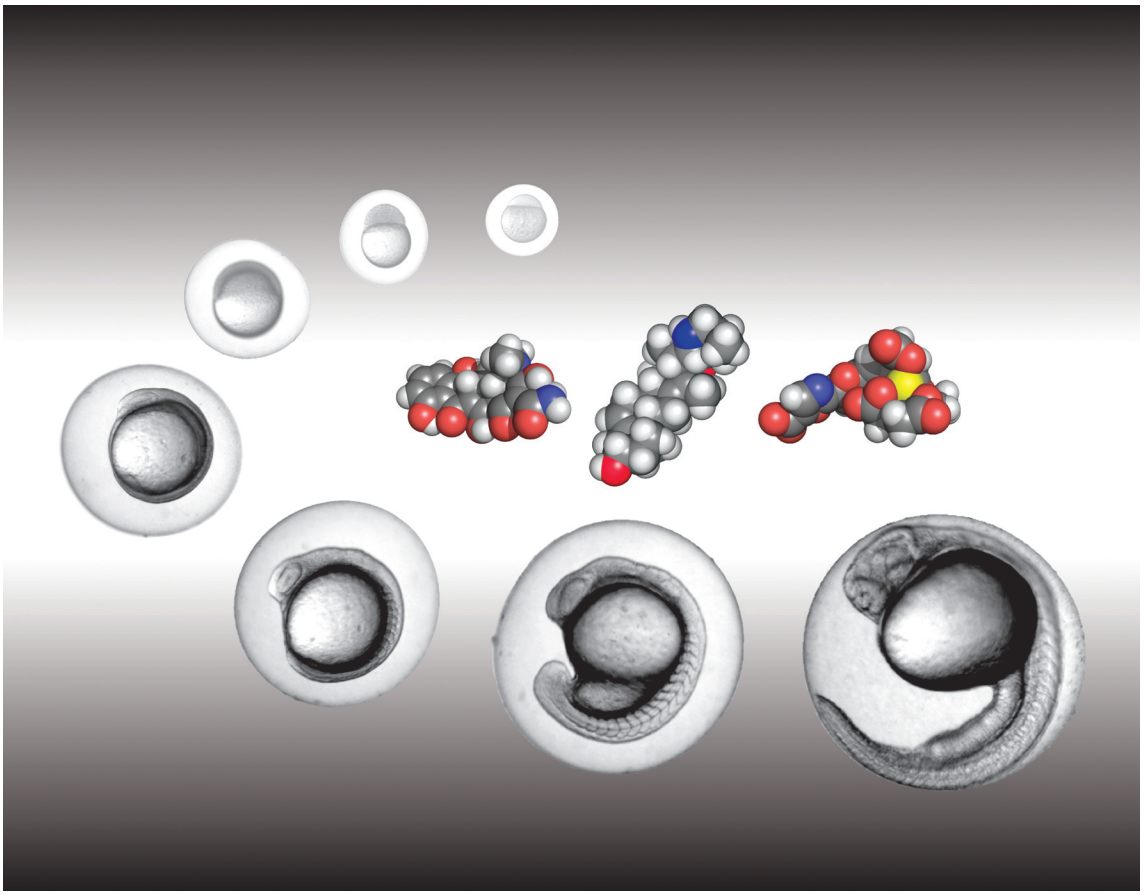
Chem Soc Rev

This article was published as part of the

2008 Chemistry–Biology Interface Issue

Reviewing research at the interface where chemistry meets biology

Please take a look at the full table of contents to access the other papers in this issue



Chemoselective modification of proteins: hitting the target†

Isaac S. Carrico*

Received 16th May 2008
First published as an Advance Article on the web 5th June 2008
DOI: 10.1039/b703364h

The use of synthetic molecules to modulate and track biological events is a central component of chemical biology. As a result, the precise, covalent modification of biomolecules is a key goal for this field. Several strategies have emerged that allow specific tailoring of polypeptides through either endogenous residues or introduced functionality. This *tutorial review* discusses these recent advances in the context of *in vitro* and *in vivo* applications.

Introduction

The twentieth century was witness to tremendous effort by chemists to develop tools for the synthesis and specific tailoring of small molecules. As a result, synthetic chemistry has advanced to a stage where virtually any small molecule can be made, yielding significant dividends in terms of both basic science and medicine. In an analogous fashion, developments in genetics provide precise control over not only the nature of protein products, but also timing and localization within organisms. This exquisite control has driven biological science over the last half-century. However, these approaches generally lack the flexibility of chemistry and do not necessarily allow sufficient control over post-translational modification of proteins. The preponderance of synthetic methodology, largely developed in organic solvents and with the exclusion of significant competing functionality, is unsuitable for modification of proteins. As a result, there remains a need for the precise chemical modification of biopolymers.

Department of Chemistry, Stony Brook University, 1 John S. Toll Road, Stony Brook, NY 11790-3400, USA.
E-mail: isaac.carrico@sunysb.edu; Fax: +1 631-632-7960;
Tel: +1 631-632-7935
† Part of a thematic issue examining the interface of chemistry with biology.

Isaac Carrico received his PhD from the California Institute of Technology where he worked under the direction of David Tirrell in the area of protein engineering. As a postdoctoral fellow, he worked under the guidance of Carolyn Bertozzi at the University of California, Berkeley. In 2006, he accepted a position at Stony Brook University, where he is an Assistant Professor of Chemistry. His current interests involve developing new methods for glycoproteomics and the optimization of protein and viral based therapeutics.

Discrimination against the natural 20 amino acids is necessary for the selective modification of proteins. However, this level of selectivity is not sufficient for chemistry in physiological contexts; these reactions also must tolerate ambient thiols, electrophiles and redox active metals. In addition, to ensure that chemical modification proceeds rapidly under biological conditions, high concentrations of reagents are often needed to offset low concentrations of the target, leading to unwanted perturbation of the system under study. Finally, the stability of the introduced reactants, as well as that of the resultant conjugates, is critical. Inherently difficult to achieve, this combination of speed, stability, small size and limited toxicity has limited the number of accessible chemoselective ligations. This tutorial review is designed to introduce the reader to reactions that satisfy these criteria and the approaches developed to install selective chemical functionality. This material is covered in three sections. The first details the chemoselective modification of natural proteinaceous functionality. While most of these reactions do not reach the level of specificity required for use *in vivo*, they are instrumental in constructing the majority of man-made protein conjugates. The next section describes the chemistry of introduced non-native functionality, which can be used in the contexts of cells. Finally, methods for the introduction of these chemoselective handles is detailed.

