

Site-selective protein-modification chemistry for basic biology and drug development

Nikolaus Krall,^a Filipa P. da Cruz,^b Omar Boutureira^c and Gonçalo J. L. Bernardes^{*b,c}

^a*Department of Chemistry and Applied Biosciences, Institute of Pharmaceutical Sciences, Swiss Federal Institute of Technology Zürich (ETH Zürich), Vladimir Prelog Weg 1-5/10, CH-8093 Zürich (Switzerland)*

^b*Instituto de Medicina Molecular, Faculdade de Medicina da Universidade de Lisboa, Av. Prof. Egas Moniz, 1649-028 Lisboa (Portugal)*

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

Correspondence should be addressed to G.J.L.B.:
Tel. +44 (0) 1223 336305; E-mail: gb453@cam.ac.uk

ABSTRACT (154 WORDS)

Nature has produced an intricate machinery to covalently diversify the structure of proteins after their synthesis in the ribosome. In an attempt to mimic nature, synthetic chemists have developed a large set of reactions that enables post-expression modification of proteins at pre-determined sites. These reactions are now used to selectively install particular modifications on proteins for many biological and therapeutic applications. For example, they provide an opportunity to install post-translational modifications (PTMs) on proteins to dissect their exact biological roles. Labelling of proteins in live cells with fluorescent dyes allows to track protein uptake and intracellular trafficking and to optically measure physiological parameters. Through the conjugation of potent cytotoxics to antibodies, novel anti-cancer drugs with improved efficacy and reduced side-effects spectra may be obtained. In this Perspective we highlight the most exciting current and future applications of chemical site-selective protein-modification and consider which hurdles still need to be overcome for a more widespread use.

Over billions of years of evolution, nature has developed an intricate machinery to covalently attach diverse functional groups to proteins after their synthesis by the ribosome, a process called post-translational modification (PTM). Naturally occurring protein PTMs are now understood to play important roles in tuning the physicochemical properties of proteins, modulating enzymatic activity, controlling protein–protein recognition and imparting enzymes with chemical functionalities that are not covered by the standard proteinogenic amino acids.¹ For example, the attachment of carbohydrates to proteins — a process known as glycosylation — can improve the solubility of proteins and modulate ligand–receptor interactions.¹ Through the reversible addition of phosphate groups at serine (Ser), threonine (Thr) or tyrosine (Tyr) residues, enzymes can be switched between their active and inactive states. In a similar manner, the formation of complexes between two or more proteins can be controlled.¹ Covalently bound co-factors (so-called prosthetic groups) such as flavins or inorganic metal clusters can impart metabolic enzymes with the ability to perform redox chemistry.¹ Collectively, these examples highlight how, through the introduction of small modifications, nature often re-designs the structure and function of existing proteins to impart immense biological diversity.

In an attempt to follow nature's example, chemists have developed tools to covalently modify proteins with diverse functionalities in the laboratory. Initial methodologies typically offered poor control over the precise position of modification (i.e., they exhibited poor site-selectivity). Water-stable electrophilic carboxylic acid derivatives such as N-hydroxysuccinimidyl (NHS) esters were amongst the first reagents used for covalent protein-modification applications.² NHS esters react with diverse nucleophiles on protein surfaces, most importantly the ϵ -amine of lysine (Lys) residues. Given the abundance of surface Lys on most proteins, the reaction typically results in a mixture of products modified at different protein positions. This in turn makes functional characterisation of modified proteins difficult, because one is dealing with a heterogeneous protein sample rather than a single, homogenous molecule. Therefore, reactions that enable chemists to selectively install modifications on protein surfaces at pre-defined positions would facilitate both the chemical characterization of these modifications and investigation of their exact biological

function. In the context of this Perspective, we will refer to methods that permit the formation of a covalent bond between a protein and a synthetic organic molecule at a pre-defined residue as site-selective protein-modification reactions.

The challenges associated with the development of site-selective protein-modification reactions are manifold. Reactions have to be chemoselective for one amino acid side chain over all others on the protein surface. Amino acids that are lowly abundant on protein surfaces and have unique reactivity lend themselves particularly well to site-selective modification chemistries. The most important natural amino acid fulfilling these criteria has proven to be surface exposed cysteine (Cys) residues.³ Certain natural amino acids may also have unique properties when introduced at the N-termini position of a protein. For instance, an N-termini serine (Ser) can be oxidized to generate an aldehyde that can subsequently be used for site-selective modification.⁴ Also, at the N-termini, Cys can react selectively with thioester derivatives to give a native peptide bond in a process called native chemical ligation (NCL).⁵ Site-selective protein-modification reactions must furthermore be compatible with aqueous buffered systems, not interfere with proper protein folding and proceed at near neutral pH and moderate temperatures (i.e., 20-37 °C). Because the chances of encountering multiple proteins that carry even lowly abundant natural amino acids on their surface increases considerably in complex protein mixtures, it is typically only possible to site-selectively label purified recombinant proteins on naturally occurring amino acids (**Fig. 1**).

One method of achieving site-selective modification of a particular protein target within a complex mixture of proteins, relies on recent advances in genetic code expansion techniques.^{6,7} These techniques allow for the ribosomal incorporation of non-canonical amino acids into proteins that display functionalities which are normally not present in living cells.⁷ These include side chains featuring azide, alkyne, ketone, alkene or tetrazine functional groups.⁸ For each of these functionalities, reactions can be envisioned that are highly chemoselective or even mechanistically specific for the non-canonical amino acid over all other functionalities present in living cells. This concept is referred to as bioorthogonal

chemistry and is now permitting chemists to site-selectively install modifications on individual proteins in complex biological mixtures including in living organisms (**Fig. 1**).⁹⁻¹¹ Bioorthogonal ligation reactions may furthermore be used on purified protein preparations when bond stability or labelling accuracy going beyond that achievable with native amino acids is required (e.g., construction of antibody-drug conjugates (ADCs)).

Chemists and biologists are now increasingly using site-selective protein-modification tools to study the function of naturally occurring PTMs. Additionally, by attaching synthetically prepared modifications that are typically not found in nature they are attempting to confer novel functionality to proteins of bacterial or eukaryotic origin. The aim of this Perspective is to discuss key developments in the field of chemical site-selective protein-modification in the context of their biological and therapeutic applications, rather than exhaustively present a large number of reactions concerning this topic.^{8,12} The methods covered in this Perspective involve ligation reactions between the side-chain of a natural or non-canonical amino acid and a suitable small molecule derivative (**Fig. 1**).

Study of post-translational protein-modifications (PTMs)

Elucidating the effect of PTMs on protein function can greatly be assisted by studying pure protein preparations with precise modifications *in vitro*. Access to modified proteins could permit raising modification-specific antibodies,¹³ studying enzymatic activities *in vitro* or potentially affinity-purifying modification-specific interaction partners from cell lysates, much like established affinity purification protocols.¹⁴ However, isolation of single isoforms of modified proteins from natural sources in sufficient quantities is often challenging. Efficient chemical means to prepare proteins with defined PTMs *in vitro* are thus required. Achieving high precision with respect to the site of the modification is typically the prime concern.

