. Finbloom, J. A., Han, K., Slack, C. C., Furst, A. L. & Francis, M. B. Cucurbit[6]uril-
1226 Promoted Click Chemistry for Protein Modification. *J. Am. Chem. Soc.* **139**, 9691–9697
1227 (2017).
- 1228 134. Yang, M., Li, J. & Chen, P. R. Transition metal-mediated bioorthogonal protein chemistry
1229 in living cells. *Chem. Soc. Rev.* **43**, 6511–6526 (2014).
- 1230 135. Wallace, S. & Chin, J. W. Strain-promoted sydnone bicyclo[6.1.0]nonyne cycloaddition.
1231 *Chem. Sci.* **5**, 1742–1744 (2014).
- 1232 136. Darko, A. *et al.* Conformationally strained trans-cyclooctene with improved stability and
1233 excellent reactivity in tetrazine ligation. *Chem. Sci.* **5**, 3770–3776 (2014).
- 1234 137. Kamber, D. N. *et al.* 1,2,4-Triazines Are Versatile Bioorthogonal Reagents. *J. Am. Chem.*
1235 *Soc.* **137**, 8388–8391 (2015).
- 1236 138. Borrmann, A. *et al.* Strain-Promoted Oxidation-Controlled Cyclooctyne–1,2-Quinone
1237 Cycloaddition (SPOCQ) for Fast and Activatable Protein Conjugation. *Bioconjugate*
1238 *Chem.* **26**, 257–261 (2015).

- 1239 139. Lampkowski, J. S., Villa, J. K., Young, T. S. & Young, D. D. Development and
1240 Optimization of Glaser–Hay Bioconjugations. *Angew. Chem., Int. Ed.* **54**, 9343–9346
1241 (2015).
- 1242 140. Kwan, T. T. L. *et al.* Protein modification via alkyne hydrosilylation using a
1243 substoichiometric amount of ruthenium(II) catalyst. *Chem. Sci.* **8**, 3871–3878 (2017).
- 1244 141. Kölmel, D. K. & Kool, E. T. Oximes and Hydrazones in Bioconjugation: Mechanism and
1245 Catalysis. *Chem. Rev.* **117**, 10358–10376 (2017).
- 1246 142. Spears, R. J. *et al.* Site-selective C–C modification of proteins at neutral pH using
1247 organocatalyst-mediated cross aldol ligations. *Chem. Sci.* **9**, 5585–5593 (2018).
- 1248 143. Kudirka, R. *et al.* Generating site-specifically modified proteins via a versatile and stable
1249 nucleophilic carbon ligation. *Chem. Biol.* **22**, 293–298 (2015).
- 1250 144. Kudirka, R. A. *et al.* Site-Specific Tandem Knoevenagel Condensation-Michael Addition
1251 to Generate Antibody-Drug Conjugates. *ACS Med. Chem. Lett.* **7**, 994–998 (2016).
- 1252 145. Wang, P. *et al.* Site-Specific Chemical Modification of Peptide and Protein by
1253 Thiazolidinediones. *Org. Lett.* **17**, 1361–1364 (2015).
- 1254 146. Lee, Y.-J., Kurra, Y. & Liu, W. R. Phospha-Michael Addition as a New Click Reaction
1255 for Protein Functionalization. *ChemBioChem* **17**, 456–461 (2016).
- 1256 147. Tomlin, F. M. *et al.* Site-specific incorporation of quadricyclane into a protein and
1257 photocleavage of the quadricyclane ligation adduct. *Bioorg. Med. Chem.* (2018).
1258 doi:<https://doi.org/10.1016/j.bmc.2018.04.009>
- 1259 148. Row, R. D., Shih, H.-W., Alexander, A. T., Mehl, R. A. & Prescher, J. A.
1260 Cyclopropanones for Metabolic Targeting and Sequential Bioorthogonal Labeling. *J. Am.*
1261 *Chem. Soc.* **139**, 7370–7375 (2017).
- 1262 149. Shih, H.-W. & Prescher, J. A. A Bioorthogonal Ligation of Cyclopropanones Mediated by
1263 Triarylphosphines. *J. Am. Chem. Soc.* **137**, 10036–10039 (2015).
- 1264 150. Bernardes, G. J. L., Chalker, J. M., Errey, J. C. & Davis, B. G. Facile Conversion of
1265 Cysteine and Alkyl Cysteines to Dehydroalanine on Protein Surfaces: Versatile and
1266 Switchable Access to Functionalized Proteins. *J. Am. Chem. Soc.* **130**, 5052–5053 (2008).
- 1267 151. Wright, T. H. *et al.* Posttranslational mutagenesis: A chemical strategy for exploring
1268 protein side-chain diversity. *Science* **354**, 597 (2016).
- 1269 152. Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nat. Commun.* **5**,
1270 4740 (2014).
- 1271 153. Freedy, A. M. *et al.* Chemoselective Installation of Amine Bonds on Proteins through
1272 Aza-Michael Ligation. *J. Am. Chem. Soc.* **139**, 18365–18375 (2017).

