A molecular engineering toolbox for the structural biologist

Galia T. Debelouchina and **Tom W. Muir**[*]

Department of Chemistry, Princeton University, Princeton, NJ 08540, USA

Abstract

Exciting new technological developments have pushed the boundaries of structural biology, and have enabled studies of biological macromolecules and assemblies that would have been unthinkable not long ago. Yet, the enhanced capabilities of structural biologists to pry into the complex molecular world have also placed new demands on the abilities of protein engineers to reproduce this complexity into the test tube. With this challenge in mind, we review the contents of the modern molecular engineering toolbox that allow the manipulation of proteins in a site-specific and chemically well-defined fashion. Thus, we cover concepts related to the modification of cysteines and other natural amino acids, native chemical ligation, intein and sortase-based approaches, amber suppression, as well as chemical and enzymatic bio-conjugation strategies. We also describe how these tools can be used to aid methodology development in X-ray crystallography, nuclear magnetic resonance, cryo-electron microscopy and in the studies of dynamic interactions. It is our hope that this monograph will inspire structural biologists and protein engineers alike to apply these tools to novel systems, and to enhance and broaden their scope to meet the outstanding challenges in understanding the molecular basis of cellular processes and disease.

1. Introduction

Since the first crystal structure of myoglobin (Kendrew *et al.* 1958), the three-dimensional (3D) reconstruction image of T4 bacteriophage tails by electron microscopy (EM) (De Rosier & Klug, 1968), and the solution nuclear magnetic resonance (NMR) structure of proteinase inhibitor IIA (Williamson *et al.* 1985), structural biology has made tremendous strides toward revealing intimate atomic level details that guide the function of biological molecules. We live at a time when we know the structures of more than 120 000 (and counting – Source: http://www.rcsb.org/pdb/statistics/holdings.do) biological macromolecules, when we can visualize the inner workings of the ribosome (Ben-Shem *et al.* 2011), or the nucleosome interactions that preserve the integrity and identity of our genome (Luger *et al.* 1997). At the same time, advances in instrumentation engineering have pushed the frontiers of structural biology methodologies and have allowed experiments and accomplishments that would have been unthinkable 30 years ago. Thus, it is now possible to record high-resolution movies of fast protein motions using X rays (Tenboer *et al.* 2014), obtain cryo-EM electron density maps at sub 3 Å resolution (Campbell *et al.* 2015; Merk *et*

[*]Author for correspondence: T. W. Muir, Department of Chemistry, Princeton University, Princeton, NJ 08540, USA. muir@princeton.edu.

Author Manuscript   Author Manuscript   Author Manuscript   Author Manuscript

al. 2016), or record multidimensional NMR spectra of protein crystals (Igumenova *et al.* 2004). Yet, the task in front of the structural biologist is getting harder and harder. The wealth of structural, biochemical and biological data has revealed that many mammalian cellular proteins are very large (>50 kD) (Brocchieri & Karlin, 2005), that they are often part of complex assemblies composed of many interchangeable molecular players, and that their function is often defined and regulated by an intricate layer of post-translational modifications (PTMs). In addition, many disease-related biological macromolecules do not have a defined secondary or tertiary structure at all, and function, instead, through intrinsic disorder and numerous weak, transient interactions (Hyman *et al.* 2014; Tompa, 2012). To make sense of this complicated, multilayered and often chaotic biological world, the structural biologist will become more and more dependent on the ability of protein engineers to faithfully and efficiently reproduce this complexity in the test tube.

Analogous to the advances in instrumentation design and engineering that have allowed structural biology to travel far, the tools of protein engineering have also become much more sophisticated, efficient and ultimately broader in scope over time. It is now possible to routinely synthesize polypeptide chains that are 50 amino acids long, to stitch them together into much longer chains without leaving any chemical scars (Dawson *et al.* 1994), and to decorate them with PTMs, biophysical probes and chemical moieties that perturb or enhance their function. It is also possible to 'persuade' the cellular protein synthesis machinery to produce polypeptide chains incorporating completely unnatural amino acids, thus expanding the genetic code of engineered living organisms (Wang *et al.* 2001). The current protein engineering toolbox contains many biocompatible chemical reactions, proteins with unique polypeptide 'stitching' abilities, and concepts and ideas that might ultimately prove essential in solving the interesting and relevant structural biology problems of today (Fig. 1). As the structural biologist might not be aware of all the current developments in protein chemistry, we intend this review as a resource that describes the state-of-the-art protein engineering tools, keeping an eye on the past and future to provide context for their limitations and the exciting new possibilities that undoubtedly lie ahead. We start with a very brief overview of recent advances in X-ray crystallography, cryo-EM and NMR, and outline challenges where the tools of protein engineering might be the most impactful. We then describe the contents of the molecular engineering toolbox that allow the construction of large modified proteins and complex macromolecular assemblies. We continue with a discussion of the concepts and ideas that directly concern structural biology methodology development. Our monograph ends with an outlook toward emerging trends in structural and chemical biology and exciting new developments that will guide the two fields in the future.

2. Advances and challenges in structural biology

2.1 X-ray crystallography

The workhorse of structural biology, X-ray crystallography, is more than 100 years old and has contributed nearly 90% of the macromolecular structures deposited in the PDB (http://www.rcsb.org/pdb/statistics/holdings.do). Protein engineering has long been part of the everyday life of the crystallographer as mutations, truncations and fusion proteins are often required to 'trick' proteins into adopting a crystal form. There are many crystal structures of

proteins containing PTMs or their analogs, and chemical approaches are often used to trap interesting functional states, stabilize dynamic interactions or aid the formation of crystals (e.g. racemic crystallography (Yeates & Kent, 2012)). Yet, the voracious need to test hundreds if not thousands of single crystal growth conditions has certainly challenged the protein chemist to optimize her tools and deliver relevant samples with much greater yields. Recent instrumentation developments such as X-ray free-electron lasers may potentially alleviate this need as these sources allow the acquisition of room-temperature data from easier to obtain micro-, nano- and 2D crystals (Neutze *et al.* 2015). Currently, there is also a growing demand for the construction of homogeneous, chemically well defined and stable complex biological assemblies such as those relevant for chromatin biology, for example.

2.2 Cryo-EM

Exciting developments in the last few years have propelled cryo-EM into the spotlight and turned this method into a mainstream and vital structural biology technique that can achieve crystallographic resolution (Cheng, 2015; Nogales, 2016). The commercialization of direct electron-detection cameras has allowed the acquisition of images with higher contrast and fast readouts that can overcome beam-induced motion and radiation damage (Brilot *et al.* 2012; McMullan *et al.* 2009). On the other hand, improvements in data analysis approaches have made it possible to characterize heterogeneous samples and even rare structural states (Fernandez *et al.* 2013; Scheres, 2012). Coupled with other cryo-EM advantages (no need for crystallization and only small amounts of sample required), these advances have made it possible to obtain subnanometer (and in some cases <3 Å) resolution maps of integral membrane proteins (Matthies *et al.* 2016), biological polymers (von der Ecken *et al.* 2015), chromatin (Song *et al.* 2014), as well as biological assemblies such as the transcription and translation initiation complexes (Fernandez *et al.* 2013; He *et al.* 2013; Plaschka *et al.* 2016). In this context, protein chemistry and engineering can have a tremendous benefit for the cryo-EM structural biologist in the design and construction of relevant biological samples such as post-translationally modified proteins. Perhaps more importantly, however, chemical biology approaches such as cross-linking can allow the preparation of samples that are more robust and do not fall apart during sample vitrification. The addition of cross-linkers can also be extremely useful in integrated cryo-EM/mass-spec structural approaches for samples where the high-resolution identification of protein–protein interfaces might not be possible (Leitner *et al.* 2016).

2.3 NMR spectroscopy

NMR spectroscopy detects the magnetic properties of nuclei in molecules, which in turn provide a window into their surrounding chemical environment. Uniquely suited to probe molecular structure and dynamics in solution at physiologically relevant conditions (temperature, pH and salts) and ultimately non-destructive in its readout, NMR spectroscopy has long been battered by its intrinsically low sensitivity. The introduction of 'NMR-visible' isotopic labels into biological macromolecules has become a standard practice in the field, and efficient molecular engineering approaches that allow the installation of nuclear isotopes at specific positions within the polypeptide or polynucleotide chain are highly desirable. Recent advances such as (methyl)-TROSY and dark-state exchange saturation transfer experiments have pushed the molecular size limits of solution NMR into the MDa regime

(Fawzi *et al.* 2011; Pervushin *et al.* 1997; Tugarinov *et al.* 2003), while the rapid instrumentation and pulse sequence developments in magic angle spinning NMR have made it possible to pursue the structures of large biological polymers such as amyloid fibrils (Fitzpatrick *et al.* 2013; Lu *et al.* 2013; Wasmer *et al.* 2008), bacterial secretion needles (Loquet *et al.* 2012), membrane proteins embedded in their native lipid environments (Cady *et al.* 2010; Wang *et al.* 2013), or even the molecular composition of bones (Chow *et al.* 2014). Thus, molecular engineering approaches can have a profound impact on the assembly of homogeneous, isotopically labeled and yet native substrates for structural investigation by *in vitro* NMR. Also, uniquely suited to probe structure and dynamics in the cellular milieu (Burz *et al.* 2006; Frederick *et al.* 2015; Inomata *et al.* 2009; Sakakibara *et al.* 2009), NMR spectroscopy can benefit tremendously from chemical and molecular biology techniques that allow the specific isotopic labeling of macromolecules in the cell.

3. Molecular engineering toolbox for complex biological samples

Before we delve into the chemistry, it is important to note that the methods described below complement the well-established molecular biology framework that allows the manipulation of protein sequences at the genetic level. Such manipulations can now be achieved in several different organisms ranging from bacteria (*Escherichia coli, Lactococcus lactis*), yeast (*Saccharomyces cerevisiae, Pichia pastoris*), insect cells, and stable mammalian expression cell lines such as HEK293 and CHO. Thus, we will start with review of the methodologies that can selectively modify natural amino acids introduced at specific protein positions with site-directed mutagenesis. We will then describe tools that can be used to ligate modified peptides and proteins into longer polypeptide chains, including native chemical ligation (NCL), inteins and transpeptidases. We will then discuss the molecular engineering toolbox afforded by incorporation of unnatural amino acids by amber suppression. These chemical and genetic tools give chemists the ability to position bio-orthogonal reactive handles into polypeptide chains with extraordinary precision and control, and we will end this section with discussion of bioconjugation approaches that take advantage of this rapidly evolving expertise.

3.1 Cysteine chemistry

The field of protein chemistry would have been very different in a world without cysteines. Indeed, many of the protein chemistry tools that exist today have been developed to exploit the reactivity of the cysteine sulfhydryl group that uniquely stands out among a sea of other protein side-chains. Cysteines are relatively rare in Nature (<2% abundance), and their high nucleophilicity makes them good candidates for the development of selective chemical reactions that work well in protein compatible conditions (aqueous solution, physiological pH and temperature). Generally, these reactions can be divided into three types: (1) cysteine alkylations, (2) oxidations and (3) desulfurization reactions, each providing a unique way to exploit the reactivity of this amino acid and a pathway to build proteins with distinct and desirable properties (reviewed in (Chalker *et al.* 2009; Spicer & Davis, 2014)) (Fig. 2).

Alkylation reactions have long been used to modify cysteine containing proteins and rely on familiar reagents such as iodoacetamide and maleimide. By careful control of the buffer pH,

these electrophiles can selectively react with the cysteine sulfhydryl group and tether additional functionalities to the protein of interest. Iodoacetamides and other α-halocarbonyls, for example, have been used to attach carbohydrate moieties to proteins and create glycoprotein mimics (Macmillan *et al.* 2001; Tey *et al.* 2010). Maleimides, on the other hand, are commercially available and user-friendly means to add spectroscopic probes to proteins, including fluorescent, electron paramagnetic resonance (EPR) and paramagnetic relaxation enhancement (PRE) labels. Aminoethylation of cysteines also provides a cheap and convenient way to generate lysine methylation analogs, available in the mono-, di- and tri-methylation states (Simon *et al.* 2007). In recent years, the chemist's attention has turned toward the development of more advanced cysteine bioconjugation protocols that circumvent, for example, the reversibility of maleimide additions in the presence of external bases and thiols (Lyon *et al.* 2014), or provide the opportunity to conjugate additional functionalities (such as aryl groups) under an expanded range of reaction conditions (Vinogradova *et al.* 2015).

Cysteine oxidation in the protein context is usually associated with the formation of disulfide bonds, a unique structural transformation with often dramatic consequences to protein function. In the test tube, disulfide bonds are very easy to build – all that is required is basic pH and exposure to air. Furthermore, the reaction is fast and does not require large excess of reagents. Therefore, it comes as no surprise that protein chemists often exploit disulfide bonds as means to attach useful functionalities to proteins of interest. The challenging aspect of such protocols is to ensure that only the desired disulfide bonds are formed and the final product is not a mixture of homo- and heterodimer species. A common strategy used to alleviate this problem relies on controlled cysteine activation and disulfide exchange based on the lower pKa of aromatic thiols. In this case, a selected cysteine side-chain can be activated (and protected) at low pH with the aromatic thiol, and upon addition of the other thiol-containing component and increase in pH, the aromatic disulfide bond is exchanged with the desired connectivity (Pollack & Schultz, 1989; Rabanal *et al.* 1996). This strategy has found diverse applications ranging from the design of cytochrome peptides (Rabanal *et al.* 1996) to the tethering of ubiquitin moieties to various proteins (Chatterjee *et al.* 2010; Chen *et al.* 2010a; Meier *et al.* 2012). Other activating molecules include methanethiosulfonates, glycosyl and allylic thiols with applications in protein glycosylation, prenylation and sulfenylation (Gamblin *et al.* 2008; Grayson *et al.* 2005; van Kasteren *et al.* 2007).

While the formation of disulfide bonds is fast and facile, they are just as easily destroyed in the presence of common biochemical reducing reagents or in the cellular environment where glutathione is present at high concentrations. If compatible linkages are desired, it is possible to convert the disulfide to a thioether bond, which is stable under reducing conditions. This is essentially a desulfurization reaction that proceeds with the formation of a dehydroalanine intermediate. Dehydroalanines, on the other hand, can be useful stepping stones to a vast number of protein PTMs, reactions that will be discussed in more detail in Section 3.7. Another key desulfurization reaction involves the conversion of cysteine to alanine. This transformation, discussed in more detail in Section 3.3, is particularly important for NCL, as it allows the construction of polypeptide chains without cysteine 'scars'.

Cysteines are relatively rare in polypeptide chains and are usually essential for protein function. Thus, the outstanding challenge for the protein chemist is to find new approaches and reaction conditions that target only the desired residues in a polypeptide chain. Many of these efforts are focused on the search of suitable peptide sequences that can provide the necessary amino acid context for tuning side-chain reactivity. For example, Tsien and co-workers have developed tetra-cysteine motifs that can be selectively targeted with biarsenic reagents, including in living cells (Griffin *et al.* 1998). Pentelute and co-workers, on the other hand, have reported the so-called 'π-clamp' sequence (Phe–Pro–Cys–Phe) that reacts preferentially with perfluoroaromatic moieties in aqueous solvents (Zhang *et al.* 2016a). Excellent selectivity can be obtained by replacing cysteine with selenocysteine (reviewed in (Metanis *et al.* 2009; Yoshizawa & Bock, 2009), a rare natural amino acid that shares many of the desirable properties of the cysteine side-chain, yet is more acidic and substantially more reactive at lower pH (selenol pKa is ~5·2 *versus* 8·3 for thiols).

3.2 Chemical modification of other amino acids

In addition to cysteine, several other natural amino acids present functional groups that can be targeted for protein modification (reviewed in (Basle *et al.* 2010; Spicer & Davis, 2014)). These include lysine, tyrosine, arginine, glutamate, aspartate, serine, threonine, methionine, histidine and tryptophan side-chains, as well as N-terminal amines or C-terminal carboxyls (Basle *et al.* 2010; Hu *et al.* 2016; Lin *et al.* 2017). The modification of these residues is less precise as they are significantly more abundant than cysteine, yet, some selectivity can be achieved in a context-dependent manner. The primary amines of lysine side-chains are often popular targets due to favorable reaction kinetics that can be achieved with activated esters such as N-hydroxysuccinimide (Kalkhof & Sinz, 2008) (Fig. 3*a*), isothiocyanates (Nakamura *et al.* 2009), or aldehydes in a reductive alkylation reaction with sodium cyanoborohydride (Jentoft & Dearborn, 1979; McFarland & Francis, 2005). Since these reactions usually modify all accessible lysine side-chains, they can be used in applications that require multiple modifications (e.g. therapeutic protein conjugates) or in protein cross-linking for mass spectrometry analysis (Holding, 2015). An example of a more discriminating lysine-based modification strategy involves the 6π-aza-electrocyclization reaction with unsaturated aldehyde esters that targets solvent-accessible lysine residues with excellent selectivity and reaction kinetics, and has been used for the attachment of fluorescent or positron emission tomography probes (Tanaka *et al.* 2011). Yet, further selectivity can be achieved by discriminating the lower pKa of N-terminal amino groups (~8) from the pKa of the ε-amine of a lysine side-chain (~10·5). The identity of the N-terminal amino acid may further change the reactivity of the α-amine, although more general modification strategies have been developed. For example, functionalized ketenes can preferentially react with the α-amine in the context of 13 different N-terminal amino acids (Chan *et al.* 2012), while 2-pyridinecarboxyaldehydes provide efficient and specific N-terminal labeling for all amino acids except proline (MacDonald *et al.* 2015) (Fig. 3*b*). The unique reactivity of the N-terminus is also central to the mechanism of native chemical ligation, discussed in Section 3.3.

A set of chemical reactions has also been developed for the specific modification of the aromatic electron-rich tyrosine side-chain. The selectivity of tyrosine-focused reactions

usually exploits the low exposure of this residue on native protein surfaces and its susceptibility to oxidation (ElSohly & Francis, 2015; Seim *et al.* 2011). In particular, modification of the *ortho*-position can be achieved with diazonium salts (Schlick *et al.* 2005) or the three-component Mannich-type reaction with aldehydes and anilines (Joshi *et al.* 2004; McFarland *et al.* 2008) (Fig. 3*c*). While some side-reactions have been known to occur, it has been possible to optimize the selectivity of the Mannich-type reaction to achieve efficient modification of a single tyrosine residue with an EPR spin label in the presence of disulfide and tryptophan functionalities (Mileo *et al.* 2013). Proteins containing few surface exposed tyrosine residues can also be modified at low concentrations (5 μM) in aqueous solvents with π-allylpalladium reagents (Tilley & Francis, 2006).

Carboxyl groups in glutamate and aspartate side-chains can be targeted with water-soluble carbodiimides such as *N*-ethyl-3-*N′*-*N′*-dimethylaminopropylcarbodiimide (EDC) (Fig. 3*d*). In what is essentially a standard peptide coupling reaction, this reagent pairs carboxyls and amines in a covalent amide bond under aqueous conditions in a pH-dependent manner (Gilles *et al.* 1990). Commonly employed as a cross-linking reagent for the identification of protein–protein interactions in biochemical assays and mass spectrometry, EDC can also be used to modify protein assemblies such as viral capsids with a variety of functional and biophysical probes (Schlick *et al.* 2005). While EDC-based reactions do not discriminate between side-chain and C-terminal carboxyl groups, unique reactivity at the C-terminus can be generated by replacing the carboxyl functionality with a thioester, as discussed in the sections below.

3.3 Native chemical ligation

The unique reactivity of the cysteine side-chain and the positional control afforded by the protein N- and C-termini are at the heart of a simple, yet powerful, chemical reaction that allows the construction of native polypeptide chains from peptide-building blocks. This methodology, called NCL (Dawson *et al.* 1994), grants unprecedented control over the chemical functionalities that can be introduced in a polypeptide chain and is a fundamental tool in the repertoire of protein chemists. NCL simply requires two components with the following properties: (1) a peptide with a C-terminal thioester, and (2) a peptide with an N-terminal cysteine residue (or functional equivalent). When these components are mixed, the cysteine side-chain attacks the thioester of the other peptide, resulting in the formation of an intermolecular thioester (Fig. 4). This intermediate rapidly and irreversibly rearranges to form a native peptide bond, thus ligating the two fragments. Since both peptides can be produced by solid-phase peptide synthesis, virtually any modified natural or unnatural amino acid, spectroscopic probe, cross-linker or isotopic label can be incorporated at a well-defined position into the ligated sequence (Dawson & Kent, 2000). Moreover, in contrast to biosynthetic approaches such as amber suppression (see Section 3.6), there is no practical restriction on the number, or type, of unnatural amino acids that can be introduced – this point is driven home by the total synthesis of enantiomeric proteins composed of all D-amino acids (Mandal *et al.* 2012). The remarkable chemoselectivity of the NCL reaction means that ligations can be performed under aqueous conditions in the presence of internal cysteine residues. NCL is also compatible with protein denaturants or detergents allowing the construction of aggregation prone polypeptide chains or even polytopic membrane

proteins (Hejjaoui *et al.* 2012; Kwon *et al.* 2015; Valiyaveetil *et al.* 2006). The fragment bearing the N-terminal cysteine can also be produced recombinantly, usually preceded by a cleavable tag or fusion protein to avoid N-terminal cysteine processing complications in bacteria. Alternatively, the thioester component can be generated recombinantly with the help of proteins called inteins (Muir *et al.* 1998) (see Section 3.4). It is also possible to perform sequential NCL reactions, allowing three or more building blocks to be assembled in a regioselective fashion (Mandal *et al.* 2012; Torbeev & Hilvert, 2013). These advances thus allow the construction of considerably larger polypeptide chains that would be accessible from two synthetic peptides alone.

While NCL results in a native peptide bond, the required cysteine may still produce an unwanted 'scar' at the ligation junction – i.e. in cases where the target does not contain a native cysteine at an appropriate ligation point. If the final ligation product does not contain other cysteine residues, then the undesirable cysteine can be converted to alanine in a subsequent desulfurization step. Options for desulfurization reactions include reduction with metals or radical desulfurization mediated by the familiar tris(2-carboxyethyl)phosphine (TCEP) reagent (Dawson, 2011; Wan & Danishefsky, 2007; Yan & Dawson, 2001). Selective desulfurization reactions are also possible in the presence of other cysteine residues, provided that compatible protecting groups are used on the non-targeted cysteine side-chains (Ficht *et al.* 2007; Pentelute & Kent, 2007). Alternatively, selective ligation and desulfurization of selenocysteine can provide additional sequence positional control (Hondal *et al.* 2001; Reddy *et al.* 2016). Today, desulfurization has become a routine part of NCL protocols and the abundant alanine residue is a commonplace choice for a ligation site. To expand the junction amino acid set, a more advanced strategy involves the incorporation of β- and γ-thio amino acids. These moieties replace the cysteine-like residue at the N-terminal position, and provide a reactive thiol for *trans*-thioesterification. Desulfurization protocols can then produce the desired native side-chain. While ligations at such sites proceed more slowly due to increased steric hindrance, NCL can now be performed at phenylalanine (Crich & Banerjee, 2007), valine (Chen *et al.* 2008; Haase *et al.* 2008), leucine (Harpaz *et al.* 2010; Tan *et al.* 2010), threonine (Chen *et al.* 2010b), lysine (El Oualid *et al.* 2010; Kumar *et al.* 2009; Yang *et al.* 2009), proline (Shang *et al.* 2011), glutamine (Siman *et al.* 2012), arginine (Malins *et al.* 2013), tryptophan (Malins *et al.* 2014), aspartate (Thompson *et al.* 2013), glutamate (Cergol *et al.* 2014) and asparagine (Sayers *et al.* 2015). Work also continues on more streamlined ligation/desulfurization approaches that remove purification steps and increase the yield of ligated products (Moyal *et al.* 2013; Thompson *et al.* 2014).

The preceding discussion highlights just a few of the many refinements and extensions the NCL strategy has undergone since its introduction over 20 years ago (Harmand *et al.* 2014; Malins & Payne, 2015). As a consequence of this massive effort, the technique has become a central tool in protein science, having been applied to literally hundreds of protein targets. Of particular relevance here, it has provided the raw materials for numerous structural biology studies that have employed a broad range of spectroscopic or crystallographic methods (Fig. 5) (Grosse *et al.* 2011; Kent *et al.* 2012; Muralidharan & Muir, 2006). In general, the ability to modify any atom in the protein of interest with the precision afforded by synthetic organic chemistry is enormously powerful for dissecting protein function, especially when combined with high-resolution structural approaches such as NMR

spectroscopy and X-ray crystallography. Thus, we imagine that NCL will continue to evolve as an approach and be integrated into structural biology campaigns.

3.4 Inteins

Inteins (*in*tervening pro*teins*) are a peculiar group of proteins that can excise themselves from a larger precursor polypeptide chain, a process that leads to the formation of a native peptide bond between the flanking extein (*ex*ternal pro*tein*) fragments. This auto-processing event, called protein splicing, is analogous to the self-splicing of RNA introns and is spontaneous, i.e. it does not require external factors or ATP. Since they were first discovered in the early 1990s (Hirata *et al.* 1990; Kane *et al.* 1990), thousands of putative intein domains have been identified in the genomes of many unicellular organisms and viruses, with some containing multiple inteins in their genomes or even within the same gene (Perler, 2002; Shah & Muir, 2014). A small fraction of the known inteins has an even more curious property – the intein is split into two fragments, each fused to a separate extein fragment (the N- and C-exteins) (Wu *et al.* 1998). These intein fragments, called split inteins, are transcribed and translated separately, and upon a spontaneous and non-covalent association in the cellular milieu, they carry out protein splicing *in trans* to unite the extein fragments into a single polypeptide chain. While it is known that many inteins are embedded within essential protein genes (such as DNA or RNA polymerase, ribonucleotide reductase or metabolic enzymes), their evolutionary origins and biological significance remain mysterious, and only a small percentage of the identified intein domains have been carefully characterized (Pietrokovski, 2001; Shah & Muir, 2014). Despite these big gaps in our knowledge, the unique reactivity of inteins has turned them into a versatile and transformative tool in protein chemistry and chemical biology. For a detailed overview of intein applications, we refer the interested reader elsewhere (Shah & Muir, 2014; Topilina & Mills, 2014; Volkmann & Mootz, 2013; Wood & Camarero, 2014). Here, we will focus on aspects of intein function that would be of use to the structural biologist looking to install site-specific PTMs, segmentally label proteins with NMR isotopes, or aid the purification of recombinant polypeptides. Inteins have come a long way since their first applications in structural biology (Xu *et al.* 1999; Yamazaki *et al.* 1998), so we will end this section with a discussion of the current members of the intein toolbox and research directions taken to circumvent their limitations.

3.4.1 The intein splicing mechanism—Despite the low-sequence homology of known intein domains, they share a common protein splicing mechanism that relies on several conserved residues in the intein/extein polypeptide (Fig. 6*a*) (Volkmann & Mootz, 2013). One of these key residues is a cysteine (or in some cases a serine) at position 1 of the intein sequence. This nucleophilic side-chain attacks the amide carbon of the N-extein at position −1 (Fig. 6*b*) resulting in an N to S(O) acyl shift and the formation of a linear thio(oxy)ester intermediate. This intermediate is subject to a nucleophilic attack by a side-chain (cysteine, serine or threonine) at position +1 on the C-extein leading to *trans*-(thio)esterification and the generation of a branched intermediate. The branched intermediate is resolved through the cyclization of the C-terminal asparagine of the intein and results in intein excision from the polypeptide chain. Next, the spliced exteins quickly undergo an S(O) to N acyl shift to form a native peptide bond (i.e. identical to the last step in NCL). The protein splicing

mechanism is facilitated by several conserved threonine and histidine residues occupying strategic positions in the intein structural fold (Frutos *et al*. 2010). The efficiency and kinetics of the splicing mechanism may also depend on the identity of the residues immediately flanking the intein placing important constraints on the choice of ligation junction (Cheriyan *et al*. 2013; Iwai *et al*. 2006; Shah *et al*. 2012).