Modification of native functionality

Traditional protein chemistry, as defined by classic nucleophilic/electrophilic modification of the natural 20 amino acids, remains instrumental in modern biochemistry. Electrophilic modification of lysines, typically by NHS-esters and isocyanates, is the most common method for protein conjugation. Alternatively, the acidic side chains of glutamate and aspartate residues can be activated, generally with water-soluble carbodiimides, to produce activated esters. Subsequent nucleophilic addition yields stable bioconjugates. The density of surface exposed lysine, glutamic acid, and aspartic acid side chains generally limits their utility for the production of well-defined bioconjugates. Selective modification of cysteines with maleimides and α-halo reagents results in stable adducts. High specificity can be achieved due to the relatively low pK_a (~8) and potent nature of the thiolate anion. When used in

conjunction with site-specific cysteine introduction this approach is particularly powerful.

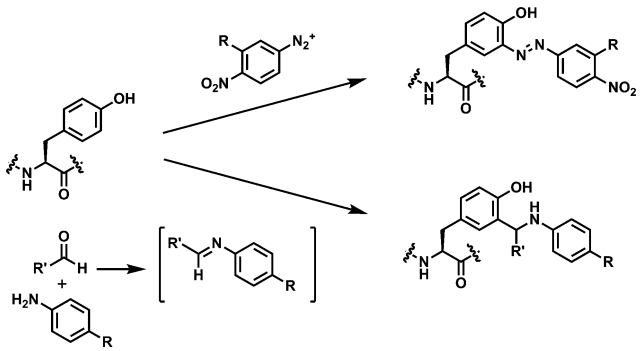
Alternatively, the Lewis basic character of several natural amino acids can be exploited to create selective coordination environments. Oligohistidine motifs, popular for protein purification, have been used to direct NTA-dye probes to target proteins. While this interaction is of relatively low affinity, it has been used intracellularly and extracellularly to visualize trafficking and discern structure (*via* fluorescence resonance energy transfer). The FlAsH tag, a higher affinity alternative, utilizes the introduction of a tetracysteine motif developed to specifically ligate a biarsinical ligand.[1] Genetic introduction of an optimized 12 amino acid sequence, incorporating four cysteines, provides an attachment point for a fluorescein based-dye containing two arsenic ligands. Coordination to the peptide motif generates a considerable increase in quantum yield, resulting in ligation-induced green fluorescence. Two alternative chromophores, ReAsH and CHoXAsH, can chelate the same amino acid motif as FlAsH and yield red and blue fluorophores, respectively.[2]

FlAsH and ReAsH have been used in combination to examine the maintenance of connexin-43 in the gap junctions of HeLa cells.[3] Incubation with ReAsH followed initial labeling with FlAsH. This allowed tracking of genetically tagged connexin as a function of lifetime, which indicated that new connexin is sent to the periphery of the junction, while older proteins are endocytosed from the center. Such studies would not be possible with fluorescent proteins as their maturation time is longer than that of protein recycling. An apparent drawback of this labeling scheme is arsenic based toxicity. Such toxicity can be largely ameliorated by the addition of 1,2-dithiols. While the optimized motif increases efficiency, native cysteine motifs can still generate significant background.

Selective tyrosine modification

As an alternative to lysine and cysteine labeling, chemistry that allows selective tyrosine labeling has been explored. Particularly notable are two sets of electrophilic aromatic substitution reactions specifically targeting tyrosine residues optimized by Francis and co-workers. The first involves coupling of the *ortho* position with diazonium salts resulting in the generation of azobenzene functionality (Scheme 1).[4] Electron deficient aromatics promote this reaction. 4-Nitrodiazonium compounds are particularly effective, demonstrating high coupling efficiencies (>90%) at 4 °C on a very short time scale (15 min). However, this reaction is not efficient at physiological pH values, optimal reactivity is obtained at pH 9.

The second method is a Mannich modification of tyrosines *via* exposure to activated imines (Scheme 1).[5] The nature of both the aniline and aldehyde coupling partners impacts the degree of coupling. Predictably, anilines with electron donating substituents incapable of delocalization (*i.e.* aliphatic) provide the best substrates. While formaldehyde generally generates the highest degree of tyrosine modification, pyruvaldehyde and glycoxylic acid have also demonstrated effectiveness, whereas enolizable aldehydes typically do not give the desired product. Advantageously, this methodology demonstrates high selectivity and proceeds rapidly at near physiological pH (6.5); however coupling efficiency is significantly lower than modification with diazonium salts. The authors point out that aldehyde-mediated crosslinking experiments are typically done at higher temperatures with significantly higher aldehyde concentrations. In addition, aldehyde lysine modification has not been detected under these conditions.



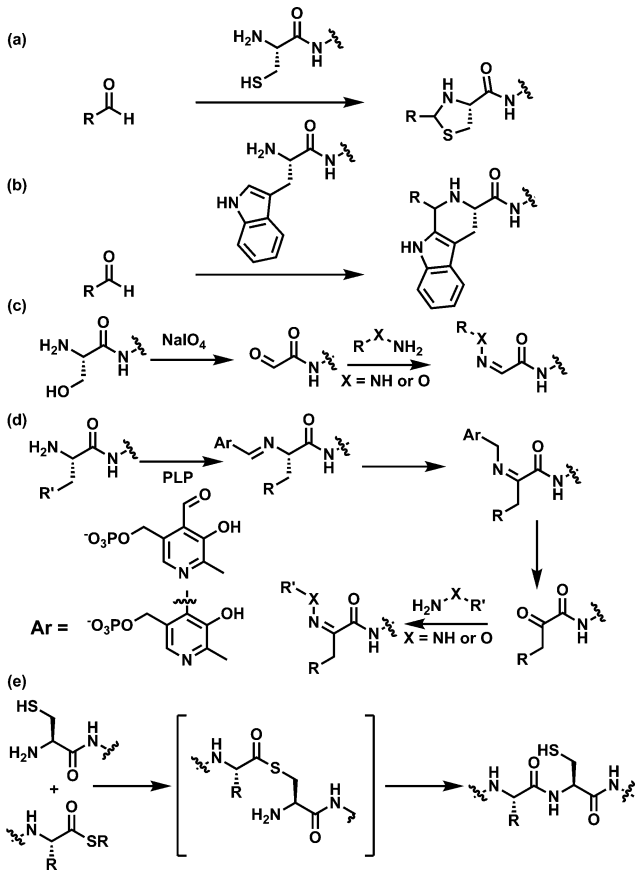
Scheme 1 Selective electrophilic aromatic ligation of tyrosine residues.