- 1273 154. Dadova, J. *et al.* Precise probing of residue roles by post-translational β,γ -C,N aza-
1274 Michael mutagenesis in enzyme active sites. *ACS Cent. Sci.* **3**, 1168–1173 (2017).
- 1275 155. Yang, A. *et al.* A chemical biology route to site-specific authentic protein modifications.
1276 *Science* **354**, 623–626 (2016).
- 1277 156. Galan, S. R. G. *et al.* Post-translational site-selective protein backbone α -deuteration. *Nat.*
1278 *Chem. Biol.* **14**, 955-963 (2018).
- 1279 157. Addy, P. S., Erickson, S. B., Italia, J. S. & Chatterjee, A. A Chemoselective Rapid Azo-
1280 Coupling Reaction (CRACR) for Unclickable Bioconjugation. *J. Am. Chem. Soc.* **139**,
1281 11670–11673 (2017).
- 1282 158. Zegota, M. M., Wang, T., Seidler, C., Wah Ng, D. Y., Kuan, S. L. & Weil, T. “Tag and
1283 Modify” Protein Conjugation with Dynamic Covalent Chemistry. *Bioconjugate Chem.* **29**,
1284 2665-2670 (2018).
- 1285 159. Row, R. D. & Prescher, J. A. Constructing New Bioorthogonal Reagents and Reactions.
1286 *Acc. Chem. Res.* **51**, 1073-1081 (2018).
- 1287 160. Beck, A., Goetsch, L., Dumontet, C. & Corvaia, N. Strategies and challenges for the next
1288 generation of antibody-drug conjugates. *Nat. Rev. Drug Discov.* **16**, 315–337 (2017).
- 1289 161. Chudasama, V., Maruani, A. & Caddick, S. Recent advances in the construction of
1290 antibody-drug conjugates. *Nat. Chem.* **8**, 114–119 (2016).
- 1291 162. Akkapeddi, P. *et al.* Construction of homogeneous antibody-drug conjugates using site-
1292 selective protein chemistry. *Chem. Sci.* **7**, 2954–2963 (2016).
- 1293 163. Kelkar, S. S. & Reineke, T. M. Theranostics: Combining Imaging and Therapy.
1294 *Bioconjugate Chem.* **22**, 1879–1903 (2011).
- 1295 164. Li, X. *et al.* Site-Specific Dual Antibody Conjugation via Engineered Cysteine and
1296 Selenocysteine Residues. *Bioconjugate Chem.* **26**, 2243–2248 (2015).
- 1297 165. Pelegri-O’Day, E. M., Lin, E.-W. & Maynard, H. D. Therapeutic Protein–Polymer
1298 Conjugates: Advancing Beyond PEGylation. *J. Am. Chem. Soc.* **136**, 14323–14332
1299 (2014).
- 1300 166. Cini, E. *et al.* Antibody drug conjugates (ADCs) charged with HDAC inhibitor for
1301 targeted epigenetic modulation. *Chem. Sci.* **9**, 6490–6496 (2018).
- 1302 167. Lee, B.-C. *et al.* FRET Reagent Reveals the Intracellular Processing of Peptide-Linked
1303 Antibody–Drug Conjugates. *Bioconjugate Chem.* **29**, 2468–2477 (2018).
- 1304 168. Wang, R. E. *et al.* An Immunosuppressive Antibody-dasatinib conjugate and its Lck
1305 inhibitory activity. *J. Am. Chem. Soc.* **137**, 3229–3232 (2015).

