Protein ligation: an enabling technology for the biophysical analysis of proteins

Vasant Muralidharan & Tom W Muir

Biophysical techniques such as fluorescence spectroscopy and nuclear magnetic resonance (NMR) spectroscopy provide a window into the inner workings of proteins. These approaches make use of probes that can either be naturally present within the protein or introduced through a labeling procedure. In general, the more control one has over the type, location and number of probes in a protein, then the more information one can extract from a given biophysical analysis. Recently, two related approaches have emerged that allow proteins to be labeled with a broad range of physical probes. Expressed protein ligation (EPL) and protein *trans*-splicing (PTS) are both intein-based approaches that permit the assembly of a protein from smaller synthetic and/or recombinant pieces. Here we provide some guidelines for the use of EPL and PTS, and highlight how the dovetailing of these new protein chemistry methods with standard biophysical techniques has improved our ability to interrogate protein function, structure and folding.

An intimate understanding of protein structure and function remains a principal goal of molecular biology. The seemingly byzantine structure-activity relationships underlying protein function make this endeavor extremely challenging and one that requires ever more sophisticated approaches as we attempt to move toward a complete physical-chemical description of how proteins work.

Standard bioconjugation and isotopic-labeling procedures used to introduce probes into proteins have limitations. For example, relatively subtle modifications are often not possible, such as replacement of a single atom in a protein with an isotope thereof, or another atom. In addition, it is generally not possible to incorporate multiple, different spectroscopic or isotopic probes into discrete sites or regions of a protein using any standard procedure. These and other restrictions continue to fuel the development of new protein-chemistry approaches. The last decade, in particular, has seen the introduction of several techniques that allow the covalent structure of proteins to be manipulated with an unprecedented level of control[1]. Among these newer approaches, the protein-ligation methods EPL and PTS have emerged as being especially powerful for the site-specific incorporation of physical probes for studying protein structure and function *in vitro*. The goal of this review is to provide an overview of these protein-engineering approaches, with a particular eye toward the nonspecialist interested in using these techniques to generate labeled proteins for biophysical studies.

Overview of EPL and PTS

EPL and PTS are protein semisynthesis approaches that allow the assembly of a target protein from smaller unprotected polypeptide building blocks (**Figs. 1** and **2**). The techniques are related in that they both involve the use of a class of auto-processing proteins called inteins (**Box 1**). EPL is an extension of the immensely successful native chemical ligation (NCL) method[2] (**Box 2**) in which two unprotected synthetic peptides are chemically ligated together (**Fig. 1**). In EPL, one or more of the peptides is of recombinant origin, but the actual ligation step is still a chemical process and can be performed under a wide range of reaction conditions[3]. By contrast, the ligation step in PTS must be performed under conditions compatible with protein folding because the process involves the functional reconstitution of a split intein[4] (**Fig. 2**).

Laboratory of Synthetic Protein Chemistry, The Rockefeller University, New York, New York 10021, USA. Correspondence should be addressed to T.W.M. (muirt@rockefeller.edu).

PUBLISHED ONLINE 23 MAY 2006; DOI:10.1038/NMETH886

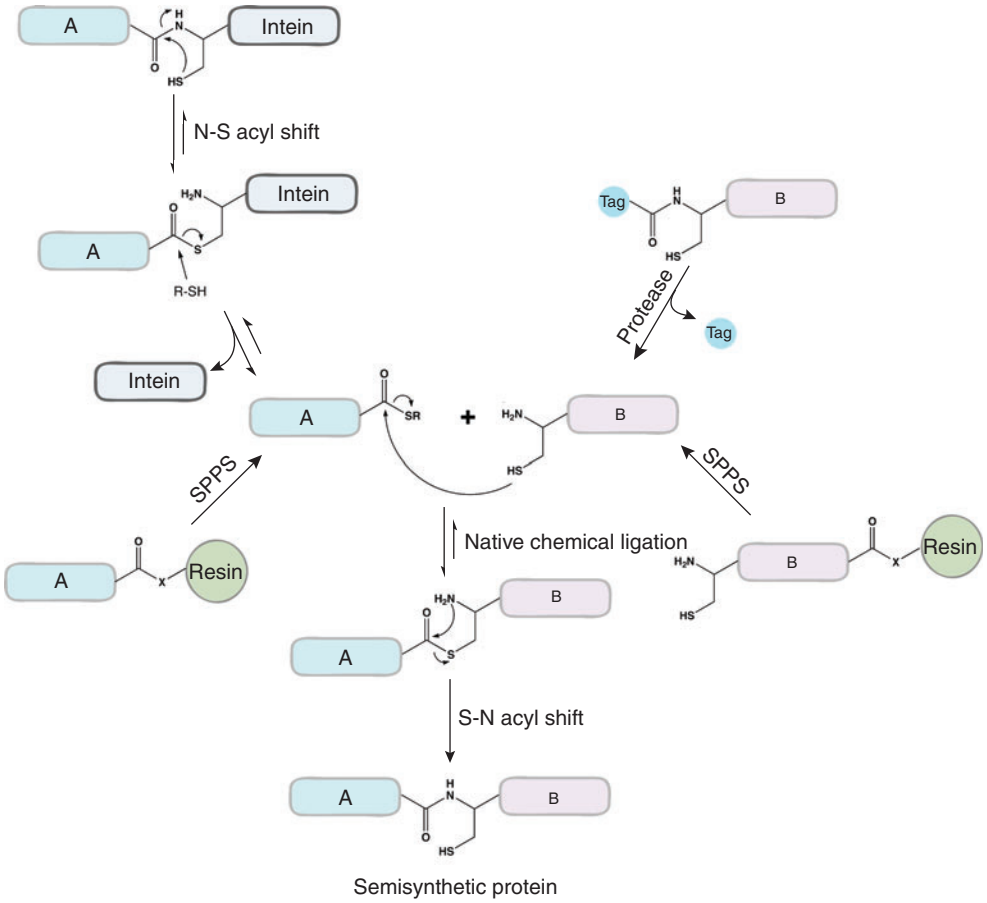

Figure 1 | Principle of expressed protein ligation (EPL). EPL is a semisynthetic version of NCL in which synthetic and recombinant polypeptides are chemically ligated together. Proteins (A) expressed as intein fusions can be cleaved from the intein with a variety of thiols to give the corresponding α-thioester derivative. Proteins (B) containing N-Cys can be made recombinantly by masking the cysteine with a protease tag that can be later removed. The two protein building blocks, A and B, can also be synthesized by SPPS on a solid support (resin) containing an appropriate linker (X).