One approach towards the chemical preparation of post-translationally modified proteins that meets the requirement for precision involves the total or semi-synthesis of proteins with defined PTMs *in vitro*.⁵ Typically, peptide fragments

bearing PTMs are prepared by solid-phase peptide synthesis, joined through native chemical ligation (NCL) reactions and are folded *in vitro* to give the full-length protein.⁵ In one example, through the use of NCL, Danishefsky and co-workers assembled a single glycoform of human erythropoietin with the consensus glycosylation motifs placed at their native sites.¹⁵ While this approach is feasible for relatively short polypeptide chains, it becomes excessively cumbersome for large proteins. Alternatively, recombinantly expressed protein fragments with N-terminal Cys or C-terminal protein thioesters obtained from split-intein fusions can be directly conjugated to small peptide fragments using NCL. This technique termed expressed protein ligation (EPL)¹⁶ is particularly useful, when the modification is relatively close to the N- or C-terminus of a protein. EPL has been widely employed by Muir and co-workers for the study of the effect of histone tail methylation and acetylation on chromatin structure and function.¹⁷ In one example, a semi-synthetic strategy was used to generate a dimethylated arginine (Arg) at position 42 of histone H3. Specifically, the sequence from alanine (Ala) 1 to valine (Val) 46 of histone H3 was prepared by total synthesis and then joined to a recombinantly expressed fragment representing Cys 47 to Ala 135 (**Fig. 2a**). Dimethylation of Arg 42 was shown to stimulate transcription *in vitro*.¹⁸ In many other cases, however, PTMs are located at sites in the inner regions of the protein, and a totally synthetic or semi-synthetic approach for the preparation of the modified protein becomes particularly difficult. In such cases, tools that permit the site-selective installation of PTMs on fully folded proteins may allow easier access to the desired modified protein product.

Site-selective protein-modification chemistry promises to allow the facile preparation of proteins with defined PTMs from full-length recombinantly produced precursors whilst still meeting the requirement for introducing the PTM at a precise position. So far, however, most chemical site-selective protein-modification approaches only allow the introduction of PTM mimics. For many applications, e.g., raising antibodies, mimics may, however, be fully sufficient.¹⁹ Davis and co-workers have recently reported a study in which a genetically engineered Cys on the surface of kinase p38 α was first converted to dehydroalanine (Dha) via a bis-alkylation elimination²⁰ procedure with α,α' -di-

bromo-adipyl(bis)amide **1** and subsequently modified via Michael-addition with sodium thiophosphate (**Fig. 2b**) to mimic a native phosphate group.²¹ In this case, the phospho-Cys variant of p38 α was sufficiently similar to the natural phospho-Tyr to switch the kinase into its active state. The introduction of phosphorylated Tyr analogues has also been achieved through the Staudinger-Phosphite reaction of azides.²²

Furthermore, other PTM mimics such as acetylated or methylated Lys, can be installed at Dha by reacting the Michael acceptor with a suitable thiol precursor (**Fig. 2c**).²³ Other approaches to prepare histones with thioether mimics of Lys methylation and acetylation involve direct Cys alkylation with suitable electrophiles,^{24,25} mixed disulfide formation,²⁶ and radical-initiated thiol-ene reactions (**Fig. 2c**).²⁷

The chemical introduction of native PTMs into full-length proteins at defined positions is more difficult, but has found particular success to site-selective installation of ubiquitin into recombinant substrate proteins. One strategy consisted of site-directed incorporation of a δ -thiol-Lys derivative into a recombinant protein that reacts with a C-terminal thioester of ubiquitin through NCL; subsequent desulfurization yields a naturally occurring isopeptide bond between substrate proteins (e.g., small ubiquitin-like modifier (SUMO)) and ubiquitin (**Fig. 2d**).²⁸ Alternatively, a genetically encoded orthogonal protection and activated ligation strategy consisting of genetic encoding of a protected Lys residue has been developed. All the other Lys residues present on the protein are then chemically protected followed by orthogonal deprotection of the Lys that was genetically introduced. The free single Lys then reacts with a C-terminal thioester of ubiquitin through NCL to yield a natural isopeptide bond. Final global deprotection affords the ubiquitinated protein.²⁹ This approach has been used to prepare, for example, mono-ubiquitinated human Dvl2 DIX domain at Lys54 and Lys58 providing insight into the effect of ubiquitination into oligomerization of these domains.³⁰ Protein substrates may contain multiple ubiquitin chains. Polyubiquitination of target proteins at specific sites through defined chemical linkages has been achieved, for example, by Cu(I)-catalysed azide-alkyne cycloaddition-based polymerization of a ubiquitin containing two

orthogonal non-canonical amino acids which led to the formation of a protease resistant polyubiquitinated protein.³¹ Alternatively, Brik and co-workers have introduced suitable electrophiles at the proximal end on ubiquitin for chemoselective reaction with a Cys-tagged α -globin protein substrate via disulfide or thioether bonds.³²

To add further complexity, nature often introduces more than one distinct PTM at pre-determined sites on a protein. Yet, with all of the abovementioned approaches it is currently difficult to site-selectively install two distinct modifications *in vitro* when modifying only natural amino acids. Therefore, further work is needed to develop reactions that allow for the facile installation of native PTMs on recombinantly produced full-length proteins and the installation of two different PTMs on the same protein substrate.

The ability to genetically encode unnatural amino acids into pre-defined sites on proteins can help in achieving site-selective protein-modification at multiple positions. As an example of such a strategy, Davis and co-workers have built a reconstituted and fully functional mimic of P-selectin glycoprotein ligand 1 (PSGL-1) containing both a sulfo-Tyr mimic and a trisaccharide moiety. This was possible through the placement of a Cys residue and an unnatural azidohomoalanine (Aha) amino acid which allowed for sequential mixed disulfide and Cu(I)-catalysed azide-alkyne cycloaddition reactions on a bacterially expressed protein scaffold.³³

As a possible alternative, techniques are now emerging which allow for the ribosomal incorporation of amino acids that already carry PTMs into proteins.^{7,34} These approaches could allow for the expression of precisely post-translationally modified proteins in living cells and have great potential to facilitate the study of PTMs in their native environment.

Site-selective protein-modification for imaging

Modern biology would be impossible without the ability to label biological macromolecules with visible light dyes or radionuclides for diverse imaging

applications. Very often, purified proteins are conjugated to fluorophores or radionuclides *in vitro* and then used to track the localisation of the labelled protein or its binding partners in an experimental system *in vitro* (e.g., immunofluorescence) or *in vivo* (e.g., nuclear imaging techniques). *In vitro*, site-selective labelling is not a strict requirement. Labelled protein reagents are typically used in excess and non-binding species are removed in washing steps. For *in vivo* applications, on the other hand, it has been shown that the position in which a modification is attached to a protein can have an influence on pharmacokinetics and biodistribution.^{35,36} It is so far poorly understood, however, how much or if at all site-selectivity would improve the specificity or sensitivity of an imaging agent *in vivo* and further work is required to address this question.