3.4.2 Applications in protein engineering—Expressed protein ligation (EPL) is an extension of NCL that employs a contiguous intein to recombinantly generate a protein bearing a C-terminal thioester (Muir *et al*. 1998). In this case, the N-extein is fused to a modified intein construct lacking the ability to perform *trans*-thioesterification. Instead, this step is performed by an exogenously added thiol, resulting in cleavage of the N-extein *α*-thioester intermediate (Fig. 7*a*). The resultant thioester can be used in NCL reactions as described in Section 3.3, while the recombinant origin of this fragment allows the construction of much larger semi-synthetic proteins as compared to total chemical synthesis by NCL. The efficiency of thioester generation rests on the propensity of the intein to avoid unwanted side reactions that result in premature N-extein cleavage and hydrolyzed products, problems that can be alleviated by the use of streamlined EPL protocols (Vila-Perello *et al*. 2013). Alternatively, hydrolysis of the N-extein under slightly basic conditions can be exploited in the so-called tagless protein purification protocols (Batjargal *et al*. 2015; Guan *et al*. 2013; Southworth *et al*. 1999) (Fig. 7*b*). In this case, the protein of interest is fused to an intein carrying the appropriate mutations and a suitable purification tag. The tag can be used for affinity column enrichment of the construct, followed by increase in the buffer pH. This results in the release of the tagless protein while the intein remains on the column.

Harnessing the protein *trans*-splicing (PTS) process mediated by split inteins offers an alternative approach to the ligation of polypeptide building blocks (Fig. 7*c*). Natural split inteins are especially attractive in this regard due to the extremely high affinity between the fragments (Shah *et al*. 2011, 2013) – this renders the ligation reaction less dependent on reagent concentration as compared to strictly chemical processes like NCL/EPL. Using orthogonal split intein pairs, it is also possible to perform one-pot three-piece ligations (Carvajal-Vallejos *et al*. 2012; Shah *et al*. 2011; Shi & Muir, 2005) resulting in the regiospecific assembly of the associated extein building blocks. While most natural split inteins have N- and C-fragments that are relatively large, it is possible to generate artificially split inteins that are as short as six or eleven residues (Appleby *et al*. 2009; Ludwig *et al*. 2006). There is also an efficient natural split intein pair (AceL–TerL) where the N-intein fragment is only 25 amino acids long (Thiel *et al*. 2014). Thus, it is now possible to use PTS with both synthetic or recombinant intein fragments and to install a wide range of N- or C-terminal chemical modification, including biophysical probes. One of the most important applications of PTS in structural biology is protein segmental isotopic labeling discussed in Section 4.2. Other applications include the cyclization of proteins and peptides (Lennard & Tavassoli, 2014; Scott *et al*. 1999), conditional protein splicing (Mootz *et al*. 2003; Schwartz *et al*. 2007), and protein semi-synthesis in cells (David *et al*. 2015).

3.4.3 Toward fast and promiscuous inteins—The first intein tools were introduced in the mid-to-late 1990s. One of the 'early' inteins, the 198-residue gyrase A intein from

Mycobacterium xenopi (Mxe GyrA) (Southworth *et al.* 1999), still used today, exhibits many desirable properties for applications in protein engineering – it is relatively small, can be efficiently expressed in *E. coli*, works in moderate concentrations of denaturants and its activity can be controlled with temperature. The first 'natural' split intein was discovered in 1998 in the cyanobacterium *Synechocystis* sp. strain PCC6803 (Ssp) where it was found to ligate two fragments of the catalytic subunit of DNA polymerase III (DnaE) (Wu *et al.* 1998). This discovery paved the road to more efficient protein *trans*-splicing and opened the way to using split inteins in other applications such as the cyclization of proteins and peptides and the creation of large cyclized libraries for potential therapeutic applications (Scott *et al.* 1999).

While these first generation intein tools were certainly enabling (Vila-Perello & Muir, 2010), they were not without their limitations – in retrospect, one could even say they were rather fussy and slow. For example, depending on the fusion partners, splicing (or thiolysis) could take hours to days and was often inefficient (Muralidharan & Muir, 2006). These 'idiosyncrasies' constrained the application of inteins in structural biology and fueled the search for faster, more promiscuous and efficient inteins. Several important discoveries in the mid-2000s challenged the view that all natural inteins are inefficient and slow. A genomic study of cyanobacterial genes expanded the DnaE intein family (Caspi *et al.* 2003) and the characterization of one newly discovered member, the DnaE intein from *Nostoc punctiforme* (Npu) revealed a few surprises. This split intein could perform protein *trans*-splicing reactions *in vitro* on a minute timescale and was much more tolerant to sequence deviations on the attached exteins than Ssp ((Iwai *et al.* 2006; Zettler *et al.* 2009). Now we know that many members of the DnaE family are fast (Shah *et al.* 2012), thus greatly expanding the choice of natural intein tools for the efficient generation of protein α-thioesters for EPL or the ligation of protein fragments in PTS (Table 1). Furthermore, efficient natural split inteins that are not part of the DnaE family have also been discovered and these include the gp41-1 and gp41-8 inteins (with insertion sites in the gp41 DNA gyrase gene), the IMPDH-1 intein (splitting a gene coding for inosine-5′-monophosphate dehydrogenase), the NrdJ intein (splitting the gene coding for the ribonucleotide reductase subunit NrdJ) (Carvajal-Vallejos *et al.* 2012), and the AceL–TerL pair (discovered in metagenomics data from the antarctic permanently stratified saline Ace Lake) (Thiel *et al.* 2014). These proteins bring more diversity to the intein molecular engineering toolbox, including splicing rates that are an order of magnitude faster than those for NpuDnaE; N- and C-intein fragments that are relatively short and can be made by peptide synthesis rather than recombinantly; the option of utilizing serine instead of cysteine at the +1 position; and the possibility of exploiting orthogonality in one-pot multi-piece ligations.

Careful biochemical characterization of newly discovered inteins has provided insights into the principles governing fast splicing and extein tolerance. For example, a batch mutagenesis approach that compared the slow split intein DnaE family member Ssp and the fast Npu split intein revealed that speed is determined by a handful of 'accelerator' residues located in the second shell of the folded protein, adjacent to the intein active site (Stevens *et al.* 2016). These residues were used as a filter in an informatics analysis of the DnaE sequence database, leading to the identification of several dozens of other split inteins predicted to support ultrafast splicing. A consensus split intein sequence, termed Cfa, was then derived

from this putative fast set and was found to possess quite remarkable properties; in addition to splicing faster than Npu at ambient conditions, Cfa is extremely robust, maintaining efficient activity at 80 °C or in the presence of up to 4 M guanidinium chloride or 8 M urea. As a result of these attributes, Cfa was found to be a superior tool for several PTS applications (Stevens *et al.* 2016).

3.5 Sortases

Sortases are a class of cysteine transpeptidases responsible for the attachment of virulence proteins to the cell wall of Gram-positive bacteria (Mazmanian *et al.* 1999). They are also involved in the polymerization of pilin subunits to form the pilus structures responsible for bacterial attachment to the host and biofilm formation (Mandlik *et al.* 2008). As important players in bacterial virulence, they have evolved to recognize a specific sorting sequence (LPXTG in the case of *Staphylococcus aureus*) and to attach the virulence factor to the cell wall using a pentaglycine cross-bridge (Ton-That *et al.* 2000). Naturally, sortases are of a considerable interest as drug targets, but they have also become an important and versatile protein engineering tool (Mao *et al.* 2004). Most sortase-based applications utilize the soluble fragment of wild-type or modified sortase A from *S. aureus.* These enzymes recognize the LPXTG motif and use their catalytic cysteine residue to cleave between the threonine and glycine backbone within the recognition sequence (Fig. 8). The cleavage reaction involves a thioacyl intermediate similar to the intermediates generated by cysteine proteases (Aulabaugh *et al.* 2007). Unlike the water molecules employed by proteases, however, sortases use a nucleophilic attack from the N-terminus of an oligoglycine motif to create a peptide bond between the acyl donor and acceptor. This results in the ligation of polypeptide chains that are subsequently connected with a $LPXT(G)_5$ linker. The sortase mechanism also requires binding of Ca^{2+} to a dynamic loop of the enzyme, an event that slows down the loop motion and allows enough time for the substrate to find the catalytic site (Naik *et al.* 2006). Another peculiarity of sortase-based ligations is their reversibility: the generated ligation site has the recognition sequence LPXTG and can serve as an acyl donor, while the released fragment contains an aminoglycine acyl receptor. Thus, to obtain efficient ligations, the donor or acceptor polypeptide chain typically has to be added in large excess (Guimaraes *et al.* 2013).

Analogous to intein technology development, sortases have considerably improved as protein engineering tools since their introduction in 2004 (Antos *et al.* 2016). The sortase-based protein engineering toolbox now contains evolved variants that exhibit much faster kinetics or that eliminate Ca^{2+}-dependence (albeit at the cost of slightly reduced enzyme activity) (Chen *et al.* 2011; Hirakawa *et al.* 2015; Wuethrich *et al.* 2014). There are also alternatives based on *S. aureus* sortase A or homologs from other organisms that can recognize variations of the LPXTG motif and/or allow non-glycine amino acids as the acyl acceptor (Antos *et al.* 2016; Dorr *et al.* 2014; Glasgow *et al.* 2016). To increase the yields of the ligation reaction, several clever strategies have been employed. In situations where the released aminoglycine peptide fragment is relatively small, it can be removed by dialysis or centrifugation while the reaction is proceeding (Freiburger *et al.* 2015). Affinity immobilization strategies or flow-based platforms have also been used for the selective removal of reaction components (Policarpo *et al.* 2014; Warden-Rothman *et al.* 2013).

Alternatively, the equilibrium of the reaction can be controlled by ligation product or by-product deactivation. In the first case, a WTWTW motif was added to the donor and acceptor, and upon ligation this sequence promoted a stable hairpin at the ligation junction, rendering the site inaccessible for cleavage (Yamamura *et al.* 2011). In the latter case, the acyl donor glycine was chemically modified such that upon release, chemical rearrangements occurred on the by-product transforming it into a poor nucleophile (Liu *et al.* 2014a; Williamson *et al.* 2014).

One important advantage of sortase-based ligations is that the acyl donor and acceptor polypeptide chains can be very short (only the LPXTG tag is required on the donor, and the oligoglycine motif is necessary on the acceptor) and thus are easily accessible by solid-phase peptide synthesis. Therefore, N- and C-terminal labeling reactions of large proteins are relatively straightforward (assuming by-products are efficiently removed) (Guimaraes *et al.* 2013; Theile *et al.* 2013). Larger polypeptides, on the other hand, can be expressed recombinantly with the appropriate donor and acceptor tags and the ligation reaction unites them in a single polypeptide chain with an $LPXT(G)_n$ 'scar'. In such cases it is recommended that the ligation junction is chosen on an unstructured region where it will not affect the function and/or fold of the protein and will be accessible to the sortase catalytic site (Guimaraes *et al.* 2013). Such reactions can be applied to create polymers and cyclized polypeptides (van 't Hof *et al.* 2015), or to stitch together domains into bifunctional or segmentally labeled proteins (Matsumoto *et al.* 2016; Williams *et al.* 2016; Witte *et al.* 2012). Sortase-based labeling has also been used in the functionalization of solid supports, nanoparticles, antibodies or cell surfaces, as well as in the labeling of proteins *in vivo*. We refer the interested reader to several comprehensive reviews on the subject (Popp & Ploegh, 2011; Ritzefeld, 2014; Schmohl & Schwarzer, 2014).

The success of sortase-based ligations has stimulated efforts to discover other protein ligases with expanded capabilities. A promising candidate is butelase-1, which was isolated from the plant *Clitoria ternatea* (Nguyen *et al.* 2014). Butelase-1 is the fastest ligase known with catalytic efficiencies as high as $542\,000\ \mathrm{M}^{-1}\ \mathrm{s}^{-1}$. Furthermore, it only requires the recognition sequence NHV on the acyl donor and produces ligation junctions with a minimal 'scar' (NX). Currently, the major limitation of this technology is that the enzyme is not available in recombinant form, and therefore has to be extracted and purified from the native plant (Nguyen *et al.* 2015). There are, however, evolutionary related ligases that may be more amenable to protein engineering approaches (Yang *et al.* 2017).

3.6 Genetic code expansion

After billions of years of evolution, Nature has engineered extraordinary functional diversity into proteins with only 20 amino acid building blocks. Yet, many of these building blocks are often modified post-translationally, clearly indicating the need of living organisms to enhance and modulate their protein repertoire with additional chemical functionalities. There are also organisms from all domains of life that can produce and incorporate other building blocks into their proteins. This includes selenocysteine, often called the 21st amino acid that provides a unique reactive site for precise tuning of biological function in cells. Interestingly, this amino acid is incorporated into proteins by a natural reassignment of the UGA stop

codon coupled with the recognition of a specific structural element on the mRNA transcript known as the selenocysteine-insertion sequence (reviewed in (Metanis *et al.* 2009; Yoshizawa & Bock, 2009)). Similarly, there are methane producing Archaea species that have evolved a specialized tRNA/aminoacyl-tRNA synthetase (tRNA/aaRS) pair to exploit the UAG stop codon and insert pyrrolysine site-specifically into certain methyltransferase proteins (Srinivasan *et al.* 2002). Exploiting the natural translation machinery, protein engineers have worked hard to 'persuade' living organisms to incorporate additional building blocks into polypeptide chains (reviewed in (Liu & Schultz, 2010)). One approach involves the use of cell lines that are auxotrophic for one of the 20 amino acids, for example methionine, and that will only grow when the missing amino acid is included in the culture medium. Replacing this amino acid with a close structural analog that is utilized by the wild-type aminoacyl-tRNA synthetase, results in incorporation of the unnatural amino acid (UAA) into overexpressed proteins. Since this results in global incorporation of the UAA, this approach is often used for the replacement of rare amino acids with their structural analogs. For example, methionine can be substituted with selenomethionine to provide a heavy atom for phasing crystallographic data (Barton *et al.* 2006; Yang *et al.* 1990).

A second approach involves the semi-synthesis of tRNAs that are pre-loaded with the UAA of interest (Hecht *et al.* 1978; Noren *et al.* 1989). These tRNAs have been used in *in vitro* translation systems that bypass the need for a matching aaRS, and since the identity of the UAA is decoupled from the information content of the tRNA, any coding or blank codon can be used for reassignment (Cornish *et al.* 1994; Judice *et al.* 1993; Koh *et al.* 1997). While the semi-synthesis of acylated tRNAs can be technically challenging, more efficient production strategies have been developed. This includes flexizymes, flexible tRNA acylation ribozymes that accept a versatile range of aminoacyl substrates and tRNAs with different sequences (Goto *et al.* 2011). Pre-loaded tRNAs can also be injected or transfected directly into living cells (England *et al.* 1999; Kohrer *et al.* 2003), although the success of such strategies has been limited by their short lifetimes in the cellular environment and challenges associated with in-cell delivery.

Today, UAA incorporation in living cells is almost exclusively performed following the strategy introduced by Peter Schultz and co-workers in 2001 (Wang *et al.* 2001). This methodology, commonly referred to as amber suppression, relied on the development of an orthogonal tRNA/aaRS pair that could be expressed in *E. coli* and was used to incorporate *O*-methyl-L-tyrosine into dihydrofolate reductase with 99% fidelity. In the 15 years since this landmark study, the unnatural building block palette for genetic incorporation has grown more than 100 amino acids strong (Lang & Chin, 2014; Liu & Schultz, 2010; Neumann-Staubitz & Neumann, 2016). This includes amino acids carrying natural modifications (e.g. phosphoserine or acetyllysine), biophysical and structural probes, cross-linkers, reactive handles for bio-orthogonal reactions, and site-specific protein engineering functionalities that can modify their attendant proteins upon a specific cellular or chemical cue. Here, we describe the basic principles of the technology, review UAAs of particular interest to the structural biologist, and discuss current limitations and efforts to improve the efficiency of UAA incorporation.

3.6.1 Amber codon suppression in living cells—The successful incorporation of an UAA into a protein synthesized by a living cell requires several important considerations and components (Fig. 9*a*). First and foremost, the UAA of interest must be chemically and metabolically stable, cell permeable or otherwise biosynthetically accessible in the cellular environment. It also must be tolerated by the ribosome and the cellular elongation factors without being recognized as a substrate by any of the endogenous synthetases. The UAA then requires its own unique codon, with the amber stop codon (UAG) being a popular choice due to its low occurrence in both prokaryotic and eukaryotic systems. The successful site-specific incorporation of the UAA, however, rests on the presence of a dedicated tRNA/ aaRS pair that is highly specific for the UAA of interest, yet orthogonal in the context of all endogenous tRNA/aaRS pairs. Developing such pairs for a chemically diverse set of UAA is currently one of the time consuming and difficult steps of this technology. Since tRNA recognition by aaRS is often species specific, it is sometimes possible to import a heterologous pair into the cell of interest and use it as a starting point to build orthogonality and specificity into the system. For example, many UAA incorporation systems in *E. coli* are based on the heterologous tRNA^{Tyr}/TyrRS pair from *Methanococcus jannaschii*, while the tRNA^{Tyr}/TyrRS and tRNA^{Leu}/LeuRS pairs from *E. coli* have been used in eukaryotic cells (reviewed in (Chin, 2014)). The tRNA^{Pyr}/PyrRS pair from methanogenic bacteria that can incorporate pyrrolysine has also been a very useful tool, as it is orthogonal in *E. coli*, yeast and mammalian cell lines, and has allowed the incorporation of many lysine-based UAAs, including acetyllysine (Neumann *et al.* 2008). While these systems provide a useful starting point, it is usually necessary to use mutagenesis and rounds of negative and positive selection to improve on the selectivity and orthogonality of the pair. Directed evolution approaches can also be used for the generation of *de novo* tRNA/aaRS pairs, or to expand the function of other components of the translational machinery (reviewed in (Chin, 2014)). Once an appropriate tRNA/aaRS pair is developed, however, the practical implementation of amber suppression for the UAA is relatively straightforward. *E. coli* cells, for example, can be transformed with two plasmids: (1) a plasmid encoding the protein of interest and an appropriate point mutation with the amber TAG codon, and (2) a plasmid carrying the appropriate DNA sequence to produce the optimized tRNA/aaRS pair. After addition of UAA to the media, gene expression is induced for both plasmids and the UAA is incorporated into the protein of interest by the bacterial translational machinery. To separate the full-length protein from prematurely truncated species, often a purification tag is added to the protein C-terminus – notably, these can involve 'silent' intein- or sortase-based purification tags (Batjargal *et al.* 2015; Warden-Rothman *et al.* 2013). Well-established protocols for UAA incorporation are now available for yeast (Hancock *et al.* 2010), mammalian (Chen *et al.* 2009) and insect cells (Koehler *et al.* 2016; Mukai *et al.* 2010b) and amber suppression can even be performed in multicellular organisms including *C. elegans* (Greiss & Chin, 2011), *D. melanogaster* (Bianco *et al.* 2012), mice (Ernst *et al.* 2016; Kang *et al.* 2013) and plants (Li *et al.* 2013b).

3.6.2 The amber suppression toolbox—The amber suppression toolbox contains many UAAs designed with structural biology applications in mind (Fig. 9*b–e*). For example, heavy atoms can be incorporated site-specifically for solving the phase problem in X-ray crystallography – appropriate UAAs include *p*-iodo-L-phenylalanine (Xie *et al.* 2004) and 3-

iodo-L-tyrosine (Sakamoto *et al.* 2009), as well as metal-ion chelating amino acids (Lee *et al.* 2009b). For NMR spectroscopists, amber suppression offers a relatively cheap and efficient way to install site-specific isotopic labels in otherwise unlabeled proteins. Many of the NMR 'friendly' UAAs are fluorinated derivatives that exploit the unique spectroscopic properties of ^{19}F as a reporter of global protein folding and dynamics (Jones *et al.* 2010; Yang *et al.* 2015). Similarly, amber suppression has been used to install nitroxide spin labels for distance measurements by EPR, thus overcoming the problems often associated with cysteine-based approaches (Park *et al.* 2015; Schmidt *et al.* 2015). More importantly, however, amber suppression is a living cell protein engineering tool; thus, it is ideally suited for NMR or EPR studies designed to follow the structural fate of proteins in the cellular milieu.

Of particular interest to the structural biologist are UAAs carrying natural PTMs. Currently, amber suppression can directly incorporate the following modifications: phosphoserine (Rogerson *et al.* 2015), acetyllysine (Neumann *et al.* 2008), several lysine acylations (Gattner *et al.* 2013; Kim *et al.* 2012), phosphotyrosine (Fan *et al.* 2016) and sulfonated tyrosine (Liu *et al.* 2007). To expand this toolbox, however, amber suppression can be used to install a reactive handle at the position of interest, and then the appropriate modification can be chemically generated after protein purification. Using this strategy, for example, the UAA *δ*-thiol-lysine can be incorporated, followed by traceless attachment of a ubiquitin moiety with NCL (Virdee *et al.* 2011). Similarly, methylated lysines can be generated by the incorporation of a suitable pre-cursor UAA (Nguyen *et al.* 2010; Wang *et al.* 2017). Phosphoserine, on the other hand, can serve as a starting point for the generation of dehydroalanine, which in turn can be converted into a number of modified side-chains (Wright *et al.* 2016).

Amber suppression also allows the facile installation of site-specific cross-linkers, a valuable tool for the identification of protein–protein and protein–ligand interactions both *in vitro* and *in vivo*. There are several options for UV-activatable cross-linkers that exploit different cross-linking mechanisms, and afford temporal and spatial control of the reaction. The oldest members of this toolbox are *p*-benzophenylalanine (Chin *et al.* 2002a) and *p*-azido-L-phenylalanine (Chin *et al.* 2002b), both available for bacterial and eukaryotic systems, and extensively used for cross-linking experiments of purified proteins or in the cellular environment. More recently, diazirine-modified lysine-based amber suppression systems have been developed (Ai *et al.* 2011; Chou *et al.* 2011; Zhang *et al.* 2011), exhibiting superior cross-linking efficiency, smaller structural perturbation effects and ideally suited for cross-linking experiments of lysine-rich proteins such as histones. UAAs that can cross-link proteins to nucleic acids are also available, and these include *p*-benzophenylalanine (Winkelman *et al.* 2016), and a furan-based cross-linker activated upon red-light irradiation (Schmidt & Summerer, 2013). Interestingly, *p*-azido-L-phenylalanine can also serve as an infrared (IR) spectroscopy probe, and as such can be used to report on the structural transitions of a protein along its activation pathway (Ye *et al.* 2010). Similarly, *para*-cyanophenylalanine can be installed as a site-specific and sensitive vibrational probe of ligand binding (Schultz *et al.* 2006).

A significant fraction of the amber suppression UAA library has been designed for imaging and fluorescence-based biophysical applications. Some fluorescent UAAs can be incorporated directly into polypeptide chains and these include coumarin-and dansyl-based modifications (Kuhn *et al.* 2011; Summerer *et al.* 2006), as well as environmentally sensitive aminonaphthalene-based probes available for both yeast and mammalian applications (Chatterjee *et al.* 2013a; Lee *et al.* 2009a). More commonly, however, optical probes are attached site-specifically using bio-orthogonal reactions. In this case, amber suppression is used to install a reactive handle at a site in the polypeptide chain, while the fluorescent probe is modified with a compatible reactive functionality. The conjugation reaction can be carried out *in vitro* with purified components, or the optical label can be added to the media and/or delivered into cells for bio-orthogonal reaction chemistry within the cellular milieu. While this places important UAA-fluorescent label design constraints with respect to cell permeability, chemical stability and reaction kinetics, this approach allows the incorporation of optical labels that work at a variety of wavelengths, amid reduced background fluorescence. Currently, UAAs with a diverse set of functionalities for bio-orthogonal reactions are available (reviewed in (Lang & Chin, 2014)), and it is likely that this list will become much more expansive in the future as new fast and specific bio-compatible approaches are developed (Section 3.7).

3.6.3 Limitations and future directions—The remarkable plasticity of the natural and evolved cellular protein synthesis machinery has allowed chemical biologists to create a large and diverse set of UAAs that can be incorporated into biological polymers assembled *in vivo*. Yet, the incorporation of many of these UAAs is inefficient, context dependent and essentially limited to the inclusion of a single modification per protein. Since UAA incorporation relies on the reassignment of natural stop codons, the suppressor tRNAs compete for binding sites with the endogeneous release factors that terminate translation. Therefore, truncations of the desired protein are often produced, resulting in compromised yields, complicated purification protocols and potential toxicity for the recombinant cell. In *E. coli*, translation termination involves release factor protein 1 (RF1) responsible for recognition of ochre (UAA) and amber (UAG) stop codons, and release factor protein 2 (RF2) that identifies ochre (UAA) and opal (UGA) stop codons. Thus, it has been possible to engineer bacterial strains that lack RF1 to enhance amber codon translation efficiency, and in particular improve the incorporation of the same UAA at multiple positions in the protein sequence (Johnson *et al.* 2011; Mukai *et al.* 2010a). These strains, however, can be problematic as misincorporation of glutamine and codon skipping have also been reported (George *et al.* 2016). More dramatically, organisms can be genomically recoded to replace the UAG codon completely with the synonomous UAA stop codons, and free this new unique codon for more efficient amber suppression (Lajoie *et al.* 2013).

Protein yields are also affected by the lower efficiency of the evolved synthetases and suboptimal interactions of the tRNA with the wild-type elongation factors and ribosomes. These problems can be partially alleviated by the design of more efficient plasmid systems that have optimized promoters and can produce higher copy numbers of the synthetase and the tRNA. For example, the pEVOL and the pUltra plasmids have significantly increased incorporation efficiency in *E. coli* (Chatterjee *et al.* 2013b; Young *et al.* 2010). For

mammalian cells, protein expression is further affected by the transfection efficiency of the delivered constructs, thus it is usually desirable to incorporate all of the necessary genetic components (multiple copies of the synthetase and tRNA, gene of interest, engineered release factor, etc.) on the same plasmid ((Cohen & Arbely, 2016). Baculovirus vectors that can deliver a large cargo of genetic material (>30 kb) to a variety of mammalian cells have also been developed for more efficient UAA incorporation (Chatterjee *et al.* 2013c; Zheng *et al.* 2016). To avoid the heterogeneous expression levels associated with transient transfection and viral transduction altogether, the creation of stable mammalian cell lines capable of defined UAA incorporation is highly desirable and efforts have already been undertaken in this regard (Elsasser *et al.* 2016; Tian *et al.* 2014).

The incorporation of several different modifications into the same polypeptide chain not only requires optimized suppression of the most commonly reassigned amber stop codon, but also the availability of other unique or rare codons that can be used to develop orthogonal tRNA/aaRS pairs. Not surprisingly, the ochre and opal stop codons are often combined with the amber codon for dual incorporation of UAAs, primarily with fluorescence-type applications in mind (Chatterjee *et al.* 2013b; Wan *et al.* 2010; Xiao *et al.* 2013). The use of two stop codons, however, suffers from and amplifies many of the drawbacks described above and, if successful, results in very low protein yields. To circumvent this problem, translation machinery engineered to recognize quadruplet codons has been developed (Neumann *et al.* 2010; Wang *et al.* 2014a). The use of quadruplet codons can in principle provide 256 blank codons, thus dramatically expanding the capabilities of the natural 64-codon based genomes. Such systems, however, require the engineering of orthogonal ribosomes that can efficiently recognize quadruplet messages (Neumann *et al.* 2010). Alternatively, the genomes of organisms such as *E. coli* can be reprogrammed to create compressed codon schemes and free up unique codons for reassignment (Ostrov *et al.* 2016; Wang *et al.* 2016). In the future, it may also be possible to create new coding schemes by expanding the genetic alphabet of living organisms and utilizing synthetic DNA bases in new synthetic organisms (Malyshev *et al.* 2014).

3.7 Chemical modification of unnatural amino acids

The ability to install unnatural amino acids in polypeptide chains with exquisite specificity has opened widely the doors to a post-cysteine world of bio-orthogonal reactions. There is now a rapidly expanding collection of reactions that work efficiently and selectively with low concentrations of reactants in physiological buffers or the cellular milieu. Here, we will review a small sample of these tools focusing primarily on reactions that might be used to install site-specific PTMs or their mimics, as well as probes for structural investigations. For a comprehensive discussion of the bio-orthogonal literature, we refer the interested reader to several excellent reviews on the subject (Lang & Chin, 2014; Shieh & Bertozzi, 2014; Spicer & Davis, 2014; Stephanopoulos & Francis, 2011).