- 1306 169. Lehar, S. M. *et al.* Novel antibody-antibiotic conjugate eliminates intracellular *S. aureus*.
1307 *Nature* **527**, 323–328 (2015).
- 1308 170. Balintová, J., Welter, M. & Marx, A. Antibody–nucleotide conjugate as a substrate for
1309 DNA polymerases. *Chem. Sci.* **9**, 7122–7125 (2018).
- 1310 171. Adumeau, P., Davydova, M. & Zeglis, B. M. Thiol-Reactive Bifunctional Chelators for
1311 the Creation of Site-Selectively Modified Radioimmunoconjugates with Improved
1312 Stability. *Bioconjugate Chem.* **29**, 1364–1372 (2018).
- 1313 172. Lerchen, H.-G. *et al.* Antibody–Drug Conjugates with Pyrrole-Based KSP Inhibitors as
1314 the Payload Class. *Angew. Chem., Int. Ed.* **57**, 15243–15247 (2018).
- 1315 173. Kelemen, R. E. *et al.* A Precise Chemical Strategy To Alter the Receptor Specificity of the
1316 Adeno-Associated Virus. *Angew. Chem., Int. Ed.* **55**, 10645–10649 (2016).
- 1317 174. Greineder, C. F. *et al.* Site-Specific Modification of Single-Chain Antibody Fragments for
1318 Bioconjugation and Vascular Immunotargeting. *Bioconjugate Chem.* **29**, 56–66 (2018).
- 1319 175. Wu, T. *et al.* Transglutaminase mediated PEGylation of nanobodies for targeted nano-
1320 drug delivery. *J. Mater. Chem. B* **6**, 1011–1017 (2018).
- 1321 176. Brasino, M. *et al.* Anti-EGFR Affibodies with Site-Specific Photo-Cross-Linker
1322 Incorporation Show Both Directed Target-Specific Photoconjugation and Increased
1323 Retention in Tumors. *J. Am. Chem. Soc.* **140**, 11820–11828 (2018).
- 1324 177. Kuan, S. L. *et al.* Boosting Antitumor Drug Efficacy with Chemically Engineered
1325 Multidomain Proteins. *Adv. Sci.* **5**, 1701036–1701036 (2018).
- 1326 178. Jones, L. H. Recent advances in the molecular design of synthetic vaccines. *Nat. Chem.* **7**,
1327 952 (2015).
- 1328 179. Schoonen, L., Pille, J., Borrmann, A., Nolte, R. J. M. & van Hest, J. C. M. Sortase A-
1329 Mediated N-Terminal Modification of Cowpea Chlorotic Mottle Virus for Highly
1330 Efficient Cargo Loading. *Bioconjugate Chem.* **26**, 2429–2434 (2015).
- 1331 180. Kobayashi, T., Hoppmann, C., Yang, B. & Wang, L. Using Protein-Confined Proximity
1332 To Determine Chemical Reactivity. *J. Am. Chem. Soc.* **138**, 14832–14835 (2016).
- 1333 181. Strumillo, M. & Beltrao, P. Towards the computational design of protein post-
1334 translational regulation. *Bioorg. Med. Chem.* **23**, 2877–2882 (2015).
- 1335 182. Wagner, J. A., Mercadante, D., Nikić, I., Lemke, E. A. & Gräter, F. Origin of
1336 Orthogonality of Strain-Promoted Click Reactions. *Chem. - Eur. J.* **21**, 12431–12435
1337 (2015).
- 1338 183. Narayanam, M. K., Liang, Y., Houk, K. N. & Murphy, J. M. Discovery of new mutually
1339 orthogonal bioorthogonal cycloaddition pairs through computational screening. *Chem. Sci.*

- 1340 7, 1257–1261 (2016).
- 1341 184. Knall, A.-C., Hollauf, M., Saf, R. & Slugovc, C. A trifunctional linker suitable for
1342 conducting three orthogonal click chemistries in one pot. *Org. Biomol. Chem.* **14**, 10576–
1343 10580 (2016).
- 1344 185. Bryden, F. *et al.* Assembly of High-Potency Photosensitizer–Antibody Conjugates
1345 through Application of Dendron Multiplier Technology. *Bioconjugate Chem.* **29**, 176–181
1346 (2018).
- 1347 186. Cazzamalli, S., Dal Corso, A., Widmayer, F. & Neri, D. Chemically Defined Antibody–
1348 and Small Molecule–Drug Conjugates for in Vivo Tumor Targeting Applications: A
1349 Comparative Analysis. *J. Am. Chem. Soc.* **140**, 1617–1621 (2018).