N-terminal protein modification

Derivatization of N-termini has presented an attractive target as the pK_a of the terminal amine (~8) is significantly lower than that of lysine (~10). As a result, Electrophilic modification at near neutral pH can afford selective functionalization of the N-terminus. However, in practice this selectivity can be difficult to achieve. Alternatively, functionality vicinal to the N-terminal amine can be used to further distinguish it from lysine residues. In the presence of aldehydes, N-terminal cysteines readily form thiazolidines, which are stable under physiological conditions (Scheme 2(a)). Similarly, Pictet–Spengler cyclizations can be accessed by N-terminal tryptophans in the presence of suitable aldehydes (Scheme 2(b)). Alternatively, useful two-step N-terminal modifications have been developed. Such methods require initial activation of the N-terminus followed by labeling with a suitably reactive tag. Generation of N-terminal aldehydes *via* oxidation of N-terminal serines by periodate has been known for decades (Scheme 2(c)).[6] This β-amino hydroxyl group is unique within proteins and is efficiently oxidized by periodate. The use of periodate has been demonstrated in a number of protein settings, however completely selective oxidation is often difficult to achieve.

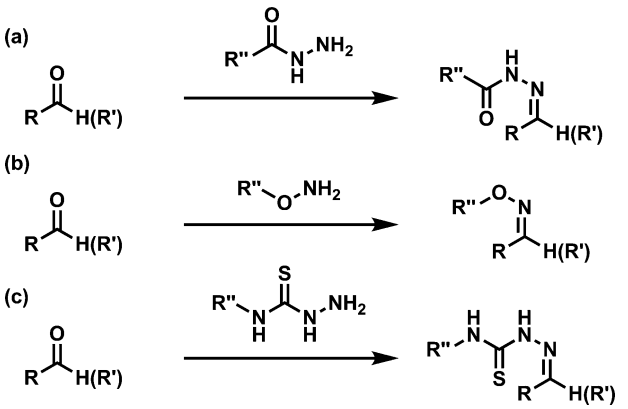
As an alternative, N-terminal aldehydes can be generated through a biomimetic approach. Addition of pyridoxal phosphate (PLP) yields imines with available amines. However, only at the N-terminus is there a proton α to the imine of sufficient acidity to allow enolization resulting in the formation of a new imine (Scheme 2(d)). Hydrolysis results in formation of a ketone and release of pyridoxal amine.[7] This modification is stable and avoids the use of oxidants, but is context dependent. Currently, this dependence is not well-understood and investigations are underway to determine the source of discrimination.

Native chemical ligation (NCL) is the most well developed class of N-terminal chemical modification. Initially,



Scheme 2 Highly selective N-terminal protein modifications. (a–b) aldehyde modification of N-terminal cysteine and tryptophan resides. (c–d) N-terminal generation of backbone linked aldehydes *via* sodium periodate or PLP treatment. (e) Native chemical ligation.

C-terminal thioesters transesterify in the presence of N-terminal cysteines (Scheme 2(e)). The resultant conjugates spontaneously undergo an S- to N-acyl shift, yielding a stable amide linkage. NCL has been widely used to create fully or partially synthetic proteins. Thioester and cysteine terminating peptides can be generated synthetically, however the necessity for either very long synthetic peptides or several sites of ligation, all at an internal cysteines, inherently limit this fully synthetic approach for large proteins.[8] In order to access larger targets, either fragment can be biosynthetically generated. Programmed proteolysis sites have been used to produce expressed protein fragments with N-terminal cysteines. To access protein domains with C-terminal thioesters, proteins can be expressed as target–intein fusions, which will spontaneously undergo N- to S-acyl transfer to generate internal thioester linkages between the target proteins and the attached intein. Exposure of this intermediate to high concentrations of small molecule thiols allows transthioesterification yielding modifiable, C-terminal thioesters. Using the latter approach, termed expressed protein ligation (EPL), Cole and co-workers were able to synthesize the protein tyrosine phosphotase SHP-2 with non-hydrolyzable phosphorylated tyrosine mimics in place of two naturally occurring tyrosines.[9] The sites of substitution had been highlighted as important in stimulation of the mitogen activated kinase pathway, however the role of each site on activity was not known. Introduction of these semi-synthetic constructs into cells allowed determination of the regulatory roles of each of these post-translationally modified residues with respect to stimulation of the mitogen activated kinase pathway.



Scheme 3 Aldehyde/ketone modification with hydrazide (a), aminooxy (b) and thiosemicarbazide (c) reagents yield hydrazones, oximes and thiosemicarbazones, respectively.

Modification of non-native functionality

Ketone and aldehyde modification

While first utilized in a biological setting decades ago, nucleophilic modification of aldehydes and ketones is still in wide use. In this case selectivity is a result of the privileged nature of these electrophiles within proteins. Aldehydes and ketones can be placed into proteins *via* a variety of methods including chemical oxidation of terminal residues (*vide supra*), introduction of unnatural amino acids, sugars and enzymatic modification of peptide motifs (*vide infra*). While free amines can readily react with aldehydes and ketones, equilibrium generally favors the free carbonyl. In contrast hydrazide, aminooxy and thiosemicarbazide compounds, react rapidly to form stable Schiff bases, hydrazone, oxime and thiosemicarbazones, respectively (Scheme 3). These reactions proceed in high yield in completely aqueous conditions at ambient temperatures. Modification at a pH above 6 generally yields sluggish reactions, which is a primary hurdle for these reactions in the context of biological systems. As a result of their reversible nature, such reactions generally tolerate a wide range of competing functionality. Alternative nucleophiles (*i.e.* β-amino thiols) can access these electrophiles, but they are efficiently displaced by hydrazine, semithiocarbazide and aminooxy compounds. Importantly, the resultant hydrazones, thiosemicarbazones and oximes demonstrate excellent stability under physiological conditions.[10] Intracellularly, formation of stabilized Schiff bases is compromised by the presence of significant amounts of aldehydes and ketones. However, cell surface labeling can proceed readily and has been demonstrated in several formats detailed below.

For some time ketone and aldehyde modification represented the best method for peptide ligation without the need for side chain protection. As a result, several early attempts to create fully synthetic proteins relied upon this ligation.[6] Eventually, these efforts were replaced by native chemical ligation, which allows the creation of an amide bond and thus the

production of completely native synthetic proteins. In the context of this new development, ketone ligation has been used to further modify fully synthetic proteins. Kent and co-workers have combined these two techniques to create a fully synthetic mimic of human erythropoietin, a glycosylated protein hormone that stimulates the production of erythroid cells.[11] NCL was used to create the full-length polypeptide product with specifically placed ketones. These ketones were used as attachment points for aminooxy terminated polymers designed to mimic natural glycans, which are critical for biological activity. Notably, these analogs have excellent activity, demonstrating that, fully native structure is not a requirement for biological activity.