The azido functional group is one of the most versatile modifications that can be installed on peptides and proteins by amber suppression, total chemical synthesis and/or peptide ligation. In addition to serving as a light-activatable cross-linker or infrared probe, this small moiety is also at the heart of several important bio-orthogonal reactions. For example, in an

extension of the classical Staudinger reduction reaction, azides react with triarylphosphines bearing an electrophilic trap (usually an ester) to form a stable amide bond (Saxon & Bertozzi, 2000). While the first version of this methodology generated a residual phosphine oxide at the ligation junction, 'traceless' Staudinger ligation variants are now available (Nilsson *et al.* 2000; Saxon *et al.* 2000) (Fig. 10*a*), as well as reactions based on azobenzene or caged phosphine reagents that can be activated with light (Shah *et al.* 2016; Szymanski *et al.* 2013). This reaction can be used to ligate a peptide fragment containing a C-terminal thioester to another fragment carrying an N-terminal azide resulting in an extended native polypeptide chain and bypassing the need for a cysteine residue at the ligation junction (Nilsson *et al.* 2000). While displaying slower reaction kinetics compared with some other bio-orthogonal reactions, the Staudinger ligation can be impactful in protein engineering applications where versatility and orthogonality are desired. For example, it has been used to create native isopeptide linkages between ubiquitin and other proteins (Andersen & Raines, 2015), install multiple probes into proteins such as RNA polymerase or GPCRs (Chakraborty *et al.* 2015; Huber *et al.* 2013), prepare glycoprotein or phosphoprotein mimics (Bernardes *et al.* 2011; Serwa *et al.* 2009), and turn on the fluorescence of optical probes upon protein labeling (Lemieux *et al.* 2003).

Another popular and versatile bio-orthogonal reaction is the azide-alkyne 1,3-dipolar cycloaddition (Rostovtsev *et al.* 2002; Tornoe *et al.* 2002), commonly referred to as 'click chemistry' (Kolb *et al.* 2001). The cycloaddition can be promoted either by the presence of Cu(I) ligands or the use of highly strained cyclooctyne systems (Agard *et al.* 2004). Both versions are specific, exhibit relatively fast reaction kinetics and are easy to use, which has led to numerous applications in the protein conjugation area. Recent examples include the incorporation of dual EPR or FRET probes for distance measurements in proteins or RNA (Kucher *et al.* 2016; Lavergne *et al.* 2016), and the attachment of rigid lanthanide tags to large proteins for paramagnetic relaxation experiments (Mallagaray *et al.* 2016). *In vivo* applications of the Cu(I) version have been limited by copper toxicity, although extracellular labeling of cells has been reported (Link *et al.* 2004; Uttamapinant *et al.* 2012). The strained cyclooctyne systems, on the other hand, are compatible with living systems and improved versions that exhibit faster labeling kinetics are available for genetic incorporation through amber suppression (Dommerholt *et al.* 2010; Lang *et al.* 2012*b*; Plass *et al.* 2011) (Fig. 10*b*). These copper-free cycloadditions have thus been valuable in imaging and proteomic applications in living cells and organisms (Laughlin *et al.* 2008; Smits *et al.* 2016; Xie *et al.* 2014), and can be useful for the structural biologist interested in understanding the dynamics of macromolecules in the cellular environment either by FRET or EPR (Kucher *et al.* 2016).

The success of click chemistry bioconjugation has inspired the search for faster and more efficient reactions for labeling in the cellular milieu. An important development along this direction involves the use of inverse-electron demand Diels-Alder reactions between strained dienophiles and tetrazine dienes (Blackman *et al.* 2008; Devaraj *et al.* 2008). The dienophile system in this case can be a *trans*-cyclooctene or a norbornene, both exhibiting significantly faster reaction kinetics as compared to the 1,3-dipolar cycloadditions described above, and thus allowing labeling reactions of proteins at much lower concentrations (Fig. 10*c*). These bio-orthogonal reactive handles are also available for genetic incorporation by amber suppression (Lang *et al.* 2012a, 2012b; Seitchik *et al.* 2012) and provide an expanded and

rapidly evolving toolbox for bio-orthogonal protein manipulations and modifications in the cellular environment.

While most of the reactions described above are incredibly useful for the conjugation of small molecules and biophysical probes to proteins *in vitro* and *in vivo*, they generally leave a relatively large chemical 'scar' at the conjugation site to be practical in the generation of PTMs or their mimics. A versatile approach for the site-specific installation of PTMs involves the UAA dehydroalanine. This UAA can be generated from a cysteine residue installed by site-directed mutagenesis (Chalker *et al.* 2012) or from an *O*-phosphoserine precursor incorporated by amber suppression (Yang *et al.* 2016). Under biocompatible conditions, dehydroalanine can then be reacted with a variety of alkyl halides, via a radical-mediated process, to produce an impressive list of more than 50 modified functionalities at the site of interest (Wright *et al.* 2016; Yang *et al.* 2016). These include methylated lysine and arginine residues, as well as fluorinated, glycosylated, phosphorylated, alkylated or isotopically labeled natural and UAA side chains. It should be noted that this methodology currently leads to the generation of both D and L configurations at the mutation site (i.e. the reaction lacks stereochemical control), a potential limitation for some applications.

3.8 Enzymatic bioconjugation approaches

The site-specific incorporation of bio-orthogonal handles on proteins can also be achieved enzymatically. These enzymes choose their targets based on the recognition of a specific amino acid sequence and can perform various chemical modifications of cysteine, lysine, serine, glutamine or glycine residues within their target sequences. This toolbox, for example, includes the formylglycine generating enzyme (FGE) that recognizes the CXPXR sequence motif and converts the cysteine residue to formylglycine, thus introducing an aldehyde functional group (Wu *et al.* 2009) (Fig. 11*a*). The aldehyde handle can be subsequently elaborated with various probes via bio-orthogonal transformations such as oximation and Hydrazino-Pictet-Spengler reactions (Agarwal *et al.* 2013; Dirksen & Dawson, 2008). Another popular tool is lipoic acid ligase, an enzyme that modifies a lysine side-chain within the 13-residue target sequence (Uttamapinant *et al.* 2010). Engineered versions of this enzyme can accommodate lipoic acid analogs and have been used to introduce bio-orthogonal handles, including azides (Plaks *et al.* 2015; Uttamapinant *et al.* 2012) (Fig. 11*b*), aryl aldehydes and hydrazines (Cohen *et al.* 2012), *p*-iodophenyl derivatives (Hauke *et al.* 2014), norbornenes (Best *et al.* 2015) an *trans*-cyclooctenes (Liu *et al.* 2012). It is also possible to engineer fluorescent lipoic acid analogs that can be installed directly on the protein of interest (Uttamapinant *et al.* 2010). Other members of the enzymatic toolbox include biotin ligase, farnesyltransferase, transglutaminase and *N*-myristoyltransferase (reviewed in (Rashidian *et al.* 2013)).

Enzymatic bioconjugation is an increasingly useful tool for cell imaging applications, functionalization of therapeutic proteins, immobilization of proteins on solid supports or the preparation of protein—polymer or protein—nanoparticle conjugates (Hu *et al.* 2016; Slavoff *et al.* 2011; Walper *et al.* 2015). For the structural biologist and biophysicist, it offers an orthogonal strategy for protein modification that is based on genetically encodable peptide tags (Stephanopoulos & Francis, 2011). For example, it provides a conceptually

straightforward way to introduce a second modification to a construct that already contains an UAA incorporated by amber suppression. Enzymatic bioconjugation can also be useful in the multiplex labeling of complex protein mixtures, and the modification of constructs that are large, difficult to purify or not easily amenable to the approaches outlined above. Since the recognition sequence is fused to the protein of interest, the modification site is usually limited in location to the N- and C-termini, or to a surface exposed and flexible loop of the protein. Other important considerations include the reaction kinetics, the stability and solubility of the enzymes, and the relatively large size of the modifications that are installed and that can perturb the function of the protein target.

4. Protein engineering approaches for tackling outstanding challenges in structural biology

4.1 X-ray crystallography

The molecular engineering toolbox presented here can aid crystallographers in all stages of the structure determination process. For example, when phase values cannot be determined by direct methods or molecular replacement, protein crystallographers can incorporate heavy atoms (e.g. Se, I or Br) into the protein crystal and use their anomalous diffraction patterns to solve the structure. Heavy atom incorporation can be accomplished in several ways (reviewed in (Pike *et al.* 2016). For example, cysteine residues can be derivatized with mercury by pre-labeling, co-crystallization or soaking of the protein crystals in mercury salts (Martinez *et al.* 1993). If necessary, cysteine accessibility can be pre-evaluated with Ellman's reagent (5,5′-dithiobis-2-nitrobenzoic acid) (Li *et al.* 2015). Another popular heavy atom labeling methodology relies on the incorporation of the modified amino acid selenomethionine (Hendrickson *et al.* 1990). In this case, selenomethionine is added to the culture of a methionine-auxotrophic bacterial strain resulting in substitution of all methionine residues in the protein. Selenomethionine incorporation is also possible in other non-bacterial expression systems, albeit with somewhat lower substitution efficiency (Cronin *et al.* 2007; Nettleship *et al.* 2010). Site-specific incorporation of appropriately modified unnatural amino acids (e.g. *p*-iodo-L-phenylalanine, *p*-bromo-L-phenylalanine and 3-iodo-L-tyrosine) can be performed with amber suppression or chemical synthesis (Kwon *et al.* 2006; Sakamoto *et al.* 2009; Xie *et al.* 2004; Yeung *et al.* 2016). Solving the phase problem can also be aided by the development of metal-binding peptide tags (e.g. Tb^{3+}) that can be genetically fused to the protein of interest (Silvaggi *et al.* 2007).

Total chemical synthesis of proteins is an important tool for racemic and quasi-racemic crystallography (Yeates & Kent, 2012) (Fig. 12). These methods are based on the observation that mixing D- and L-forms of proteins can aid crystallization as macromolecules have access to a much larger set of crystallographic space groups, including those that contain mirror or center of inversion operations (Wukovitz & Yeates, 1995). D-enantiomers of proteins can be made using solid-phase peptide synthesis and NCL, with synthesis efficiencies that have allowed the structure determination of constructs in the 200-amino acid range (Mandal *et al.* 2012; Pan *et al.* 2016). The synthetic origin of such polypeptides also allows the facile co-incorporation of amino acids containing heavy atoms

(Yeung *et al.* 2016), or chemically well-defined post-translational modifications (Okamoto *et al.* 2014).

For structure characterization of larger biological macromolecules and complexes, it is often more practical to use mimics rather than the native PTM. For example, methylation mimics are accessible through the alkylation of cysteine residues, and produce constructs with high yields and chemical purity in all possible methylation states (Simon *et al.* 2007). Such approaches have been invaluable in elucidating the impact of histone methylation on the nucleosome surface, a structural problem that requires the efficient crystallization of a biological assembly comprised of four different proteins and DNA (Lu *et al.* 2008). Sometimes, non-native linkages and modifications are essential in trapping an important functional state of a protein complex, for example between an ubiquitylated protein and the corresponding deubiquitinating enzyme (Morgan *et al.* 2016). The ongoing optimization of amber suppression systems for in-cell and cell-free protein synthesis has also made it easier to produce certain native modifications in high-enough yield for protein crystallization. There are, for example, several crystal structures of acetylated and phosphorylated proteins prepared by amber suppression in *E. coli* (Arbely *et al.* 2011; Huguenin-Dezot *et al.* 2016; Kuhlmann *et al.* 2016; Lammers *et al.* 2010). This technology has also enabled the structural analysis of ubiquitin chains (Virdee *et al.* 2010). Crystallographic studies of proteins containing multiple modifications have been successfully addressed with cell-free protein synthesis (Wakamori *et al.* 2015) and EPL (Wang *et al.* 2014b).

Beyond the incorporation of PTMs, the modern protein engineering toolbox provides the means to precisely alter the covalent structure of proteins, importantly allowing access to both the amino acid side-chains and the polypeptide backbone. Crystallographic analysis of such modified proteins can be extremely powerful, yielding insights into processes as diverse as ion conduction (Grosse *et al.* 2011; Valiyaveetil *et al.* 2006), enzyme catalysis (Torbeev *et al.* 2011; Wang *et al.* 2014b) and protein–protein interactions (Lu *et al.* 1999; Morgan *et al.* 2016).

4.2 Nuclear magnetic resonance

The site-specific incorporation of NMR probes into biological macromolecules presents two challenging requirements to the protein chemist. First, similarly to X-ray crystallography, large amounts of sample are required, and the incorporation methods should be robust and efficient. Second, the isotopic labeling precursors are often prohibitively expensive. It therefore comes as no surprise that most NMR labeling strategies exploit the metabolic pathways in bacteria to produce uniform or sparse isotopic labeling schemes, or incorporate isotopic labels in an amino acid specific fashion (for a comprehensive series of reviews on isotopic labeling of biomolecules, the interested reader is referred to volume 565 of Methods in Enzymology, 2015). Here, we will focus on labeling methodologies that allow more control over the placement of spectroscopic probes in a polypeptide chain, and that have historically been demonstrated to produce sufficient protein amounts for NMR analysis.

4.2.1 Segmental isotopic labeling—Segmental isotopic labeling is a powerful protein engineering approach that allows the generation of an intact and typically natively folded

and functional protein where only a certain segment of the polypeptide chain is 'visible' by NMR (Xu *et al.* 1999; Yamazaki *et al.* 1998). This strategy is particularly impactful if the protein of interest is large and/or has a degenerate amino acid sequence, thus simplifying assignment and interpretation of crowded and overlapped NMR spectra. Such constructs can be crucial in generating unambiguous structural constraints in large proteins and assemblies, and have been exploited in structural investigations both by solution and solid-state NMR (Mehler *et al.* 2015; Schubeis *et al.* 2015; Tremblay *et al.* 2015; Williams *et al.* 2016). Several approaches can be used to achieve segmental labeling of polypeptide chains. For example, a protein domain can be fused to a contiguous intein and produced recombinantly in labeled media (e.g. supplemented with ^{15}N and ^{13}C-enriched nutrient sources). Following thiolysis of the intein, the thioester derivative of the labeled domain can then be ligated, via EPL, to a protein fragment that does not contain NMR isotopes (Xu *et al.* 1999). In principle, the NMR silent segment can be a synthetic peptide carrying a PTM or a paramagnetic tag, thus allowing the facile incorporation of C-terminal protein modifications. For example, this approach has been used to introduce a phosphorylated tyrosine residue at position 125 of the amyloid-related protein a-synuclein, while residues 1–106 were prepared recombinantly and isotopically labeled for NMR analysis (Hejjaoui *et al.* 2012). In practice, however, the second fragment is usually produced recombinantly due to cost and yield considerations.

PTS offers an alternative intein-based approach for segmental isotopic labeling. The high affinity of naturally split intein fragments means that the splicing reactions can be conducted at very low concentrations of reactants (low μM), a capability that distinguishes the PTS strategy from EPL where reactant concentrations in the mM range are typically required. In practice, the target protein is divided into two segments, for example at domain boundaries, each fused to an appropriate intein fragment (Fig. 13*a*). These constructs are expressed separately, allowing the differential incorporation of NMR-active isotopes. After purification, the two extein–intein fragments are mixed, protein *trans*-splicing takes place and the leftover intein components can be purified using an appropriate affinity tag or chromatographic separation. Notably, the splicing reaction can also be performed directly in the cellular environment. In this case, both constructs are transformed into the same cell, but with different promoters so that the expression of each construct can be controlled separately, and therefore timed with the addition of isotopically enriched nutrients to the media (Muona *et al.* 2010) (Fig. 13*b*). This strategy can be particularly helpful for labeling proteins that are hard to refold *in vitro*, e.g. membrane proteins (Mehler *et al.* 2015). The first segmental labeling protocols relied on artificially split or natural inteins that were not very efficient, thus limiting the utility of the technology to few favorable cases (Yamazaki *et al.* 1998). Today, however, the toolbox of useful split inteins has greatly expanded, and it is now possible to choose from inteins with diverse properties, e.g. those giving better expression yields for the protein of interest, capable of carrying out ligations at higher temperatures or denaturant concentrations, or more forgiving to the choice of extein junction site (see Section 3.4.3).

Another option for segmental isotopic labeling is provided by sortase-mediated ligation (Fig. 13*c*). As discussed in Section 3.5, this method relies on the use of Sortase A, a transpeptidase that can ligate two constructs provided that one contains an LPXTG

recognition sequence, and the other, a glycine repeat motif. Since it results in an LPXTG 'scar' at the ligation junction, sortase-mediated ligation is best suited for the segmental labeling of protein domains separated by a flexible, mutation-tolerant linker (Williams *et al.* 2016). The reversibility of the sortase ligation also necessitates the removal of cleaved peptide byproducts during the reaction course for optimal yields (Freiburger *et al.* 2015). On the other hand, segmental labeling using sortase works well with low μM concentrations of reagents, and is a good alternative for constructs that do not express well as intein fusions.

The segmental labeling approaches described here have also been used to generate protein constructs with NMR silent solubility enhancement tags (Kobashigawa *et al.* 2009; Zuger & Iwai, 2005). To keep proteins more stable in solution during a long multidimensional experiment, NMR spectroscopists can resort to fusing the protein of interest to a soluble domain such as SUMO, GB1, MBP or thioredoxin. If prepared as a single construct, however, the solubility tag will be isotopically labeled and will contribute to the NMR spectrum, a complication that can be avoided with segmental labeling.

4.2.2 Site-specific incorporation of magnetic resonance probes—The introduction of a single isotopic label at a well-defined position in a polypeptide chain provides NMR spectroscopists with a benign and unambiguous reporter of protein structure, dynamics and ligand binding. Strategically positioned isotopes can also be crucial in disentangling higher order structural interactions in complex biological assemblies such as amyloid fibrils (Debelouchina *et al.* 2013; Petkova *et al.* 2006). The incorporation of such probes is relatively straightforward in short polypeptides prepared by solid-phase peptide synthesis, and many protected isotopically labeled amino acids are available commercially at a reasonable cost. Longer polypeptide chains containing specific labels can be prepared by NCL as exemplified by the synthesis of the membrane protein M2 containing five $^{13}C,^{15}N$-labeled amino acids dispersed throughout the sequence (Kwon *et al.* 2015). A serendipitously positioned native cysteine residue in this case provided a convenient ligation site without the need for desulfurization.

In recent years, the biomolecular NMR field has seen a surge in the popularity of ^{19}F as a tool to investigate large and complex systems both by solution and solid-state NMR spectroscopy. This nucleus is bio-orthogonal, has magnetic properties that ensure high sensitivity (~80% of the sensitivity of ^1H nuclei) and a wide chemical shift range (~100-fold larger than the range of ^1H). It can therefore serve as a highly sensitive probe of protein and peptide structure, folding and aggregation, dynamics and ligand binding (reviewed in (Chen *et al.* 2013; Sharaf & Gronenborn, 2015). Specific ^{19}F labels can be installed by peptide synthesis, conjugation or fluorination of cysteine and lysine residues, or by supplementing the culture media with fluorinated amino acids (which results in amino acid specific labeling). Increasingly, amber suppression is becoming the site-specific incorporation method of choice ((Marsh & Suzuki, 2014; Sharaf & Gronenborn, 2015). There are several orthogonal tRNA/synthetase systems and fluorinated UAAs available, each imparting different spectroscopic signatures with respect to chemical shift anisotropy and relaxation. These include 3,5-difluorotyrosine (Li *et al.* 2013a; Yang *et al.* 2015), *p*-trifluoromethoxyphenylalanine (Cellitti *et al.* 2008) and *p*-trifluoromethylphenylalanine (Jackson *et al.* 2007; Loscha *et al.* 2012). While most protein ^{19}F NMR applications rely on

amber suppression in *E. coli*, magnetic resonance imaging studies are already exploiting the unique spectroscopic properties of ^{19}F in mammalian systems or whole organisms (Yu *et al.* 2013). Site-specific ^{13}C and ^{15}N probes can also be installed by amber suppression, including ^{13}C/^{15}N-labeled *p*-methoxyphenylalanine and ^{15}N-labeled *o*-nitrobenzyl-tyrosine (Cellitti *et al.* 2008). Ligand-binding studies of large molecular systems may also benefit from incorporation of *o-tert*-butyltyrosine, where rapid bond rotations of the *tert*-butyl group ensure a narrow linewidth of the characteristic peak in solution NMR spectra (Chen *et al.* 2015).

Another problem of particular relevance to the magnetic resonance spectroscopist concerns the site-directed installation of EPR and PRE probes. These probes carry stable nitroxide radicals or chelated metal ions such as Mn^{2+}, Cu^{2+} or Gd^{3+}, and can be used to measure distances by EPR (up to 80 Å), or change the relaxation properties of nuclear spins in their vicinity (up to 35 Å). Usually, paramagnetic tags are incorporated at available cysteine residues using alkylation with maleimides, and many appropriately functionalized reagents are available commercially. As discussed in Section 3.1, however, cysteine-based approaches often have limited selectivity, the generated maleimide-based linker can be unstable, and is generally not applicable to cellular investigations. The development of amber suppression and bio-orthogonal reactions has certainly provided more options for EPR and PRE label incorporation both *in vitro* and *in vivo*. For example, orthogonal tRNA/tRNA synthetase systems are available for direct incorporation of nitroxide-containing (Jones *et al.* 2010; Schmidt *et al.* 2015) and metal chelating UAAs (Park *et al.* 2015). Another amber suppression incorporation route involves the UAA *p*-acetyl-L-phenylalanine that can react with paramagnetic tags functionalized with a hydroxylamine moiety (Fleissner *et al.* 2009). An NMR-specific and often prohibitive requirement for the incorporation of PRE tags with amber suppression, however, is the concurrent need to introduce isotopic labels in the protein of interest. Thus, general protocols that provide efficient UAA incorporation under suboptimal bacterial growth conditions (minimal ^{13}C/^{15}N-supplemented or perdeuterated media, for example) are desperately needed (Evans & Millhauser, 2015; Venditti *et al.* 2012).

Another demanding problem concerns the site-specific installation of isotopically labeled PTMs such as different methylation states, glycosylation, ubiquitylation and various acylations (phosphorylation being somewhat of an exception as the naturally abundant ^{31}P is the NMR active isotope). Coupled with the often problematic behavior of the modified proteins (e.g. aggregation prone intrinsically disordered and amyloidogeneic peptides and proteins, insoluble transmembrane domains or chromatin effectors), PTM installation in itself is a challenge even without the extra complication and associated cost of isotopic labeling. Not surprisingly, therefore, most NMR studies involving PTMs are performed with constructs that have been modified enzymatically. Enzymatic modification, however, is often incomplete and can result in a heterogeneous set of modifications, thus complicating the interpretation of NMR data. Well-established non-enzymatic site-directed protocols are available for ubiquitylation, taking advantage of the possibility to prepare and isotopically label ubiquitin recombinantly. The labeled ubiquitin moiety can then be installed at a specific position of the modified polypeptide using a variety of strategies including an asymmetric disulfide (Debelouchina *et al.* 2017), a condensation reaction of a C-terminal

ubiquitin thioester and the ε-amine of a targeted lysine residue (Castaneda *et al.* 2011a), or a combined recombinant expression–chemical synthesis approach (Castaneda *et al.* 2011b). Site-specific phosphorylation can be incorporated by solid-phase peptide synthesis, followed by EPL to construct segmentally labeled and modified proteins (Hejjaoui *et al.* 2012). Studies of glycosylated proteins also require the attachment of homogeneous glycans that are often isotopically labeled to facilitate resonance assignments and structure determination (Skrisovska *et al.* 2010). Cysteine-based ^{13}C-labeled methyllysine analogs are a convenient and affordable way to generate methylated constructs, and have been used in the NMR structural analysis of proteins such as HP1 and p53 (Cui *et al.* 2012; Munari *et al.* 2012).

4.3 Studies of dynamic interactions

4.3.1 Incorporation of optical probes—The advances in single-molecule optical spectroscopy and imaging have propelled the search for better fluorescent probes and the development of more efficient and targeted strategies for their incorporation *in vitro* and in cells. The information obtained from such experiments can complement high-resolution structural studies and provide data not easily accessible by other methods, including dynamic protein–protein interactions, conformational states and protein folding pathways. There are, of course, many excellent reviews on the subject (Dimura *et al.* 2016; Haney *et al.* 2015; Minoshima & Kikuchi, 2017; Nikic & Lemke, 2015). Therefore, we will limit our discussion to the problem of incorporating two small FRET probes for biophysical and structural investigations, still a surprisingly challenging task for the protein chemist.

A protein construct prepared for FRET experiments can be labeled with two different fluorophores (A and B) by cysteine mutagenesis at two appropriately selected sites. Incubation with the maleimide, iodoacetamide or methyl bromide derivatives of the two probes results in a statistical mixture of labeled constructs even if labeling is efficient (A–A, A–B, B–A, B–B). Since the mixture of products can complicate data analysis (Husada *et al.* 2015), more selective strategies for labeling are highly desirable and have been explored in the literature. A popular approach, for example, combines cysteine labeling with amber suppression that can introduce a second fluorescent moiety either directly or through bio-orthogonal chemistries (Brustad *et al.* 2008; Haney *et al.* 2016; Milles *et al.* 2012; Ratzke *et al.* 2014). Suitable UAAs include *p*-acetylphenylalanine for labeling with commercially available hydroxylamine probe derivatives (Fig. 14*a*), as well as azido- and alkyne-functionalized phenylalanine, tyrosine and lysine residues for click reactions. Cysteine labeling can also be combined with the site-specific incorporation of 1,2-aminothiols that react selectively with cyanobenzothiazole fluorescent-probe derivatives (Nguyen *et al.* 2011). A conceptually different cysteine-based labeling approach involves the FlAsH system that targets biarsenic reagents to the genetically encoded tetracysteine motif CCPGCC (Griffin *et al.* 1998). This technology has the advantage that the small molecule ligands are cell permeable and virtually non-fluorescent until they bind their recognition motif, and are therefore well-suited for in-cell applications. In the context of dual labeling, FlAsH has been combined with amber suppression (Perdios *et al.* 2017) (Fig. 14*b*) or fluorescent proteins such as CFP (Hoffmann *et al.* 2005) for the investigation of protein conformational states *in vitro* or in the cellular environment.

When the protein construct of interest contains functionally or structurally important cysteines, dual labeling can be attempted with NCL, EPL, the incorporation of genetically encodable UAAs and peptide tags for bioenzymatic conjugation, or a combination of these approaches. For example, a fluorescent probe and a thioamide quencher can be installed using a combination of amber suppression and protein ligation, where the quencher is a backbone modification that is only accessible through protein synthesis (Wissner *et al.* 2013) (Fig. 14*c*). As discussed in Section 3.6.3, incorporation of two UAAs in a polypeptide chain by genetic means is still difficult; however, the growing numbers of successful applications are almost exclusively developed for FRET-labeling. For example, using cells containing the orthogonal ribosome ribo-Q1, two fluorescent labels could be incorporated efficiently into calmodulin, one in response to the amber TAG codon, the other at the newly assigned AGTA quadruplet codon (Sachdeva *et al.* 2014) (Fig. 14*d*). Rapid, quantitative, one-pot labeling was achieved by using mutually orthogonal cycloadditions at UAAs bearing a terminal alkyne and a cyclopropene. Alternatively, amber–ochre or amber–opal dual incorporation schemes can be used (Chatterjee *et al.* 2013b; Wan *et al.* 2010; Xiao *et al.* 2013).