In the simplest case, EPL and PTS allow the semisynthesis of a target protein from two polypeptide pieces. This permits the incorporation of probes within the flanking regions of a protein. More sophisticated approaches are available for ligating together three building blocks in a regioselective fashion, thereby permitting internal regions of proteins to be specifically modified[5–7]. Both EPL and PTS can be used to ligate a recombinant polypeptide to a synthetic peptide. EPL is the more generally useful approach for this purpose[3], largely because of the robust nature of the ligation reaction and the relative ease of preparing the reactive synthetic peptides compared to PTS, in which the synthetic fragment must also include part of the intein. Indeed, EPL has been widely used to incorporate unnatural amino acids, post-translational modifications and spectroscopic probes into many classes of protein[3,8].

An underappreciated aspect of EPL and PTS is that they can be used to ligate together two recombinant polypeptides. The rationale for doing this might at first seem unclear, as this does not allow the introduction of synthetic probes into a protein. This capability, however, has proven extremely useful for the production of cytotoxic proteins from innocuous fragments[9,10] and for the segmental isotopic labeling of proteins for NMR spectroscopy studies[11,12]. For these applications, EPL and PTS each have their advantages and disadvantages. For example, EPL reactions can be performed in the presence of many different additives, but high concentrations of purified protein reactants (ideally in the millimolar range) are usually required for efficient ligation. By contrast, PTS reactions can be performed using crude preparations of protein reactants at relatively low concentrations (low micromolar), but the reactions must be performed under physiological conditions.

Both EPL and PTS can be used to generate circular polypeptides both *in vitro* and *in vivo*[13–23]. This involves an intramolecular ligation reaction in which the N and C termini of a linear recombinant polypeptide precursor end up linked by a peptide bond. Circularization has been used to stabilize proteins[14,19,23] and to prepare libraries of constrained bioactive peptides[15,21,22]. PTS is advantageous over EPL for the circularization of unstructured peptides. This is because complementation of the split intein fragments can help overcome the unfavorable entropic energy associated with bringing the ends of the linear peptide precursor together. Thus, PTS is the method of choice for the generation of circular peptide libraries for medicinal chemistry and chemical biology type applications[15,21,22].

Performing EPL

General considerations. Several factors must be taken into account when designing an optimal semisynthesis. First and foremost, one should consider the effect of the ligation procedure on the protein fold. Some applications will require that the protein be split within a structured domain, meaning that folding will be necessary after liga-

tion. Furthermore, EPL (and PTS) reactions are performed under conditions that will most likely reduce any disulfides in the protein. Thus, an oxidation step may be required after ligation if the protein contains native disulfides. Clearly, in such cases it is critical to establish that the native protein can be reversibly unfolded before beginning a semisynthesis. Whenever possible, the ligation site should be located between independently folded domains. EPL is most readily applied when the desired modifications are restricted to a short stretch of sequence located within the flanking regions of the protein, as this allows the use of two building blocks and a single ligation step. Moreover, if the application involves a synthetic peptide, then the closer the ligation site can be made to the native N or C terminus of the protein, the better, as this reduces the length of the synthetic peptide; generally speaking, it is difficult to prepare peptides longer than ~50 residues using solid-phase peptide synthesis (SPPS)[24]. For applications that require the selective introduction of probes into the middle of a protein, or the incorporation of probes within two or more remote parts of the sequence, it will be necessary to assemble the protein from a minimum of three peptide segments. This adds extra steps to the semisynthesis and is often technically challenging. Nonetheless, approaches to this have been developed and rely on the use of N-terminal cysteine (N-Cys) protection schemes that permit sequential ligation reactions[6]. The most robust of these involves protection of the N-Cys as a thiazolidine derivative[25,26].

Ligation sites. As EPL is an extension of NCL, the ligation product will have a cysteine residue at the ligation junction (**Fig. 1**). Although some proteins will contain a native cysteine residue relatively close to the site of probe incorporation, many will not, particularly because cysteine is the second least common of the 20 amino acids in proteins[27]. In this situation one has two options. The first, and simplest, solution is to introduce a cysteine at an appropriate location by mutagenesis. In fact, this is the strategy used in the majority of EPL studies reported to date[3,8]. Care should be taken when introducing a non-native cysteine into a protein. Thus, the mutation should be designed to be as conservative as possible in terms of sequence (for example, alanine or serine to cysteine) and structure (for example, in a linker or loop region). Moreover, one should avoid mutating highly conserved residues known to be involved in function. Whenever possible it is advisable to evaluate the effect of a cysteine mutation on protein function by making the site-mutant before beginning a semisynthesis[28]. If a cysteine mutation is not tolerated, it is possible to perform the ligation reaction in a traceless manner by desulfurization with Raney nickel and hydrogen[29]. This converts cysteine to alanine, but will do so uniformly throughout the protein. Alternatively, if the C-terminal fragment is synthetic it may be possible to use cysteine auxiliaries[30,31] or the Staudinger ligation approach[32,33], both of which show great promise for traceless ligation.

Another factor in choosing a ligation site is the identity of the amino acid immediately upstream of the intein; this residue will be at the C terminus in the α-thioester fragment. Individual inteins have their own set of preferences at this position[34–36], with certain residues associated with increased levels in premature, *in vivo* cleavage (for example, aspartic acid), whereas other residues are associated with no cleavage at all (for example, proline). A related issue is the effect of varying the identity of this residue on the kinetics of NCL. Not

BOX 1 PROTEIN SPLICING AND INTEINS

Protein splicing is a post-translational process in which a precursor protein undergoes self-catalyzed intramolecular rearrangements that result in the removal of an internal protein domain, termed an intein, and the ligation of the two flanking polypeptides, referred to as the N and C exteins. Over 170 members of the intein family are currently cataloged (http://www.neb.com/neb/inteins.html), being characterized by several conserved sequence motifs[85]. Inteins are autocatalytic and some are remarkably promiscuous with respect to the sequences of the two flanking exteins[4]. Biochemical studies have revealed the basic chemical steps in protein splicing and have identified the key conserved residues within the intein family required for these steps[34,76,86]. Although most protein splicing reactions occur in *cis* and involve intact inteins, there are several split inteins that participate in a protein *trans*-splicing reaction. In this process, complementation of the intein fragments precedes the normal splicing reaction (**Fig. 2**).