Site-selectivity becomes strictly important when optical imaging techniques are used to obtain a more complex functional readout beyond the mere localisation of a single protein. For example, measuring the association of two proteins by Förster Resonance Energy Transfer (FRET) will require specific placement of the FRET donor and acceptor in sites that are not involved in protein-protein interaction yet are close enough for FRET to take place. Site-selective labelling of purified protein preparations with FRET-dyes can nowadays routinely be achieved using commercially available maleimide, sulfone Michael-acceptor and haloacetamide reagents *in vitro*.

Few approaches, however, exist so far to site-selectively label the same protein with two different dyes on two different native amino acids.³⁷ Double labelling is required if one for example wants to follow conformational changes of a protein by FRET. To achieve this aim, it is very often necessary to first introduce an unnatural amino acid into the protein of choice that can bioorthogonally react with a suitably modified dye. The second label is then site-selectively introduced using, for example, the reaction of maleimides with free Cys.³⁸ Alternatively, the introduction of two non-canonical amino acids that are suitable precursors for protein double-labelling with a FRET pair is now possible using genetic code expansion techniques (**Fig. 3d**).³⁹

Bioorthogonal chemistry also allows for the site-selective labelling of proteins in complex mixtures including live cells and animals. Using an electron demand inverse Diels–Alder cycloaddition reaction, Chin and co-workers were able to site-selectively fluorescently label the non-native amino acid norbornene, which was genetically encoded into a protein expressed on the membrane of a mammalian cell (**Fig. 3a**).⁴⁰ As an additional benefit, fluorophores can also be engineered such that their fluorescence markedly increases upon bioorthogonal reaction with the protein bearing the unnatural amino acid handle. It is thought that this “turn-on” mechanism may give better signal to background ratios than those found when labelling proteins with constitutively fluorescent dyes. Bioorthogonal labelling of specific proteins may also be used to detail protein uptake and internalization in the context of disease. Using an inverse electron demand Diels–Alder cycloaddition strategy it was possible to label in a quantitative manner anthrax lethal factor (LF) with a small fluorophore.⁴¹ This enabled the time-lapse monitoring of anthrax LF internalisation and membrane translocation processes in living cells. The abovementioned and other recent studies⁸ provide exciting examples of the localisation and tracking of bioorthogonally labelled proteins in living cells or *in vivo* with minimal disruption of the native protein and with an unprecedented molecular precision.

Weissleder and co-workers recently extended the use of bioorthogonal chemistry on proteins to nuclear imaging applications. They demonstrated that a tumour-homing antibody bearing a *trans*-cyclooctene (TCO) ring on its surface could selectively and rapidly react with a radiolabelled variant of a tetrazine ring in an inverse electron demand Diels–Alder reaction *in vivo* in a mouse xenograft tumour, thus making the antibody visible in a PET-scanner (**Fig. 3b**).⁴² More recently, ¹⁸F-labeled antibody fragments for positron emission tomography (PET) imaging were generated by installing TCO into an anti-Class II MHC single domain antibody followed by reaction with a tetrazine-labeled-¹⁸F-2-deoxyfluoroglucose (FDG).⁴³ The method was found to be particularly rapid and efficient with a radiochemical yield of >25% enabling imaging of pancreatic tumours *in vivo* with an increased level of specificity compared to conventional ¹⁸F-FDG imaging. Furthermore, Robillard and colleagues used a similar approach where, prior to injection of the radiolabelled tetrazine, residual

amounts of antibody-TCO in circulation were first removed by injection of a tetrazine-functionalized clearing agent.⁴⁴ This led to a significant increase in tumour to healthy tissue ratios, and a predicted 8-fold higher tumour dose compared with the directly labeled antibody. Whilst for most protein-modification reactions that occur *in vitro*, speed is not the primary objective, the extremely fast kinetics of the inverse electron demand Diels–Alder reaction makes it particularly suitable for *in vivo* applications where reagents may rapidly be excreted through liver or kidney, however an improved understanding of the stability of the trans double bond of TCO that may undergo isomerization *in vivo* will be required. Alternatively, it may also be possible to encode tetrazine motifs into the antibody followed by ligation with strained alkene probes avoiding long exposure of the trans double bond to *in vivo* conditions.⁴⁵

In the coming years, one of the most exciting applications of bioorthogonal protein-labelling chemistry may well be the ability to construct functional optical sensors to measure diverse physiological parameters inside the dynamic environment of living cells and even in animals. For example, Chen and co-workers have recently described the modification of the acid-sensitive chaperone HdeA with an environment-sensitive fluorophore inside the periplasm of *E. Coli* to create a biosensor for extremely low pH values (**Fig. 3c**).⁴⁶ Upon lowering the pH from 7 to 2, HdeA adopts a highly plastic structure making hydrophobic residues accessible to the fluorophore, which is reflected in an increase in fluorescence. Even bioorthogonal dual modification of a single protein with two distinct dyes has recently been achieved. Impressively, Chin's group has developed an optimised translational system for proteins containing non-natural amino acids in *E. coli* that enable the placement of two distinct non-native amino acids at two pre-determined sites that could be modified with mutually and bioorthogonal reactions (**Fig. 3d**). Site-selective, bioorthogonal dual labelling of calmodulin (CaM) with a FRET pair made it possible to follow alterations in calmodulin conformation inside the living cell in response to changes in Ca²⁺ concentrations.³⁹

These examples demonstrate that bioorthogonal protein-modification chemistry is a very promising tool for the construction of protein biosensors *in vivo*. Protein

biosensors have the unique potential to give an optical readout in response to physiological changes inside living cells and organisms. These sensors will allow biologists to use imaging techniques to study complex processes inside living cells and even whole organisms in a spatiotemporally controlled manner.

Modulating the properties of protein therapeutics

Therapeutic proteins are a rapidly growing class of drugs.⁴⁷ Unfavourable pharmacokinetics and immunogenicity, however, often hamper the efficacy of recombinant protein therapeutics *in vivo*. Therapeutic proteins may be quickly degraded by proteases or excreted by the kidneys leading to a rapid drop in plasma concentration after administration of the drug. A strong immune reaction may also be a life threatening complication akin to an allergic shock. Alternatively, anti-drug antibodies may inactivate the therapeutic protein and lead to treatment resistance. One of the first applications of protein-modification chemistry in the context of drug development was the conjugation of protein therapeutics with long polyethylene glycol (PEG) chains to extend their plasma circulation half-life and attenuate their immunogenicity *in vivo*.⁴⁸ By attaching a long unstructured polymer chain the hydrodynamic radius of the macromolecule is increased thus reducing renal filtration. At the same time, the PEG polymer physically shields the protein from degradation by proteases and recognition by the immune system.