4.3.2 Incorporation of vibrational probes—Two-dimensional infrared spectroscopy (2D IR) has emerged as a powerful technique to characterize the dynamic states and interactions of biological macromolecules (Baiz *et al.* 2013; Le Sueur *et al.* 2015). Dynamic and structural information is encoded in the vibrational modes of functional groups such as backbone amides and carbonyls, as well as side-chains with aromatic, carbonyl and guanidinium groups. Since even polypeptides of moderate length contain many overlapping signals, site-specific incorporation of vibrational reporters with distinct spectroscopic signatures is essential for data interpretation. The installation of such probes can take advantage of several of the protein engineering approaches described above (Zhang *et al.* 2016b). Segmental isotopic labeling, for example, can be used to distinguish the spectroscopic behavior of protein domains (Moran *et al.* 2012). In this case, one domain is labeled with ^{13}C, while the other domain remains at natural abundance. The higher mass of the ^{13}C isotope changes the vibrational modes of the backbone carbonyls, and results in a downward shift of the vibrational frequency, effectively decoupling the spectroscopic signatures of the two domains. Higher resolution in linear and multidimensional IR spectra can be achieved by the site-specific installation of isotopic labels such as ^{13}C–^{18}O pairs by solid-phase peptide synthesis, EPL, NCL or the incorporation of appropriately labeled methionine during recombinant protein expression (Courter *et al.* 2014; Davis *et al.* 2015; Dhayalan *et al.* 2016; Marecek *et al.* 2007; Zhang *et al.* 2016b). Resolution and specificity can also be achieved by the introduction of orthogonal vibrational probes such as nitrile, cyano, azido and thiocyanate functional groups. To this end, several UAAs suitable for amber suppression are available, including 4-cyano-, 4-azido- and 4-azidomethyl-phenylalanine (Bazewicz *et al.* 2013; Schultz *et al.* 2006; Ye *et al.* 2010). ^{13}C, ^{15}N-labeled thiocyanates, on the other hand, can be obtained by cyanylation of cysteine residues (van Wilderen *et al.* 2014).

5. Outlook

Our goal in this review has been to examine the contents of the modern protein engineering toolbox, with the particular needs of the structural biologist in mind. As is hopefully evident from the preceding sections, there are now many highly versatile strategies to manipulate and decorate protein constructs, allowing the creation of molecules and assemblies that faithfully represent the complexity of biological systems. We hope that this resource will serve as an inspiration to the structural biologist looking to improve on sample preparation protocols, construct interesting and relevant samples, or devise new strategies that push the boundaries of modern structural biology methods. Undoubtedly, these tools will continue to improve and evolve, expanding the range of options for both *in vitro* and *in vivo* applications. Yet, many challenges still lie ahead. For example, better tools are desperately needed to reign in obstinate systems such as membrane proteins, intrinsically disordered domains or amyloidogeneic polypeptides (Butterfield *et al.* 2012; Uversky, 2015; Zuo *et al.* 2015). These constructs, traditionally recalcitrant to protein engineering and structural biology analysis alike, beg for more efficient (semi)-synthetic, genetic or intein/ sortase-based approaches that circumvent low expression yields and poor solubility issues. Next, the site-specific manipulation of large proteins and assemblies (especially those prepared in eukaryotic systems) is still an overwhelmingly difficult task, and new chemical tools and ideas are needed to supplement the existing genetic and enzymatic approaches. At the same time, the demand for constructs bearing multiple modifications will continue to increase. For example, dissecting the molecular basis of biological cross-talks requires access to homogeneous samples of proteins carrying multiple chemically diverse PTMs (Allis & Muir, 2011; Bah & Forman-Kay, 2016). Many biological molecules and assemblies are also too complex for structural analysis by one technique alone (Cramer, 2016; McGinty & Tan, 2015; Tynan *et al.* 2016), and the increasing need for method integration will necessitate the construction of samples suitable for multi-modal studies.

Concurrently, structural biologists are devising new strategies to improve the sensitivity of their respective methodologies and reduce the amounts of precious biological materials required for structural analysis. New technological developments such as X-ray free-electron lasers (Neutze *et al.* 2015), polarization enhancement strategies for NMR spectroscopy (Ardenkjaer-Larsen *et al.* 2015; Maly *et al.* 2008), and the enhanced capabilities of cryo-EM instrumentation (Nogales, 2016) will certainly be central to these efforts. The rise of mass spectrometry based structural approaches, including hydrogen–deuterium exchange, ion mobility-mass spectrometry or cross-linking based analysis, also holds the bright promise to reveal structural information from samples in the picomole regime (Lossl *et al.* 2016). At the same time, structural biologists have started to turn to the cellular milieu as the future arena of structural endeavors (Beck & Baumeister, 2016; Freedberg & Selenko, 2014). In this undoubtedly daunting task, structural biologists are not alone. Protein engineers, with their long track record of successes in the selective manipulation of complex systems both *in vitro* and *in vivo*, are ready to meet these exciting new challenges through innovations that will continue to push the boundaries of chemical biology.

Acknowledgments

References

Agard NJ, Prescher JA, Bertozzi CR. A strain-promoted [3 + 2] azide-alkyne cycloaddition for covalent modification of biomolecules in living systems. Journal of the American Chemical Society. 2004; 126(46):15046–15047. [PubMed: 15547999]

Agarwal P, Kudirka R, Albers AE, Barfield RM, De Hart GW, Drake PM, Jones LC, Rabuka D. Hydrazino-Pictet-Spengler ligation as a biocompatible method for the generation of stable protein conjugates. Bioconjugate Chemistry. 2013; 24(6):846–851. [PubMed: 23731037]

Ai HW, Shen W, Sagi A, Chen PR, Schultz PG. Probing protein–protein interactions with a genetically encoded photo-crosslinking amino acid. ChemBioChem. 2011; 12(12):1854–1857. [PubMed: 21678540]

Allis CD, Muir TW. Spreading chromatin into chemical biology. ChemBioChem. 2011; 12(2):264–279. [PubMed: 21243714]

Andersen KA, Raines RT. Creating site-specific isopeptide linkages between proteins with the traceless Staudinger ligation. Methods in Molecular Biology. 2015; 1248:55–65. [PubMed: 25616325]

Antos JM, Truttmann MC, Ploegh HL. Recent advances in sortase-catalyzed ligation methodology. Current Opinion in Structural Biology. 2016; 38:111–118. [PubMed: 27318815]

Appleby JH, Zhou K, Volkmann G, Liu XQ. Novel split intein for trans-splicing synthetic peptide onto C terminus of protein. Journal of Biological Chemistry. 2009; 284(10):6194–6199. [PubMed: 19136555]

Arbely E, Natan E, Brandt T, Allen MD, Veprintsev DB, Robinson CV, Chin JW, Joerger AC, Fersht AR. Acetylation of lysine 120 of p53 endows DNA-binding specificity at effective physiological salt concentration. Proceedings of the National Academy of Sciences United States of America. 2011; 108(20):8251–8256.

Ardenkjaer-Larsen JH, Boebinger GS, Comment A, Duckett S, Edison AS, Engelke F, Griesinger C, Griffin RG, Hilty C, Maeda H, Parigi G, Prisner T, Ravera E, Van Bentum J, Vega S, Webb A, Luchinat C, Schwalbe H, Frydman L. Facing and overcoming sensitivity challenges in biomolecular NMR spectroscopy. Angewandte Chemie (International edition in English). 2015; 54(32):9162–9185.

Aulabaugh A, Ding W, Kapoor B, Tabei K, Alksne L, Dushin R, Zatz T, Ellestad G, Huang X. Development of an HPLC assay for *Staphylococcus aureus* sortase: evidence for the formation of the kinetically competent acyl enzyme intermediate. Analytical Biochemistry. 2007; 360(1):14–22. [PubMed: 17107653]

Bah A, Forman-Kay JD. Modulation of intrinsically disordered protein function by post-translational modifications. Journal of Biological Chemistry. 2016; 291(13):6696–6705. [PubMed: 26851279]

Baiz CR, Reppert M, Tokmakoff A. An introduction to protein 2D IR spectroscopy. Ultrafast Infrared Vibrational Spectroscopy. 2013:361–403.

Barton WA, Tzvetkova-Robev D, Erdjument-Bromage H, Tempst P, Nikolov DB. Highly efficient selenomethionine labeling of recombinant proteins produced in mammalian cells. Protein Science. 2006; 15(8):2008–2013. [PubMed: 16823037]

Basle E, Joubert N, Pucheault M. Protein chemical modification on endogenous amino acids. Chemistry & Biology. 2010; 17(3):213–227. [PubMed: 20338513]

Batjargal S, Walters CR, Petersson EJ. Inteins as traceless purification tags for unnatural amino acid proteins. Journal of the American Chemical Society. 2015; 137(5):1734–1737. [PubMed: 25625321]

Bazewicz CG, Liskov MT, Hines KJ, Brewer SH. Sensitive, site-specific, and stable vibrational probe of local protein environments: 4-azidomethyl-L-phenylalanine. Journal of Physical Chemistry B. 2013; 117(30):8987–8993.

Beck M, Baumeister W. Cryo-electron tomography: can it reveal the molecular sociology of cells in atomic detail? Trends in Cell Biology. 2016; 26(11):825–837. [PubMed: 27671779]

Ben-Shem A, Garreau De Loubresse N, Melnikov S, Jenner L, Yusupova G, Yusupov M. The structure of the eukaryotic ribosome at 3·0 A resolution. Science. 2011; 334(6062):1524–1529. [PubMed: 22096102]

Bernardes GJ, Linderoth L, Doores KJ, Boutureira O, Davis BG. Site-selective traceless Staudinger ligation for glycoprotein synthesis reveals scope and limitations. ChemBioChem. 2011; 12(9): 1383–1386. [PubMed: 21598371]

Best M, Degen A, Baalmann M, Schmidt TT, Wombacher R. Two-step protein labeling by using lipoic acid ligase with norbornene substrates and subsequent inverse-electron demand Diels–Alder reaction. ChemBioChem. 2015; 16(8):1158–1162. [PubMed: 25900689]

Bianco A, Townsley FM, Greiss S, Lang K, Chin JW. Expanding the genetic code of *Drosophila melanogaster*. Nature Chemical Biology. 2012; 8(9):748–750. [PubMed: 22864544]

Blackman ML, Royzen M, Fox JM. Tetrazine ligation: fast bioconjugation based on inverse-electron-demand Diels-Alder reactivity. Journal of the American Chemical Society. 2008; 130(41):13518–13519. [PubMed: 18798613]

Brilot AF, Chen JZ, Cheng A, Pan J, Harrison SC, Potter CS, Carragher B, Henderson R, Grigorieff N. Beam-induced motion of vitrified specimen on holey carbon film. Journal of Structural Biology. 2012; 177(3):630–637. [PubMed: 22366277]

Brocchieri L, Karlin S. Protein length in eukaryotic and prokaryotic proteomes. Nucleic Acids Research. 2005; 33(10):3390–3400. [PubMed: 15951512]

Brustad EM, Lemke EA, Schultz PG, Deniz AA. A general and efficient method for the site-specific dual-labeling of proteins for single molecule fluorescence resonance energy transfer. Journal of the American Chemical Society. 2008; 130(52):17664–17665. [PubMed: 19108697]

Burz DS, Dutta K, Cowburn D, Shekhtman A. Mapping structural interactions using in-cell NMR spectroscopy (STINT-NMR). Nature Methods. 2006; 3(2):91–93. [PubMed: 16432517]

Butterfield S, Hejjaoui M, Fauvet B, Awad L, Lashuel HA. Chemical strategies for controlling protein folding and elucidating the molecular mechanisms of amyloid formation and toxicity. Journal of Molecular Biology. 2012; 421(2–3):204–236. [PubMed: 22342932]

Cady SD, Schmidt-Rohr K, Wang J, Soto CS, Degrado WF, Hong M. Structure of the amantadine binding site of influenza M2 proton channels in lipid bilayers. Nature. 2010; 463(7281):689–692. [PubMed: 20130653]

Campbell MG, Veesler D, Cheng A, Potter CS, Carragher B. 2·8 A resolution reconstruction of the thermoplasma acidophilum 20S proteasome using cryo-electron microscopy. Elife. 2015; 4:e06380.

Carvajal-Vallejos P, Pallisse R, Mootz HD, Schmidt SR. Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. Journal of Biological Chemistry. 2012; 287(34):28686–28696. [PubMed: 22753413]

Caspi J, Amitai G, Belenkiy O, Pietrokovski S. Distribution of split DnaE inteins in cyanobacteria. Molecular Microbiology. 2003; 50(5):1569–1577. [PubMed: 14651639]

Castaneda C, Liu J, Chaturvedi A, Nowicka U, Cropp TA, Fushman D. Nonenzymatic assembly of natural polyubiquitin chains of any linkage composition and isotopic labeling scheme. Journal of the American Chemical Society. 2011a; 133(44):17855–17868. [PubMed: 21962295]

Castaneda CA, Spasser L, Bavikar SN, Brik A, Fushman D. Segmental isotopic labeling of ubiquitin chains to unravel monomer-specific molecular behavior. Angewandte Chemie (International edition in English). 2011b; 50(47):11210–11214.

Cellitti SE, Jones DH, Lagpacan L, Hao X, Zhang Q, Hu H, Brittain SM, Brinker A, Caldwell J, Bursulaya B, Spraggon G, Brock A, Ryu Y, Uno T, Schultz PG, Geierstanger BH. *In vivo* incorporation of unnatural amino acids to probe structure, dynamics, and ligand binding in a large protein by nuclear magnetic resonance spectroscopy. Journal of the American Chemical Society. 2008; 130(29):9268–9281. [PubMed: 18576636]

Cergol KM, Thompson RE, Malins LR, Turner P, Payne RJ. One-pot peptide ligation-desulfurization at glutamate. Organic Letters. 2014; 16(1):290–293. [PubMed: 24294973]

Chakraborty A, Mazumder A, Lin M, Hasemeyer A, Xu Q, Wang D, Ebright YW, Ebright RH. Site-specific incorporation of probes into RNA polymerase by unnatural-amino-acid mutagenesis and Staudinger–Bertozzi ligation. Methods in Molecular Biology. 2015; 1276:101–131. [PubMed: 25665560]

Chalker JM, Bernardes GJ, Lin YA, Davis BG. Chemical modification of proteins at cysteine: opportunities in chemistry and biology. Chemistry, an Asian Journal. 2009; 4(5):630–640.

Chalker JM, Lercher L, Rose NR, Schofield CJ, Davis BG. Conversion of cysteine into dehydroalanine enables access to synthetic histones bearing diverse post-translational modifications. Angewandte Chemie (International edition in English). 2012; 51(8):1835–1839.

Chan AO, Ho CM, Chong HC, Leung YC, Huang JS, Wong MK, Che CM. Modification of N-terminal alpha-amino groups of peptides and proteins using ketenes. Journal of the American Chemical Society. 2012; 134(5):2589–2598. [PubMed: 22288779]

Chatterjee A, Guo J, Lee HS, Schultz PG. A genetically encoded fluorescent probe in mammalian cells. Journal of the American Chemical Society. 2013a; 135(34):12540–12543. [PubMed: 23924161]

Chatterjee A, Sun SB, Furman JL, Xiao H, Schultz PG. A versatile platform for single- and multiple-unnatural amino acid mutagenesis in *Escherichia coli.* Biochemistry. 2013b; 52(10):1828–1837. [PubMed: 23379331]

Chatterjee A, Xiao H, Bollong M, Ai HW, Schultz PG. Efficient viral delivery system for unnatural amino acid mutagenesis in mammalian cells. Proceedings of the National Academy of Sciences United States of America. 2013c; 110(29):11803–11808.

Chatterjee C, Mcginty RK, Fierz B, Muir TW. Disulfide-directed histone ubiquitylation reveals plasticity in hDot1L activation. Nature Chemical Biology. 2010; 6(4):267–269. [PubMed: 20208522]

Chen H, Viel S, Ziarelli F, Peng L. 19F NMR: a valuable tool for studying biological events. Chemical Society Reviews. 2013; 42(20):7971–7982. [PubMed: 23864138]

Chen I, Dorr BM, Liu DR. A general strategy for the evolution of bond-forming enzymes using yeast display. Proceedings of the National Academy of Sciences United States of America. 2011; 108(28):11399–11404.

Chen J, Ai Y, Wang J, Haracska L, Zhuang Z. Chemically ubiquitylated PCNA as a probe for eukaryotic translesion DNA synthesis. Nature Chemical Biology. 2010a; 6(4):270–272. [PubMed: 20208521]

Chen J, Wan Q, Yuan Y, Zhu JL, Danishefsky SJ. Native chemical ligation at valine: a contribution to peptide and glycopep-tide synthesis. Angewandte Chemie (International Edition in English). 2008; 47(44):8521–8524.

Chen J, Wang P, Zhu JL, Wan Q, Danishefsky SJ. A program for ligation at threonine sites: application to the controlled total synthesis of glycopeptides. Tetrahedron. 2010b; 66(13):2277–2283. [PubMed: 20798898]

Chen PR, Groff D, Guo J, Ou W, Cellitti S, Geierstanger BH, Schultz PG. A facile system for encoding unnatural amino acids in mammalian cells. Angewandte Chemie (International edition in English). 2009; 48(22):4052–4055.

Chen WN, Kuppan KV, Lee MD, Jaudzems K, Huber T, Otting G. O-tert-Butyltyrosine, an NMR tag for high-molecular-weight systems and measurements of submicromolar ligand binding affinities. Journal of the American Chemical Society. 2015; 137(13):4581–4586. [PubMed: 25789794]

Cheng Y. Single-particle cryo-em at crystallographic resolution. Cell. 2015; 161(3):450–457. [PubMed: 25910205]

Cheriyan M, Pedamallu CS, Tori K, Perler F. Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. Journal of Biological Chemistry. 2013; 288(9): 6202–6211. [PubMed: 23306197]

Chin JW. Expanding and reprogramming the genetic code of cells and animals. Annual Review of Biochemistry. 2014; 83:379–408.

Chin JW, Martin AB, King DS, Wang L, Schultz PG. Addition of a photocrosslinking amino acid to the genetic code of *Escherichia coli*. Proceedings of the National Academy of Sciences United States of America. 2002a; 99(17):11020–11024.

Chin JW, Santoro SW, Martin AB, King DS, Wang L, Schultz PG. Addition of p-azido-L-phenylalanine to the genetic code of *Escherichia coli*. Journal of the American Chemical Society. 2002b; 124(31):9026–9027. [PubMed: 12148987]

Chou CJ, Uprety R, Davis L, Chin JW, Deiters A. Genetically encoding an aliphatic diazirine for protein photocrosslinking. Chemical Science. 2011; 2(3):480–483.

Chow WY, Rajan R, Muller KH, Reid DG, Skepper JN, Wong WC, Brooks RA, Green M, Bihan D, Farndale RW, Slatter DA, Shanahan CM, Duer MJ. NMR spectroscopy of native and *in vitro* tissues implicates polyADP ribose in biomineralization. Science. 2014; 344(6185):742–746. [PubMed: 24833391]

Cohen JD, Zou P, Ting AY. Site-specific protein modification using lipoic acid ligase and bis-aryl hydrazone formation. ChemBioChem. 2012; 13(6):888–894. [PubMed: 22492621]

Cohen S, Arbely E. Single-plasmid-based system for efficient noncanonical amino acid mutagenesis in cultured mammalian cells. ChemBioChem. 2016; 17(11):1008–1011. [PubMed: 27120490]

Cornish VW, Benson DR, Altenbach CA, Hideg K, Hubbell WL, Schultz PG. Site-specific incorporation of biophysical probes into proteins. Proceedings of the National Academy of Sciences United States of America. 1994; 91(8):2910–2914.

Courter JR, Abdo M, Brown SP, Tucker MJ, Hochstrasser RM, Smith AB III. The design and synthesis of alanine-rich alpha-helical peptides constrained by an S,S-tetrazine photochemical trigger: a fragment union approach. Journal of Organic Chemistry. 2014; 79(2):759–768. [PubMed: 24359446]

Cramer P. Structure determination of transient transcription complexes. Biochemical Society Transactions. 2016; 44(4):1177–1182. [PubMed: 27528766]

Crich D, Banerjee A. Native chemical ligation at phenylalanine. Journal of the American Chemical Society. 2007; 129(33):10064. [PubMed: 17658806]

Cronin CN, Lim KB, Rogers J. Production of selenomethionyl-derivatized proteins in baculovirus-infected insect cells. Protein Science. 2007; 16(9):2023–2029. [PubMed: 17660253]

Cui G, Park S, Badeaux AI, Kim D, Lee J, Thompson JR, Yan F, Kaneko S, Yuan Z, Botuyan MV, Bedford MT, Cheng JQ, Mer G. PHF20 is an effector protein of p53 double lysine methylation that stabilizes and activates p53. Nature Structural & Molecular Biology. 2012; 19(9):916–924.

David Y, Vila-Perello M, Verma S, Muir TW. Chemical tagging and customizing of cellular chromatin states using ultrafast trans-splicing inteins. Nature Chemistry. 2015; 7(5):394–402.

Davis CM, Cooper AK, Dyer RB. Fast helix formation in the B domain of protein A revealed by site-specific infrared probes. Biochemistry. 2015; 54(9):1758–1766. [PubMed: 25706439]

Dawson PE. Native chemical ligation combined with desulfurization and deselenization: a general strategy for chemical protein synthesis. Israel Journal of Chemistry. 2011; 51(8–9):862–867.

Dawson PE, Kent SB. Synthesis of native proteins by chemical ligation. Annual Review of Biochemistry. 2000; 69:923–960.

Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Synthesis of proteins by native chemical ligation. Science. 1994; 266(5186):776–779. [PubMed: 7973629]

Debelouchina GT, Bayro MJ, Fitzpatrick AW, Ladizhansky V, Colvin MT, Caporini MA, Jaroniec CP, Bajaj VS, Rosay M, Macphee CE, Vendruscolo M, Maas WE, Dobson CM, Griffin RG. Higher order amyloid fibril structure by MAS NMR and DNP spectroscopy. Journal of the American Chemical Society. 2013; 135(51):19237–19247. [PubMed: 24304221]

Debelouchina GT, Gerecht K, Muir TW. Ubiquitin utilizes an acidic surface patch to alter chromatin structure. Nature Chemical Biology. 2017; 13(1):105–110. [PubMed: 27870837]

De Rosier DJ, Klug A. Reconstruction of three dimensional structures from electron micrographs. Nature. 1968; 217(5124):130–134. [PubMed: 23610788]

Devaraj NK, Weissleder R, Hilderbrand SA. Tetrazine-based cycloadditions: application to pretargeted live cell imaging. Bioconjugate Chemistry. 2008; 19(12):2297–2299. [PubMed: 19053305]

Dhayalan B, Fitzpatrick A, Mandal K, Whittaker J, Weiss MA, Tokmakoff A, Kent SB. Efficient total chemical synthesis of (13) C=(18) O isotopomers of human insulin for isotope-edited FTIR. ChemBioChem. 2016; 17(5):415–420. [PubMed: 26715336]

Dimura M, Peulen TO, Hanke CA, Prakash A, Gohlke H, Seidel CA. Quantitative FRET studies and integrative modeling unravel the structure and dynamics of biomolecular systems. Current Opinion in Structural Biology. 2016; 40:163–185. [PubMed: 27939973]

Dirksen A, Dawson PE. Rapid oxime and hydrazone ligations with aromatic aldehydes for biomolecular labeling. Bioconjugate Chemistry. 2008; 19(12):2543–2548. [PubMed: 19053314]

Dommerholt J, Schmidt S, Temming R, Hendriks LJ, Rutjes FP, Van Hest JC, Lefeber DJ, Friedl P, Van Delft FL. Readily accessible bicyclononynes for bioorthogonal labeling and three-dimensional imaging of living cells. Angewandte Chemie (International edition in English). 2010; 49(49): 9422–9425.

Dorr BM, Ham HO, An C, Chaikof EL, Liu DR. Reprogramming the specificity of sortase enzymes. Proceedings of the National Academy of Sciences United States of America. 2014; 111(37): 13343–13348.

El Oualid F, Merkx R, Ekkebus R, Hameed DS, Smit JJ, De Jong A, Hilkmann H, Sixma TK, Ovaa H. Chemical synthesis of ubiquitin, ubiquitin-based probes, and diubiquitin. Angewandte Chemie (International edition in English). 2010; 49(52):10149–10153.

Elsasser SJ, Ernst RJ, Walker OS, Chin JW. Genetic code expansion in stable cell lines enables encoded chromatin modification. Nature Methods. 2016; 13(2):158–164. [PubMed: 26727110]

Elsohly AM, Francis MB. Development of oxidative coupling strategies for site-selective protein modification. Accounts of Chemical Research. 2015; 48(7):1971–1978. [PubMed: 26057118]

England PM, Zhang Y, Dougherty DA, Lester HA. Backbone mutations in transmembrane domains of a ligand-gated ion channel: implications for the mechanism of gating. Cell. 1999; 96(1):89–98. [PubMed: 9989500]

Ernst RJ, Krogager TP, Maywood ES, Zanchi R, Beranek V, Elliott TS, Barry NP, Hastings MH, Chin JW. Genetic code expansion in the mouse brain. Nature Chemical Biology. 2016; 12(10):776–778. [PubMed: 27571478]

Evans EG, Millhauser GL. Genetic incorporation of the unnatural amino acid p-acetyl phenylalanine into proteins for site-directed spin labeling. Methods in Enzymology. 2015; 563:503–527. [PubMed: 26478497]

Fan C, Ip K, Soll D. Expanding the genetic code of *Escherichia coli* with phosphotyrosine. FEBS Letters. 2016; 590(17):3040–3047. [PubMed: 27477338]

Fawzi NL, Ying J, Ghirlando R, Torchia DA, Clore GM. Atomic-resolution dynamics on the surface of amyloid-beta pro-tofibrils probed by solution NMR. Nature. 2011; 480(7376):268–272. [PubMed: 22037310]

Fernandez IS, Bai XC, Hussain T, Kelley AC, Lorsch JR, Ramakrishnan V, Scheres SH. Molecular architecture of a eukaryotic translational initiation complex. Science. 2013; 342(6160):1240585. [PubMed: 24200810]

Ficht S, Payne RJ, Brik A, Wong CH. Second-generation sugar-assisted ligation: a method for the synthesis of cysteine-containing glycopeptides. Angewandte Chemie (International edition in English). 2007; 46(31):5975–5979.

Fitzpatrick AW, Debelouchina GT, Bayro MJ, Clare DK, Caporini MA, Bajaj VS, Jaroniec CP, Wang L, Ladizhansky V, Muller SA, Macphee CE, Waudby CA, Mott HR, De Simone A, Knowles TP, Saibil HR, Vendruscolo M, Orlova EV, Griffin RG, Dobson CM. Atomic structure and hierarchical assembly of a cross-beta amyloid fibril. Proceedings of the National Academy of Sciences United States of America. 2013; 110(14):5468–5473.

Fleissner MR, Brustad EM, Kalai T, Altenbach C, Cascio D, Peters FB, Hideg K, Peuker S, Schultz PG, Hubbell WL. Site-directed spin labeling of a genetically encoded unnatural amino acid. Proceedings of the National Academy of Sciences United States of America. 2009; 106(51): 21637–21642.

Frederick KK, Michaelis VK, Corzilius B, Ong TC, Jacavone AC, Griffin RG, Lindquist S. Sensitivity-enhanced NMR reveals alterations in protein structure by cellular milieus. Cell. 2015; 163(3):620–628. [PubMed: 26456111]

Freedberg DI, Selenko P. Live cell NMR. Annual Review of Biophysics. 2014; 43:171–192.

Freiburger L, Sonntag M, Hennig J, Li J, Zou P, Sattler M. Efficient segmental isotope labeling of multi-domain proteins using Sortase A. Journal of Biomolecular NMR. 2015; 63(1):1–8. [PubMed: 26319988]

Frutos S, Goger M, Giovani B, Cowburn D, Muir TW. Branched intermediate formation stimulates peptide bond cleavage in protein splicing. Nature Chemical Biology. 2010; 6(7):527–533. [PubMed: 20495572]

Gamblin DP, Van Kasteren S, Bernardes GJ, Chalker JM, Oldham NJ, Fairbanks AJ, Davis BG. Chemical site-selective prenylation of proteins. Molecular BioSystems. 2008; 4(6):558–561. [PubMed: 18493653]

Gattner MJ, Vrabel M, Carell T. Synthesis of epsilon-N-propionyl-, epsilon-N-butyryl-, and epsilon-N-crotonyl-lysine containing histone H3 using the pyrrolysine system. Chemical Communications (Cambridge). 2013; 49(4):379–381.