Protein splicing has been used extensively in the areas of biotechnology and chemical biology. The methods fall into two categories, those that exploit *cis*-splicing inteins and those that use split inteins. A number of mutant *cis*-splicing inteins have been generated that can participate in only the first[35,45,76,78,86,87] or last[34,35,45,46] steps of protein splicing. In each case, a protein of interest is expressed as an in-frame N- or C-terminal fusion to the mutant intein, which is usually linked to an affinity tag[76]. After purification, the protein of interest can be cleaved off the intein by addition of thiols (C-terminal cleavage, that is, the protein corresponds to the N extein) or by changing the pH and temperature of the solution (N-terminal cleavage, that is, the protein corresponds to the C extein). These intein-based expression systems can be used for the 'traceless' purification of recombinant proteins[76]. The approach can be coupled with most bioseparation strategies including, most recently, the use of mechanical separations based on phase transitions[88,89]. Notably, these intein fusions provide a convenient route to recombinant proteins containing an N-terminal cysteine or a C-terminal thioester. These reactive building blocks can then be used in EPL strategies[9,40].

Trans-splicing inteins come in two flavors, those that are artificially generated by splitting a normal *cis*-splicing intein[11,52,53] and those that are naturally split[55,90], the best characterized of which is the Ssp DnaE intein[91]. The key functional distinction relates to the affinity of the complementary fragments; the DnaE split intein fragments associate with high affinity[51], whereas artificially split intein fragments typically do not[11,52,53]. For this reason, the DnaE split intein is preferable for most protein ligation applications *in vitro* and *in vivo* because of the spontaneous nature of PTS. Conversely, artificially split inteins are useful for two- and three-hybrid applications as PTS only occurs when the fragments are at high local concentration[5,11,51,57,58]. Another very useful feature of PTS is that different pairs of split inteins do not crossreact, allowing multiple protein ligation reactions to be performed simultaneously[5,51].

surprisingly, increasing the steric bulk of the side chain (particularly at the β position) slows down the reaction[37]. Thus, one should avoid using Thr-Cys, Ile-Cys and Val-Cys ligation junctions.

Generation of peptide α-thioesters. Peptide α-thioesters can be prepared synthetically by SPPS or biosynthetically via intein fusion strategies (**Fig. 1**). Several solid-phase methods are available for the chemical synthesis of peptide α-thioesters (**Table 1**). The most general of these involves the use of *tert*-butoxycarbonyl (Boc) SPPS that uses acid-based deprotections to which thioesters are stable. This should be contrasted with 9-fluorenylmethoxycarbonyl (Fmoc) SPPS that requires the use of repeated base treatments, rendering the strategy incompatible with thioester groups. But Fmoc SPPS is more commonly used by peptide chemists and, notably, allows the incorporation of acid-sensitive groups (for example, phosphates, carbohydrates) into the peptide that are not compatible with Boc SPPS. Consequently, several technologies have been developed to allow the Fmoc synthesis of peptide α-thioesters (**Table 1**). Although none of these Fmoc strategies are as robust as Boc SPPS (still the method of choice) for the synthesis of peptide α-thioesters, the recently introduced 'masked thioester' methods show some promise[38,39].

Recombinant protein α-thioesters can be produced by thiolysis

Table 1 | Strategies for generating reactive polypeptides for EPL

EPL building blocks	Method of preparation	Comments
Recombinant thioesters[a]	**Intein-mediated preparation**	
	Mxe GyrA intein[35]	Mini-intein (198 aa); refolds efficiently; works in the presence of denaturants, lipids, detergents and other additives.
	Sce Vma intein[76]	Contains internal homing endonuclease domain (overall size 455 aa); does not refold efficiently.
	Mtb RecA intein[46,77]	Mini-intein (229 aa).
	Mth RIR1 intein[78]	Smallest known mini-intein (134 aa); splicing occurs at 15 °C but not at 37 °C.
	Ssp DnaB intein[45]	Mini-intein (154 aa).
Synthetic thioesters	**Boc SPPS**	Method of choice for the synthesis of peptide α-thioesters; not suitable for synthesis of phospho- and glyco-peptides.
	Mercaptopropionamide linker[79]	Widely used; compatible with several polymeric resins (PEGA, poly(ethylene glycol)-poly(*N*,*N*-dimethylacrylamide; MBHA, methylbenzhydrylamine, and others); direct synthesis of peptide α-thioester.
	Fmoc SPPS	Widely used; allows incorporation of acid-labile groups; α-thioesters are sensitive to base treatments used in Fmoc deprotection.
	Sulfonamide linker[80,81]	Safety-catch linker activated by alkylation after chain assembly; efficiency of thiolysis can be low; nonspecific alkylations can occur.
	Backbone amide linker[82]	Thioester installed after chain assembly; racemization of the penultimate residue can be a problem.
	Lewis acid–activated cleavage[83]	Uses commercially available resin-linkers; α-thioester is introduced upon cleavage from resin with alkylaluminium and thiols; racemization and other side reactions can occur.
	Mercaptocarboxyethylester linker[38]	Peptide is synthesized as C-terminal ester; peptide thioester derivate is generated *in situ* during ligation; hydrolysis of thioester can be a problem.
	Aryl hydrazine linker[84]	Peptide is synthesized as C-terminal acyl diazene; α-thioester is introduced upon cleavage from resin with α-amino acid S-alkyl thioester under mildly basic conditions; requires solution synthesis of α-amino acid S-alkyl thioester.
	S-protected oxazolidinone linker[39]	Peptide is synthesized as C-terminal S-protected oxazolidinone; treatment with trimethylsilyl bromide and thioanisole in trifluoroacetic acid results in S-deprotected oxazolidinone; α-thioester is generated *in situ* through an N–S acyl transfer; racemization of C-terminal thioester amino acid is a problem.
Recombinant N-Cys	**Using proteolytic leader sequence**[b]	
	N-terminal methionine[14,18]	Depends on endogenous methionyl aminopeptidase; N-terminal Cys residue may be modified *in vivo*.
	Factor Xa protease[42]	Commercially available; usually efficient; nonspecific cleavage can be a problem; requires a four-residue recognition sequence; protease is sensitive to thiols.
	TEV protease[43]	Can be overexpressed in *E. coli*; also available commercially; requires a seven-residue recognition sequence; very specific.
	Using N-terminal intein fusion[c]	
	Mxe GyrA intein[35]	Cleavage induced by temperature shift from 37 °C to 16–25 °C.
	Mth RIR1 intein[78]	Cleavage is induced by addition of thiols.
	Ssp DnaB intein[45]	Cleavage is most efficient at 15 °C. Cleavage only occurs between pH 6.0 and 7.5.
Synthetic N-Cys	**Direct synthesis**	Straightforward; standard procedure; can be prepared by either Boc or Fmoc SPPS.