1350

1351 **Acknowledgements**

1352 We thank FCT Portugal (FCT Investigator to G.J.L.B., IF/00624/2015 and Postdoctoral
1353 Fellowship to P.M.S.D.C.), the Herchel Smith Fund (PhD Studentship to E.A.H.) and the EU
1354 Horizon 2020 programme (Marie Skłodowska-Curie ITN grant agreement No. 675007 to G.J.L.B.
1355 and Marie Skłodowska-Curie IEF grant agreement No. 702574 to B.L.O.) for funding. G.J.L.B. is
1356 a Royal Society University Research Fellow (UF110046 and URF\R\180019) and the recipient of
1357 a European Research Council Starting Grant (TagIt, grant agreement No. 676832).

1358

1359 **Author contributions**

1360 All authors made substantial contributions to the discussion and organization of the content as well
1361 as reviewed and/or edited the manuscript before submission. Additionally, E.A.H. conducted the
1362 research, wrote the main body of the paper, edited the figures, and put the complete manuscript
1363 together; P.M.S.D.C. played an integral role in writing the introduction, researching, and providing
1364 content guidance at all stages of manuscript preparation; and B.L.O. designed and created the
1365 figures. G.J.L.B. coordinated the research and writing.

1366

1367 **Competing interests**

1368 The authors declare no competing interests.

1369

1370 **Publisher's note**

1371

1372 Springer Nature remains neutral with regard to jurisdictional claims in published maps and
1373 institutional affiliations.

1374

1375

1376

1377 **Figure captions**

1378

1379 **Figure 1 | Juxtaposition of classical and modern protein modification methods.** On the left:
1380 classical methods focused primarily on the modification of Cys and Lys sidechains. The
1381 reactions most commonly focused on and depicted here include: thiol-exchange (i), alkylation of
1382 α -halocarbonyl electrophiles (ii), maleimide Michael addition (iii), NHS-ester amidation (iv),
1383 isothiocyanate or isocyanate addition (v), and reductive amination (vi).^{6,7} On the right: modern
1384 methods focused on the improvement of selectivity, reaction efficiency, and generality of
1385 application. The general modification classifications discussed in this review, as shown by
1386 specific examples, are represented: native protein/endogenous AA sidechain modification,
1387 engineered canonical AA insertion, engineered ncAA insertion, and motif/tag insertion.¹³⁻¹⁶
1388

1389 **Figure 2 | N-/C-terminal selective protein modification.** Above: C-terminal modification
1390 based on a decarboxylative strategy facilitated by a photocatalyst (i).¹⁶ Below: N-terminal
1391 modification techniques based on oxidative (ii), direct (iii), reductive (iv), or enzymatic type
1392 reactions (v).^{24-26,29} To mention a few more specific points on each reaction: (i) While not tested
1393 with glycine, phenylalanine, or proline as C-terminal residues, desired products were observed
1394 for all other AAs with lower yields for histidine (His), tyrosine (Tyr), and lysine (Lys) terminal
1395 AAs.¹⁶ (ii) Best with proline in the terminal position (only residue allowing for high yields at the
1396 protein level) and not performed with cysteine in the terminal position unless only cysteine in
1397 protein sequence as method modifies cysteine residues regardless of terminal position.²⁶ (iii)
1398 This method has been exhibited with all AAs in terminal positions but cannot be performed on
1399 proteins with proline in the second position.²⁵ (iv) With the exception of cysteine as the terminal
1400 residue due to thiazolidine side products, all types of terminal AAs achieve high yields.²⁴ (v)
1401 Butelase 1 can interact with a variety of terminal AA types aside from proline and prefers the
1402 second position to be either, isoleucine, valine, leucine, or cysteine. Most of the terminal AA
1403 type compatibility types were performed on peptide platforms.²⁹
1404