Azide modification

Staudinger ligation. While the natural abundance of ketones and aldehydes limit their intracellular applicability, azides represent an important step towards the ideal chemical handle. Intrinsically azides are small, kinetically stable under physiological conditions, easily approachable synthetically, have significant accessible potential energy and are well tolerated by biosynthetic processes. As a result, azides are largely introduced by metabolic incorporation of unnatural amino acids and monosaccharides (*vide infra*). Reduction of aryl azides is generally slow in the presence of monothiols (glutathione, β-mercaptoethanol) but significant in the context of dithiothreitol.[12] However, no evidence of reduction of *alkyl* azides under physiological conditions has been found.

In 2000, Bertozzi and Saxon introduced the first chemoselective ligation that utilized azides as reactive partners under physiological conditions.[13] This reaction is a variant of the classical Staudinger reaction involving reduction of azides with tryarylphosphines that progresses through an azaylide intermediate. The Staudinger *ligation* utilizes a triaryl phosphine reagent with a proximal ester, which efficiently traps the generated azaylide intermediate (Scheme 4(a)). The resultant amide is stable and produced in good yields. Critically, neither the azide nor the modified triarylphosphine demonstrate cross reactivity with biological molecules. Substituted triarylphosphine reagents can be produced relatively easily, allowing appendage of a variety of reporter tags.

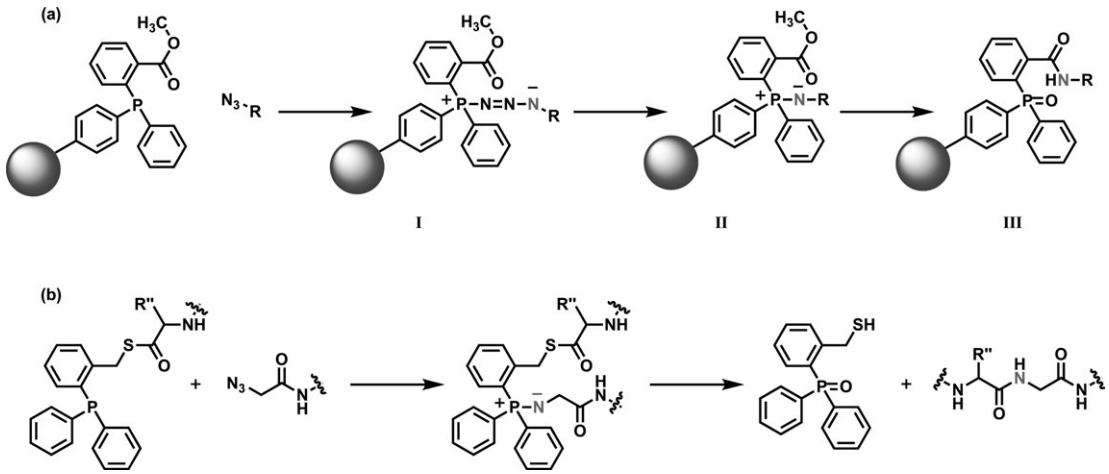
An interesting derivation, known as the traceless Staudinger ligation, uses a phosphinothioester to trap the azaylide intermediate (Scheme 4(b)). This results in the formation of an amide bond and expulsion of nitrogen as well as the oxidized phosphine reagent.[14] This is an attractive alternative to native chemical ligation as it can produce a fully native polypeptide without the need for cysteine at the site of ligation. However, a glycine is generally required on either the azide or thioester for efficient coupling.

Copper catalyzed azide–alkyne [3 + 2] cycloaddition (CuAAC)—"Click". Electrocyclizations constitute an alternative set of reactions accessible with azides. Exposure of these 1,3-dipoles to suitable dipolarophiles can result in highly exothermic conjugations. Terminal alkynes react with azides, formally a Huisgen electrocyclization, to produce substituted 1,2,3-triazoles in good yield. Notably, both reactants are stable under physiological conditions and are rarely found naturally. These reactions proceed readily in water and tolerate a wide range of functionality. However, despite the remarkably exothermic nature of these reactions, they require elevated temperatures to proceed at an appreciable rate. Further, a mixture of regioisomers is typically produced. While proximity accelerated variants of this reaction demonstrate significant rate enhancements, this requires molecular recognition and localization of both reactants, diminishing general utility.[15]

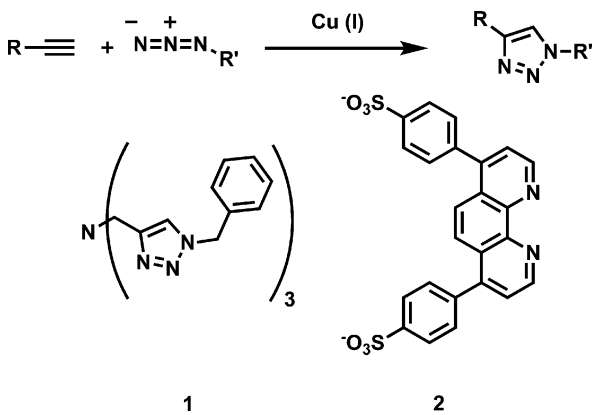
In 2002, two labs simultaneously reported dramatic acceleration of azide–alkyne [3 + 2] electrocyclizations by addition of a Cu(I) catalyst.[16,17] The catalyzed cycloaddition, thought to progress through a series of copper acetylide intermediates, produces the 1,4-regioisomer exclusively. However, the precise nature of catalysis is still under investigation. This reaction, dubbed Cu(I) accelerated azide–alkyne cycloaddition (CuAAC) but commonly called the "click" reaction, demonstrates excellent selectivity and has remarkable functional group tolerance. Akin to Staudinger ligations, CuAAC reactions proceed rapidly in water and are fairly insensitive to pH, although optimal kinetics are obtained between pH 7 and 9. Initially, active Cu(I) catalysts were generated *in situ* from Cu(II) salts and a reductant, typically copper wire and ascorbic acid or tris(2-carboxyethylphosphine) (TCEP). More recently bioconjugation reactions have used Cu(I) salts in the presence of a ligand (Scheme 5), which have been shown to significantly increase the rate of reaction. Easily accessible reagents, high conversion, and selectivity have made the CuAAC remarkably popular in a wide number of chemistry related fields.

While the use of CuAAC in proteome-wide settings has highlighted the excellent specificity associated with this reaction, complications involving Cu(I) catalysts have also been brought to light. Residual protein bound Cu(I) can interfere with mass spectrometric analysis. However, treatment with divalent metal chelators has been demonstrated to ameliorate this effect. In addition, ligands needed to stabilize the Cu(I) oxidation state are problematic. The most commonly used ligand to date, the tris-triazole ligand (**1**), is only marginally water soluble. A newer alternative, the sulfonated bathophenanthroline ligand (**2**),[18] demonstrates excellent water solubility and higher modification, but requires oxygen-free coupling conditions. More importantly, reliance on cytotoxic Cu(I) presents a significant hurdle for applications within biological systems.