Sce, *Saccharomyces cerevisiae*; Mtb, *Mycobacterium tuberculosis*; Mth, *Methanobacterium thermoautotrophicum*. [a]The C-terminal asparagine residue of the intein is mutated to an alanine. Thus the intein can only catalyze the first step of protein splicing. [b]A leader sequence is located N-terminally to the cysteine, and is cleaved off with a specific protease. [c]The N-terminal cysteine of the intein is mutated to an alanine. Thus the mutant intein can only catalyze the last step of protein splicing.

of the corresponding intein fusions[9,40]. Several modified inteins are commonly used for this purpose (**Table 1**) and many are commercially available as *E. coli* expression vectors. In our experience the most generally useful of these is the *Mycobacterium xenopi* DNA gyrase A (Mxe GyrA) intein. This intein has several nice characteristics: (i) It is relatively small (198 amino acids); (ii) It can be cleaved efficiently with a variety of thiols to give the corresponding α-thioesters[7,28,35,41]; (iii) Thiolysis can be performed in the presence of detergents[28,41], moderate concentrations of denaturants[41] and mixed aqueous/organic solvents[28], and; (iv) It can be efficiently refolded[28]. The latter point is especially useful since it allows Mxe GyrA fusions to be recovered from bacterial inclusion bodies[28].

Generation of N-Cys peptides. N-Cys peptides can be prepared synthetically by routine SPPS using either Boc or Fmoc methodologies (**Fig. 1**). Several methods are available for generating recombinant proteins containing an N-Cys (**Table 1**). All rely on processing of a leader sequence from the N terminus of the polypeptide, thereby revealing the N-Cys. This leader sequence can be as simple as the initiating methionine[14,18], or can be one of several protease recognition sequences[42,43]. Additionally, several engineered inteins are available that will cleave themselves off the protein of interest leaving behind an N-Cys[35,44–46]. These methods each have their pros and cons (see **Table 1**), and in our experience it is difficult to predict which strategy will work best for a given protein. Thus, an empirical approach to this usually has to be taken.

Ligation conditions. EPL reactions can be performed in two ways: either the thiolysis and NCL reactions are carried out in one 'pot', or the recombinant-protein α-thioester is isolated first. The former approach limits, to some extent, the type of additives that can be present in the reaction mixture because the intein must remain folded during the ligation reaction. Prior isolation of the protein α-thioester allows the NCL step to be performed in the presence of a much broader range of additives, including conditions that fully denature the protein[7,28,47,48]. The advantage of using additives such as chaotropes or detergents is that they allow high concentrations (millimolar) of the reactant polypeptides to be achieved, thereby improving the ligation yields. Note that lower reactant concentrations can be used if the peptide reactants have an innate affinity for one another as the ligation reaction essentially becomes a unimolecular process[41,49]. Notably, EPL reactions should always be performed

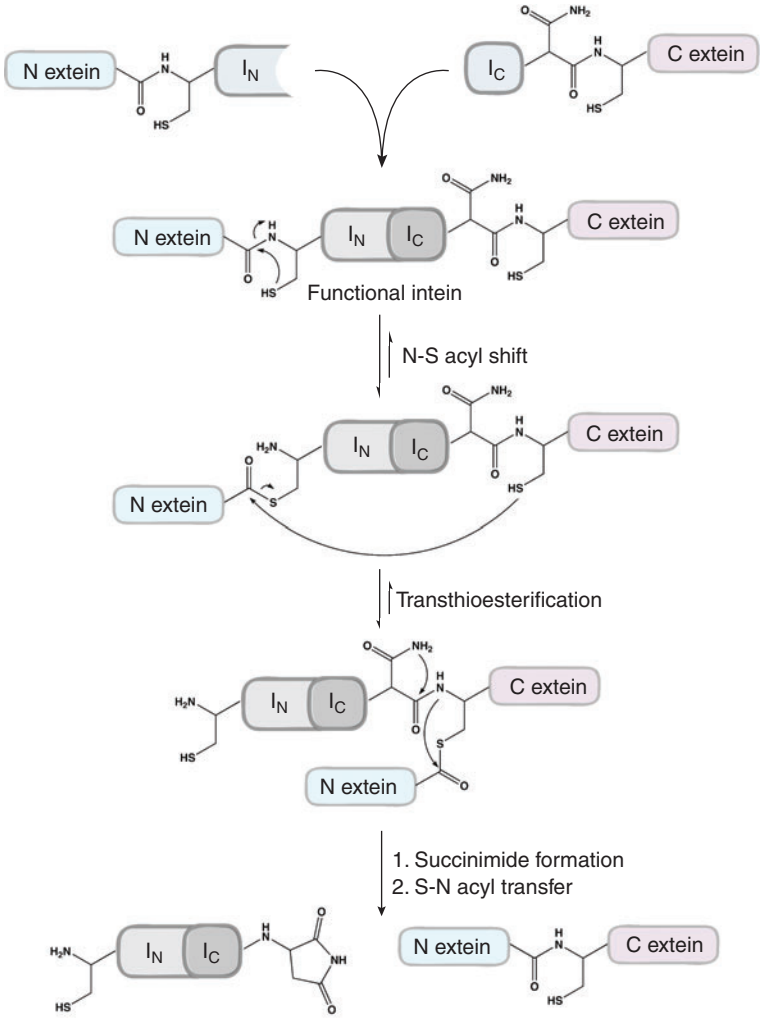

Figure 2 | Principle of protein *trans*-splicing (PTS). The two halves of the split intein, labeled as I_N and I_C, associate and fold to form a functional intein. This functional intein can then undergo a pseudo-intramolecular protein splicing reaction, wherein the flanking polypeptides, termed the N and C exteins, are ligated together and the intein excises itself.

in the pH range of 7–8 and in the presence of a moderate concentration (10–100 mM) of a thiol cofactor such as mercaptoethanesulfonic acid or thiophenol. Use of this pH window ensures the ligation reaction is chemoselective, whereas the addition of thiols helps to suppress competing and unproductive oxidation and hydrolysis reactions[2]. Note that the presence of trace amounts of aldehydes or ketones (such as glyceraldehyde found in glycerol preparations) can react with the N-Cys residue to form a 'dead-end' thiazolidine derivative. This type of side reaction has also been observed when the N-Cys is generated *in vivo*, where intracellular pyruvic acid reacts with the cysteine residue[50]. The thiazolidine modification of the N-Cys can be reversed using methoxylamine and a reducing agent such as tris(2-carboxy-ethyl)phosphine[25].