1405 **Figure 3 | In-chain endogenous sidechain modification.** Methods for endogenous sidechain
1406 modification of in-chain residues. The recent discoveries and reports of successful strategies
1407 have followed three different trends: modification based on selection of conditions and reagents
1408 to target the most reactive instance of a repeated sidechain, site-selective modification based on
1409 the direction of metals and ligands to native binding sites, and modification via disulfide
1410 rebridging. Each trend has several valuable examples that have been established in the last five
1411 years and are represented. Reagent- or condition-based targeting of the most reactive instance of
1412 an AA: selective trifluoromethylation of tryptophan residues (i), sulfonyl acrylate modification
1413 of most reactive lysine (ii), a three-component reaction for the modification of a single lysine
1414 (iii).^{12,33,34} Ligand- or metal-binding site directed methods: Cys arylation based on proximity to
1415 Asp-regulated binding site (iv), selective modification of Tyr residues proximal to the SH3
1416 binding domain (v), diazotransfer to lysines proximal to ligand site (vi), metallopeptide targeted
1417 Asp modification in antibodies using Fc-binding peptides (vii).^{19,32,37,39} Disulfide rebridging: by
1418 oxetanes (viii), by water-soluble allyl sulfones (ix), by dibromide-based moieties (x), and by
1419 thiol-yne coupling (xi).^{43,45-50} The colored highlights represent different reactive handles or
1420 functionalities added to the protein by the conjugation reactions.
1421

1422 **Figure 4 | Modification methods for engineered Cys residues.** Out of all of the canonical AAs
1423 for the installation of chemical handles in proteins, Cys residues have the broadest reactivity
1424 profile as represented by these examples. The methods have been separated into three classifying
1425 groups: metal-free, metal-assisted, and reversible type reactions. As the eventual application for
1426 the bioconjugate determines the type of chemistry linking the added functionality and the
1427 protein, these three categories all represent different mechanistic strategies that would allow
1428 these methods to be useful under various circumstances (e.g. cleavable linkers for ADCs to
1429 prevent reliance on release of the attached drug by native cell processes). Metal-free methods:
1430 addition of alkyl bromide electrophilic handles via an isobutylene and oxetane type chemical
1431 handle (i and ii), amine functionalization followed by addition of CBTF (iii), S-arylation by
1432 fluorobenzene derivatives (iv), addition of carbonylacrylic reagents (v), allenamide addition (vi),
1433 cyclopropenyl ketone addition (vii), chlorotetrazine addition (viii), and 2-azidoacrylate addition
1434 (ix).^{15,59–61,63–67} Metal-assisted methods: S-arylation by way of Au(III) (x) or Pd(II) catalyst
1435 (xi).^{70,72} Reversible methods: addition of 5MP derivatives (xii), addition of 4-acetoxy
1436 cyclopentenone (xiii), 2-FPBA addition at the N-terminus (xiv), and the addition of NQMPs
1437 (xv).^{73–76}
1438

1439 **Figure 5 | Insertion of canonical amino acids aside from Cys.** Site-selective protein
1440 modification based on the genetic insertion of canonical AAs aside from Cys. Forming unique
1441 reactive handles when inserted due to low abundance or being unlikely to be expressed
1442 endogenously in a solvent accessible position, these inserted AAs (Trp, Tyr, His, and Met) allow
1443 modifications that will lead to products with higher levels of homogeneity. Trp modification: a
1444 metal-assisted reaction with TIPS-EBX to install a protected alkyne reactive handle
1445 regioselectively at the C2 position on the indole rings of installed Trp residues (i);⁸³ using an
1446 organoradical-driven mechanism, keto-ABNO derivatives add to the indole ring of Trp residues
1447 with elevated conversion in acidic environments (ii);⁷⁹ in the presence of H₂O₂ and hemin, luminol
1448 derivatives are added to Tyr *ortho* positions (iii);⁸⁴ Rh(III)-mediated reaction for the addition of
1449 arene complexes to Tyr *ortho* positions (iv).⁸⁶ Pt(II)- and Ru(II)-driven selective complexation
1450 with His residues (v and vi);^{81,82} and a metal-free, redox-based reaction targeting Met residues with
1451 oxaziridine derivatives (vii).⁷⁸
1452

1453 **Figure 6 | Insertion of ncAAs.** The most common method and novel chemistries made available
1454 by ncAA insertion over the last five years: a) Depiction of the groundbreaking use of orthogonal
1455 aminoacyl tRNA synthetase/tRNA pairs for the insertion of ncAAs.^{90–93} 1 - Binding of tRNA and
1456 ncAA to aminoacyl tRNA synthetase (aaRS). 2 - Attachment of ncAA to tRNA by aaRS. 3 –
1457 Recognition of the amber codon by the ncAA-equipped tRNA. 4 – Incorporation of ncAA into
1458 the protein sequence by a native ribosome. b) examples of ncAAs that have been synthesized and
1459 inserted into proteins for the first time over the last five years. The colored highlights on the
1460 ncAA examples represent different chemical handles or functionalities added to the protein when
1461 the respective ncAAs are inserted. The red-highlight for HibK is to signify the PTM that this
1462 ncAA directly inserts.¹⁰¹ Orange highlights represent protection groups that must be removed to
1463 reveal a phosphorylation PTM or an allysine residue that can be hydrolyzed to an aldehyde
1464 functionality (pTyr and AcdK respectively).^{102,103} Blue highlights shows the photo-reactive
1465 moieties that allow for further functionalization of the protein or crosslinking.^{13,107–109} While the
1466 green highlights different electrophilic handles incorporated by these ncAAs, the purple highlight
1467 signifies a nucleophilic site for subsequent reactions.^{104–106,108,109}