Strain promoted alkyne–azide [3 + 2] electrocyclization. While high affinity ligands could diminish toxicity of CuAAC reactions, an alternative is activation of the dipolarophile without the need for copper. Electron deficient alkynes can dramatically accelerate this reaction, but such systems will likely act as Michael acceptors, limiting specificity. Alternatively, strain introduced into the alkyne can activate it for electrocyclization (Scheme 6). This was effectively demonstrated with variants of the smallest stable cycloalkyne, cyclooctyne. The strain in this system is significant ($\sim$18 kcal mol^{-1}) and partially relieved in the transition state. The first derivative made (**3**) demonstrated selective labeling of azides with comparable efficiency to the Staudinger ligation.[19] Poor



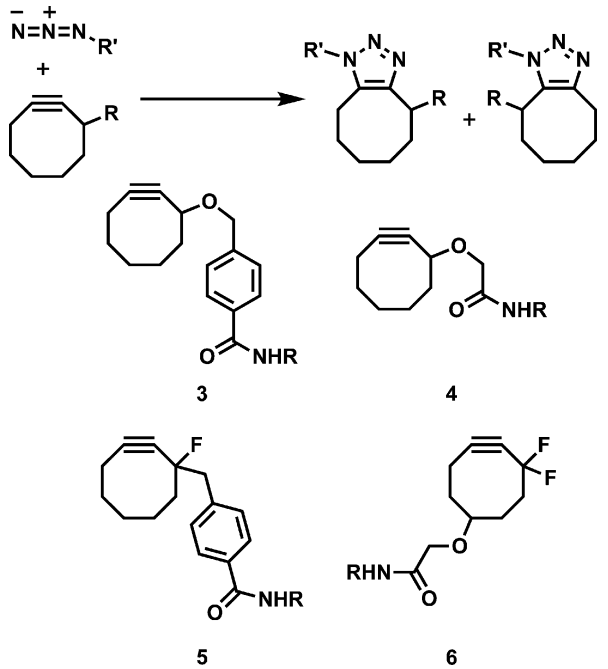
Scheme 4 Staudinger ligations. (a) Traditional Staudinger ligations proceed *via* initial nucleophilic attack on azides to form intermediate **I**. Loss of nitrogen yields the azaylide intermediate **II**, which is trapped by the neighboring methyl ester to form the final amide product **III**. (b) Traceless Staudinger ligations proceed in a similar fashion, but the azaylide can be trapped by an intramolecular thioester, resulting in loss of the oxidized triarylphosphine byproduct and formation of a native amide backbone.



Scheme 5 CuAAC (''click'') reaction and Cu(I) ligands.

solubility and slow kinetics stimulated the exploration of alternative cyclooctyne reagents. These derivatives removed the phenyl ring in the side chain and/or added fluorines designed to lower the energy of the LUMO and promote electrocyclization (Scheme 6, **4**–**6**). The best analog yet synthesized, dubbed difluorocycloalkyne or DIFO (**6**), has good solubility and demonstrates significantly faster kinetics in protein labeling experiments (*vide infra*). Unfortunately, all of the cycloalkyne analogs are somewhat difficult to access synthetically. DIFO in the free acid form, without fluorescent or epitope tags, requires 10 linear steps resulting in $\sim$2% overall yield.[20] However, this reaction is very attractive, in terms of chemoselective attributes and tunability, and as a result, is currently being synthetically optimized.

Contrasting azide ligations. All three reactions demonstrate excellent selectivity. Further, both of the resulting linkages (amide and triazole) are exceptionally stable. Significant differences are present in the synthetic accessibility of reagents. Terminal alkynes and azides are easily accessible synthetically. In addition, a number of these reagents are commercially available. Synthesis of Staudinger reagents is relatively facile, requiring only two steps to generate the free acid. It should be



Scheme 6 Strain promoted alkyne–azide [3 + 2] electrocyclizations.

noted that this reagent, as well as the epitope enabled FLAG-Phosphine, is commercially available. As discussed above the cyclooctyne reagents are the most difficult to produce. However, as the strain promoted process is the newest of the three, it is anticipated that the synthesis will be streamlined and that reagents will become commercially available.

Free Cu(I) associated toxicity is a significant concern with CuAAC ligations. In contrast, ligations with the strain promoted click and Staudinger reagents are amenable to cell culture and have even been demonstrated *in vivo*. While a full toxicity analysis has not yet been reported, it has been noted that mice injected with these reagents demonstrate no observable adverse effects. Live cell labeling with CuAAC chemistry has been reported, however cell viability after labeling was not explored.[21,22]

Kinetic comparisons were made between the Staudinger ligation and strain promoted click reaction. First- and second-generation cyclooctynes (Scheme 6: **3**, **4** and **5**) demonstrated similar kinetics to the Staudinger reaction, $k = 1.2–4.3 \times 10^{-3}$ M^{-1} s^{-1}, with benzyl azide in CH_3CN, whereas DIFO was significantly faster (20–60 fold).[23] Although it is difficult to make a direct comparison, as they are different order reactions, it is clear that CuAAC is significantly faster than the Staudinger ligation. DIFO demonstrates comparable reactivity to CuAAC under reasonable biological labeling conditions.[20] Electron-rich derivatives of the Staudinger reagent were generated to accelerate phosphine attack on the azide, the rate-determining step. However, these derivatives also significantly increase non-productive phosphine oxidation. In contrast, tuning the strain and electronics of the cyclooctyne derivatives shows promise in terms of generating faster electrocyclizations. Similarly, new catalyst design should aid solubility, oxidative stability, speed and may help limit the toxicity of the CuAAC reaction. However, rational design of new catalysts has been hampered by lack of a fully consistent mechanistic hypothesis. While these reactions are generally robust, it should be noted that urea and ionic detergents inhibit CuAAC and Staudinger ligations, respectively.

Introduction of non-native functionality

Unnatural amino acid incorporation

Unnatural amino acids can serve to tune protein function or serve as handles for subsequent modification. The two main incorporation methods have complementary strengths and weaknesses. Residue-specific incorporation is the older and more easily accessed method, however it lacks the control of site-specific introduction. This method relies on competition with the natural amino acids during the step at which they are appended to *t*RNA. The ribosome pairs mRNA codons with anticodons from charged *t*RNA with high fidelity; however it is largely insensitive to the nature of the amino acid appended to the 3′-end of the *t*RNA, leaving responsibility of correctly charging *t*RNA to a family of enzymes known as the aminoacyl *t*-RNA synthetases (aaRSs). While this family of enzymes has excellent fidelity against the other 19 natural amino acids, they demonstrate remarkable promiscuity *vs.* unnatural analogs. As a result, unnatural amino acids that are structurally and electronically similar to natural analogs can be charged to *t*RNA and are accordingly placed into growing polypeptides.[24] Furthermore, aaRSs with altered substrate specificity have been generated using both rational design and directed evolution,[25] leading to the introduction of analogs with larger structural and electronic deviations from the natural amino acid. A large number or residues have been introduced *via* these methods, but particularly pertinent to this review is the ability to incorporate azides, ketones and terminal alkynes as phenylalanine or methionine surrogates.[26] Notably, residue-specific incorporation can easily produce target proteins in gram quantities.