Many FDA-approved PEGylated protein drugs are still produced using non-site-selective modification chemistry,⁴⁹ which can potentially produce inactive protein species. Indeed, many marketed PEGylated drugs are less active than their unmodified counterparts.⁵⁰ For this reason, site-selective PEGylation should be preferred and site-selective protein-modification chemistry is increasingly used for the PEGylating of proteins.⁵¹

In one of the first examples of site-selective PEGylation, a N-terminal serine residue was oxidized to the corresponding aldehyde using sodium periodate followed by subsequent oxime ligation with an aminoxy PEG derivative (**Fig. 4a**).⁴ The modified proteins, interleukin (IL)-8, G-CSF, and IL-1 α , fully retained

their activity after PEGylation. Alternatively, Francis and co-workers reported a pyridoxal-phosphate (PLP) mediated transamination reaction that introduces reactive aldehydes at the N-termini (**Fig. 4a**).⁵² In a similar manner, subsequent reaction aminoxy-PEG derivative at acidic pH 6.5 provides the corresponding oxime product. These methods require however chemical manipulation of the N-terminal groups to introduce a reactive aldehyde functionality. More recently, the same research group reported a direct site-selective method that enable PEGylation of native proteins.⁵³ The method uses 2-pyridinecarboxylaldehydes that selective react at the N-termini at neutral pH. The reaction proceeds via the formation of the imine followed by the attack of the neighbouring amide nitrogen in the protein backbone leading to the formation of a stable imidazolidinone conjugate (**Fig. 4b**). This methodology may be applied to a wide range of native proteins with the exception of those where the N-terminus is acylated, those with proline in the second position or those where the N-termini are hindered. However, in some instances the modification of internal sites on the protein may be preferred. Site-selective PEGylation of proteins may also be achieved using genetic encoding non-canonical amino acids (ncAA) bearing unique reactive handles. Incorporation of *para*-acetylphenylalanine (*pAF*) into human growth hormone (hGH) followed by reaction with aminoxy-PEG derivatives at pH 6 enabled the construction of hGH variants with PEG at 20 different pre-determined sites (**Fig. 4c**).⁵⁴ The PEGylated construct of hGH modified at site 35 demonstrated to possess improved pharmacokinetics in rats and comparable efficacy in clinical studies performed in GH-deficient adults but with reduced injection frequency. This example is illustrative of the advantage of using site-selective PEGylation methods that allow to perform structure-activity relationships of modified proteins for optimal pharmacokinetics while retaining their biological activity and improving clinical application.

Antibody conjugation for delivery of cytotoxic drugs

More recently, the conjugation of monoclonal antibodies directed against tumour marker proteins with highly potent cytotoxic drugs is emerging as a strategy to create antibodies with improved killing potential towards malignant cells or even turn antibodies lacking any cytotoxic activity into potent

antineoplastic agents.⁴⁷ Initial work on antibody-drug conjugates (ADCs) relied on poorly site-selective conjugation strategies to attach drugs to antibodies yielding a heterogeneous mixture of antibodies with different drug loading at different sites.⁵⁵⁻⁵⁷

It was soon found that both the number of drugs per antibody and the attachment site could have a profound effect on the pharmacokinetics, efficacy and toxicity of an ADC.^{36,58} A high ratio of drug to antibody (typically > 2) may lead to reduced thermostability of the ADC⁵⁶ and result in a faster clearance from the bloodstream.⁵⁷ ADCs with high drug to antibody ratio (DAR) were also found to have a smaller therapeutic window than ADCs with low DARs.⁵⁷ Additionally the site of drug attachment can have a profound effect on clearance kinetics and off-target ADC toxicity^{36,58} as well as in plasma stability and efficacy.⁵⁹ These observations demonstrate the importance of using site-selective protein-modification chemistries for the synthesis of homogenous protein conjugate therapeutics.

Currently, the most commonly used approach for the site-selective generation of antibody-conjugate therapeutics is the Michael-addition reaction of the thiol side-chain of genetically engineered Cys residues on the antibody's surface with maleimides (**Fig. 5a**).³ Protocols have been established which allow for the selective reduction of engineered Cys whilst leaving native disulfide bonds on the antibody intact.³⁶ Using carefully optimized conditions, ADC preparations containing >90% ADCs with exactly two drugs per antibody could be prepared.³⁶

A potential drawback of maleimide chemistry is related to the fact that the thioether bond formed between the protein's Cys and the maleimide may not be fully stable in plasma (**Fig. 5a**).⁵⁸ Conjugates can undergo a retro-Michael addition releasing the maleimide from the carrier protein under physiological conditions.⁶⁰ In the case of ADCs, the product of this reaction would be a highly potent cytotoxic drug that can kill cells in healthy tissues. Indeed, off-target toxicity is a pressing problem in the clinical development of ADCs. Different strategies have thus been pursued to improve the stability of maleimide

conjugates. Conjugation sites have been placed in close proximity to positively charged amino acid residues, which promote the hydrolysis and ring-opening of maleimides to give more stable species.⁵⁸ Linkers containing primary amines in close proximity to a maleimide that undergo rapid hydrolysis to the more stable ring-opened form have also been proposed (**Fig. 5a**).^{61,62} The resulting ADCs were demonstrated to be more stable in plasma and more efficacious *in vivo* while reducing the side effects observed when using traditional maleimides.

A number of reactions for the modification of Cys that inherently give more stable thioether bonds, such as the Julia–Kocienski type modification developed by Barbas and co-workers (**Fig. 5a**)⁶³ or the Michael-addition of thiol nucleophiles to Dha developed by Davis and co-workers (**Fig. 5a**),⁶⁴ may also be exploited for the synthesis of ADCs. These technologies are fairly novel, however, and we are not aware of any industrial ADC discovery programs using them thus far.

As an alternative to the chemical modification of genetically incorporated Cys residues on monoclonal antibody (mAb) surfaces, one can incorporate unnatural amino acids into mAbs and modify them using bioorthogonal reactions. This approach has the potential advantage of yielding highly homogeneous protein conjugate preparations and antibody-drug linkages with high stability.

Schultz and co-workers have genetically incorporated *para*-acetylphenylalanine (*p*AF) into mAbs and reacted the side chain ketone with an aminoxy-modified version of the potent cytotoxic drug monomethyl auristatin E (MMAE) **3** for the preparation of ADCs (**Fig. 5b i**).⁶⁵ The same antibody was expressed with two Cys in the same positions as the *p*AF and conjugated with a maleimide-containing version of MMAE. The *p*AF-modified ADC had a drug-to-antibody ratio (DAR) of exactly 2.0 whereas attempts to selectively modify the two genetically introduced Cys on the mAb using a maleimide-containing derivative of MMAE gave a DAR of 2.3. The homogenous *p*AF-modified ADCs were found to be more stable than conventional maleimide linkages *in vivo*.⁶⁶ These

characteristics resulted in an improved efficacy of the bioorthogonally-modified ADC in xenograft-tumour-bearing mice.

Bertozzi and co-workers have reported a site-selective protein-ligation where aldehyde-containing proteins rapidly react with 3-methylhydroxylamino indoles in a Pictet–Spengler type mechanism.⁶⁷ A variation of the methodology, the hydrazino-*iso*-Pictet–Spengler (HIPS) ligation, yields highly stable C–C linkages between aldehyde-containing antibodies and drugs containing 3-methylhydrazino indole-groups **4 (Fig. 5b ii)**. As such, this methodology has been used for the construction of ADCs.⁶⁸ By performing HIPS conjugations at different locations on an IgG1 backbone, it was shown that the site of conjugation influences dramatically the *in vivo* efficacy and pharmacokinetic parameters.⁶⁸ In addition, the stable conjugates built using HIPS conjugation strategy exhibited an improved safety profile in rats when compared with those conjugates prepared using a non-site-selective lysine conjugation strategy.