George S, Aguirre JD, Spratt DE, Bi Y, Jeffery M, Shaw GS, O'donoghue P. Generation of phospho-ubiquitin variants by orthogonal translation reveals codon skipping. FEBS Letters. 2016; 590(10): 1530–1542. [PubMed: 27096575]

Gilles MA, Hudson AQ, Borders CL Jr. Stability of water-soluble carbodiimides in aqueous solution. Analytical Biochemistry. 1990; 184(2):244–248. [PubMed: 2158246]

Glasgow JE, Salit ML, Cochran JR. *In vivo* site-specific protein tagging with diverse amines using an engineered sortase variant. Journal of the American Chemical Society. 2016; 138(24):7496–7499. [PubMed: 27280683]

Goto Y, Katoh T, Suga H. Flexizymes for genetic code reprogramming. Nature Protocols. 2011; 6(6): 779–790. [PubMed: 21637198]

Grayson EJ, Ward SJ, Hall AL, Rendle PM, Gamblin DP, Batsanov AS, Davis BG. Glycosyl disulfides: novel glyco-sylating reagents with flexible aglycon alteration. Journal of Organic Chemistry. 2005; 70(24):9740–9754. [PubMed: 16292802]

Greiss S, Chin JW. Expanding the genetic code of an animal. Journal of the American Chemical Society. 2011; 133(36):14196–14199. [PubMed: 21819153]

Griffin BA, Adams SR, Tsien RY. Specific covalent labeling of recombinant protein molecules inside live cells. Science. 1998; 281(5374):269–272. [PubMed: 9657724]

Grosse W, Essen LO, Koert U. Strategies and perspectives in ion-channel engineering. ChemBioChem. 2011; 12(6):830–839. [PubMed: 21472911]

Guan D, Ramirez M, Chen Z. Split intein mediated ultra-rapid purification of tagless protein (SIRP). Biotechnology and Bioengineering. 2013; 110(9):2471–2481. [PubMed: 23568256]

Guimaraes CP, Witte MD, Theile CS, Bozkurt G, Kundrat L, Blom AE, Ploegh HL. Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. Nature Protocols. 2013; 8(9):1787–1799. [PubMed: 23989673]

Haase C, Rohde H, Seitz O. Native chemical ligation at valine. Angewandte Chemie (International edition in English). 2008; 47(36):6807–6810.

Hancock SM, Uprety R, Deiters A, Chin JW. Expanding the genetic code of yeast for incorporation of diverse unnatural amino acids via a pyrrolysyl-tRNA synthetase/tRNA pair. Journal of the American Chemical Society. 2010; 132(42):14819–14824. [PubMed: 20925334]

Haney CM, Wissner RF, Petersson EJ. Multiply labeling proteins for studies of folding and stability. Current Opinion in Chemical Biology. 2015; 28:123–130. [PubMed: 26253346]

Haney CM, Wissner RF, Warner JB, Wang YJ, Ferrie JJ, Covell DJ, Karpowicz RJ, Lee VM, Petersson EJ. Comparison of strategies for non-perturbing labeling of alpha-synuclein to study amyloidogenesis. Organic & Biomolecular. 2016; 14(5):1584–1592.

Harmand TJ, Murar CE, Bode JW. New chemistries for chemoselective peptide ligations and the total synthesis of proteins. Current Opinion in Chemical Biology. 2014; 22:115–121. [PubMed: 25299573]

Harpaz Z, Siman P, Kumar KSA, Brik A. Protein synthesis assisted by native chemical ligation at leucine. ChemBioChem. 2010; 11(9):1232–1235. [PubMed: 20437446]

Hauke S, Best M, Schmidt TT, Baalmann M, Krause A, Wombacher R. Two-step protein labeling utilizing lipoic acid ligase and Sonogashira cross-coupling. Bioconjugate Chemistry. 2014; 25(9): 1632–1637. [PubMed: 25152073]

He Y, Fang J, Taatjes DJ, Nogales E. Structural visualization of key steps in human transcription initiation. Nature. 2013; 495(7442):481–486. [PubMed: 23446344]

Hecht SM, Alford BL, Kuroda Y, Kitano S. 'Chemical aminoacylation' of tRNA's. Journal of Biological Chemistry. 1978; 253(13):4517–4520. [PubMed: 248056]

Hejjaoui M, Butterfield S, Fauvet B, Vercruysse F, Cui J, Dikiy I, Prudent M, Olschewski D, Zhang Y, Eliezer D, Lashuel HA. Elucidating the role of C-terminal post-translational modifications using protein semisynthesis strategies: alpha-synuclein phosphorylation at tyrosine 125. Journal of the American Chemical Society. 2012; 134(11):5196–5210. [PubMed: 22339654]

Hendrickson WA, Horton JR, Lemaster DM. Selenomethionyl proteins produced for analysis by multiwavelength anomalous diffraction (MAD): a vehicle for direct determination of three-dimensional structure. EMBO Journal. 1990; 9(5):1665–1672. [PubMed: 2184035]

Hirakawa H, Ishikawa S, Nagamune T. Ca^{2+} -independent sortase-A exhibits high selective protein ligation activity in the cytoplasm of *Escherichia coli*. Biotechnol Journal. 2015; 10(9):1487–1492.

Hirata R, Ohsumk Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y. Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. Journal of Biological Chemistry. 1990; 265(12):6726–6733. [PubMed: 2139027]

Hoffmann C, Gaietta G, Bunemann M, Adams SR, Oberdorff-Maass S, Behr B, Vilardaga JP, Tsien RY, Ellisman MH, Lohse MJ. A FlAsH-based FRET approach to determine G protein-coupled receptor activation in living cells. Nature Methods. 2005; 2(3):171–176. [PubMed: 15782185]

Holding AN. XL-MS: protein cross-linking coupled with mass spectrometry. Methods. 2015; 89:54–63. [PubMed: 26079926]

Hondal RJ, Nilsson BL, Raines RT. Selenocysteine in native chemical ligation and expressed protein ligation. Journal of the American Chemical Society. 2001; 123(21):5140–5141. [PubMed: 11457362]

Hu QY, Berti F, Adamo R. Towards the next generation of biomedicines by site-selective conjugation. Chemical Society Reviews. 2016; 45(6):1691–1719. [PubMed: 26796469]

Huber T, Naganathan S, Tian H, Ye S, Sakmar TP. Unnatural amino acid mutagenesis of GPCRs using amber codon suppression and bioorthogonal labeling. Methods in Enzymology. 2013; 520:281–305. [PubMed: 23332705]

Huguenin-Dezot N, De Cesare V, Peltier J, Knebel A, Kristaryianto YA, Rogerson DT, Kulathu Y, Trost M, Chin JW. Synthesis of isomeric phosphoubiquitin chains reveals that phosphorylation controls deubiquitinase activity and specificity. Cell Reports. 2016; 16(4):1180–1193. [PubMed: 27425610]

Husada F, Gouridis G, Vietrov R, Schuurman-Wolters GK, Ploetz E, De Boer M, Poolman B, Cordes T. Watching conformational dynamics of ABC transporters with single-molecule tools. Biochemical Society Transactions. 2015; 43(5):1041–1047. [PubMed: 26517920]

Hyman AA, Weber CA, Julicher F. Liquid-liquid phase separation in biology. Annual Review of Cell and Developmental Biology. 2014; 30:39–58.

Igumenova TI, Mcdermott AE, Zilm KW, Martin RW, Paulson EK, Wand AJ. Assignments of carbon NMR resonances for microcrystalline ubiquitin. Journal of the American Chemical Society. 2004; 126(21):6720–6727. [PubMed: 15161300]

Inomata K, Ohno A, Tochio H, Isogai S, Tenno T, Nakase I, Takeuchi T, Futaki S, Ito Y, Hiroaki H, Shirakawa M. High-resolution multi-dimensional NMR spectroscopy of proteins in human cells. Nature. 2009; 458(7234):106–109. [PubMed: 19262675]

Iwai H, Zuger S, Jin J, Tam PH. Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. FEBS Letters. 2006; 580(7):1853–1858. [PubMed: 16516207]

Jackson JC, Hammill JT, Mehl RA. Site-specific incorporation of a (19)F-amino acid into proteins as an NMR probe for characterizing protein structure and reactivity. Journal of the American Chemical Society. 2007; 129(5):1160–1166. [PubMed: 17263397]

Jentoft N, Dearborn DG. Labeling of proteins by reductive methylation using sodium cyanoborohydride. Journal of Biological Chemistry. 1979; 254(11):4359–4365. [PubMed: 571437]

Johnson DB, Xu J, Shen Z, Takimoto JK, Schultz MD, Schmitz RJ, Xiang Z, Ecker JR, Briggs SP, Wang L. RF1 knockout allows ribosomal incorporation of unnatural amino acids at multiple sites. Nature Chemical Biology. 2011; 7(11):779–786. [PubMed: 21926996]

Jones DH, Cellitti SE, Hao X, Zhang Q, Jahnz M, Summerer D, Schultz PG, Uno T, Geierstanger BH. Site-specific labeling of proteins with NMR-active unnatural amino acids. Journal of Biomolecular NMR. 2010; 46(1):89–100. [PubMed: 19669620]

Joshi NS, Whitaker LR, Francis MB. A three-component Mannich-type reaction for selective tyrosine bioconjugation. Journal of the American Chemical Society. 2004; 126(49):15942–15943. [PubMed: 15584710]

Judice JK, Gamble TR, Murphy EC, De Vos AM, Schultz PG. Probing the mechanism of staphylococcal nuclease with unnatural amino acids: kinetic and structural studies. Science. 1993; 261(5128):1578–1581. [PubMed: 8103944]

Kalkhof S, Sinz A. Chances and pitfalls of chemical cross-linking with amine-reactive N-hydroxysuccinimide esters. Analytical and Bioanalytical Chemistry. 2008; 392(1–2):305–312. [PubMed: 18724398]

Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebl M, Stevens TH. Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. Science. 1990; 250(4981):651–657. [PubMed: 2146742]

Kang JY, Kawaguchi D, Coin I, Xiang Z, O'leary DD, Slesinger PA, Wang L. *In vivo* expression of a light-activatable potassium channel using unnatural amino acids. Neuron. 2013; 80(2):358–370. [PubMed: 24139041]

Kendrew JC, Bodo G, Dintzis HM, Parrish RG, Wyckoff H, Phillips DC. A three-dimensional model of the myoglobin molecule obtained by x-ray analysis. Nature. 1958; 181(4610):662–666. [PubMed: 13517261]

Kent S, Sohma Y, Liu S, Bang D, Pentelute B, Mandal K. Through the looking glass–a new world of proteins enabled by chemical synthesis. Journal of Peptide Science. 2012; 18(7):428–436. [PubMed: 22674813]

Kim CH, Kang M, Kim HJ, Chatterjee A, Schultz PG. Site-specific incorporation of epsilon-N-crotonyllysine into histones. Angewandte Chemie (International edition in English). 2012; 51(29):7246–7249.

Kobashigawa Y, Kumeta H, Ogura K, Inagaki F. Attachment of an NMR-invisible solubility enhancement tag using a sortase-mediated protein ligation method. Journal of Biomolecular NMR. 2009; 43(3):145–150. [PubMed: 19140010]

Koehler C, Sauter PF, Wawryszyn M, Girona GE, Gupta K, Landry JJ, Fritz MH, Radic K, Hoffmann JE, Chen ZA, Zou J, Tan PS, Galik B, Junttila S, Stolt-Bergner P, Pruneri G, Gyenesei A, Schultz C, Biskup MB, Besir H, Benes V, Rappsilber J, Jechlinger M, Korbel JO, Berger I, Braese S, Lemke EA. Genetic code expansion for multiprotein complex engineering. Nature Methods. 2016; 13:997–1000. [PubMed: 27749839]

Koh JT, Cornish VW, Schultz PG. An experimental approach to evaluating the role of backbone interactions in proteins using unnatural amino acid mutagenesis. Biochemistry. 1997; 36(38): 11314–11322. [PubMed: 9298950]

Kohrer C, Yoo JH, Bennett M, Schaack J, Rajbhandary UL. A possible approach to site-specific insertion of two different unnatural amino acids into proteins in mammalian cells via nonsense suppression. Chemistry & Biology. 2003; 10(11):1095–1102. [PubMed: 14652077]

Kolb HC, Finn MG, Sharpless KB. Click chemistry: diverse chemical function from a few good reactions. Angewandte Chemie (International edition in English). 2001; 40(11):2004–2021.

Kucher S, Korneev S, Tyagi S, Apfelbaum R, Grohmann D, Lemke EA, Klare JP, Steinhoff HJ, Klose D. Orthogonal spin labeling using click chemistry for *in vitro* and *in vivo* applications. Journal of Magnetic Resonance. 2016; 275:38–45. [PubMed: 27992783]

Kuhlmann N, Wroblowski S, Knyphausen P, De Boor S, Brenig J, Zienert AY, Meyer-Teschendorf K, Praefcke GJ, Nolte H, Kruger M, Schacherl M, Baumann U, James LC, Chin JW, Lammers M.

Structural and mechanistic insights into the regulation of the fundamental rho regulator rhogdialpha by lysine acetylation. Journal of Biological Chemistry. 2016; 291(11):5484–5499. [PubMed: 26719334]

Kuhn SM, Rubini M, Muller MA, Skerra A. Biosynthesis of a fluorescent protein with extreme pseudo-Stokes shift by introducing a genetically encoded non-natural amino acid outside the fluorophore. Journal of the American Chemical Society. 2011; 133(11):3708–3711. [PubMed: 21341705]

Kumar KSA, Haj-Yahya M, Olschewski D, Lashuel HA, Brik A. Highly efficient and chemoselective peptide ubiquitylation. Angewandte Chemie (International edition in English). 2009; 48(43): 8090–8094.

Kwon B, Tietze D, White PB, Liao SY, Hong M. Chemical ligation of the influenza M2 protein for solid-state NMR char-acterization of the cytoplasmic domain. Protein Science. 2015; 24(7): 1087–1099. [PubMed: 25966817]

Kwon I, Wang P, Tirrell DA. Design of a bacterial host for site-specific incorporation of p-bromophenylalanine into recombinant proteins. Journal of the American Chemical Society. 2006; 128(36):11778–11783. [PubMed: 16953616]

Lajoie MJ, Rovner AJ, Goodman DB, Aerni HR, Haimovich AD, Kuznetsov G, Mercer JA, Wang HH, Carr PA, Mosberg JA, Rohland N, Schultz PG, Jacobson JM, Rinehart J, Church GM, Isaacs FJ. Genomically recoded organisms expand biological functions. Science. 2013; 342(6156):357–360. [PubMed: 24136966]

Lammers M, Neumann H, Chin JW, James LC. Acetylation regulates cyclophilin A catalysis, immunosuppression and HIV isomerization. Nature Chemical Biology. 2010; 6(5):331–337. [PubMed: 20364129]

Lang K, Chin JW. Cellular incorporation of unnatural amino acids and bioorthogonal labeling of proteins. Chem Rev. 2014; 114(9):4764–4806. [PubMed: 24655057]

Lang K, Davis L, Torres-Kolbus J, Chou C, Deiters A, Chin JW. Genetically encoded norbornene directs site-specific cellular protein labelling via a rapid bioorthogonal reaction. Nature Chemistry. 2012a; 4(4):298–304.

Lang K, Davis L, Wallace S, Mahesh M, Cox DJ, Blackman ML, Fox JM, Chin JW. Genetic encoding of bicyclono-nynes and trans-cyclooctenes for site-specific protein labeling *in vitro* and in live mammalian cells via rapid fluorogenic Diels-Alder reactions. Journal of the American Chemical Society. 2012b; 134(25):10317–10320. [PubMed: 22694658]

Laughlin ST, Baskin JM, Amacher SL, Bertozzi CR. *In vivo* imaging of membrane-associated glycans in developing zebrafish. Science. 2008; 320(5876):664–667. [PubMed: 18451302]

Lavergne T, Lamichhane R, Malyshev DA, Li Z, Li L, Sperling E, Williamson JR, Millar DP, Romesberg FE. FRET characterization of complex conformational changes in a large 16S ribosomal RNA fragment site-specifically labeled using unnatural base pairs. ACS Chemical Biology. 2016; 11(5):1347–1353. [PubMed: 26942998]

Lee HS, Guo J, Lemke EA, Dimla RD, Schultz PG. Genetic incorporation of a small, environmentally sensitive, fluorescent probe into proteins in *Saccharomyces cerevisiae*. Journal of the American Chemical Society. 2009a; 131(36):12921–12923. [PubMed: 19702307]

Lee HS, Spraggon G, Schultz PG, Wang F. Genetic incorporation of a metal-ion chelating amino acid into proteins as a biophysical probe. Journal of the American Chemical Society. 2009b; 131(7): 2481–2483. [PubMed: 19193005]

Leitner A, Faini M, Stengel F, Aebersold R. Crosslinking and mass spectrometry: an integrated technology to understand the structure and function of molecular machines. Trends in Biochemical Sciences. 2016; 41(1):20–32. [PubMed: 26654279]

Lemieux GA, De Graffenried CL, Bertozzi CR. A fluorogenic dye activated by the staudinger ligation. Journal of the American Chemical Society. 2003; 125(16):4708–4709. [PubMed: 12696879]

Lennard KR, Tavassoli A. Peptides come round: using SICLOPPS libraries for early stage drug discovery. Chemistry. 2014; 20(34):10608–10614. [PubMed: 25043886]

Le Sueur AL, Horness RE, Thielges MC. Applications of two-dimensional infrared spectroscopy. Analyst. 2015; 140(13):4336–4349. [PubMed: 26007625]

Li D, Pye VE, Caffrey M. Experimental phasing for structure determination using membrane-protein crystals grown by the lipid cubic phase method. Acta Crystallographica D, Biological Crystallography. 2015; 71(Pt 1):104–122. [PubMed: 25615865]

Li F, Shi P, Li J, Yang F, Wang T, Zhang W, Gao F, Ding W, Li D, Li J, Xiong Y, Sun J, Gong W, Tian C, Wang J. A genetically encoded 19F NMR probe for tyrosine phosphorylation. Angewandte Chemie (International edition in English). 2013a; 52(14):3958–3962.

Li F, Zhang H, Sun Y, Pan Y, Zhou J, Wang J. Expanding the genetic code for photoclick chemistry in *E. coli*, mammalian cells, and *A. thaliana*. Angewandte Chemie (International edition in English). 2013b; 52(37):9700–9704.

Lin S, Yang X, Jia S, Weeks AM, Hornsby M, Lee PS, Nichiporuk RV, Iavarone AT, Wells JA, Toste FD, Chang CJ. Redox-based reagents for chemoselective methionine bioconjugation. Science. 2017; 355(6325):597–602. [PubMed: 28183972]

Link AJ, Vink MK, Tirrell DA. Presentation and detection of azide functionality in bacterial cell surface proteins. Journal of the American Chemical Society. 2004; 126(34):10598–10602. [PubMed: 15327317]

Liu CC, Brustad E, Liu W, Schultz PG. Crystal structure of a biosynthetic sulfo-hirudin complexed to thrombin. Journal of the American Chemical Society. 2007; 129(35):10648–10649. [PubMed: 17685615]

Liu CC, Schultz PG. Adding new chemistries to the genetic code. Annual Review of Biochemistry. 2010; 79:413–444.

Liu DS, Tangpeerachaikul A, Selvaraj R, Taylor MT, Fox JM, Ting AY. Diels-Alder cycloaddition for fluorophore targeting to specific proteins inside living cells. Journal of the American Chemical Society. 2012; 134(2):792–795. [PubMed: 22176354]

Liu F, Luo EY, Flora DB, Mezo AR. Irreversible sortase A-mediated ligation driven by diketopiperazine formation. Journal of Organic Chemistry. 2014a; 79(2):487–492. [PubMed: 24378034]

Liu Z, Frutos S, Bick MJ, Vila-Perello M, Debelouchina GT, Darst SA, Muir TW. Structure of the branched intermediate in protein splicing. Proceedings of the National Academy of Sciences United States of America. 2014b; 111(23):8422–8427.

Loquet A, Sgourakis NG, Gupta R, Giller K, Riedel D, Goosmann C, Griesinger C, Kolbe M, Baker D, Becker S, Lange A. Atomic model of the type III secretion system needle. Nature. 2012; 486(7402):276–279. [PubMed: 22699623]

Loscha KV, Herlt AJ, Qi R, Huber T, Ozawa K, Otting G. Multiple-site labeling of proteins with unnatural amino acids. Angewandte Chemie (International edition in English). 2012; 51(9):2243–2246.

Lossl P, Van De Waterbeemd M, Heck AJ. The diverse and expanding role of mass spectrometry in structural and molecular biology. EMBO Journal. 2016; 35(24):2634–2657. [PubMed: 27797822]

Lu JX, Qiang W, Yau WM, Schwieters CD, Meredith SC, Tycko R. Molecular structure of beta-amyloid fibrils in Alzheimer's disease brain tissue. Cell. 2013; 154(6):1257–1268. [PubMed: 24034249]

Lu W, Randal M, Kossiakoff A, Kent SB. Probing intermolecular backbone H-bonding in serine proteinase-protein inhibitor complexes. Chemistry & Biology. 1999; 6(7):419–427. [PubMed: 10381402]

Lu X, Simon MD, Chodaparambil JV, Hansen JC, Shokat KM, Luger K. The effect of H3K79 dimethylation and H4K20 trimethylation on nucleosome and chromatin structure. Nature Structural & Molecular Biology. 2008; 15(10):1122–1124.

Ludwig C, Pfeiff M, Linne U, Mootz HD. Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing. Angewandte Chemie (International edition in English). 2006; 45(31):5218–5221.

Luger K, Mader AW, Richmond RK, Sargent DF, Richmond TJ. Crystal structure of the nucleosome core particle at 2·8 A resolution. Nature. 1997; 389(6648):251–260. [PubMed: 9305837]

Lyon RP, Setter JR, Bovee TD, Doronina SO, Hunter JH, Anderson ME, Balasubramanian CL, Duniho SM, Leiske CI, Li F, Senter PD. Self-hydrolyzing maleimides improve the stability and

pharmacological properties of antibody-drug conjugates. Nature Biotechnology. 2014; 32(10): 1059–1062.

Macdonald JI, Munch HK, Moore T, Francis MB. One-step site-specific modification of native proteins with 2-pyridinecarboxyaldehydes. Nature Chemical Biology. 2015; 11(5):326–331. [PubMed: 25822913]

Macmillan D, Bill RM, Sage KA, Fern D, Flitsch SL. Selective *in vitro* glycosylation of recombinant proteins: semi-synthesis of novel homogeneous glycoforms of human erythropoietin. Chemistry & Biology. 2001; 8(2):133–145. [PubMed: 11251288]

Malins LR, Cergol KM, Payne RJ. Peptide ligation-desulfurization chemistry at Arginine. ChemBioChem. 2013; 14(5):559–563. [PubMed: 23426906]

Malins LR, Cergol KM, Payne RJ. Chemoselective sulfenylation and peptide ligation at tryptophan. Chemical Science. 2014; 5(1):260–266.

Malins LR, Payne RJ. Modern extensions of native chemical ligation for chemical protein synthesis. Topics in Current Chemistry. 2015; 362:27–87. [PubMed: 25518971]

Mallagaray A, Dominguez G, Peters T, Perez-Castells J. A rigid lanthanide binding tag to aid NMR studies of a 70 kDa homodimeric coat protein of human norovirus. Chemical Communications (Cambridge). 2016; 52(3):601–604.

Maly T, Debelouchina GT, Bajaj VS, Hu KN, Joo CG, Mak-Jurkauskas ML, Sirigiri JR, Van Der Wel PC, Herzfeld J, Temkin RJ, Griffin RG. Dynamic nuclear polarization at high magnetic fields. Journal of Chemical Physics. 2008; 128(5):052211. [PubMed: 18266416]

Malyshev DA, Dhami K, Lavergne T, Chen T, Dai N, Foster JM, Correa IR Jr, Romesberg FE. A semi-synthetic organism with an expanded genetic alphabet. Nature. 2014; 509(7500):385–388. [PubMed: 24805238]

Mandal K, Uppalapati M, Ault-Riche D, Kenney J, Lowitz J, Sidhu SS, Kent SB. Chemical synthesis and X-ray structure of a heterochiral {D-protein antagonist plus vascular endothelial growth factor} protein complex by racemic crystallography. Proceedings of the National Academy of Sciences United States of America. 2012; 109(37):14779–14784.

Mandlik A, Swierczynski A, Das A, Ton-That H. Pili in Gram-positive bacteria: assembly, involvement in colonization and biofilm development. Trends in Microbiology. 2008; 16(1):33–40. [PubMed: 18083568]

Mao H, Hart SA, Schink A, Pollok BA. Sortase-mediated protein ligation: a new method for protein engineering. Journal of the American Chemical Society. 2004; 126(9):2670–2671. [PubMed: 14995162]

Marecek J, Song B, Brewer S, Belyea J, Dyer RB, Raleigh DP. A simple and economical method for the production of 13C,18O-labeled Fmoc-amino acids with high levels of enrichment: applications to isotope-edited IR studies of proteins. Organic Letters. 2007; 9(24):4935–4937. [PubMed: 17958432]

Marsh EN, Suzuki Y. Using (19)F NMR to probe biological interactions of proteins and peptides. ACS Chemical Biology. 2014; 9(6):1242–1250. [PubMed: 24762032]

Martinez C, De Geus P, Stanssens P, Lauwereys M, Cambillau C. Engineering cysteine mutants to obtain crystallographic phases with a cutinase from *Fusarium solani pisi*. Protein Engineering. 1993; 6(2):157–165. [PubMed: 8475042]

Matsumoto T, Furuta K, Tanaka T, Kondo A. Sortase A-mediated metabolic enzyme ligation in *Escherichia coli*. ACS Synthetic Biology. 2016; 5(11):1284–1289. [PubMed: 27700053]

Matthies D, Dalmas O, Borgnia MJ, Dominik PK, Merk A, Rao P, Reddy BG, Islam S, Bartesaghi A, Perozo E, Subramaniam S. Cryo-EM structures of the magnesium channel CorA reveal symmetry break upon gating. Cell. 2016; 164(4):747–756. [PubMed: 26871634]

Mazmanian SK, Liu G, Ton-That H, Schneewind O. *Staphylococcus aureus* sortase, an enzyme that anchors surface proteins to the cell wall. Science. 1999; 285(5428):760–763. [PubMed: 10427003]

Mcfarland JM, Francis MB. Reductive alkylation of proteins using iridium catalyzed transfer hydrogenation. Journal of the American Chemical Society. 2005; 127(39):13490–13491. [PubMed: 16190700]

Mcfarland JM, Joshi NS, Francis MB. Characterization of a three-component coupling reaction on proteins by isotopic labeling and nuclear magnetic resonance spectroscopy. Journal of the American Chemical Society. 2008; 130(24):7639–7644. [PubMed: 18498164]

Mcginty RK, Tan S. Nucleosome structure and function. Chemical Reviews. 2015; 115(6):2255–2273. [PubMed: 25495456]

Mcmullan G, Chen S, Henderson R, Faruqi AR. Detective quantum efficiency of electron area detectors in electron microscopy. Ultramicroscopy. 2009; 109(9):1126–1143. [PubMed: 19497671]

Mehler M, Eckert CE, Busche A, Kulhei J, Michaelis J, Becker-Baldus J, Wachtveitl J, Dotsch V, Glaubitz C. Assembling a correctly folded and functional heptahelical membrane protein by protein trans-splicing. Journal of Biological Chemistry. 2015; 290(46):27712–27722. [PubMed: 26405032]

Meier F, Abeywardana T, Dhall A, Marotta NP, Varkey J, Langen R, Chatterjee C, Pratt MR. Semisynthetic, site-specific ubiquitin modification of alpha-synuclein reveals differential effects on aggregation. Journal of the American Chemical Society. 2012; 134(12):5468–5471. [PubMed: 22404520]

Merk A, Bartesaghi A, Banerjee S, Falconieri V, Rao P, Davis MI, Pragani R, Boxer MB, Earl LA, Milne JL, Subramaniam S. Breaking Cryo-EM resolution barriers to facilitate drug discovery. Cell. 2016; 165(7):1698–1707. [PubMed: 27238019]

Metanis, N., Beld, J., Hilvert, D. The chemistry of selenocysteine. In: Rappoport, Z., editor. PATAI'S Chemistry of Functional Groups. John Wiley & Sons, Ltd; 2009.