1468
1469 **Figure 7 | Motif and enzymatic tag insertion.** Rather than the insertion of single residues and
1470 reliance only on the chemistry of the added functional group to drive selectivity, the insertion of
1471 tags allows for manipulation of the microenvironment around specific AAs. Such manipulation
1472 can lead to elevated reactivity of the targeted AA or to the enzymatic recognition of the inserted
1473 tag. In either case, the site-specific modification of the targeted residue within the motif or tag
1474 occurs. Ideal motif insertion methods allow for site-specific modification at either in-chain or
1475 terminal sites as well as cause minimal disruption of the protein structure, even with multiple
1476 instances of the tag inserted. Methods discussed here: fluorett fusion for fluorophore
1477 functionalization (i), modification mediated by TTL (ii), modification mediated by trypsiligase
1478 (iii), DBCO tag for Cys modification (iv), tag for CBT modification of Cys (v), π -clamp for Cys
1479 modification (vi), Lys activation by HSAdI (vii), noncanonical function of sortase A allows for
1480 Lys-specific modification of inserted pilin domain (viii).^{14,114–116,118–122}

1481
1482 **Figure 8 | Downstream functionalization methods.** Clean and efficient modification of unique,
1483 bioorthogonal chemical handles. A general scheme of the concept of “click chemistry” is included
1484 at the top of the diagram.^{131–133} The methods highlighted here fall under four classifications:
1485 alkyne-based reactions, aldehyde-based reactions, methods aside from alkyne- and aldehyde-based
1486 reactions, and Dha functionalization. Alkyne-based reactions: SPOCQ cycloaddition (i), phenyl
1487 sydnone [3+2] cycloaddition with BCN (ii), Glaser-Hay coupling (iii), Ru(II)-catalyzed alkyne
1488 hydrosilylation (iv).^{135,138–140} Aldehyde-based reactions: trapped Knoevenagel-type condensation
1489 (v), tandem Knoevenagel-type condensation (vi), aldol reaction with 2,4-thiazolidinediones and
1490 an N-terminal aldehyde (vii), OPAL with an inserted aldehyde functionality (viii).^{142–145} Methods
1491 aside from alkyne- and aldehyde-based strategies: phospho-Michael addition (ix), QC ligation (x),
1492 triarylphosphine-mediated addition to cyclopropanone (xi).^{146–148} Methods for Dha modification:
1493 aza-Michael addition to Dha (xii), radical-based reactions for the formation of C–C bonds with
1494 Dha (xiii).^{151,153–155}

1495
1496
1497

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); <i>o</i> -aminophenols; 2-PCA; aldehydes; thiopeptides (mediated by butelase 1)	<p>Advantages:</p> <ul style="list-style-type: none"> -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 <p>Limitations:</p> <ul style="list-style-type: none"> -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	<p>Reactive-residue targeting: Sodium trifluoromethanesulfinate (Trp modification); sulfonyl acrylates (Lys modification); multicomponent reaction with aldehydes, alkynes, and copper(I) iodide (Lys modification)</p> <p>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</p>	<p>Advantages:</p> <ul style="list-style-type: none"> -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 <p>Limitations:</p> <ul style="list-style-type: none"> -Necessary to either have a distinct microenvironment to enhance reactivity of a specific residue or have a native metal- or ligand-