Collectively, these observations highlight the fact that site-selective protein-conjugation reactions can vastly improve the therapeutic efficacy of protein-conjugate therapeutics when compared to protein conjugates that were synthesised using non-selective bioconjugation chemistry, at least within animal models of disease. Although clinical proof is not likely within the next few years, we expect that protein conjugate therapeutics synthesised using site-selective modification reactions that yield homogenous and more stable products will also be more efficacious and tolerable to patients than currently used heterogeneous and potentially unstable conjugates.

The future of bioorthogonal chemistry on protein surfaces

Evidence is now accumulating that chemistry on protein surfaces will in the future go far beyond bioconjugation. In particular the combination of genetic code expansion for the incorporation of non-canonical amino acids into proteins with bioorthogonal chemistry could enable tantalizing new applications. For example, the genetic incorporation of caged amino acids that can be deprotected using bioorthogonal reactions holds great potential for the precise

spatiotemporal control of protein function *in vivo*. Indeed, a palladium-mediated depropargylation has recently been achieved within a living cell and could be used to restore the function of a propargyl-protected protein (**Fig. 6a**).⁶⁹ In another example, a tetrazine-triggered protein decaging method based on the inverse electron-demand Diels–Alder reaction was successfully employed for the bioorthogonal deprotection of a caged Lys within living cells (**Fig. 6b**).⁷⁰ The inverse electron-demand Diels–Alder reaction elimination occurs through the conversion of the dihydropyridazine adduct, formed from the [4+2] cycloaddition between a 1,2,4,5-tetrazine and TCO, to a pyridazine with elimination of CO₂ and an amine derivative.

This inverse electron demand Diels–Alder elimination reaction also was used by Robillard and co-workers in a small molecule pro-drug strategy. A carbamate between 1-hydroxy-*trans*-cyclooctene (HO-TCO) and the amino sugar of doxorubicin (DOXO), a chemotherapeutic agent commonly used for the treatment of cancer, was prepared.⁷¹ When reacting the DOXO derivative with a tetrazine reagent in an inverse electron demand Diels–Alder fashion, the dihydropyridazine product undergoes conversion into a conjugated pyridazine with consequent elimination of the NH₂-substituted DOXO and CO₂.⁷¹ It remains to be demonstrated that this strategy can be expanded to antibodies for *in vivo* bioorthogonal elimination and delivery of drugs. For example, by incorporating a DOXO HO-TCO carbamate to a tumour-homing antibody, one could selectively transport the inactivated drug into the tumour and release it selectively by injecting tetrazine as a trigger substance. This in turn would decrease exposure of healthy tissues to DOXO and reduce treatment-associated side effects. Bioorthogonal chemistry on protein surfaces may thus permit the development of novel medicines with a spatially and temporally defined action for the treatment of cancer and other debilitating diseases.

Alternatively, genetic encoding of photocaged variants of natural amino acids (e.g., Cys and Tyr) could be used to the same aim. In one case, photodeprotection of photocaged Cys results in rapid activation of TEV protease within living cells in their native environment (**Fig. 6c**).⁷² Similarly, deprotection of a key photocaged Tyr residue was used to control phosphorylation and signal

transduction.⁷³ It will emerge over the coming years how far and in which settings these techniques are superior to currently used methods for the control of protein function, for example, at the transcription level with doxocyclin inducible promoters, and for in situ pro-drug antibody-based delivery strategies.

Conclusions and outlook

Site-selective protein-modification chemistry can enable the *in vitro* and *in vivo* study of PTMs, construction of protein-based sensors for biological applications or the preparation of powerful new protein therapeutics. Other applications that were not addressed in detail here may be found in the fields of material sciences or regenerative medicine where the synthesis of chemically defined scaffolds containing protein building blocks (e.g., factors for cell attachment and growth) may play an important role.

Site-selective protein-modification chemistry comes in two flavours: reactions with naturally occurring but lowly abundant amino acids is typically limited to the modification of isolated proteins *in vitro*. Methods that allow the modification of proteins with diverse synthetic molecules ranging from PTMs or their mimics to fluorophores and drugs with high accuracy with respect to the site of modification are now available. The use of low reagent's concentration and the stability of the bond between protein and the modification are primary concerns for the modification of proteins *in vitro* (e.g., for ADCs) and methods that satisfy this requirement are now becoming available. The advantage of modifying naturally occurring amino acids lies in the fact that standard methods of protein engineering can be easily employed to generate the protein starting material.

Bioorthogonal ligation reactions, on the other hand, are a type of site-selective protein-modification chemistry directed at unnatural functionalities introduced into pre-determined sites on proteins often using genetic code expansion techniques. Bioorthogonal ligation reactions are generally more versatile than classical site-selective protein-modification reactions directed at natural proteinogenic amino acids. In particular, bioorthogonal ligation chemistry may allow for the site-selective modification of proteins even in complex protein

mixtures, where rare amino acids traditionally used for site-selective protein-modification (i.e., most commonly Cys) are present in multiple copies. It now even allows for the site-selective modification of individual proteins in cells and living organisms where high reaction speeds on top of high selectivity may be required.

Bioorthogonal protein-modification chemistry may also be particularly suited to site-selectively introduce two or even more modifications on the same protein opening new doors to interrogate protein structure, dynamics and function in their native environment. The main drawback of bioorthogonal approaches remains the need for introducing non-natural amino acids into the target protein prior to modification but efficient methods even for the large-scale production of proteins incorporating non-native amino acids are now becoming available.^{66,74} Together with efforts to integrate optimised genetic encoding expression systems for researchers including the Unnatural Protein Facility, located at Oregon State University (OSU),⁷⁵ such techniques will hopefully find increased implementation in laboratories.

Modern site-selective, bioorthogonal protein-modification techniques have shown potential to provide molecular insights in numerous proof-of-concept studies and enabled the construction of more efficient and safer protein therapeutics. As the number of reactions at our disposal for the site-selective and bioorthogonal modification of proteins increases, we anticipate that the widespread use of bioorthogonal protein-modification will generate fundamental new basic biology knowledge, imaging and therapeutic applications with unprecedented precision. It is, however, key to make these methodologies available in an easily usable format to as wide a research community as possible.

Acknowledgements

We thank FCT Portugal (FCT Investigator to G.J.L.B.), the EU (Marie-Curie CIG to G.J.L.B. and Marie-Curie IEF to O.B.) and the EPSRC for funding. G.J.L.B. is a Royal Society University Research Fellow. Due to space limitations, many

primary and historical publications have not been cited, in particular in those cases where topical reviews are available.

Author contributions

N.K. and G.J.L.B. developed the concept, researched and wrote the manuscript. F.P.C. designed and produced the figures, and F.P.C. and O.B. assisted with writing the manuscript.

Competing financial interests

The authors declare no competing financial interests.