In contrast, biosynthetic site-specific unnatural amino acid incorporation generally depends upon charging a *t*RNA capable of suppressing a stop codon. In this way, stop codons introduced into mRNA transcripts dictate the placement of analogs within the target protein. The primary hurdle for this methodology is production of the modified suppressor *t*RNA. Two main solutions have been explored, chemical acylation of suppressor *t*RNA and development of a 21st *t*RNA/aaRS pair.[27] The most successful method for chemical acylation involves first chemically modifying the 3′-terminal CCA of the acceptor stem, followed by enzymatic ligation to form the full, modified suppressor *t*RNA. Target protein production is limited by the amount of conjugated *t*RNA that can be produced. As a result, this technique has been limited to applications in which small amounts of protein can be used effectively, *i.e.* ion channels.[28,29]

The second solution involves generation of charged suppressor *t*RNA *via* introduction of new metabolic machinery, namely an orthogonal *t*RNA/aaRS, into cells. In this context, orthogonality implies that the suppressor *t*RNA is not a substrate for any of the natural aaRSs and the introduced aaRS utilizes neither natural *t*RNA nor any natural amino acids. As both the *t*RNA and aaRS are engineered from natural homologs, this represents a significant challenge. However, significant success has been achieved in *E. coli*, particularly using exogeneous tyrosyl-*t*RNA synthetases. As a result, a large number of analogs have been incorporated, including chemoselectively available azides, alkynes and ketones. Furthermore, *t*RNA is not stoichiometric, as in the case of chemical acylation, and as a result generally useful quantities of proteins are routinely obtained (micrograms–milligrams).[30] It is possible to use this method within eukaryotic cells, however obtaining significant quantities of protein has been a challenge.[31,32] Recently, Wang *et al.* have increased the efficiency of yeast expression to *E. coli*-like levels (10–20 mg L^{-1}) by increasing the production of the engineered *t*RNA and reducing the rate of non-sense mediated mRNA decay.[33]

Metabolic incorporation of unnatural monosaccharides

Chemical functionality can be metabolically introduced into oligosaccharides in an analogous fashion to polypeptides. Unnatural monosaccharides can compete with their natural counterparts within metabolic pathways to produce activated unnatural sugars that can be used by glycosyltransferases. This strategy has been used to place chemoselective functionality into glycoproteins containing sialic acid (Sia), fucose (Fuc), *N*-acetylglucosamine (GlcNAc) and *N*-acetylgalactosamine (GalNAc). Mahal *et al.* first demonstrated this principle by introduction of a ketone containing sialic acid analog, SiaLev (Fig. 1(a)).[34] Incorporation of SiaLev was analyzed by modification of ketones exposed on the cell surface with a hydrazide probe. As a result, changes in global sialylation could easily be tracked. While this represented a powerful new tool to interrogate cells for changes in a post-translational modification, it was technically hampered by the need for chemical labeling at lower than physiological pH and low levels of incorporated ketone. Saxon and co-workers extended this methodology by introducing an azido analog, SiaNAz (Fig. 1(b)). A primary benefit of this approach is that azides tend to be better tolerated, metabolically.[35] This is illustrated by the substantially higher incorporation of SiaNAz in

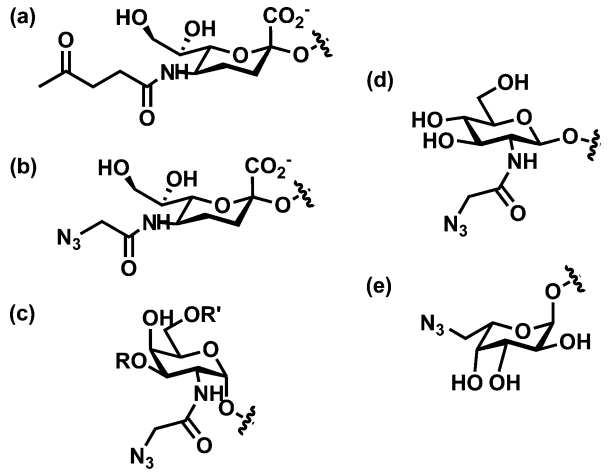


Fig. 1 Metabolically introduced unnatural monosaccharides amenable to chemoselective modification: (a) SiaLev, (b) SiaAz, (c) GalNAz, (d) GlcNAz, (e) FucAz.

comparison to SiaLev. Biosynthetic tolerance has allowed this strategy to be extended to new monosaccharide analogs, including azido analogs of *N*-acetylgalactosamine, *N*-acetylglucosamine, and fucose variants (Fig. 1(c)–(e)). Expansion of the accessible monomer pool has enabled delineation of different classes of glycans *via* their monomer composition.[36] Further, since introduction of the Staudinger ligation, novel azido amino acids and azido lipids[37] have been metabolically incorporated into proteins. It is notable that these new research initiatives were driven by the advent of the novel azide based reactions.

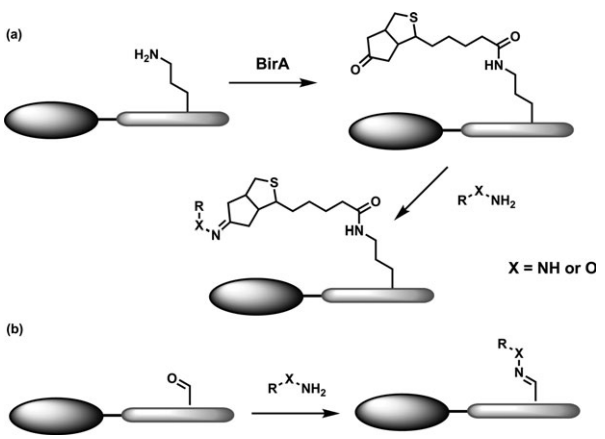
As a result of the excellent specificity and non-toxic nature of the triarylphosphine reagents, this reaction has been explored *in vivo*.[38] Mice treated with either ManNAz or GalNAz demonstrated dose responsive modification with Staudinger reagents. Initially metabolic labeling was carried out *in vivo* and chemical labeling was performed on tissue lysates to maximize triarylphosphine concentration and control for pharmacokinetic issues. Marked labeling was seen in a variety of tissues. Remarkably, *in vivo* Staudinger labeling provided equivalent signals to those samples labeled *ex vivo*. The efficiency of the completely *in vivo* labeling, both metabolic and chemical, opens up new avenues in imaging. As changes in glycosylation are known to be a hallmark of oncogenic transformation, *in vivo* imaging of these post-translational modifications is a significant development. Recently, Laughlin *et al*. metabolically labeled developing zebrafish embryos with GalNAz. Visualization of introduced azides was accomplished by strain promoted click ligation with fluorescent cyclooctyne derivatives. As a result, patterns of glycosylation development, previously unknown, could be monitored in real-time.[39]