Mileo E, Etienne E, Martinho M, Lebrun R, Roubaud V, Tordo P, Gontero B, Guigliarelli B, Marque SR, Belle V. Enlarging the panoply of site-directed spin labeling electron paramagnetic resonance (SDSL-EPR): sensitive and selective spin-labeling of tyrosine using an isoindoline-based nitroxide. Bioconjugate Chemistry. 2013; 24(6):1110–1117. [PubMed: 23642211]

Milles S, Tyagi S, Banterle N, Koehler C, Vandelinder V, Plass T, Neal AP, Lemke EA. Click strategies for single-molecule protein fluorescence. Journal of the American Chemical Society. 2012; 134(11):5187–5195. [PubMed: 22356317]

Minoshima M, Kikuchi K. Photostable and photoswitching fluorescent dyes for super-resolution imaging. Journal of Biological Inorganic Chemistry. 2017; doi: 10.1007/s00775-016-1435-y

Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW. Conditional protein splicing: a new tool to control protein structure and function *in vitro* and *in vivo*. Journal of the American Chemical Society. 2003; 125(35):10561–10569. [PubMed: 12940738]

Moran SD, Woys AM, Buchanan LE, Bixby E, Decatur SM, Zanni MT. Two-dimensional IR spectroscopy and segmental C-13 labeling reveals the domain structure of human gamma D-crystallin amyloid fibrils. Proceedings of the National Academy of Sciences of the United States of America. 2012; 109(9):3329–3334. [PubMed: 22328156]

Morgan MT, Haj-Yahya M, Ringel AE, Bandi P, Brik A, Wolberger C. Structural basis for histone H2B deubiquitination by the SAGA DUB module. Science. 2016; 351(6274):725–728. [PubMed: 26912860]

Moyal T, Hemantha HP, Siman P, Refua M, Brik A. Highly efficient one-pot ligation and desulfurization. Chemical Science. 2013; 4(6):2496–2501.

Muir TW, Sondhi D, Cole PA. Expressed protein ligation: a general method for protein engineering. Proceedings of the National Academy of Sciences United States of America. 1998; 95(12):6705–6710.

Mukai T, Hayashi A, Iraha F, Sato A, Ohtake K, Yokoyama S, Sakamoto K. Codon reassignment in the *Escherichia coli* genetic code. Nucleic Acids Research. 2010a; 38(22):8188–8195. [PubMed: 20702426]

Mukai T, Wakiyama M, Sakamoto K, Yokoyama S. Genetic encoding of non-natural amino acids in *Drosophila melanogaster* Schneider 2 cells. Protein Science. 2010b; 19(3):440–448. [PubMed: 20052681]

Munari F, Soeroes S, Zenn HM, Schomburg A, Kost N, Schroder S, Klingberg R, Rezaei-Ghaleh N, Stutzer A, Gelato KA, Walla PJ, Becker S, Schwarzer D, Zimmermann B, Fischle W, Zweckstetter M. Methylation of lysine 9 in histone H3 directs alternative modes of highly

dynamic interaction of heterochromatin protein hHP1beta with the nucleosome. Journal of Biological Chemistry. 2012; 287(40):33756–33765. [PubMed: 22815475]

Muona M, Aranko AS, Raulinaitis V, Iwai H. Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing *in vivo* and *in vitro*. Nature Protocols. 2010; 5(3):574–587. [PubMed: 20203672]

Muralidharan V, Muir TW. Protein ligation: an enabling technology for the biophysical analysis of proteins. Nature Methods. 2006; 3(6):429–438. [PubMed: 16721376]

Naik MT, Suree N, Ilangovan U, Liew CK, Thieu W, Campbell DO, Clemens JJ, Jung ME, Clubb RT. *Staphylococcus aureus* Sortase A transpeptidase. Calcium promotes sorting signal binding by altering the mobility and structure of an active site loop. Journal of Biological Chemistry. 2006; 281(3):1817–1826. [PubMed: 16269411]

Nakamura T, Kawai Y, Kitamoto N, Osawa T, Kato Y. Covalent modification of lysine residues by allyl isothiocyanate in physiological conditions: plausible transformation of isothiocyanate from thiol to amine. Chem Res Toxicol. 2009; 22(3):536–542. [PubMed: 19216492]

Nettleship JE, Assenberg R, Diprose JM, Rahman-Huq N, Owens RJ. Recent advances in the production of proteins in insect and mammalian cells for structural biology. Journal of Structural Biology. 2010; 172(1):55–65. [PubMed: 20153433]

Neumann H, Peak-Chew SY, Chin JW. Genetically encoding N(epsilon)-acetyllysine in recombinant proteins. Nature Chemical Biology. 2008; 4(4):232–234. [PubMed: 18278036]

Neumann H, Wang K, Davis L, Garcia-Alai M, Chin JW. Encoding multiple unnatural amino acids via evolution of a quadruplet-decoding ribosome. Nature. 2010; 464(7287):441–444. [PubMed: 20154731]

Neumann-Staubitz P, Neumann H. The use of unnatural amino acids to study and engineer protein function. Current Opinion in Structural Biology. 2016; 38:119–128. [PubMed: 27318816]

Neutze R, Branden G, Schertler GF. Membrane protein structural biology using X-ray free electron lasers. Current Opinion in Structural Biology. 2015; 33:115–125. [PubMed: 26342349]

Nguyen DP, Elliott T, Holt M, Muir TW, Chin JW. Genetically encoded 1,2-aminothiols facilitate rapid and site-specific protein labeling via a bio-orthogonal cyanobenzothiazole condensation. Journal of the American Chemical Society. 2011; 133(30):11418–11421. [PubMed: 21736333]

Nguyen DP, Garcia Alai MM, Virdee S, Chin JW. Genetically directing varepsilon-N, N-dimethyl-L-lysine in recombinant histones. Chemistry & Biology. 2010; 17(10):1072–1076. [PubMed: 21035729]

Nguyen GK, Kam A, Loo S, Jansson AE, Pan LX, Tam JP. Butelase 1: a versatile ligase for peptide and protein macrocyclization. Journal of the American Chemical Society. 2015; 137(49):15398–15401. [PubMed: 26633100]

Nguyen GK, Wang S, Qiu Y, Hemu X, Lian Y, Tam JP. Butelase 1 is an Asx-specific ligase enabling peptide macrocyclization and synthesis. Nature Chemical Biology. 2014; 10(9):732–738. [PubMed: 25038786]

Nikic I, Lemke EA. Genetic code expansion enabled site-specific dual-color protein labeling: superresolution microscopy and beyond. Current Opinion in Chemical Biology. 2015; 28:164–173. [PubMed: 26302384]

Nilsson BL, Kiessling LL, Raines RT. Staudinger ligation: a peptide from a thioester and azide. Organic Letters. 2000; 2(13):1939–1941. [PubMed: 10891196]

Nogales E. The development of cryo-EM into a mainstream structural biology technique. Nature Methods. 2016; 13(1):24–27. [PubMed: 27110629]

Noren CJ, Anthony-Cahill SJ, Griffith MC, Schultz PG. A general method for site-specific incorporation of unnatural amino acids into proteins. Science. 1989; 244(4901):182–188. [PubMed: 2649980]

Okamoto R, Mandal K, Sawaya MR, Kajihara Y, Yeates TO, Kent SB. (Quasi-)racemic X-ray structures of glycosylated and non-glycosylated forms of the chemokine Ser-CCL1 prepared by total chemical synthesis. Angewandte Chemie (International edition in English). 2014; 53(20):5194–5198.

Ostrov N, Landon M, Guell M, Kuznetsov G, Teramoto J, Cervantes N, Zhou M, Singh K, Napolitano MG, Moosburner M, Shrock E, Pruitt BW, Conway N, Goodman DB, Gardner CL, Tyree G,

Gonzales A, Wanner BL, Norville JE, Lajoie MJ, Church GM. Design, synthesis, and testing toward a 57-codon genome. Science. 2016; 353(6301):819–822. [PubMed: 27540174]

Pan M, Gao S, Zheng Y, Tan X, Lan H, Tan X, Sun D, Lu L, Wang T, Zheng Q, Huang Y, Wang J, Liu L. Quasi-racemic X-ray structures of K27-linked ubiquitin chains prepared by total chemical synthesis. Journal of the American Chemical Society. 2016; 138(23):7429–7435. [PubMed: 27268299]

Park SH, Wang VS, Radoicic J, De Angelis AA, Berkamp S, Opella SJ. Paramagnetic relaxation enhancement of membrane proteins by incorporation of the metal-chelating unnatural amino acid 2-amino-3-(8-hydroxyquinolin-3-yl)propanoic acid (HQA). Journal of Biomolecular NMR. 2015; 61(3–4):185–196. [PubMed: 25430059]

Pentelute BL, Kent SBH. Selective desulfurization of cysteine in the presence of Cys(Acm) in polypeptides obtained by native chemical ligation. Organic Letters. 2007; 9(4):687–690. [PubMed: 17286375]

Perdios L, Lowe AR, Saladino G, Bunney TD, Thiyagarajan N, Alexandrov Y, Dunsby C, French PM, Chin JW, Gervasio FL, Tate EW, Katan M. Conformational transition of FGFR kinase activation revealed by site-specific unnatural amino acid reporter and single molecule FRET. Science Report. 2017; 7:39841.

Perler FB. InBase: the intein database. Nucleic Acids Research. 2002; 30(1):383–384. [PubMed: 11752343]

Pervushin K, Riek R, Wider G, Wuthrich K. Attenuated T2 relaxation by mutual cancellation of dipole-dipole coupling and chemical shift anisotropy indicates an avenue to NMR structures of very large biological macromolecules in solution. Proceedings of the National Academy of Sciences United States of America. 1997; 94(23):12366–12371.

Petkova AT, Yau WM, Tycko R. Experimental constraints on quaternary structure in Alzheimer's beta-amyloid fibrils. Biochemistry. 2006; 45(2):498–512. [PubMed: 16401079]

Pietrokovski S. Intein spread and extinction in evolution. Trends in Genetics. 2001; 17(8):465–472. [PubMed: 11485819]

Pike AC, Garman EF, Krojer T, Von Delft F, Carpenter EP. An overview of heavy-atom derivatization of protein crystals. Acta Crystallographica D, Biological Crystallography. 2016; 72(Pt 3):303–318.

Plaks JG, Falatach R, Kastantin M, Berberich JA, Kaar JL. Multisite clickable modification of proteins using lipoic acid ligase. Bioconjugate Chemistry. 2015; 26(6):1104–1112. [PubMed: 25982177]

Plaschka C, Hantsche M, Dienemann C, Burzinski C, Plitzko J, Cramer P. Transcription initiation complex structures elucidate DNA opening. Nature. 2016; 533(7603):353–358. [PubMed: 27193681]

Plass T, Milles S, Koehler C, Schultz C, Lemke EA. Genetically encoded copper-free click chemistry. Angewandte Chemie (International edition in English). 2011; 50(17):3878–3881.

Policarpo RL, Kang H, Liao X, Rabideau AE, Simon MD, Pentelute BL. Flow-based enzymatic ligation by sortase A. Angewandte Chemie (International edition in English). 2014; 53(35):9203–9208.

Pollack SJ, Schultz PG. A semisynthetic catalytic antibody. Journal of the American Chemical Society. 1989; 111(5):1929–1931.

Popp MW, Ploegh HL. Making and breaking peptide bonds: protein engineering using sortase. Angewandte Chemie (International edition in English). 2011; 50(22):5024–5032.

Rabanal F, Degrado WF, Dutton PL. Use of 2,2′-dithiobis(5-nitropyridine) for the heterodimerization of cysteine containing peptides. Introduction of the 5-nitro-2-pyridinesulfenyl group. Tetrahedron Letters. 1996; 37(9):1347–1350.

Rashidian M, Dozier JK, Distefano MD. Enzymatic labeling of proteins: techniques and approaches. Bioconjugate Chemistry. 2013; 24(8):1277–1294. [PubMed: 23837885]

Ratzke C, Hellenkamp B, Hugel T. Four-colour FRET reveals directionality in the Hsp90 multicomponent machinery. Nature Communications. 2014; 5:4192.

Reddy PS, Dery S, Metanis N. Chemical synthesis of proteins with non-strategically placed cysteines using selenazolidine and selective deselenization. Angewandte Chemie (International edition in English). 2016; 55(3):992–995.

Ritzefeld M. Sortagging: a robust and efficient chemoenzymatic ligation strategy. Chemistry. 2014; 20(28):8516–8529. [PubMed: 24954404]

Rogerson DT, Sachdeva A, Wang K, Haq T, Kazlauskaite A, Hancock SM, Huguenin-Dezot N, Muqit MM, Fry AM, Bayliss R, Chin JW. Efficient genetic encoding of phosphoserine and its nonhydrolyzable analog. Nature Chemical Biology. 2015; 11(7):496–503. [PubMed: 26030730]

Rostovtsev VV, Green LG, Fokin VV, Sharpless KB. A stepwise huisgen cycloaddition process: copper(I)-catalyzed regio-selective 'ligation' of azides and terminal alkynes. Angewandte Chemie (International edition in English). 2002; 41(14):2596–2599.

Sachdeva A, Wang K, Elliott T, Chin JW. Concerted, rapid, quantitative, and site-specific dual labeling of proteins. Journal of the American Chemical Society. 2014; 136(22):7785–7788. [PubMed: 24857040]

Sakakibara D, Sasaki A, Ikeya T, Hamatsu J, Hanashima T, Mishima M, Yoshimasu M, Hayashi N, Mikawa T, Walchli M, Smith BO, Shirakawa M, Guntert P, Ito Y. Protein structure determination in living cells by in-cell NMR spectroscopy. Nature. 2009; 458(7234):102–105. [PubMed: 19262674]

Sakamoto K, Murayama K, Oki K, Iraha F, Kato-Murayama M, Takahashi M, Ohtake K, Kobayashi T, Kuramitsu S, Shirouzu M, Yokoyama S. Genetic encoding of 3-iodo-L-tyrosine in *Escherichia coli* for single-wavelength anomalous dispersion phasing in protein crystallography. Structure. 2009; 17(3):335–344. [PubMed: 19278648]

Saxon E, Armstrong JI, Bertozzi CR. A 'traceless' Staudinger ligation for the chemoselective synthesis of amide bonds. Organic Letters. 2000; 2(14):2141–2143. [PubMed: 10891251]

Saxon E, Bertozzi CR. Cell surface engineering by a modified Staudinger reaction. Science. 2000; 287(5460):2007–2010. [PubMed: 10720325]

Sayers J, Thompson RE, Perry KJ, Malins LR, Payne RJ. Thiazolidine-protected beta-thiol asparagine: applications in one-pot ligation-desulfurization chemistry. Organic Letters. 2015; 17(19):4902–4905. [PubMed: 26398220]

Scheres SH. A Bayesian view on cryo-EM structure determination. Journal of Molecular Biology. 2012; 415(2):406–418. [PubMed: 22100448]

Schlick TL, Ding Z, Kovacs EW, Francis MB. Dual-surface modification of the tobacco mosaic virus. Journal of the American Chemical Society. 2005; 127(11):3718–3723. [PubMed: 15771505]

Schmidt MJ, Fedoseev A, Bucker D, Borbas J, Peter C, Drescher M, Summerer D. EPR distance measurements in native proteins with genetically encoded spin labels. ACS Chemical Biology. 2015; 10(12):2764–2771. [PubMed: 26421438]

Schmidt MJ, Summerer D. Red-light-controlled protein-RNA crosslinking with a genetically encoded furan. Angewandte Chemie (International edition in English). 2013; 52(17):4690–4693.

Schmohl L, Schwarzer D. Sortase-mediated ligations for the site-specific modification of proteins. Current Opinion in Chemical Biology. 2014; 22:122–128. [PubMed: 25299574]

Schubeis T, Yuan P, Ahmed M, Nagaraj M, Van Rossum BJ, Ritter C. Untangling a repetitive amyloid sequence: correlating biofilm-derived and segmentally labeled curli fimbriae by solid-state NMR spectroscopy. Angewandte Chemie (International edition in English). 2015; 54(49):14669–14672.

Schultz KC, Supekova L, Ryu Y, Xie J, Perera R, Schultz PG. A genetically encoded infrared probe. Journal of the American Chemical Society. 2006; 128(43):13984–13985. [PubMed: 17061854]

Schwartz EC, Saez L, Young MW, Muir TW. Post-translational enzyme activation in an animal via optimized conditional protein splicing. Nature Chemical Biology. 2007; 3(1):50–54. [PubMed: 17128262]

Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ. Production of cyclic peptides and proteins *in vivo*. Proceedings of the National Academy of Sciences United States of America. 1999; 96(24):13638–13643.

Seim KL, Obermeyer AC, Francis MB. Oxidative modification of native protein residues using cerium(IV) ammonium nitrate. Journal of the American Chemical Society. 2011; 133(42):16970–16976. [PubMed: 21967510]

Seitchik JL, Peeler JC, Taylor MT, Blackman ML, Rhoads TW, Cooley RB, Refakis C, Fox JM, Mehl RA. Genetically encoded tetrazine amino acid directs rapid site-specific *in vivo* bioorthogonal

ligation with trans-cyclooctenes. Journal of the American Chemical Society. 2012; 134(6):2898–2901. [PubMed: 22283158]

Serwa R, Wilkening I, Del Signore G, Muhlberg M, Claussnitzer I, Weise C, Gerrits M, Hackenberger CP. Chemoselective Staudinger-phosphite reaction of azides for the phosphorylation of proteins. Angewandte Chemie (International edition in English). 2009; 48(44):8234–8239.

Shah L, Laughlin ST, Carrico IS. Light-activated staudinger-bertozzi ligation within living animals. Journal of the American Chemical Society. 2016; 138(16):5186–5189. [PubMed: 27010217]

Shah NH, Dann GP, Vila-Perello M, Liu Z, Muir TW. Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. Journal of the American Chemical Society. 2012; 134(28):11338–11341. [PubMed: 22734434]

Shah NH, Eryilmaz E, Cowburn D, Muir TW. Naturally split inteins assemble through a 'capture and collapse' mechanism. Journal of the American Chemical Society. 2013; 135(49):18673–18681. [PubMed: 24236406]

Shah NH, Muir TW. Inteins: nature's gift to protein chemists. Chemical Science. 2014; 5(1):446–461. [PubMed: 24634716]

Shah NH, Vila-Perello M, Muir TW. Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein. Angewandte Chemie (International edition in English). 2011; 50(29):6511–6515.

Shang SY, Tan ZP, Dong SW, Danishefsky SJ. An advance in proline ligation. Journal of the American Chemical Society. 2011; 133(28):10784–10786. [PubMed: 21696220]

Sharaf NG, Gronenborn AM. (19)F-modified proteins and (19)F-containing ligands as tools in solution NMR studies of protein interactions. Methods in Enzymology. 2015; 565:67–95. [PubMed: 26577728]

Shi J, Muir TW. Development of a tandem protein trans-splicing system based on native and engineered split inteins. Journal of the American Chemical Society. 2005; 127(17):6198–6206. [PubMed: 15853324]

Shieh P, Bertozzi CR. Design strategies for bioorthogonal smart probes. Organic & Biomolecular Chemistry. 2014; 12(46):9307–9320. [PubMed: 25315039]

Silvaggi NR, Martin LJ, Schwalbe H, Imperiali B, Allen KN. Double-lanthanide-binding tags for macromolecular crystallographic structure determination. Journal of the American Chemical Society. 2007; 129(22):7114–7120. [PubMed: 17497863]

Siman P, Karthikeyan SV, Brik A. Native chemical ligation at glutamine. Organic Letters. 2012; 14(6): 1520–1523. [PubMed: 22360701]

Simon MD, Chu F, Racki LR, De La Cruz CC, Burlingame AL, Panning B, Narlikar GJ, Shokat KM. The site-specific installation of methyl-lysine analogs into recombinant histones. Cell. 2007; 128(5):1003–1012. [PubMed: 17350582]

Skrisovska L, Schubert M, Allain FH. Recent advances in segmental isotope labeling of proteins: NMR applications to large proteins and glycoproteins. Journal of Biomolecular NMR. 2010; 46(1):51–65. [PubMed: 19690964]

Slavoff SA, Liu DS, Cohen JD, Ting AY. Imaging protein–protein interactions inside living cells via interaction-dependent fluorophore ligation. Journal of the American Chemical Society. 2011; 133(49):19769–19776. [PubMed: 22098454]

Smits AH, Borrmann A, Roosjen M, Van Hest JC, Vermeulen M. Click-MS: tagless protein enrichment using bioorthogonal chemistry for quantitative proteomics. ACS Chemical Biology. 2016; 11(12):3245–3250. [PubMed: 27643597]

Song F, Chen P, Sun D, Wang M, Dong L, Liang D, Xu RM, Zhu P, Li G. Cryo-EM study of the chromatin fiber reveals a double helix twisted by tetranucleosomal units. Science. 2014; 344(6182):376–380. [PubMed: 24763583]

Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB. Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. Biotechniques. 1999; 27(1):110–114. 116, 118–120. [PubMed: 10407673]

Spicer CD, Davis BG. Selective chemical protein modification. Nature Communications. 2014; 5:4740.

Srinivasan G, James CM, Krzycki JA. Pyrrolysine encoded by UAG in Archaea: charging of a UAG-decoding specialized tRNA. Science. 2002; 296(5572):1459–1462. [PubMed: 12029131]

Stephanopoulos N, Francis MB. Choosing an effective protein bioconjugation strategy. Nature Chemical Biology. 2011; 7(12):876–884. [PubMed: 22086289]

Stevens AJ, Brown ZZ, Shah NH, Sekar G, Cowburn D, Muir TW. Design of a split intein with exceptional protein splicing activity. Journal of the American Chemical Society. 2016; 138(7): 2162–2165. [PubMed: 26854538]

Summerer D, Chen S, Wu N, Deiters A, Chin JW, Schultz PG. A genetically encoded fluorescent amino acid. Proceedings of the National Academy of Sciences United States of America. 2006; 103(26):9785–9789.

Szymanski W, Wu B, Poloni C, Janssen DB, Feringa BL. Azobenzene photoswitches for Staudinger-Bertozzi ligation. Angewandte Chemie (International edition in English). 2013; 52(7):2068–2072.

Tan ZP, Shang SY, Danishefsky SJ. Insights into the finer issues of native chemical ligation: an approach to cascade ligations. Angewandte Chemie (International edition in English). 2010; 49(49):9500–9503.

Tanaka K, Kitadani M, Fukase K. Target-selective fluorescent 'switch-on' protein labeling by 6pi-azaelectrocyclization. Organic & Biomolecular. 2011; 9(15):5346–5349.

Tenboer J, Basu S, Zatsepin N, Pande K, Milathianaki D, Frank M, Hunter M, Boutet S, Williams GJ, Koglin JE, Oberthuer D, Heymann M, Kupitz C, Conrad C, Coe J, Roy-Chowdhury S, Weierstall U, James D, Wang D, Grant T, Barty A, Yefanov O, Scales J, Gati C, Seuring C, Srajer V, Henning R, Schwander P, Fromme R, Ourmazd A, Moffat K, Van Thor JJ, Spence JC, Fromme P, Chapman HN, Schmidt M. Time-resolved serial crystallography captures high-resolution intermediates of photo-active yellow protein. Science. 2014; 346(6214):1242–1246. [PubMed: 25477465]

Tey LH, Loveridge EJ, Swanwick RS, Flitsch SL, Allemann RK. Highly site-selective stability increases by glycosylation of dihydrofolate reductase. FEBS Journal. 2010; 277(9):2171–2179. [PubMed: 20412060]

Theile CS, Witte MD, Blom AE, Kundrat L, Ploegh HL, Guimaraes CP. Site-specific N-terminal labeling of proteins using sortase-mediated reactions. Nature Protocols. 2013; 8(9):1800–1807. [PubMed: 23989674]

Thiel IV, Volkmann G, Pietrokovski S, Mootz HD. An atypical naturally split intein engineered for highly efficient protein labeling. Angewandte Chemie (International edition in English). 2014; 53(5):1306–1310.

Thompson RE, Chan B, Radom L, Jolliffe KA, Payne RJ. Chemoselective peptide ligation-desulfurization at aspartate. Angewandte Chemie (International edition in English). 2013; 52(37): 9723–9727.

Thompson RE, Liu X, Alonso-Garcia N, Pereira PJ, Jolliffe KA, Payne RJ. Trifluoroethanethiol: an additive for efficient one-pot peptide ligation-desulfurization chemistry. Journal of the American Chemical Society. 2014; 136(23):8161–8164. [PubMed: 24873761]

Tian F, Lu Y, Manibusan A, Sellers A, Tran H, Sun Y, Phuong T, Barnett R, Hehli B, Song F, Deguzman MJ, Ensari S, Pinkstaff JK, Sullivan LM, Biroc SL, Cho H, Schultz PG, Dijoseph J, Dougher M, Ma D, Dushin R, Leal M, Tchistiakova L, Feyfant E, Gerber HP, Sapra P. A general approach to site-specific antibody drug conjugates. Proceedings of the National Academy of Sciences United States of America. 2014; 111(5):1766–1771.

Tilley SD, Francis MB. Tyrosine-selective protein alkylation using pi-allylpalladium complexes. Journal of the American Chemical Society. 2006; 128(4):1080–1081. [PubMed: 16433516]

Tompa P. Intrinsically disordered proteins: a 10-year recap. Trends in Biochemical Sciences. 2012; 37(12):509–516. [PubMed: 22989858]

Ton-That H, Mazmanian SK, Faull KF, Schneewind O. Anchoring of surface proteins to the cell wall of *Staphylococcus aureus*. Sortase catalyzed *in vitro* transpeptidation reaction using LPXTG peptide and NH(2)-Gly(3) substrates. Journal of Biological Chemistry. 2000; 275(13):9876–9881. [PubMed: 10734144]

Topilina NI, Mills KV. Recent advances in *in vivo* applications of intein-mediated protein splicing. Mobile DNA. 2014; 5(1):5. [PubMed: 24490831]

Torbeev VY, Hilvert D. Both the cis-trans equilibrium and isomerization dynamics of a single proline amide modulate beta2-microglobulin amyloid assembly. Proceedings of the National Academy of Sciences United States of America. 2013; 110(50):20051–20056.

Torbeev VY, Raghuraman H, Hamelberg D, Tonelli M, Westler WM, Perozo E, Kent SB. Protein conformational dynamics in the mechanism of HIV-1 protease catalysis. Proceedings of the National Academy of Sciences United States of America. 2011; 108(52):20982–20987.

Tornoe CW, Christensen C, Meldal M. Peptidotriazoles on solid phase: [1,2,3]-triazoles by regiospecific copper(i)-catalyzed 1,3-dipolar cycloadditions of terminal alkynes to azides. Journal of Organic Chemistry. 2002; 67(9):3057–3064. [PubMed: 11975567]

Tremblay ML, Xu L, Lefevre T, Sarker M, Orrell KE, Leclerc J, Meng Q, Pezolet M, Auger M, Liu XQ, Rainey JK. Spider wrapping silk fibre architecture arising from its modular soluble protein precursor. Science Report. 2015; 5:11502.

Tugarinov V, Hwang PM, Ollerenshaw JE, Kay LE. Cross-correlated relaxation enhanced 1H[bond]13C NMR spectroscopy of methyl groups in very high molecular weight proteins and protein complexes. Journal of the American Chemical Society. 2003; 125(34):10420–10428. [PubMed: 12926967]

Tynan CJ, Lo Schiavo V, Zanetti-Domingues L, Needham SR, Roberts SK, Hirsch M, Rolfe DJ, Korovesis D, Clarke DT, Martin-Fernandez ML. A tale of the epidermal growth factor receptor: the quest for structural resolution on cells. Methods. 2016; 95:86–93. [PubMed: 26484734]

Uttamapinant C, Tangpeerachaikul A, Grecian S, Clarke S, Singh U, Slade P, Gee KR, Ting AY. Fast, cell-compatible click chemistry with copper-chelating azides for biomolecular labeling. Angewandte Chemie (International edition in English). 2012; 51(24):5852–5856.

Uttamapinant C, White KA, Baruah H, Thompson S, Fernandez-Suarez M, Puthenveetil S, Ting AY. A fluorophore ligase for site-specific protein labeling inside living cells. Proceedings of the National Academy of Sciences United States of America. 2010; 107(24):10914–10919.