References

- [1] Walsh, C. T., Garneau-Tsodikova, S., & Gatto, G. J. Protein posttranslational modifications: The chemistry of proteome diversifications. *Angew. Chem. Int. Ed.* **44**, 7342–7372 (2005).
- [2] Stephanopoulos, N. & Francis, M. B. Choosing an effective protein bioconjugation strategy. *Nature Chem. Biol.* **7**, 876–884 (2011).
- [3] Chalker, J. M., Bernardes, G. J. L., Lin, Y. A., & Davis, B. G. Chemical modification of proteins at cysteine: Opportunities in chemistry and biology. *Chem. Asian J.* **4**, 630–640 (2009).
- [4] Gaertner, H. F. & Offord, R. E. Site-specific attachment of functionalized poly(ethylene glycol) to the amino terminus of proteins. *Bioconjugate Chem.* **7**, 38–44 (1996).
- [5] Dawson, P., Muir, T., Clark-Lewis, I., & Kent, S. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
- [6] Xie, J. & Schultz, P. G. A chemical toolkit for proteins – an expanded genetic code. *Nature Rev. Mol. Cell Biol.* **7**, 775–782 (2006).
- [7] Davis, L. & Chin, J. W. Designer proteins: Applications of genetic code expansion in cell biology. *Nature Rev. Mol. Cell Biol.* **13**, 168–182 (2012).
- [8] Lang, K. & Chin, J. W. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. *Chem. Rev.* **114**, 4764–4806 (2014).
- [9] Sletten, E. M. & Bertozzi, C. R. Bioorthogonal chemistry: Fishing for selectivity in a sea of functionality. *Angew. Chem. Int. Ed.* **48**, 6974–6998 (2009).
- [10] Elliott, T. S. *et al.* Proteome labeling and protein identification in specific tissues and at specific developmental stages in an animal. *Nature Biotech.* **32**, 465–472 (2014).
- [11] Schumacher, D. & Hackenberger, C. P. R. More than add-on: Chemoselective reactions for the synthesis of functional peptides and proteins. *Curr. Opin. Chem. Biol.* **22**, 62–69 (2014).
- [12] Spicer, C. D. & Davis, B. G. Selective chemical protein modification. *Nature Commun.* **5**, 4740 (2014).
- [13] Arur, S. & Schedl, T. Generation and purification of highly specific antibodies for detecting post-translationally modified proteins *in vivo*. *Nature Protoc.* **9**, 375–395 (2014).
- [14] Burckstummer, T. *et al.* An efficient tandem affinity purification procedure for interaction proteomics in mammalian cells. *Nature Methods* **3**, 1013–1019 (2006).
- [15] Wang, P. *et al.* Erythropoietin derived by chemical synthesis. *Science* **342**, 1357–1360 (2013).
- [16] Muir, T. W., Sondhi, D., & Cole, P. A. Expressed protein ligation: A general method for protein engineering. *Proc. Natl Acad. Sci. USA* **95**, 6705–6710 (1998).
- [17] Fierz, B. & Muir, T. W. Chromatin as an expansive canvas for chemical biology. *Nature Chem. Biol.*, **8**, 417–427 (2012).
- [18] Casadio, F. *et al.* H3R42me2a is a histone modification with positive transcriptional effects. *Proc. Natl Acad. Sci. USA* **110**, 14894–14899 (2013).

- [19] Kee, J.-M., Oslund, R. C., Perlman, D. H., & Muir, T. W. A pan-specific antibody for direct detection of protein histidine phosphorylation. *Nature Chem. Biol.* **9**, 416–421 (2013).
- [20] Chalker, J. M. *et al.* Methods for converting cysteine to dehydroalanine on peptides and proteins. *Chem. Sci.* **2**, 1666–1676 (2011).
- [21] Chooi, K. P. *et al.* Synthetic phosphorylation of p38 α recapitulates protein kinase activity. *J. Am. Chem. Soc.* **136**, 1698–1701 (2014).
- [22] Serwa, R. *et al.* Chemoselective staudinger-phosphite reaction of azides for the phosphorylation of proteins. *Angew. Chem. Int. Ed.* **48**, 8234–8239 (2009).
- [23] Chalker, J. M., Lercher, L., Rose, N. R., Schofield, C. J., & Davis, B. G. Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. *Angew. Chem. Int. Ed.* **51**, 1835–1839 (2012).
- [24] Simon, M. D. *et al.* The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **128**, 1003–1012 (2007).
- [25] Huang, R. *et al.* Site-specific introduction of an acetyl-lysine mimic into peptides and proteins by cysteine alkylation. *J. Am. Chem. Soc.* **132**, 9986–9987 (2010).
- [26] Chatterjee, C., McGinty, R. K., Fierz, B., & Muir, T. W. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. *Nature Chem. Biol.* **6**, 267–269 (2010).
- [27] Li, F. *et al.* A direct method for site-specific protein acetylation. *Angew. Chem. Int. Ed.* **50**, 9611–9614 (2011).
- [28] Virdee, S. *et al.* Traceless and site-specific ubiquitination of recombinant proteins. *J. Am. Chem. Soc.* **133**, 10708–10711 (2011).
- [29] Virdee, S., Ye, Y., Nguyen, D. P., Komander, D., & Chin, J. W. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an otu deubiquitinase. *Nature Chem. Biol.* **6**, 750–757 (2010).
- [30] Madrzak, J. *et al.* Ubiquitination of the dishevelled DIX domain blocks its head-to-tail polymerization. *Nature Commun.* **6**, 6718 (2015).
- [31] Schneider, T. *et al.* Dissecting ubiquitin signaling with linkage-defined and protease resistant ubiquitin chains. *Angew. Chem. Int. Ed.* **53**, 12925–12929 (2014).
- [32] Hemantha, H. P. *et al.* Nonenzymatic polyubiquitination of expressed proteins. *J. Am. Chem. Soc.* **136**, 2665–2673 (2014).
- [33] van Kasteren, S. I. *et al.* Expanding the diversity of chemical protein modification allows post-translational mimicry. *Nature* **446**, 1105–1109 (2007).
- [34] Neumann, H. *et al.* A method for genetically installing site-specific acetylation in recombinant histones defines the effects of H3 K56 acetylation. *Mol. Cell* **36**, 153–163 (2009).
- [35] Perols, A. *et al.* Influence of DOTA chelator position on biodistribution and targeting properties of ¹¹¹In-labeled synthetic anti-her2 affibody molecules. *Bioconjugate Chem.* **23**, 1661–1670 (2012).
- [36] Junutula, J. R. *et al.* Site-specific conjugation of a cytotoxic drug to an antibody improves the therapeutic index. *Nature Biotech.* **26**, 925–932 (2008).