Peptide tags

Site-specific protein modification can be accessed *via* the genetic appendage of "peptide tags" that direct functionalization within target proteins, either by direct molecular recognition of labeling reagents or *via* enzyme-mediated modification. Oligohistidine and FlAsH motifs are examples of the former, whereas enzymatic modification of peptide tags can result in the introduction of aldehydes, ketones and azides. As a result, these methods represent a complementary method to the biosynthetic incorporation of amino acids and sugars outlined above. Ideally peptide tags are small, minimally perturbing, can efficiently direct modification and allow conjugation with a wide array of functionality. Furthermore, efficiency of labeling should be independent of placement within the target protein. There are a number of excellent examples of peptide tags that do not focus on introduction of chemoselective functionality, notably tags that rely upon human *O*-alkylguanine transferase, peptidyl- and acyl-carrier protein fusions. These methods have been recently reviewed elsewhere and lie outside of the focus of this tutorial review.[40,41]

Specific enzymatic labeling can be achieved *via* biotin ligase (BirA), which transfers biotin to an internal lysine of a 15 amino acid acceptor peptide in an ATP dependent fashion. However, this approach requires recognition of the biotinylated target protein by a modified biotin binding protein (*i.e*. fluorescently labeled avidin) limiting the utility of a small peptide tag. To avoid this complication, Ting and co-workers developed a ketone containing biotin isostere tolerated by BirA (Scheme 7(a)).[42] Proteins tagged with this biotin analog allow subsequent chemical labeling with a wide range of commercially available aminooxy and hydrazide compounds. Importantly, the BirA strategy is highly specific and allows labeling with a wide array of functionality. However, it depends on an enzymatic labeling step, necessitating recombinant production of the enzyme and synthesis of the commensurate synthetic substrate.

An alternative approach makes use of the so-called aldehyde tag, which relies upon the co/post-translational oxidation of a cysteine within a targeting sequence to generate an aldehyde. The peptide sequence is derived from sulfatase active site sequences, which are known to contain a formylglycine residue as the catalytic nucleophile. This residue is produced post-translationally *via* the oxidation of a cysteine by the formyl glycine generating enzyme (FGE). Structural analysis of FGE indicates recognition of a linear sequence around the sulfatase active site,[43] which suggested that a short



Scheme 7 Chemoselective modification of peptide tags. (a) Enzymatic modification of biotin accepting peptide with ketone isostere and subsequent selective chemical modification. (b) Aldehyde-tag maturation produces an internal aldehyde, which can be subsequently chemoselectively modified.

linear peptide sequence directs oxidation. In *E. coli*, a six amino acid sequence, LCTPSR, is sufficient to generate essentially complete oxidative conversion.[44] The FGE sequence is well conserved from prokaryotes to eukaryotes suggesting that such a tag may be efficient in many cell lines, however this has yet to be explored. A limitation of this method in eukaryotic systems is the need for the target protein to transit the endoplasmic reticulum, where FGE is localized. Further, while oxidation would be expected in most cell lines, overexpression of FGE may be required to obtain full conversion. Advantageously, this is the smallest peptide motif currently available and does not introduce multiple charges or cysteines, further minimization (<6 amino acids) would be difficult to achieve without inherent loss in specificity.

Concluding remarks

Protein conjugation has moved well beyond the confines of cysteine and lysine labeling. Today, we can create complex post-translationally modified proteins. However, these targets require significant expertise. While dramatically better analytical tools have enabled investigations of inhomogeneous protein populations, in many cases synthetic homogeneous protein products are still needed for complete understanding of physiological function. This importance becomes particularly clear as we begin to fully appreciate the impact of post-translational modifications.

Over the last two decades, our ability to work within physiological settings has blossomed, however we are still limited to a small number of widely useful reactions. The need to track biological processes within living cells requires significant expansion of the repertoire of chemoselective reactions. Further, while exceptional, the azide modifying reactions still exhibit significant drawbacks. Speed, toxicity and synthetic accessibility will continue to hamper the application of these chemistries.

One particularly salient aim is the concerted imaging of multiple processes within cells. In this way multiple determinants can be analyzed in parallel, which will aid our understanding of interconnected molecular processes. In order to produce such a multidimensional experiment, new chemistries and chemical reporters need to be developed. Currently, we can access extracellular or cell surface biomolecules with ease *via* existing methodology and are beginning to make intracellular inroads. However, it is clear that in the short term, chemistry and chemical reporters will be challenged to complete this goal in isolation. Peptide tags utilizing enzymatic specificity, may be able to help satisfy this need.