Uversky VN. Biophysical methods to investigate intrinsically disordered proteins: avoiding an 'elephant and blind men' situation. Advances in Experimental Medicine and Biology. 2015; 870:215–260. [PubMed: 26387104]

Valiyaveetil FI, Leonetti M, Muir TW, Mackinnon R. Ion selectivity in a semisynthetic K+ channel locked in the conductive conformation. Science. 2006; 314(5801):1004–1007. [PubMed: 17095703]

Van Kasteren SI, Kramer HB, Jensen HH, Campbell SJ, Kirkpatrick J, Oldham NJ, Anthony DC, Davis BG. Expanding the diversity of chemical protein modification allows post-translational mimicry. Nature. 2007; 446(7139):1105–1109. [PubMed: 17460675]

Van 'T Hof W, Hansenova Manaskova S, Veerman EC, Bolscher JG. Sortase-mediated backbone cyclization of proteins and peptides. Biological Chemistry. 2015; 396(4):283–293. [PubMed: 25581753]

Van Wilderen LJ, Kern-Michler D, Muller-Werkmeister HM, Bredenbeck J. Vibrational dynamics and solvatochromism of the label SCN in various solvents and hemoglobin by time dependent IR and 2D-IR spectroscopy. PhysChemChemPhys. 2014; 16(36):19643–19653.

Venditti V, Fawzi NL, Clore GM. An efficient protocol for incorporation of an unnatural amino acid in perdeuterated recombinant proteins using glucose-based media. Journal of Biomolecular NMR. 2012; 52(3):191–195. [PubMed: 22350951]

Vila-Perello M, Liu Z, Shah NH, Willis JA, Idoyaga J, Muir TW. Streamlined expressed protein ligation using split inteins. Journal of the American Chemical Society. 2013; 135(1):286–292. [PubMed: 23265282]

Vila-Perello M, Muir TW. Biological applications of protein splicing. Cell. 2010; 143(2):191–200. [PubMed: 20946979]

Vinogradova EV, Zhang C, Spokoyny AM, Pentelute BL, Buchwald SL. Organometallic palladium reagents for cysteine bioconjugation. Nature. 2015; 526(7575):687–691. [PubMed: 26511579]

Virdee S, Kapadnis PB, Elliott T, Lang K, Madrzak J, Nguyen DP, Riechmann L, Chin JW. Traceless and site-specific ubiquitination of recombinant proteins. Journal of the American Chemical Society. 2011; 133(28):10708–10711. [PubMed: 21710965]

Virdee S, Ye Y, Nguyen DP, Komander D, Chin JW. Engineered diubiquitin synthesis reveals Lys29-isopeptide specificity of an OTU deubiquitinase. Nature Chemical Biology. 2010; 6(10):750–757. [PubMed: 20802491]

Volkmann G, Mootz HD. Recent progress in intein research: from mechanism to directed evolution and applications. Cellular and Molecular Life Science. 2013; 70(7):1185–1206.

Von Der Ecken J, Muller M, Lehman W, Manstein DJ, Penczek PA, Raunser S. Structure of the F-actin-tropomyosin complex. Nature. 2015; 519(7541):114–117. [PubMed: 25470062]

Wakamori M, Fujii Y, Suka N, Shirouzu M, Sakamoto K, Umehara T, Yokoyama S. Intra- and inter-nucleosomal interactions of the histone H4 tail revealed with a human nucleosome core particle with genetically-incorporated H4 tetra-acetylation. Science Report. 2015; 5:17204.

Walper SA, Turner KB, Medintz IL. Enzymatic bioconjugation of nanoparticles: developing specificity and control. Current Opinion Biotechnology. 2015; 34:232–241.

Wan Q, Danishefsky SJ. Free-radical-based, specific desulfurization of cysteine: a powerful advance in the synthesis of polypeptides and glycopolypeptides. Angewandte Chemie (International edition in English). 2007; 46(48):9248–9252.

Wan W, Huang Y, Wang Z, Russell WK, Pai PJ, Russell DH, Liu WR. A facile system for genetic incorporation of two different noncanonical amino acids into one protein in *Escherichia coli*. Angewandte Chemie (International edition in English). 2010; 49(18):3211–3214.

Wang K, Fredens J, Brunner SF, Kim SH, Chia T, Chin JW. Defining synonymous codon compression schemes by genome recoding. Nature. 2016; 539(7627):59–64. [PubMed: 27776354]

Wang K, Sachdeva A, Cox DJ, Wilf NM, Lang K, Wallace S, Mehl RA, Chin JW. Optimized orthogonal translation of unnatural amino acids enables spontaneous protein double-labelling and FRET. Nature Chemistry. 2014a; 6(5):393–403.

Wang L, Brock A, Herberich B, Schultz PG. Expanding the genetic code of *Escherichia coli*. Science. 2001; 292(5516):498–500. [PubMed: 11313494]

Wang S, Munro RA, Shi L, Kawamura I, Okitsu T, Wada A, Kim SY, Jung KH, Brown LS, Ladizhansky V. Solid-state NMR spectroscopy structure determination of a lipid-embedded heptahelical membrane protein. Nature Methods. 2013; 10(10):1007–1012. [PubMed: 24013819]

Wang Y, Kavran JM, Chen Z, Karukurichi KR, Leahy DJ, Cole PA. Regulation of S-adenosylhomocysteine hydrolase lysine acetylation. Journal of Biological Chemistry. 2014b; 289(45):31361–31372. [PubMed: 25248746]

Wang ZA, Zeng Y, Kurra Y, Wang X, Tharp JM, Vatansever EC, Hsu WW, Dai S, Fang X, Liu WR. A genetically encoded allysine for the synthesis of proteins with site-specific lysine dimethylation. Angewandte Chemie (International edition in English). 2017; 56(1):212–216.

Warden-Rothman R, Caturegli I, Popik V, Tsourkas A. Sortase-tag expressed protein ligation: combining protein purification and site-specific bioconjugation into a single step. Analytical Chemistry. 2013; 85(22):11090–11097. [PubMed: 24111659]

Wasmer C, Lange A, Van Melckebeke H, Siemer AB, Riek R, Meier BH. Amyloid fibrils of the HET-s(218–289) prion form beta solenoid with a triangular hydrophobic core. Science. 2008; 319(5869):1523–1526. [PubMed: 18339938]

Williams FP, Milbradt AG, Embrey KJ, Bobby R. Segmental isotope labelling of an individual bromodomain of a tandem domain BRD4 using sortase A. PLoS ONE. 2016; 11(4):e0154607. [PubMed: 27128490]

Williamson DJ, Webb ME, Turnbull WB. Depsipeptide substrates for sortase-mediated N-terminal protein ligation. Nature Protocols. 2014; 9(2):253–262. [PubMed: 24407354]

Williamson MP, Havel TF, Wuthrich K. Solution conformation of proteinase inhibitor IIA from bull seminal plasma by 1H nuclear magnetic resonance and distance geometry. Journal of Molecular Biology. 1985; 182(2):295–315. [PubMed: 3839023]

Winkelman JT, Vvedenskaya IO, Zhang Y, Zhang Y, Bird JG, Taylor DM, Gourse RL, Ebright RH, Nickels BE. Multiplexed protein-DNA cross-linking: scrunching in transcription start site selection. Science. 2016; 351(6277):1090–1093. [PubMed: 26941320]

Wissner RF, Batjargal S, Fadzen CM, Petersson EJ. Labeling proteins with fluorophore/thioamide Forster resonant energy transfer pairs by combining unnatural amino acid mutagenesis and native chemical ligation. Journal of the American Chemical Society. 2013; 135(17):6529–6540. [PubMed: 23594264]

Witte MD, Cragnolini JJ, Dougan SK, Yoder NC, Popp MW, Ploegh HL. Preparation of unnatural N-to-N and C-to-C protein fusions. Proceedings of the National Academy of Sciences United States of America. 2012; 109(30):11993–11998.

Wood DW, Camarero JA. Intein applications: from protein purification and labeling to metabolic control methods. Journal Biological Chemistry. 2014; 289(21):14512–14519.

Wright TH, Bower BJ, Chalker JM, Bernardes GJ, Wiewiora R, Ng WL, Raj R, Faulkner S, Vallee MR, Phanumartwiwath A, Coleman OD, Thezenas ML, Khan M, Galan SR, Lercher L, Schombs MW, Gerstberger S, Palm-Espling ME, Baldwin AJ, Kessler BM, Claridge TD, Mohammed S, Davis BG. Posttranslational mutagenesis: a chemical strategy for exploring protein side-chain diversity. Science. 2016; 354(6312)

Wu H, Hu Z, Liu XQ. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. *PCC6803*. Proceedings of the National Academy of Sciences United States of America. 1998; 95(16):9226–9231.

Wu P, Shui W, Carlson BL, Hu N, Rabuka D, Lee J, Bertozzi CR. Site-specific chemical modification of recombinant proteins produced in mammalian cells by using the genetically encoded aldehyde tag. Proceedings of the National Academy of Sciences United States of America. 2009; 106(9): 3000–3005.

Wuethrich I, Peeters JG, Blom AE, Theile CS, Li Z, Spooner E, Ploegh HL, Guimaraes CP. Site-specific chemoenzymatic labeling of aerolysin enables the identification of new aerolysin receptors. PLoS ONE. 2014; 9(10):e109883. [PubMed: 25275512]

Wukovitz SW, Yeates TO. Why protein crystals favour some space-groups over others. Nature Structural Biology. 1995; 2(12):1062–1067. [PubMed: 8846217]

Xiao H, Chatterjee A, Choi SH, Bajjuri KM, Sinha SC, Schultz PG. Genetic incorporation of multiple unnatural amino acids into proteins in mammalian cells. Angewandte Chemie (International edition in English). 2013; 52(52):14080–14083.

Xie J, Wang L, Wu N, Brock A, Spraggon G, Schultz PG. The site-specific incorporation of p-iodo-L-phenylalanine into proteins for structure determination. Nature Biotechnology. 2004; 22(10): 1297–1301.

Xie R, Dong L, Huang R, Hong S, Lei R, Chen X. Targeted imaging and proteomic analysis of tumor-associated glycans living animals. Angewandte Chemie (International edition in English). 2014; 53(51):14082–14086.

Xu R, Ayers B, Cowburn D, Muir TW. Chemical ligation of folded recombinant proteins: segmental isotopic labeling domains for NMR studies. Proceedings of the National Academy of Sciences United States of America. 1999; 96(2):388–393.

Yamamura Y, Hirakawa H, Yamaguchi S, Nagamune T. Enhancement of sortase A-mediated protein ligation by inducing a beta-hairpin structure around the ligation site. Chemical Communications (Cambridge). 2011; 47(16):4742–4744.

Yamazaki T, Otomo T, Oda N, Kyogoku Y, Uegaki K, Ito N, Ishino Y, Nakamura H. Segmental isotope labeling for protein NMR using peptide splicing. Journal of the American Chemical Society. 1998; 120(22):5591–5592.

Yan LZ, Dawson PE. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. Journal of the American Chemical Society. 2001; 123(4): 526–533. [PubMed: 11456564]

Yang A, Ha S, Ahn J, Kim R, Kim S, Lee Y, Kim J, Soll D, Lee HY, Park HS. A chemical biology route to site-specifi authentic protein modifications. Science. 2016; 354(6312):623–626. [PubMed: 27708052]

Yang F, Yu X, Liu C, Qu CX, Gong Z, Liu HD, Li FH, Wang HM, He DF, Yi F, Song C, Tian CL, Xiao KH, Wang JY, Sun JP. Phospho-selective mechanisms of arrestin conformations and functions revealed by unnatural amino acid incorporation and (19)F-NMR. Nature Communications. 2015; 6:8202.

Yang R, Wong YH, Nguyen GK, Tam JP, Lescar J, Wu B. Engineering a catalytically efficient recombinant protein ligase. Journal of the American Chemical Society. 2017; doi: 10.1021/jacs. 1026b12637

Yang RL, Pasunooti KK, Li FP, Liu XW, Liu CF. Dual native chemical ligation at lysine. Journal of the American Chemical Society. 2009; 131(38):13592. [PubMed: 19728708]

Yang W, Hendrickson WA, Crouch RJ, Satow Y. Structure of ribonuclease H phased at 2 A resolution by MAD analysis of the selenomethionyl protein. Science. 1990; 249(4975):1398–1405. [PubMed: 2169648]

Ye S, Zaitseva E, Caltabiano G, Schertler GF, Sakmar TP, Deupi X, Vogel R. Tracking G-protein-coupled receptor activation using genetically encoded infrared probes. Nature. 2010; 464(7293): 1386–1389. [PubMed: 20383122]

Yeates TO, Kent SB. Racemic protein crystallography. Annual Review of Biophysics. 2012; 41:41–61.

Yeung H, Squire CJ, Yosaatmadja Y, Panjikar S, Lopez G, Molina A, Baker EN, Harris PW, Brimble MA. Radiation damage and racemic protein crystallography reveal the unique structure of the GASA/Snakin Protein Superfamily. Angewandte Chemie (International edition in English). 2016; 55(28):7930–7933.

Yoshizawa S, Bock A. The many levels of control on bacterial selenoprotein synthesis. Biochimica et Biophysica Acta. 2009; 1790(11):1404–1414. [PubMed: 19328835]

Young TS, Ahmad I, Yin JA, Schultz PG. An enhanced system for unnatural amino acid mutagenesis in *E. coli*. Journal of Molecular Biology. 2010; 395(2):361–374. [PubMed: 19852970]

Yu JX, Hallac RR, Chiguru S, Mason RP. New frontiers and developing applications in 19F NMR. Progress in Nuclear Magnetic Resonance Spectroscopy. 2013; 70:25–49. [PubMed: 23540575]

Zettler J, Schutz V, Mootz HD. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Letters. 2009; 583(5):909–914. [PubMed: 19302791]

Zhang C, Welborn M, Zhu T, Yang NJ, Santos MS, Van Voorhis T, Pentelute BL. Pi-Clamp-mediated cysteine conjugation. Nature Chemistry. 2016a; 8(2):120–128.

Zhang M, Lin S, Song X, Liu J, Fu Y, Ge X, Fu X, Chang Z, Chen PR. A genetically incorporated crosslinker reveals chap-erone cooperation in acid resistance. Nature Chemical Biology. 2011; 7(10):671–677. [PubMed: 21892184]

Zhang TQO, Grechko M, Moran SD, Zanni MT. Isotope-Labeled Amyloids via Synthesis, Expression, and Chemical Ligation for Use in FTIR, 2D IR, and NMR Studies. Protein Amyloid Aggregation: Methods and Protocols. 2016b; 1345:21–41.

Zheng Y, Lewis TL Jr, Igo P, Polleux F, Chatterjee A. Virus-enabled optimization and delivery of the genetic machinery for efficient unnatural amino acid mutagenesis in mammalian cells and tissues. ACS Synthetic Biology. 2017; 6(1):13–18. [PubMed: 27482719]

Zuger S, Iwai H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nature Biotechnology. 2005; 23(6):736–740.

Zuo C, Tang S, Zheng JS. Chemical synthesis and biophysical applications of membrane proteins. Journal of Peptide Science. 2015; 21(7):540–549. [PubMed: 25449562]
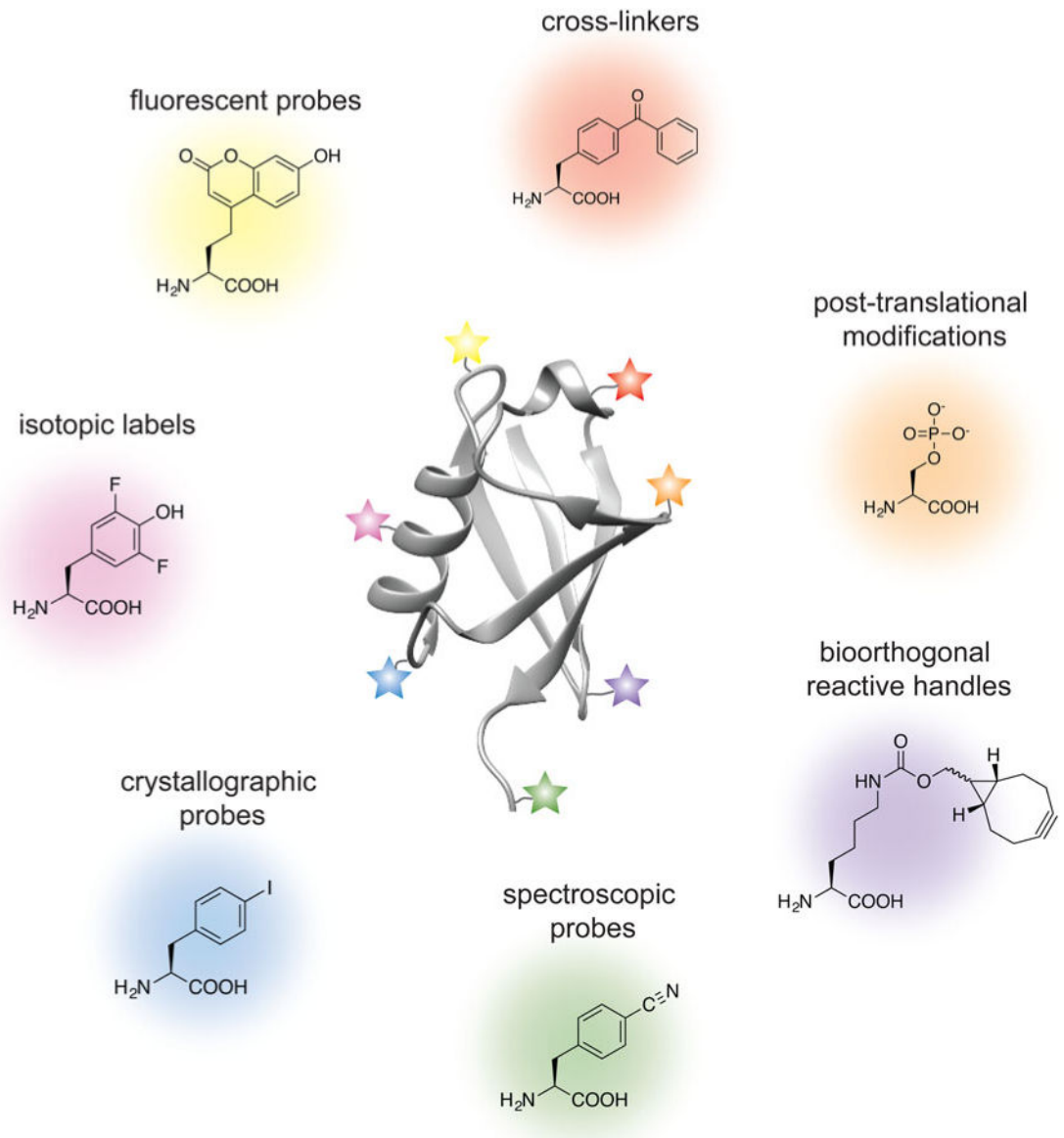
Fig. 1.
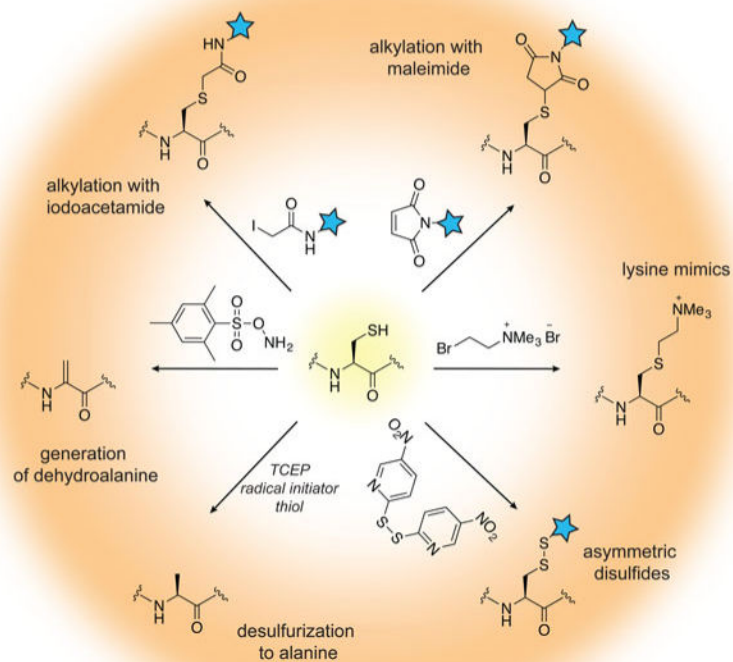Molecular engineering toolbox for the structural biologist.
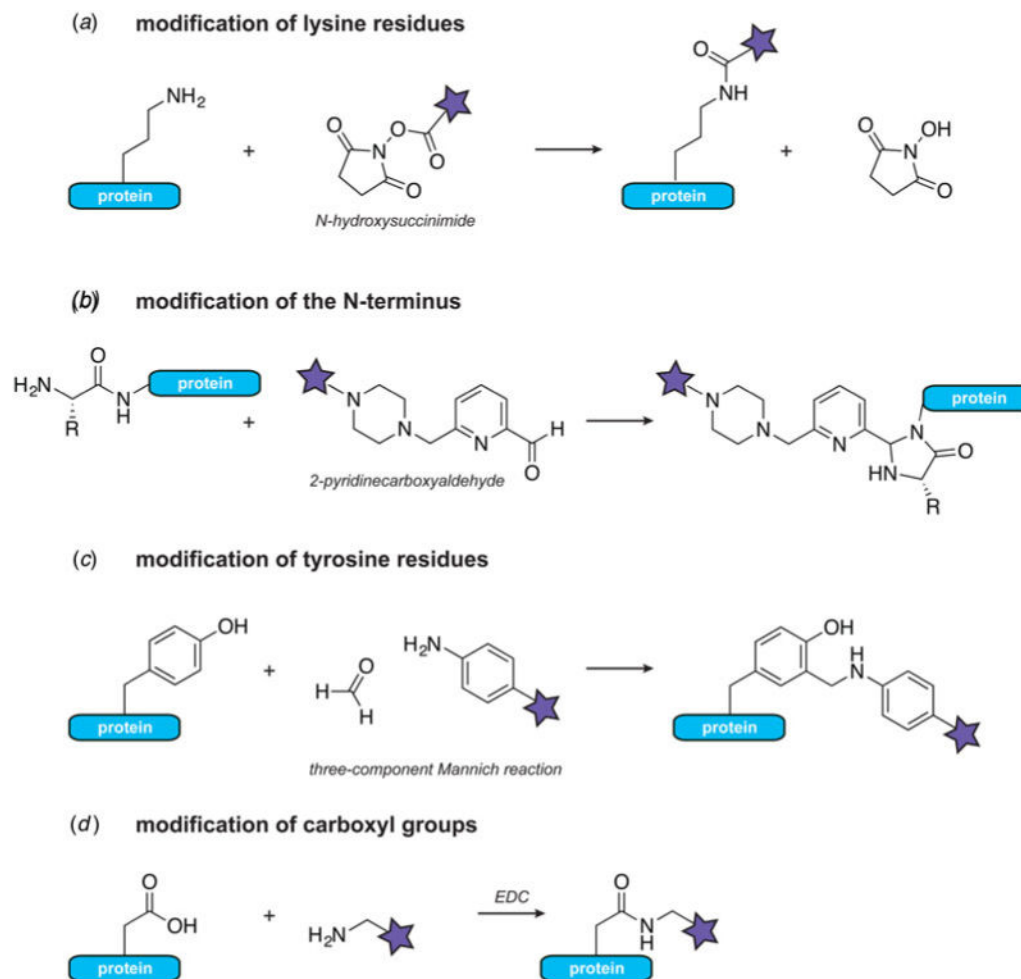
Fig. 2.
Chemical modification of cysteine residues.

Fig. 3.
Chemical modification of natural amino acids. (*a*) Modification of lysine *ε*-amines with
activated esters such as N-hydroxysuccinimide. (*b*) Modification of terminal *α*-amines with
2-pyridinecarboxaldehydes. (*c*) Three-component Mannich reaction for tyrosine
modification at the *ortho*-position. (d) Coupling of carboxyls and amines with carbodiimides
such as EDC.
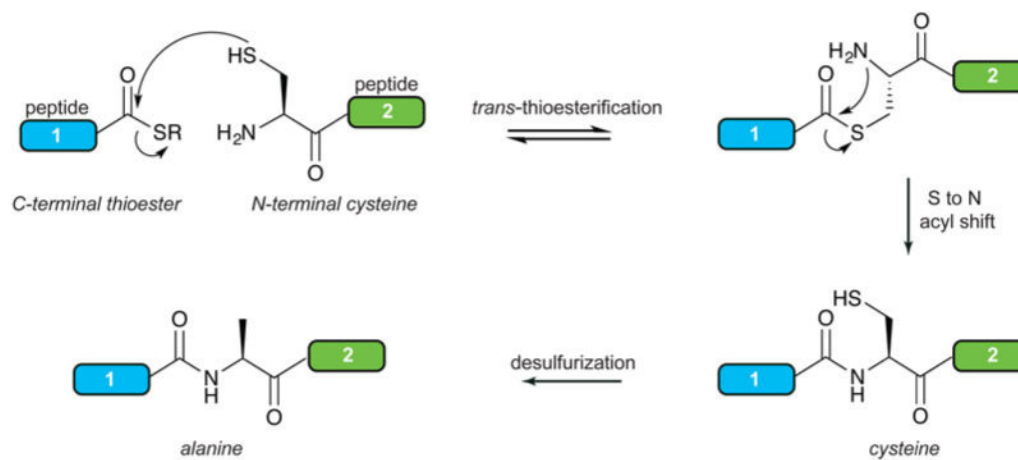
Fig. 4.