- [37] Sachdeva, A., Wang, K., Elliott, T., & Chin, J. W. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. *J. Am. Chem. Soc.* **136**, 7785–7788 (2014).
- [38] Tyagi, S. & Lemke, E. A., Chapter 9 - genetically encoded click chemistry for single-molecule fret of proteins in *Methods Cell Biol.*, edited by P. Michael Conn (Academic Press, 2013), Vol. Volume 113, pp. 169–187.
- [39] Wang, K. *et al.* Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and fret. *Nature Chem.* **6**, 393–403 (2014).
- [40] Lang, K. *et al.* Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. *Nature Chem.* **4**, 298–304 (2012).
- [41] Zheng, S., Zhang, G., Li, J., & Chen, P. R. Monitoring endocytic trafficking of anthrax lethal factor by precise and quantitative protein labeling. *Angew. Chem. Int. Ed.* **53**, 6449–6453 (2014).
- [42] Zeglis, B. M. *et al.* A pretargeted pet imaging strategy based on bioorthogonal Diels–Alder click chemistry. *J. Nucl. Med.* **54**, 1389–1396 (2013).
- [43] Rashidian, M. *et al.* Use of ¹⁸F-2-fluorodeoxyglucose to label antibody fragments for immuno-positron emission tomography of pancreatic cancer. *ACS Cent. Sci.*, **1**, 142–147 (2015).
- [44] Rossin, R., Läppchen, T., van den Bosch, S. M., Laforest, R., & Robillard, M. S. Diels–Alder reaction for tumor pretargeting: *In vivo* chemistry can boost tumor radiation dose compared with directly labeled antibody. *J. Nucl. Med.* **54**, 1989–1995 (2013).
- [45] Seitchik, J. L. *et al.* Genetically encoded tetrazine amino acid directs rapid site-specific *in vivo* bioorthogonal ligation with trans-cyclooctenes. *J. Am. Chem. Soc.* **134**, 2898–2901 (2012).
- [46] Yang, M. *et al.* Converting a solvatochromic fluorophore into a protein-based ph indicator for extreme acidity. *Angew. Chem. Int. Ed.* **51**, 7674–7679 (2012).
- [47] Chari, R. V. J., Miller, M. L., & Widdison, W. C. Antibody–drug conjugates: An emerging concept in cancer therapy. *Angew. Chem. Int. Ed.* **53**, 3796–3827 (2014).
- [48] Pasut, G. & Veronese, F. M. PEG conjugates in clinical development or use as anticancer agents: An overview. *Adv. Drug Deliv. Rev.* **61**, 1177–1188 (2009).
- [49] Alconcel, S. N. S., Baas, A. S., & Maynard, H. D. FDA-approved poly(ethylene glycol)-protein conjugate drugs. *Polym. Chem.* **2**, 1442–1448 (2011).
- [50] Keefe, A. J. & Jiang, S. Poly(zwitterionic)protein conjugates offer increased stability without sacrificing binding affinity or bioactivity. *Nature Chem.* **4**, 59–63 (2012).
- [51] Pegri-O’Day, E. M., Lin, E.-W., & Maynard, H. D. Therapeutic protein–polymer conjugates: Advancing beyond pegylation. *J. Am. Chem. Soc.* **136**, 14323–14332 (2014).
- [52] Gilmore, J. M., Scheck, R. A., Esser-Kahn, A. P., Joshi, N. S., & Francis, M. B. N-terminal protein modification through a biomimetic transamination reaction. *Angew. Chem. Int. Ed.* **45**, 5307–5311 (2006).

- [53] MacDonald, J. I., Munch, H. K., Moore, T., & Francis, M. B. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. *Nature Chem. Biol.* **11**, 326–331 (2015).
- [54] Cho, H. *et al.* Optimized clinical performance of growth hormone with an expanded genetic code. *Proc. Natl Acad. Sci. USA* **108**, 9060–9065 (2011).
- [55] Chari, R. V. J. *et al.* Immunoconjugates containing novel maytansinoids: Promising anticancer drugs. *Cancer Res.* **52**, 127–131 (1992).
- [56] Beckley, N. S., Lazzareschi, K. P., Chih, H.-W., Sharma, V. K., & Flores, H. L. Investigation into temperature-induced aggregation of an antibody drug conjugate. *Bioconjugate Chem.* **24**, 1674–1683 (2013).
- [57] Hamblett, K. J. *et al.* Effects of drug loading on the antitumor activity of a monoclonal antibody drug conjugate. *Clin. Cancer Res.* **10**, 7063–7070 (2004).
- [58] Shen, B.-Q. *et al.* Conjugation site modulates the *in vivo* stability and therapeutic activity of antibody-drug conjugates. *Nature Biotech.* **30**, 184–189 (2012).
- [59] Steiner, M. *et al.* Spacer length shapes drug release and therapeutic efficacy of traceless disulfide-linked adcs targeting the tumor neovasculature. *Chem. Sci.* **4**, 297–302 (2013).
- [60] Cal, P. M. S. D., Bernardes, G. J. L., & Gois, P. M. P. Cysteine-selective reactions for antibody conjugation. *Angew. Chem. Int. Ed.* **53**, 10585–10587 (2014).
- [61] Lyon, R. P. *et al.* Self-hydrolyzing maleimides improve the stability and pharmacological properties of antibody-drug conjugates. *Nature Biotech.* **32**, 1059–1062 (2014).
- [62] Tumey, L. N. *et al.* Mild method for succinimide hydrolysis on adcs: Impact on adc potency, stability, exposure, and efficacy. *Bioconjugate Chem.* **25**, 1871–1880 (2014).
- [63] Toda, N., Asano, S., & Barbas, C. F. Rapid, stable, chemoselective labeling of thiols with Julia–Kocięński-like reagents: A serum-stable alternative to maleimide-based protein conjugation. *Angew. Chem. Int. Ed.* **52**, 12592–12596 (2013).
- [64] Bernardes, G. J. L., Chalker, J. M., Errey, J. C., & Davis, B. G. Facile conversion of cysteine and alkyl cysteines to dehydroalanine on protein surfaces: Versatile and switchable access to functionalized proteins. *J. Am. Chem. Soc.* **130**, 5052–5053 (2008).
- [65] Axup, J. Y. *et al.* Synthesis of site-specific antibody-drug conjugates using unnatural amino acids. *Proc. Natl Acad. Sci. USA* **109**, 16101–16106 (2012).
- [66] Tian, F. *et al.* A general approach to site-specific antibody drug conjugates. *Proc. Natl Acad. Sci. USA* **111**, 1766–1771 (2014).
- [67] Agarwal, P., van der Weijden, J., Sletten, E. M., Rabuka, D., & Bertozzi, C. R. A Pictet–Spengler ligation for protein chemical modification. *Proc. Natl Acad. Sci. USA* **110**, 46–51 (2013).
- [68] Drake, P. M. *et al.* Aldehyde tag coupled with HIPS chemistry enables the production of adcs conjugated site-specifically to different antibody regions with distinct *in vivo* efficacy and PK outcomes. *Bioconjugate Chem.* **25**, 1331–1341 (2014).

- [69] Li, J. *et al.* Palladium-triggered deprotection chemistry for protein activation in living cells. *Nature Chem.* **6**, 352–361 (2014).
- [70] Li, J., Jia, S., & Chen, P. R. Diels–Alder reaction–triggered bioorthogonal protein decaging in living cells. *Nature Chem. Biol.* **10**, 1003–1005 (2014).
- [71] Versteegen, R. M., Rossin, R., ten Hoeve, W., Janssen, H. M., & Robillard, M. S. Click to release: Instantaneous doxorubicin elimination upon tetrazine ligation. *Angew. Chem. Int. Ed.* **52**, 14112–14116 (2013).
- [72] Nguyen, D. P. *et al.* Genetic encoding of photocaged cysteine allows photoactivation of TEV protease in live mammalian cells. *J. Am. Chem. Soc.* **136**, 2240–2243 (2014).
- [73] Arbely, E., Torres-Kolbus, J., Deiters, A., & Chin, J. W. Photocontrol of tyrosine phosphorylation in mammalian cells via genetic encoding of photocaged tyrosine. *J. Am. Chem. Soc.* **134**, 11912–11915 (2012).
- [74] Liu, W., Brock, A., Chen, S., Chen, S., & Schultz, P. G. Genetic incorporation of unnatural amino acids into proteins in mammalian cells. *Nature Methods* **4**, 239–244 (2007).
- [75] Goodman, C. Making sense of nonsense. *Nature Chem. Biol.* **10**, 167–167 (2014).