Performing PTS

General considerations. The successful use of PTS depends on many of the same factors discussed in relation to EPL. Thus, one should avoid splitting the target protein within a folded domain, but if this cannot be avoided then folding conditions should be established before beginning the semisynthesis. As with EPL, the most straightforward application of PTS involves the generation of the target protein from two fragments as this involves a single reaction (**Fig. 2**). More sophisticated strategies have been described which allow three polypeptides to be regioselectively joined together in a one-pot process[5,51]. These tandem PTS strategies involve the use of different pairs of split inteins, and can be performed using purified proteins[5] or in crude cell extracts[51]. For example, our laboratory recently described a tandem PTS system that uses the naturally split *Synechocystis* sp. (Ssp) DnaE intein and a chemically controlled, split *Saccharomyces cerevisiae* vacuolar ATPase subunit (VMA) intein. This technology allows three polypeptides to be joined together in the presence of a small-molecule inducer[51] and, notably, does not suffer from the obligatory refolding step associated with other PTS systems[11,52,53].

Choice of split intein. *Trans*-splicing inteins come in two types, those that are artificially generated by cutting a normal *cis*-splicing intein and those that are naturally split, such as the Ssp DnaE intein. Inteins have been split at various points along their sequence, usually between conserved motifs[11,52–54]. The Ssp DnaE intein is split disproportionately toward the C terminus; the N-terminal fragment (DnaE-N) is 123 amino acids long and is highly acidic, whereas the C-terminal fragment (DnaE-C) is 36 residues in length and is highly basic. The DnaE-N and DnaE-C fragments have remarkably high affinity for one another (K_D = ~40 nM) and there seems to be a strong electrostatic contribution to the very fast association of the fragments[51]. This property means that PTS is spontaneous even when the DnaE-N and DnaE-C fusions are mixed together at low concentrations. Thus, the split Ssp DnaE intein is preferable for many semisynthesis applications.

The efficiency of PTS depends strongly on the amino acids immediately upstream and downstream of the N-terminal and C-terminal intein fragments, respectively. In the case of the Ssp DnaE intein, there is a minimal requirement for three native C extein residues (Cys-Phe-Asn) for efficient *trans*-splicing[51,55,56]. Thus, the product of the reaction will contain these presumably non-native amino acids at the splice junction. This requirement puts substantial constraints on the location of the splicing site, with unstructured loops or linker regions being preferred[51]. The effect of these mutations on protein function should be evaluated before starting a semisynthesis.

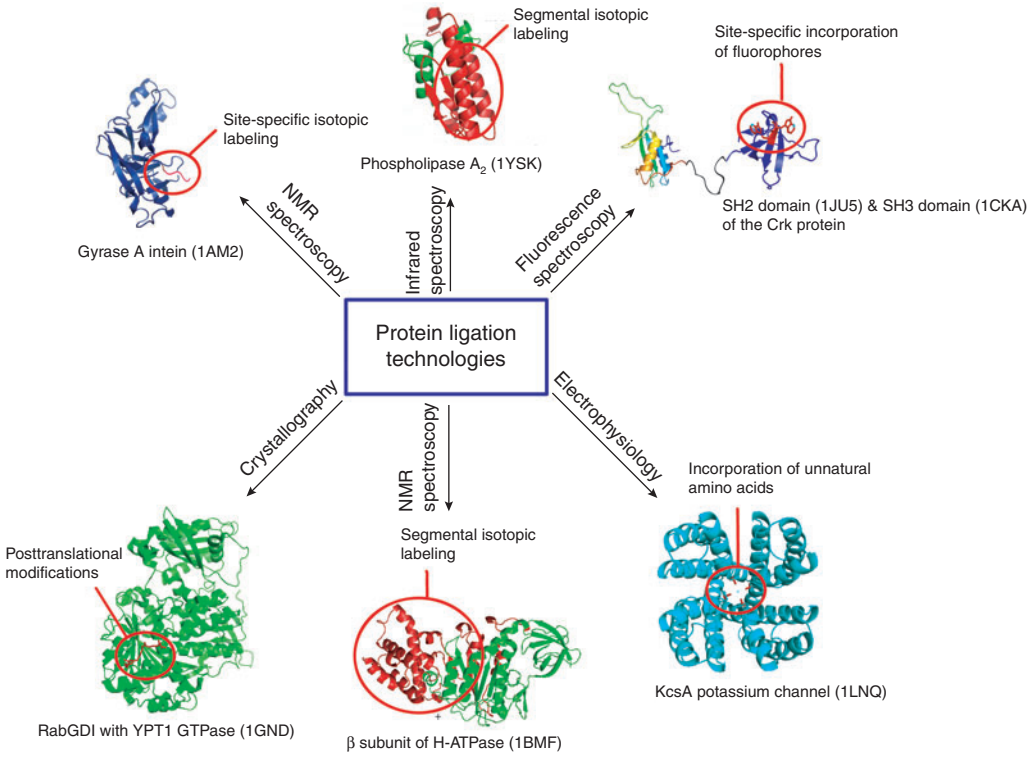

Figure 3 | Selected applications of protein ligation technologies. Protein ligation techniques offer the ability to incorporate specific labels into proteins that have been used to assay structure and function, augmenting several traditional biophysical techniques. The Protein Data Bank (PDB) codes of the structures are given in parentheses.

BOX 2 CHEMICAL LIGATION OF UNPROTECTED PEPTIDES

Over the last ~15 years, chemical ligation has emerged as a powerful technique in chemical biology, allowing unprotected molecules to be selectively linked together in an aqueous environment. Chemical ligation was initially introduced to solve the decades-old problem of how to chemically synthesize proteins from smaller, synthetically accessible, peptide building blocks[92,93]. The basic strategy, however, has found widespread use in other areas, including carbohydrate synthesis[94] and nucleic-acid synthesis[74]. All chemical ligation strategies rely on the incorporation of 'molecular velcro' at the appropriate locations within the reactive partners; that is, mutually reactive groups that are unreactive toward everything else in the molecules under the conditions used. Several chemoselective reactions are in common use in chemical biology, including oxime formation[93], the Staudinger ligation[32,33] and Huisgen cycloaddition chemistry[95].