Native chemical ligation at cysteine followed by desulfurization to alanine for the construction of larger polypeptide chains without any 'scars'.
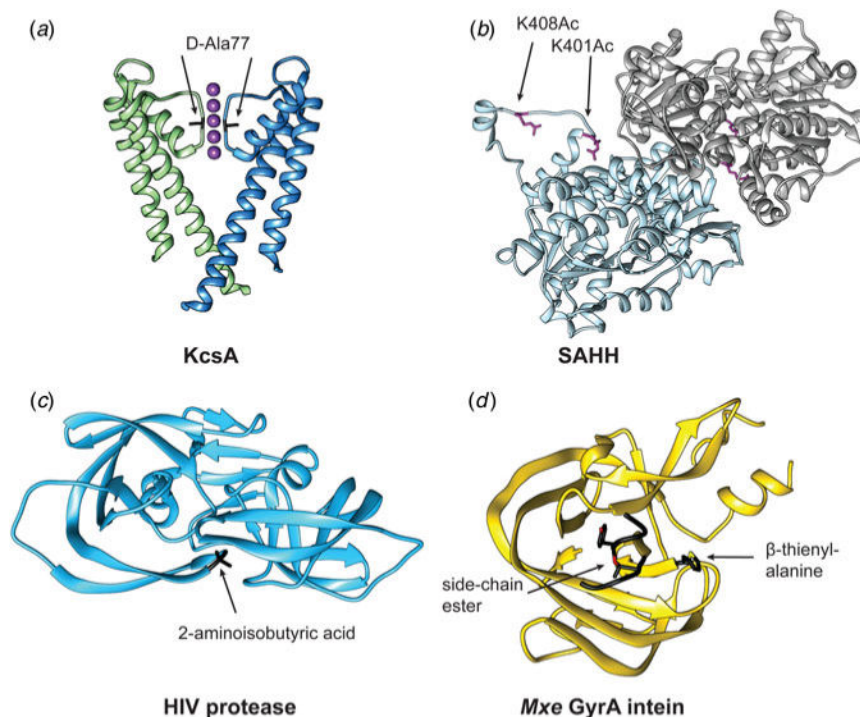
Fig. 5.
Examples of constructs prepared by NCL and EPL for X-ray crystallography studies. (*a*) D-alanine was introduced at position 77 in the sequence of the potassium channel KcsA to elucidate its ion selectivity mechanism (Valiyaveetil *et al.* 2006) (PDB ID: 2IH3). (*b*) Acetylated lysine (Ac) was incorporated at postions 401 and 408 in *S*-Adenosylhomocysteine hydrolase (SAHH) to evaluate the structural basis of enzyme inhibition (Wang *et al.* 2014b) (PDB ID: 4PFJ). (*c*) Chemical synthesis of HIV protease afforded the site-specific incorporation of unnatural amino acids such as 2-aminoisobutyric acid to modulate conformational dynamics and catalysis (Torbeev *et al.* 2011) (PDB ID: 3IAW). (*d*) Semi-synthesis of *Mxe* GyrA and the installation of β-thienyl-alanine instead of the native histidine at position 187 provided a route to trap the branched intermediate of the intein (Liu *et al.* 2014b) (PDB ID: 4OZ6).
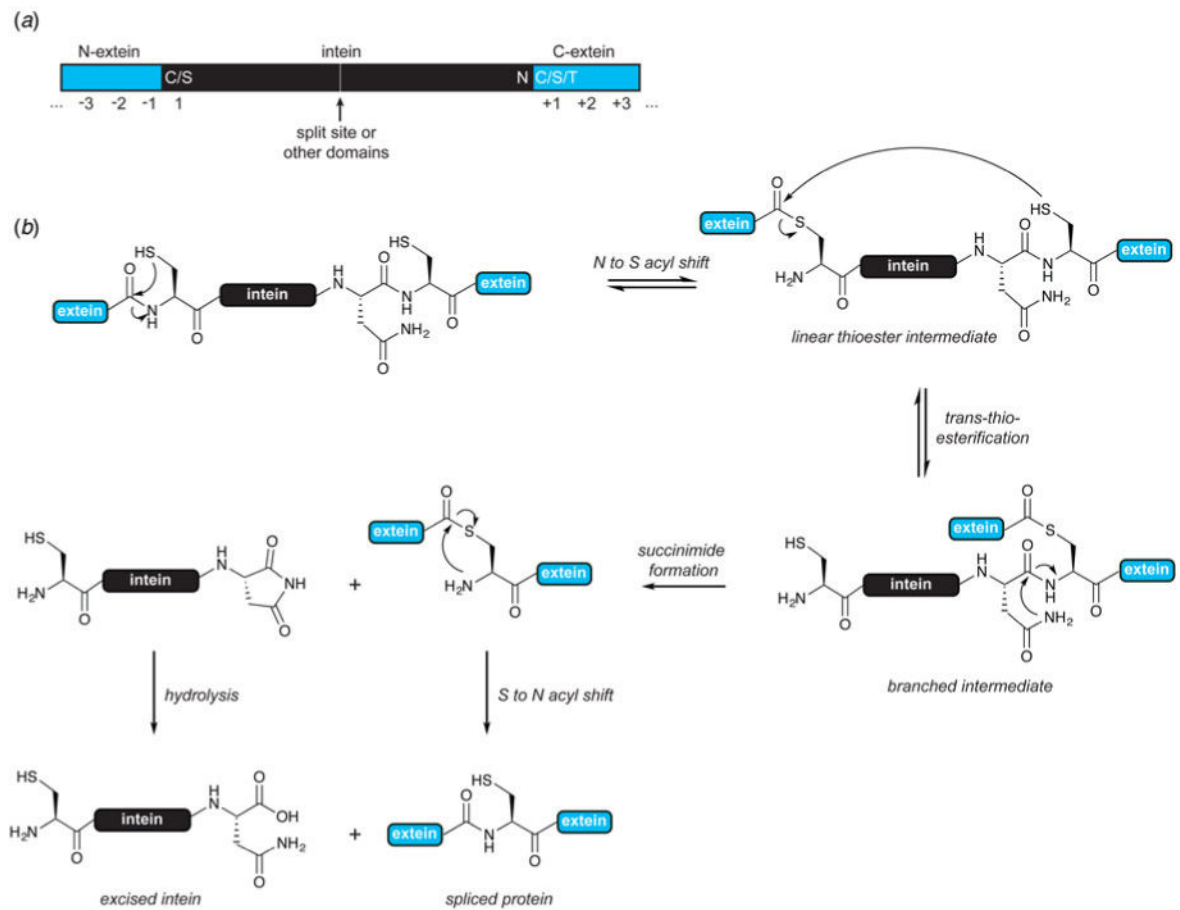
Fig. 6.

Intein structure and mechanism. (*a*) Intein/extein residues important for splicing. (*b*) Protein splicing mechanism of contiguous inteins. In some cases, the hydroxyl groups of Ser/Thr act as nucleophiles in the first two steps.
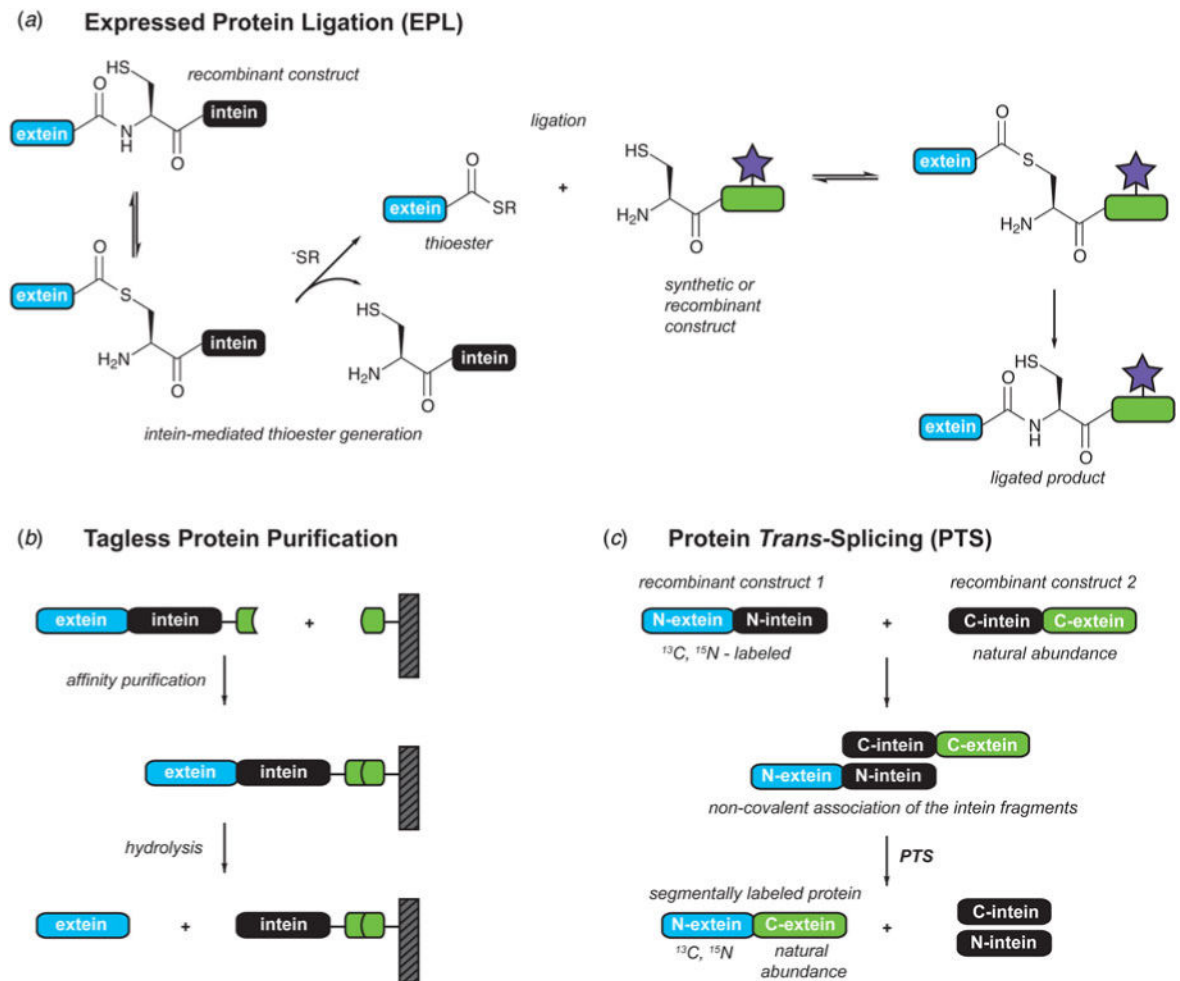
Fig. 7.
Protein engineering with inteins. (*a*) Expressed protein ligation. (*b*) Tagless protein purification. (*c*) Protein *trans*-splicing and recombinant production of segmentally isotopically labeled proteins.
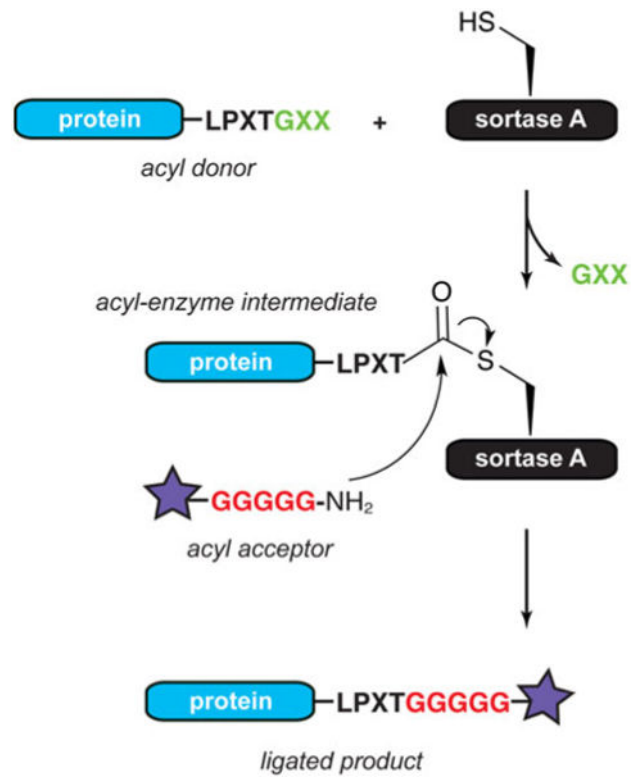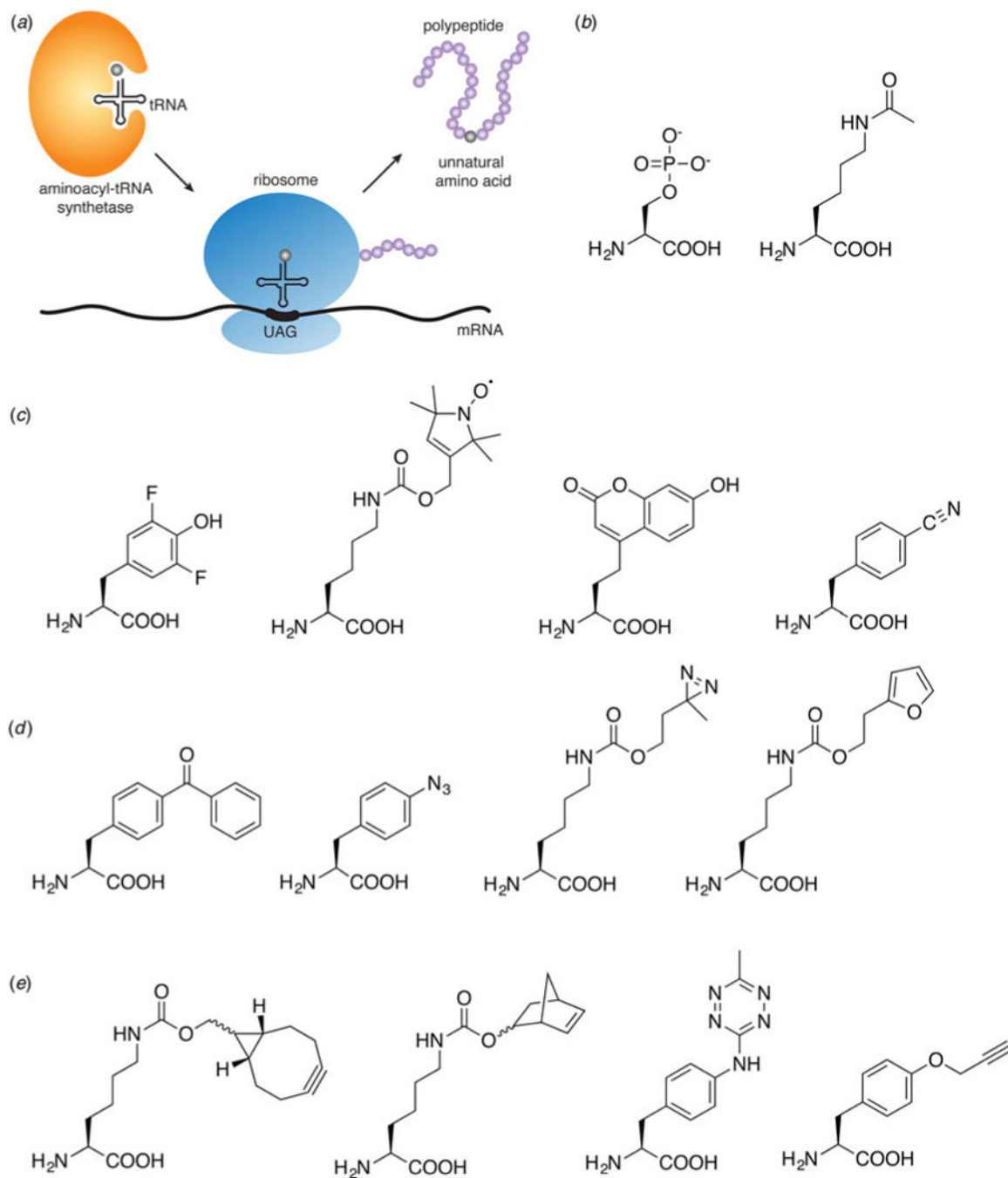
Fig. 8.
C-terminal protein labeling with sortase. The acyl donor requires the LPXTG recognition motif, while the acyl acceptor often contains a pentaglycine sequence.

Fig. 9.

Unnatural amino acid (UAA) incorporation by amber suppression. (*a*) An orthogonal aminoacyl tRNA synthetase charges a matching tRNA with the UAA of interest. The ribosome incorporates the UAA into a growing polypeptide chain by decoding the amber stop codon (UAG) on the messenger RNA. The UAA toolbox includes UAAs that represent (*b*) protein post-translational modifications, (*c*) spectroscopic probes, (*d*) cross-linkers, (*e*) bio-orthogonal reactive handles.
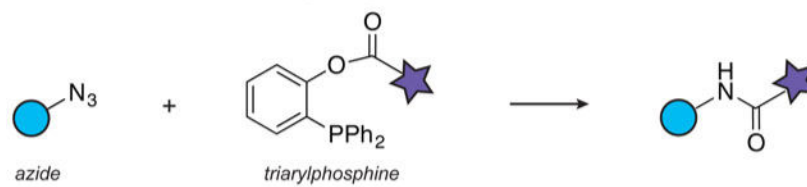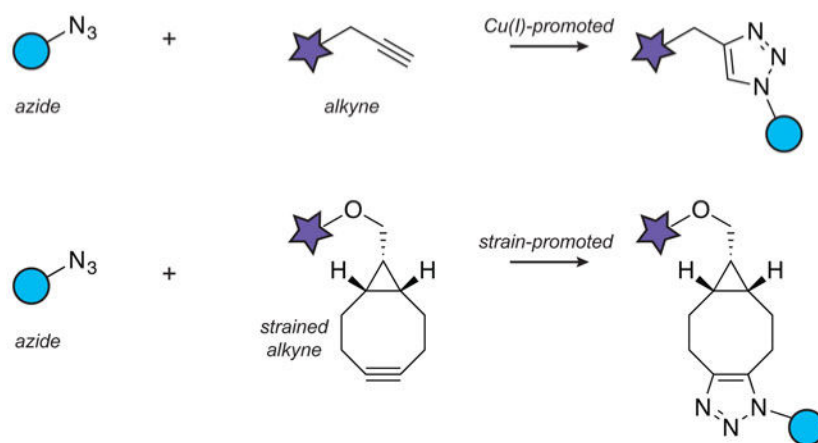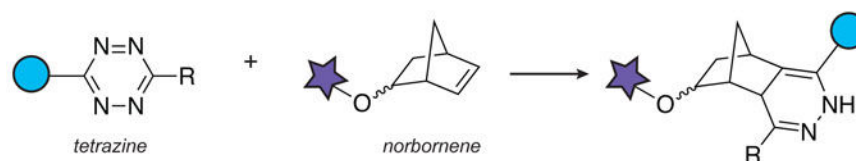
Fig. 10.
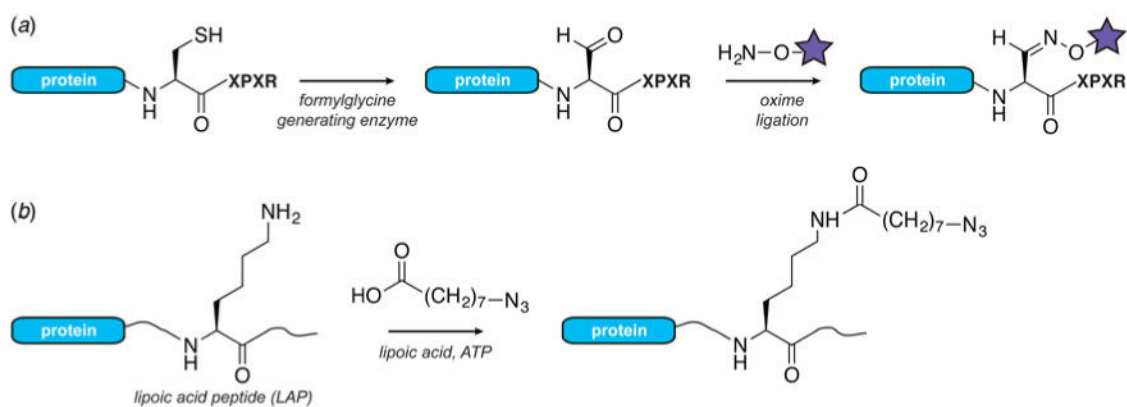Chemical modification of unnatural amino acids.

Fig. 11.
Examples of bioenzymatic conjugation approaches. (*a*) Site-specific modification of cysteine with formylglycine generating enzyme, followed by oxime ligation to attach a chemical or optical probe. (*b*) Introduction of 'click' handles into proteins using lipoic acid ligase and lipoic acid analogs.
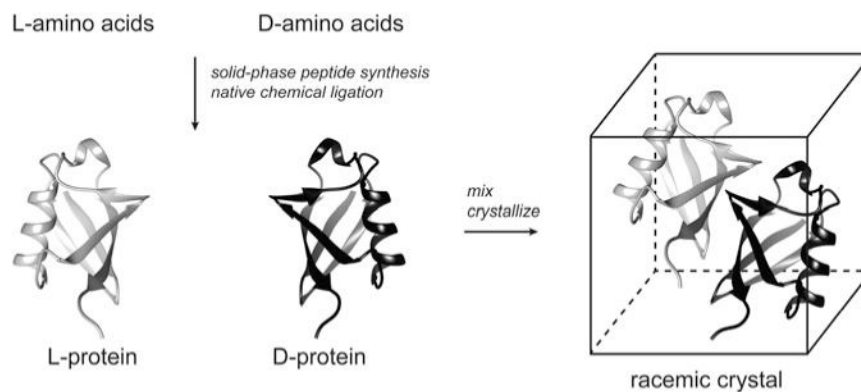
Fig. 12.
Principle of racemic crystallography. The L- and D- forms of the polypeptide chains are prepared separately by solid-phase peptide synthesis and native chemical ligation. The proteins are subsequently mixed and co-crystallized, thus gaining access to a much larger set of crystallographic space groups.
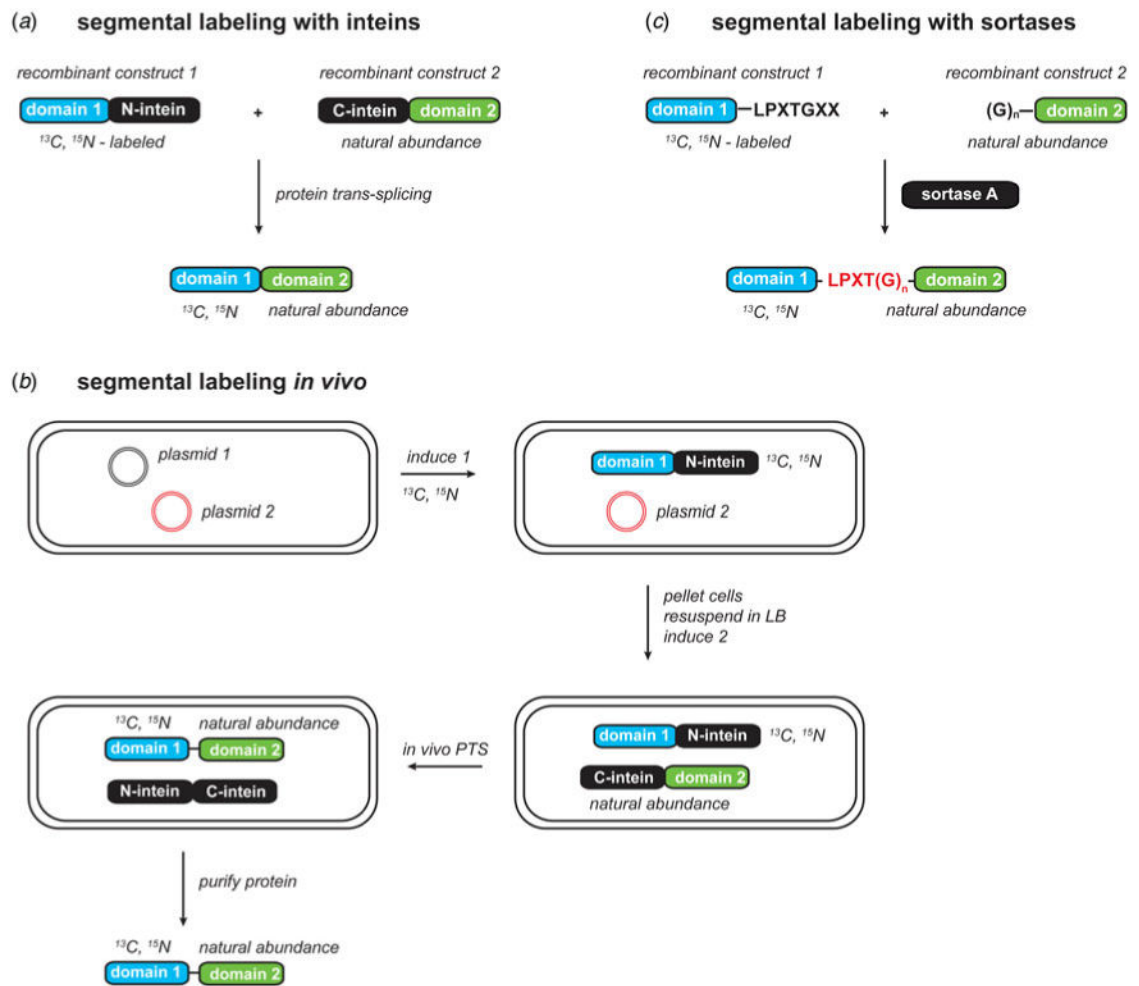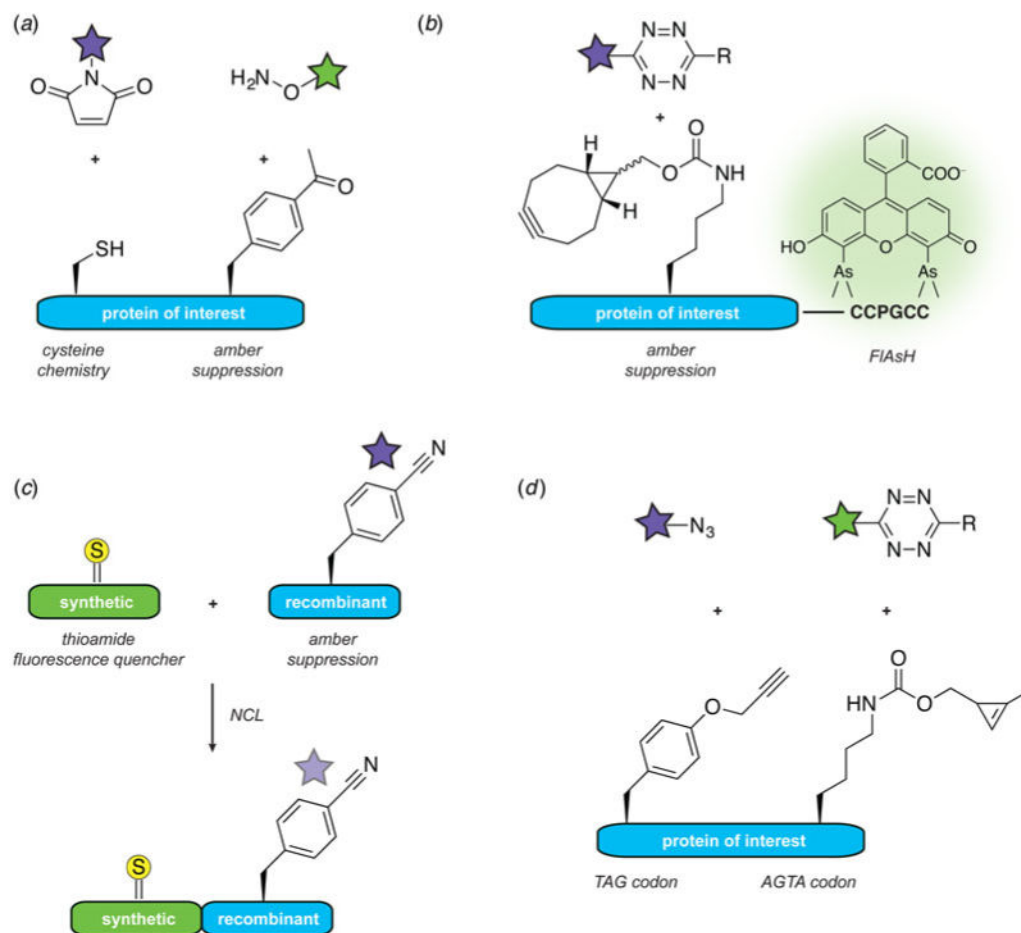
Fig. 13.

Strategies for segmental isotopic labeling of proteins for NMR analysis. (*a*) Intein-based segmental labeling. (*b*) Strategy for segmental labeling of proteins in cells. (*c*) Labeling with sortases.

Fig. 14.
Dual labeling of proteins with fluorophores. (*a*) A labeling strategy based on the combination of cysteine chemistry and amber suppression (Brustad *et al.* 2008). (*b*) The FlAsH labeling system can be used for the selective modification of a genetically encoded peptide tag, in combination with amber suppression (Perdios *et al.* 2017). (*c*) A dual labeling strategy based on native chemical ligation and amber suppression (Wissner *et al.* 2013). (*d*) Genetic incorporation of two UAAs using orthogonal ribosomes that can decode the AGTA quadruplet codon (Sachdeva *et al.* 2014).

Table 1

Intein toolbox for protein semi-synthesis

Intein	Temperature (°C)	$t_{1/2}$ [*]	Comments	References
*Mxe*GyrA	25	10 h	Commonly used contiguous intein for EPL applications	Frutos *et al.* (2010); Southworth *et al.* (1999)
*Ssp*DnaE	37	76 min	The first discovered natural split intein	Shah *et al.* (2012); Wu *et al.* (1998)
*Npu*DnaE	37	19 s	Fast split intein; synthetically accessible C-fragment	Shah *et al.* (2012); Vila-Perello *et al.* (2013)
*Ava*DnaE	37	23 s	Fast split intein; higher protein expression yields reported	David *et al.* (2015); Shah *et al.* (2012)
Cfa (consensus DnaE split intein)	30	20 s	Fast split intein; works at high temperatures and in denaturants	Stevens *et al.* (2016)
gp41-1	45	4 s	Fast and orthogonal to other split inteins	Carvajal-Vallejos *et al.* (2012)
gp41-8	37	15 s	Fast and orthogonal to other split inteins	Carvajal-Vallejos *et al.* (2012)
NrdJ-1	37	7 s	Fast and orthogonal to other split inteins.	Carvajal-Vallejos *et al.* (2012)
IMPDH-1	37	8 s	Fast and orthogonal to other split inteins.	Carvajal-Vallejos *et al.* (2012)
AceL-TerL	8	7.2 min	Synthetically accessible N-intein	Thiel *et al.* (2014)

[*] Optimal splicing kinetics in the presence of native sequences at the immediate intein-extein junctions. Variation from this sequence context can lead to less efficient splicing. See also ref. (Shah & Muir, 2014).