Figures Captions

Figure 1. Conceptual strategy to achieve site-selective protein-modification using biocompatible chemical reactions *in vitro* and *in vivo*.

Typically, site-selective chemical modification of proteins can be achieved either i) on a natural amino acid side chain with unique reactivity (e.g., Cys) of which a single copy is already present on the protein of interest or may be introduced by site-directed mutagenesis or ii) on a non-canonical amino acid featuring a side chain with a functional group that is normally not present in living systems but can be introduced using genetic encoding procedures (e.g., azide, alkyne, ketone, alkene, tetrazine). In the case of natural amino acid side chains, one is limited to the modification of purified proteins in a test tube as many other biomolecules may display the same functionality in a complex protein mixture. Using non-canonical amino acids, on the other hand, one can devise reactions that are selective over all naturally occurring amino acid side chains thus permitting site-selective modification of individual proteins in complex mixtures and living organisms. Upon choice of the functionality to be chemically targeted, reaction with a suitable modified molecule of interest (e.g., PTM, fluorescent label or cytotoxic drug) must proceed in a chemoselective manner under physiological conditions.

Figure 2. Chemical site-selective installation of PTMs at pre-determined sites on proteins. a,

A semi-synthetic strategy for the generation of a histone protein with a dimethylated Arg residue at position 42. Briefly, histone H3R42Me₂ was assembled from two synthetic fragments and one recombinant fragment through the use of NCL and EPL. The final step involves selective radical-mediated desulfurization of the two Cys residues at the ligation sites into the native Ala residues to give the final semi-synthetic H3R42Me₂. **b,** Site-selective installation of a phosphorylation mimic in the activation loop of protein kinase p38 α . This was achieved by replacing the native Tyr residue with a Cys residue, which was subsequently chemically converted into Dha. Michael-addition of sodium thiophosphate enables access to a fully functional synthetic phospho-Cys mimic capable of switching the kinase to an active state. **c,** Chemical installation of thioether mimics of acetylated or methylated Lys at pre-

determined sites on histone proteins for the study of genetic regulation. Chemically defined PTMs can be achieved directly on Cys by alkylation with suitable electrophiles, mixed disulfide formation or radical-initiated thiol-ene reactions. Alternatively, Cys can be converted to Dha followed by Michael-addition of suitable nucleophiles. Through this strategy, it is possible to create fully functional mimics of PTMs on histone proteins. **d**, Combined genetic code expansion with traceless NCL for the formation of isopeptide linked ubiquitin and protein substrates.

Figure 3. Bioorthogonal protein-modification chemistry for live cell imaging applications. **a**, Genetic encoding of a norbornene amino acid into a mammalian cell-surface protein followed by selective labelling with a tetrazine-modified version of the fluorescent dye TAMRA-X. **b**, *Trans*-cyclooctyne (TCO) was conjugated to a tumour-homing antibody and injected into a tumour-bearing mouse. Once the antibody had accumulated inside the tumour and cleared from healthy organs, a radiolabelled tetrazine was injected. The tetrazine and TCO rapidly reacted leading to retention of radioactivity at the tumour site that could be visualized by radio-imaging approaches. **c**, Encoding of an azido-amino acid into the acid-sensitive chaperone HdeA inside the *E. coli* periplasm. Bioorthogonal attachment of an alkyne derivative of the environmentally sensitive dye 4-N,N-dimethylamino-1,8-naphthalimide (4-DMN) to the protein was achieved using Cu(I)-catalyzed azide-alkyne cycloaddition. Upon pH dependent conformational changes in HdeA, 4-DMN changes its fluorescence properties enabling the measurement of changes in local pH. **d**, Simultaneous genetic encoding of two distinct non-natural amino acids into the same protein. The tetrazine and norbornene amino acids did not react with each other but the tetrazine selectively reacted with a strained cyclooctyne ring and the norbornene with a different tetrazine. This mutually orthogonal and bioorthogonal pair of reactions allowed for the site-selectively placement of two distinct fluorescent dyes onto the surface of calmodulin in *E. coli* and measure its conformational changes in response to Ca^{2+} stimuli by FRET.

Figure 4. Example of strategies to achieve site-selective PEGylation of proteins. **a**, Examples of PEGylation of the N-termini using oxime ligation. N-

terminal aldehydes can be either generated through oxidation using sodium periodate or using pyridoxal-phosphate (PLP) promoted transamination. Subsequent oxime ligation with aminoxy PEG derivatives afford the N-terminally PEGylated protein. **b**, Site-selective attachment of 2-pyridinecarboxylaldehydes PEG derivatives to native proteins at neutral pH. **c**, Genetic encoding of *pAF* into hGH at pre-determined site 35 followed by oxime ligation with a aminoxy PEG derivative at acidic pH.

Figure 5. Site-selective protein-modification reactions for the synthesis of antibody-drug conjugates (ADCs). **a**, Reactions selective for Cys. The classical chemoselective reaction of maleimides with thiols is widely used for the site-selective synthesis of ADCs. The initial product of the reaction is now known to be unstable and can either regenerate the starting materials through a retro-Michael addition or hydrolyze to form the more stable ring-opened product. Ring opening can be facilitated by placing a positively charged amino acid in close vicinity to the modified Cys on the protein surface. Alternatively, a primary amine can be placed next to the maleimide on the linker between drug and antibody. Conversion of Cys into Dha on protein surfaces using **1** followed by Michael-addition with thiol-containing drugs results in a very stable thioether bond. Similarly, the Julia–Kocienski-type reaction of a protein surface Cys with a methylsulfonylbenzothiazole reagent **2** yields a site-selective protein conjugation product that is intrinsically very stable in plasma and may be very useful for the synthesis of ADCs. **b**, Bioorthogonal reactions selective for non-natural amino acids. **i**, Genetically introduced *pAF* on the antibody surface can selectively react with hydroxylamine-containing drugs **3** to form a very stable linkage. **ii**, Genetically or enzymatically introduced aldehydes on the antibody surface can selectively react with **4** in a Pictet–Spengler type reaction to form a very stable ADC.

Figure 6. Bioorthogonal approaches to in situ protein activation. **a**, A biocompatible palladium catalyst that cleave the propargyl carbamate group of a protected Lys analogue to generate a free lysine was used to modulate the function of an intracellular protein in a gain-of-function fashion. **b**, A TCO-caged Lys on the active site of a protein renders the protein inactive. Upon reaction

with 3,6-dimethyl-1,2,4,5-tetrazine, rapid inverse electron demand Diels–Alder elimination took place enabling bioorthogonal decaging and intracellular activation of the protein target in its native cellular context. **c**, Photodeprotection of a caged Cys that was genetically encoded into TEV to reveal the native protein in live cells.