By far the most commonly used peptide ligation approach is native chemical ligation (NCL)[2]. In this reaction, two fully unprotected synthetic peptide fragments are reacted under neutral aqueous conditions with the formation of a normal (native) peptide bond at the ligation site (**Fig. 1**). The first step in NCL involves the chemoselective reaction between a peptide fragment containing an N-Cys and a second peptide fragment containing an α-thioester group. The initial transthioesterification reaction is followed by a spontaneous intramolecular S-N acyl shift to generate an amide bond at the ligation junction. NCL is compatible with all naturally occurring functional groups found in proteins, including the sulfhydryl group of cysteine. Indeed, it is worth emphasizing that internal cysteine residues are permitted in both peptide segments owing to the reversible nature of the initial transthioesterification step. NCL is a remarkably robust reaction that can be performed in the presence of chemical denaturants[2], detergents[96], lipids[97], reducing agents[2,37], organic solvents[98] and even crude cell extracts[99]. The approach has undergone several important refinements since it was introduced. These include the development of auxiliary groups that in favorable cases overcome the requirement for an N-Cys[30,31], the identification of catalytic thiol cofactors[100], and the development of solution-phase and solid-phase sequential ligation strategies that allow several peptides to be linked together in series[6,26,79]. Additionally, new SPPS approaches for the preparation of synthetic peptide α-thioesters[38,39,80–84] promise to make this class of peptide more accessible to the nonspecialist (**Table 1**).

EPL is a semisynthetic version of NCL in which one or both of the building blocks are made by recombinant DNA expression (**Fig. 1**). Intein-based fusion approaches allow the straightforward introduction of α-thioesters into recombinant proteins, whereas an N-Cys can be introduced into recombinant proteins via one of several proteolytic strategies[14,42,43], as well as through the use of intein fusions[13,35,45,78].

Split intein fragments are thought to be largely unstructured polypeptides before complementation. This can result in fusion proteins being unstable, particularly if the N or C exteins to which they are linked are themselves unstructured polypeptides. Expression of soluble split intein fusions can, however, be greatly improved by appending a well-behaved globular protein to the split end of the intein fragment[51,52,57]. For example, maltose binding protein can be fused to the C terminus of DnaE-N without affecting splicing.

***Trans*-splicing conditions.** Depending on the split intein, PTS reactions can be carried out either with or without prior purification of the fusion proteins. For example, the Ssp DnaE intein system has been used successfully in test tubes[17], cell lysates[51], cultured cells[58,59] and *in vivo*[60,61]. In contrast, most artificially split inteins require a refolding procedure before *trans*-splicing is observed[5,11,52,53]. This is thought to be due to the tendency of the intein fragments to misfold when simply combined together under physiological conditions[52,53]. But the low affinity of the intein fragments may also be a contributing factor, as some artificially split inteins will splice spontaneously when fused to interacting partners[57]. Refolding is typically performed by dialysis out of chemical denaturants in the presence of reducing agents such as dithiothreitol (DTT) or tris(2-carboxyethyl) phosphine (TCEP)[11,52,53]. The reducing agent prevents unproductive oxidation reactions involving the active-site cysteines in the intein. Note that reducing agents are also required when using the Ssp DnaE intein system *in vitro*[17,51].

Selected applications of EPL and PTS

EPL and PTS have found widespread application in the biochemical and biophysical study of proteins[3,8]. In this section, we highlight just a few examples that illustrate how these approaches have been dovetailed with standard biophysical techniques to allow the detailed analysis of protein structure and function (**Fig. 3**).

Optical spectroscopy. Fluorescence spectroscopy is perhaps the most commonly used technique in protein biophysics. Nature provides us with just two fluorescent amino acids, tryptophan and tyrosine. Tryptophan is the more sensitive fluorophore, and is thus a useful probe for monitoring protein folding transitions and ligand binding events. In general, the value of tryptophan as a fluorescent probe decreases as the size of a protein increases; most large proteins contain several tryptophan residues, thereby reducing the resolution of the analysis. Protein ligation approaches help address this by allowing the site-specific incorporation of fluorophores with spectral properties distinct from tryptophan and tyrosine. There have been a large number of fluorescent proteins prepared by EPL[3,8]. Most of these semisynthetic proteins have been used to study ligand binding events, either by monitoring changes in the fluorescence of a single probe in the protein, or by using fluorescence resonance energy transfer (FRET) between appropriately positioned donor and acceptor probes. These different approaches are nicely illustrated by the recent work of Lorsch and coworkers who, in a series of studies, used a combination of fluorescence anisotropy and FRET measurements to measure the dynamic association of fluorescent analogs of several translation initiation factors and the ribosome[62,63].

The ability to introduce noncoded fluorescent probes by protein ligation opens up new possibilities for studying protein folding. In one recent study, a combination of *in vivo* amino acid replace-

ment and EPL was used to incorporate the tryptophan analog, 7-azatryptophan (7-AW), into a single SH3 domain within the multidomain signaling protein, c-Crk[48]. The 7-AW is isosteric with tryptophan, but its fluorescence excitation and emission properties are quite distinct. Use of this noninvasive optical probe allowed the equilibrium stability and ligand-binding properties of the c-Crk SH3 domain to be unambiguously studied in the context of the full-length protein. This strategy is likely to be applicable to study the folding of discrete domains within other multidomain proteins.

Isotope-edited spectroscopy. EPL and PTS help overcome some of the size limitations for protein-structure analysis by NMR spectroscopy. Both methods allow NMR-active isotopes to be incorporated site-specifically within a protein, permitting straightforward assignment of resonances from the labeled amino acids. In one recent example, the scissile peptide bond at the N extein-intein junction of the GyrA intein was uniquely labeled with ^{13}C and ^{15}N nuclei using EPL[47]. This allowed the amide $^{1}J_{NC'}$ coupling constant to be measured on the active intein. Based on this analysis, this amide appears to be unusually polarized, most likely because of nonplanarity. This supports a 'ground-state destabilization' reaction mechanism for the first cleavage step in protein splicing. In another application of EPL, a series of ^{13}C-labeled amino acids were incorporated into the C terminus of the α subunit of a heterotrimeric G protein[64]. Analysis of the ^{13}C resonances indicated that the C terminus the Gα subunit is unstructured when the protein is bound to GDP, but adopts an ordered conformation upon activation of the subunit. The authors hypothesize that this structural change may be involved in release of the Gα subunit from the receptor.