References

1. B. A. Griffin, S. R. Adams and R. Y. Tsien, *Science*, 1998, **281**, 269–272.
2. S. R. Adams, R. E. Campbell, L. A. Gross, B. R. Martin, G. K. Walkup, Y. Yao, J. Llopis and R. Y. Tsien, *J. Am. Chem. Soc.*, 2002, **124**, 6063–6076.
3. G. Gaietta, T. J. Deerinck, S. R. Adams, J. Bouwer, O. Tour, D. W. Laird, G. E. Sosinsky, R. Y. Tsien and M. H. Ellisman, *Science*, 2002, **296**, 503–507.
4. T. L. Schlick, Z. B. Ding, E. W. Kovacs and M. B. Francis, *J. Am. Chem. Soc.*, 2005, **127**, 3718–3723.
5. N. S. Joshi, L. R. Whitaker and M. B. Francis, *J. Am. Chem. Soc.*, 2004, **126**, 15942–15943.
6. H. F. Gaertner, K. Rose, R. Cotton, D. Timms, R. Camble and R. E. Offord, *Bioconjugate Chem.*, 1992, **3**, 262–268.
7. J. M. Gilmore, R. A. Scheck, A. P. Esser-Kahn, N. S. Joshi and M. B. Francis, *Angew. Chem., Int. Ed.*, 2006, **45**, 5307–5311.
8. D. Bang, B. L. Pentelute and S. B. H. Kent, *Angew. Chem., Int. Ed.*, 2006, **45**, 3985–3988.
9. W. Lu, D. Q. Gong, D. Bar-Sagi and P. A. Cole, *Mol. Cell*, 2001, **8**, 759–769.
10. H. C. Hang and C. R. Bertozzi, *Acc. Chem. Res.*, 2001, **34**, 727–736.
11. G. G. Kochendoerfer, S. Y. Chen, F. Mao, S. Cressman, S. Traviglia, H. Y. Shao, C. L. Hunter, D. W. Low, E. N. Cagle, M. Carnevali, V. Gueriguian, P. J. Keogh, H. Porter, S. M. Stratton, M. C. Wiedeke, J. Wilken, J. Tang, J. J. Levy, L. P. Miranda, M. M. Crnogorac, S. Kalbag, P. Botti, J. Schindler-Horvat, L. Savatski, J. W. Adamson, A. Kung, S. B. H. Kent and J. A. Bradburne, *Science*, 2003, **299**, 884–887.
12. J. V. Staros, H. Bayley, D. N. Standring and J. R. Knowles, *Biochem. Biophys. Res. Commun.*, 1978, **80**, 568–572.
13. E. Saxon and C. R. Bertozzi, *Science*, 2000, **287**, 2007–2010.
14. E. Saxon, J. I. Armstrong and C. R. Bertozzi, *Org. Lett.*, 2000, **2**, 2141–2143.
15. W. L. Mock, T. A. Irra, J. P. Wepsiec and M. Adhya, *J. Org. Chem.*, 1989, **54**, 5302–5308.
16. V. V. Rostovtsev, L. G. Green, V. V. Fokin and K. B. Sharpless, *Angew. Chem., Int. Ed.*, 2002, **41**, 2596–2599.
17. C. W. Tornoe, C. Christensen and M. Meldal, *J. Org. Chem.*, 2002, **67**, 3057–3064.
18. S. Sen Gupta, J. Kuzelka, P. Singh, W. G. Lewis, M. Manchester and M. G. Finn, *Bioconjugate Chem.*, 2005, **16**, 1572–1579.
19. N. J. Agard, J. A. Prescher and C. R. Bertozzi, *J. Am. Chem. Soc.*, 2004, **126**, 15046–15047.
20. J. M. Baskin, J. A. Prescher, S. T. Laughlin, N. J. Agard, P. V. Chang, I. A. Miller, A. Lo, J. A. Codelli and C. R. Bertozzit, *Proc. Natl. Acad. Sci. U. S. A.*, 2007, **104**, 16793–16797.
21. K. E. Beatty, J. C. Liu, F. Xie, D. C. Dieterich, E. M. Schuman, Q. Wang and D. A. Tirrell, *Angew. Chem., Int. Ed.*, 2006, **45**, 7364–7367.
22. A. J. Link and D. A. Tirrell, *J. Am. Chem. Soc.*, 2003, **125**, 11164–11165.
23. N. J. Agard, J. M. Baskin, J. A. Prescher, A. Lo and C. R. Bertozzi, *ACS Chem. Biol.*, 2006, **1**, 644–648.
24. R. E. Connor and D. A. Tirrell, *Polym. Rev.*, 2007, **47**, 9–28.
25. D. Datta, P. Wang, I. S. Carrico, S. L. Mayo and D. A. Tirrell, *J. Am. Chem. Soc.*, 2002, **124**, 5652–5653.
26. K. Kirshenbaum, I. S. Carrico and D. A. Tirrell, *ChemBioChem*, 2002, **3**, 235–237.
27. M. W. Nowak, P. C. Kearney, J. R. Sampson, M. E. Saks, C. G. Labarca, S. K. Silverman, W. Zhong, J. Thorson, J. N. Abelson, N. Davidson, P. G. Schultz, D. A. Dougherty and H. A. Lester, *Science*, 1995, **268**, 439–442.
28. D. A. Dougherty, *Curr. Opin. Chem. Biol.*, 2000, **4**, 645–652.
29. M. W. Nowak, J. P. Gallivan, S. K. Silverman, C. G. Labarca, D. A. Dougherty and H. A. Lester, *Ion Channels, Part B*, 1998, **293**, 504–529.
30. J. M. Xie and P. G. Schultz, *Nat. Rev. Mol. Cell Biol.*, 2006, **7**, 775–782.
31. Z. W. Zhang, L. Alfonta, F. Tian, B. Bursulaya, S. Uryu, D. S. King and P. G. Schultz, *Proc. Natl. Acad. Sci. U. S. A.*, 2004, **101**, 8882–8887.
32. S. X. Ye, C. Kohrer, T. Huber, M. Kazmi, P. Sachdev, E. C. Y. Yan, A. Bhagat, U. L. RajBhandary and T. P. Sakmar, *J. Biol. Chem.*, 2008, **283**, 1525–1533.
33. Q. Wang and L. Wang, *J. Am. Chem. Soc.*, 2008, **130**, 6066–6067.
34. L. K. Mahal, K. J. Yarema and C. R. Bertozzi, *Science*, 1997, **276**, 1125–1128.
35. S. J. Luchansky, H. C. Hang, E. Saxon, J. R. Grunwell, C. Y. Danielle, D. H. Dube and C. R. Bertozzi, Recognition of Carbohydrates in Biological Systems: Part A: General Procedures, *Methods Enzymol.*, 2003, **362**, 249–272.
36. H. C. Hang, C. Yu, D. L. Kato and C. R. Bertozzi, *Proc. Natl. Acad. Sci. U. S. A.*, 2003, **100**, 14846–14851.

37. B. P. Duckworth, J. H. Xu, T. A. Taton, A. Guo and M. D. Distefano, *Bioconjugate Chem*., 2006, **17**, 967–974.
38. J. A. Prescher, D. H. Dube and C. R. Bertozzi, *Nature*, 2004, **430**, 873–877.
39. S. T. Laughlin, J. M. Baskin, S. L. Amacher and C. R. Bertozzi, *Science*, 2008, **320**, 664–667.
40. L. W. Miller and V. W. Cornish, *Curr. Opin. Chem. Biol*., 2005, **9**, 56–61.
41. I. Chen and A. Y. Ting, *Curr. Opin. Biotechnol*., 2005, **16**, 35–40.
42. I. Chen, M. Howarth, W. Y. Lin and A. Y. Ting, *Nat. Methods*, 2005, **2**, 99–104.
43. T. Dierks, A. Dickmanns, A. Preusser-Kunze, B. Schmidt, M. Mariappan, K. von Figura, R. Ficner and M. G. Rudolph, *Cell*, 2005, **121**, 541–552.
44. I. S. Carrico, B. L. Carlson and C. R. Bertozzi, *Nat. Chem. Biol*., 2007, **3**, 321–322.