		<p>domain-binding peptides; antibody heavy chain junction-binding protein targeted modification of lysine; LDM method for modification of His</p> <p>Disulfide rebridging: oxetanes; allyl sulfones; DVP; dibromo-PBD derivatives; alkynes for photomediated thiol-yne reactions</p>	<p>binding site as well as a proximal reactive residue</p> <ul style="list-style-type: none"> -Smaller size needed for disulfide rebridging reagents -Little control over choice of modification site
Protein modification via genetic manipulation	Canonical AA insertion: cysteine	<p>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</p>	<p>Advantages:</p> <ul style="list-style-type: none"> -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 <p>Limitations:</p> <ul style="list-style-type: none"> -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	<p>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</p>	<p>Advantages:</p> <ul style="list-style-type: none"> -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)	<p>Limitations:</p> <ul style="list-style-type: none"> -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; photolysine; FPheK; pNCSF; ThioD; ACTKs; AmAzLys; PSCaas	<p>Advantages:</p> <ul style="list-style-type: none"> -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 <p>Limitations:</p> <ul style="list-style-type: none"> -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 fluorettte 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	<p>Advantages:</p> <ul style="list-style-type: none"> -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
--	--	---

1500
1501

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 °C, 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 <i>S. aureus</i>) 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 <i>S. aureus</i> (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-Apolipoprotein 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 °C, 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 thiosuccinimide 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.

^aIn 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.

1504
1505
1506

1507 **Glossary:**
1508
1509 Site-selective:
1510 Modification methods that target a certain residue over other types of amino acids.
1511
1512 Site-specific:
1513 Modification methods that target a single occurrence of a particular type of amino acid.
1514
1515 Endogenous residues:
1516 Amino acid residues that are present in the native sequence of a protein based on the unaltered
1517 genes of the host organism.
1518
1519 Canonical amino acids:
1520 The standard 20 amino acid types encoded and inserted naturally by the genetic code and by
1521 native protein biosynthesis systems.
1522
1523 Protein microenvironment:
1524 The manipulation of amino acid sidechain properties (e.g. steric or electric characteristics) and
1525 reactivity based upon the identity of surrounding amino acids in the protein sequence.
1526
1527 Noncanonical amino acids:
1528 Amino acids that are most often synthesized and non-proteinogenic (with the exception of
1529 selenocysteine and pyrrolysine) and can either be inserted residue- or site-specifically into
1530 protein sequences.
1531
1532 Heterogenous products:
1533 Protein conjugates that have different constitutions based on the conjugation method binding
1534 differing amino acid types or various occurrences of the same amino acid type.
1535
1536 Human insulin:
1537 A protein that is made up of two separate chains of amino acids labeled as A and B, bound
1538 together by two disulfide bridges.
1539
1540 Post translational modifications:
1541 Post-translational, covalent protein modifications that have critical roles in cell signaling and
1542 control of protein activation or function.
1543
1544 SH3 domain proteins:
1545 Proteins that contain SH3 domains for the regulation of cytoplasmic signaling pathways.
1546
1547 Bioorthogonal reactions:
1548 Chemical reactions that can be executed in the complex environment of living systems (i.e. in the
1549 presence of many nucleophiles, reductants, etc.) without altering or affecting native processes.
1550
1551
1552

1553 Antibodies:
1554 Proteins that are composed of two main regions: Fc regions (constant regions for the support and
1555 stability of the antibody) and Fab regions (variable regions of the antibody that must be
1556 preserved in order to retain affinity and specificity for a corresponding antigen).
1557
1558 Click chemistry:
1559 Chemical reactions that can be defined based on high reaction and conversion rate, green solvent
1560 systems, low byproduct levels, and broad functional group applicability.
1561
1562 Conjugate payload:
1563 The chemical linker and added functionality (e.g. fluorophore, cytotoxic drug, etc.) in a protein
1564 conjugate.
1565
1566 Disulfide rebridging:
1567 Process by which two cysteine residues, revealed by disulfide reduction, reform the disrupted
1568 disulfide either through the construction of a mixed disulfide or through the introduction of a
1569 synthetic stapling molecule to connect the two residues
1570
1571 Orthogonal tRNA/RS pairs:
1572 These orthogonal pairs can use native protein biosynthesis machinery for the site-specific
1573 insertion of noncanonical amino acids and require that no native RS be able to aminoacylate the
1574 incorporated tRNA and no native tRNA be modified by the incorporated RS.
1575
1576 Fusion protein:
1577 Proteins that are produced by combining parts from different proteins or proteins with smaller
1578 amino acid sequences/tags to create one expressed entity.
1579
1580 Upconversion nanoparticles:
1581 Nanoscale particles that allow for photon upconversion (the absorption of two lower-energy
1582 photons to create one higher-energy, emitted photon) for imaging and sensors in deep tissue
1583 environments.
1584
1585