Protein ligation methods also permit the generation of segmental isotopically labeled proteins for NMR spectroscopy studies[11,12]. In this strategy, a uniformly labeled recombinant protein fragment is ligated to an unlabeled protein fragment. The end result is a 'block'-labeled protein sample that gives rise to heteronuclear NMR spectra containing a subset of the resonances that one would have observed from the corresponding uniformly labeled sample. In other words, there is substantial reduction in spectral complexity. Both EPL and PTS have been used to generate segmental isotopically labeled proteins[5,11,12]. The approach has proven particularly useful for studying conformational changes within large proteins[41]. In a recent example of this, PTS was used to generate several segmentally labeled versions of the β subunit of the F_1-ATPase[65]. The large size of this protein (52 kDa) made it an ideal target for segmental labeling. In a technical tour-de-force, a combination of segmental labeling and transverse relaxation optimized spectroscopy (TROSY) NMR experiments was used to assign most of the residues in the protein. Chemical shift perturbation analysis and residual dipolar coupling measurements were then used to show that the β subunit undergoes a conformational change upon nucleotide binding. The authors suggest that this change is an important driving force for F_1 rotation.

Technology improvements continue to be made in the segmental isotopic labeling area. In one particularly exciting development, a PTS system based on the Ssp DnaE split intein was recently reported that allows segmental labeling of proteins within *Escherichia coli* cells[56]. The system still requires some refinement, but the prospect of being able to generate segmentally labeled proteins without having to perform any *in vitro* ligation reactions is extremely appealing.

Access to isotopically labeled proteins also extends the scope of vibrational spectroscopy for the structural analysis of proteins. This is illustrated by the recent work of Tatulian and coworkers who used EPL to generate a segmental ^{13}C-labeled version of phospholipase A2 (PLA_2)[66]. PLA_2 is associated with the lipid bilayer, although the exact position of the protein within the membrane is unclear. To answer this question, the authors carried out polarized attenuated total reflection Fourier transform infrared (ATR-FTIR) spectroscopy of the semisynthetic protein bound to supported membranes. Segmental labeling was critical to this analysis because it allowed assignment of the amide I signals from two ^{13}C-labeled helices in the protein. The authors then used this information to generate geometric constraints that allowed them to determine the orientation of the protein with respect to the membrane. In principle, this approach should be quite general and, notably, should help improve our understanding of how membrane protein function is influenced by rotational and translational positioning within the bilayer.

Structural studies of modified proteins. Protein function is often regulated by posttranslational modification. Chemically defined, posttranslationally modified proteins are difficult to generate using standard protein expression approaches. Thus, protein semisynthesis has found broad application in this area[3]. It has been estimated that as many as a third of mammalian proteins are phosphorylated, thus it should not be surprising that several phospho-proteins have been prepared by semisynthesis. For example, EPL has been used to prepare phosphorylated versions of the TGF-β signaling proteins, Smad2 and Smad3[67,68]. These proteins are latent transcription factors, which, in response to phosphorylation, oligomerize and move from the cytoplasm into the nucleus where they control transcriptional programs. The availability of semisynthetic phospho-Smad proteins has allowed several high-resolution crystal structures to be determined[67–69]. These structural studies have led to a very detailed understanding of the forces and interactions underlying Smad homo- and hetero-oligomerization.

Protein ligation has made possible the detailed structural analysis of proteins bearing other types of post-translational modifications. For instance, EPL has been used to generate a monoprenylated version of the small GTPase, Ypt1 (ref. 70). The co-crystal structure of this protein in complex with Rab GDP dissociation inhibitor (RabGDI) was then determined. The structure revealed that RabGDI undergoes a conformational change upon binding Ypt1, creating a hydrophobic cavity in which the prenyl group resides. Mutations in RabGDI are associated with some forms of mental retardation in humans. The structure indicates that these mutations would disrupt the interaction with prenylated Ypt1, suggesting that aberrant Ypt1 membrane transport may be part of the underlying disease mechanism.

Electrophysiology. The analysis of single molecules is an exciting topic in protein biophysics. The prototypical method in this area is the electrophysiological analysis of single ion channels. Technical developments now allow membrane proteins such as ion channels to be modified by synthesis[71] and semisynthesis[28], thereby allowing structure-activity relationships to be studied in great detail at the single-molecule level. The power of protein ligation in this area is illustrated by stereochemical engineering of the potassium channel, KcsA[72]. The crystal structure of KcsA provided our first view of the so-called ion selectivity filter, the region of the channel that selects for K^+ over other cations[73]. The filter contains a completely conserved glycine residue (Gly77) that resides in a left-handed helical

conformation, a region of Ramachandran space normally occupied by D-amino acids. Substitution of this glycine by any L-amino acid yields an inactive channel. To explore the possibility that Gly77 is being used as a surrogate D-amino acid, EPL was used to incorporate a D-alanine residue at this position. Single-channel measurements using planar lipid bilayers revealed that the diastereomeric protein indeed retains its function as a K^+ channel. Thus it appears that glycine is conserved at this position in the filter because it is the only natural amino acid that can exist in this region of conformational space.

Summary

In this review we have tried to provide a practical guide to the protein ligation methods EPL and PTS. Although these techniques have been applied to many problems in biochemistry and biophysics, it is important to remember that the generation of modified proteins via ligation can be technically challenging. No two proteins behave the same way, and thus optimal ligation strategies have to be formulated on a case-by-case basis. Nonetheless, with sufficient effort it is usually possible to make the target protein. As we have tried to illustrate, access to semisynthetic proteins has enhanced the utility of many biophysical techniques including fluorescence and NMR spectroscopies. Undoubtedly, there are other areas of biophysics in which protein semisynthesis could make an impact, such as in the quantitative detection of small numbers of specific molecules[74], in making high-density protein microarrays for proteomic studies[75] and in making protein-nanoparticle conjugates, to name a few. Indeed, it is safe to say that we have only scratched the surface in terms of what is possible.

ACKNOWLEDGMENTS

Some of the work discussed in this review was performed by the authors and was supported by the US National Institutes of Health.

COMPETING INTERESTS STATEMENT

The authors declare that they have no competing financial interests.

Published online at http://www.nature.com/naturemethods/
Reprints and permissions information is available online at http://npg.nature.com/reprintsandpermissions/

1. Hahn, M.E. & Muir, T.W. Manipulating proteins with chemistry: a cross-section of chemical biology. *Trends Biochem. Sci.* **30**, 26–34 (2005).
2. Dawson, P.E., Muir, T.W., Clarklewis, I. & Kent, S.B.H. Synthesis of proteins by native chemical ligation. *Science* **266**, 776–779 (1994).
3. Muir, T.W. Semisynthesis of proteins by expressed protein ligation. *Annu. Rev. Biochem.* **72**, 249–289 (2003).
4. Paulus, H. Protein splicing and related forms of protein autoprocessing. *Annu. Rev. Biochem.* **69**, 447–496 (2000).
5. Otomo, T., Ito, N., Kyogoku, Y. & Yamazaki, T. NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. *Biochemistry* **38**, 16040–16044 (1999).
6. Cotton, G.J., Ayers, B., Xu, R. & Muir, T.W. Insertion of a synthetic peptide into a recombinant protein framework: a protein biosensor. *J. Am. Chem. Soc.* **121**, 1100–1101 (1999).
7. Blaschke, U.K., Cotton, G.J. & Muir, T.W. Synthesis of multi-domain proteins using expressed protein ligation: Strategies for segmental isotopic labeling of internal regions. *Tetrahedron* **56**, 9461–9470 (2000).
8. Schwarzer, D. & Cole, P.A. Protein semisynthesis and expressed protein ligation: chasing a protein's tail. *Curr. Opin. Chem. Biol.* **9**, 561–569 (2005).
9. Evans-Jr, T.C., Benner, J. & Xu, M.Q. Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci.* **7**, 2256–2264 (1998).
10. Wu, W., Wood, D.W., Belfort, G., Derbyshire, V. & Belfort, M. Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res.* **30**, 4864–4871 (2002).
11. Yamazaki, T. *et al.* Segmental isotope labeling for protein NMR using peptide splicing. *J. Am. Chem. Soc.* **120**, 5591–5592 (1998).
12. Xu, R., Ayers, B., Cowburn, D. & Muir, T.W. Chemical ligation of folded recombinant proteins: segmental isotopic labeling of domains for NMR studies. *Proc. Natl. Acad. Sci. USA* **96**, 388–393 (1999).
13. Evans, T.C., Jr, Benner, J. & Xu, M-Q. The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins. *J. Biol. Chem.* **274**, 18359–18363 (1999).
14. Iwai, H. & Pluckthun, A. Circular β-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett.* **459**, 166–172 (1999).
15. Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C. & Benkovic, S.J. Production of cyclic peptides and proteins in vivo. *Proc. Natl. Acad. Sci. USA* **96**, 13638–13643 (1999).
16. Camarero, J.A. & Muir, T.W. Biosynthesis of a head-to-tail cyclized protein with improved biological activity. *J. Am. Chem. Soc.* **121**, 5597–5598 (1999).
17. Evans, T.C., Jr *et al.* Protein *trans*-splicing and cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* Species PCC6803. *J. Biol. Chem.* **275**, 9091–9094 (2000).
18. Camarero, J.A., Fushman, D., Cowburn, D. & Muir, T.W. Peptide chemical ligation inside living cells: *in vivo* generation of a circular protein domain. *Bioorg. Med. Chem.* **9**, 2479–2484 (2001).
19. Camarero, J.A. *et al.* Rescuing a destabilized protein fold through backbone cyclization. *J. Mol. Biol.* **308**, 1045–1062 (2001).
20. Iwai, H., Lingel, A. & Pluckthun, A. Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-PfuI intein from *Pyrococcus furiosus*. *J. Biol. Chem.* **276**, 16548–16554 (2001).
21. Scott, C.P., Abel-Santos, E., Jones, A.D. & Benkovic, S.J. Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem. Biol.* **8**, 801–815 (2001).
22. Kinsella, T.M. *et al.* Retrovirally delivered random cyclic peptide libraries yield inhibitors of Interleukin-4 signaling in human B cells. *J. Biol. Chem.* **277**, 37512–37518 (2002).
23. Williams, N.K. *et al.* Stabilization of native protein fold by intein-mediated covalent cyclization. *J. Mol. Biol.* **346**, 1095–1108 (2005).
24. Kent, S.B.H. Chemical synthesis of peptides and proteins. *Annu. Rev. Biochem.* **57**, 957–989 (1988).
25. Villain, M., Vizzavona, J. & Rose, K. Covalent capture: a new tool for the purification of synthetic and recombinant polypeptides. *Chem. Biol.* **8**, 673–679 (2001).
26. Bang, D. & Kent, S.B.H. A one-pot total synthesis of crambin. *Angew. Chem. Int. Edn. Engl.* **43**, 2534–2538 (2004).
27. McCaldon, P. & Argos, P. Oligopeptide biases in protein sequences and their use in predicting protein coding regions in nucleotide sequences. *Proteins* **4**, 99–122 (1988).
28. Valiyaveetil, F.I., MacKinnon, R. & Muir, T.W. Semisynthesis and folding of the potassium channel KcsA. *J. Am. Chem. Soc.* **124**, 9113–9120 (2002).
29. Yan, L.Z. & Dawson, P.E. Synthesis of peptides and proteins without cysteine residues by native chemical ligation combined with desulfurization. *J. Am. Chem. Soc.* **123**, 526–533 (2001).
30. Canne, L.E., Bark, S.J. & Kent, S.B.H. Extending the applicability of native chemical ligation. *J. Am. Chem. Soc.* **118**, 5891–5896 (1996).
31. Offer, J., Boddy, C.N.C. & Dawson, P.E. Extending synthetic access to proteins with a removable acyl transfer auxiliary. *J. Am. Chem. Soc.* **124**, 4642–4646 (2002).
32. Saxon, E., Armstrong, J.I. & Bertozzi, C.R.A. "Traceless" Staudinger ligation for the chemoselective synthesis of amide bonds. *Org. Lett.* **2**, 2141–2143 (2000).
33. Nilsson, B.L., Kiessling, L.L. & Raines, R.T. Staudinger ligation: a peptide from a thioester and azide. *Org. Lett.* **2**, 1939–1941 (2000).
34. Chong, S., Williams, K.S., Wotkowicz, C. & Xu, M.Q. Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. *J. Biol. Chem.* **273**, 10567–10577 (1998).
35. Southworth, M.W., Amaya, K., Evans, T.C., Xu, M.Q. & Perler, F.B. Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *Biotechniques* **27**, 110–120 (1999).
36. Xu, M.Q., Paulus, H. & Chong, S. Fusions to self-splicing inteins for protein purification. *Methods Enzymol.* **326**, 376–418 (2000).
37. Hackeng, T.M., Griffin, J.H. & Dawson, P.E. Protein synthesis by native chemical ligation: expanded scope by using straightforward methodology. *Proc. Natl. Acad. Sci. USA* **96**, 10068–10073 (1999).
38. Botti, P., Villain, M., Manganiello, S. & Gaertner, H. Native chemical ligation through in situ O to S acyl shift. *Org. Lett.* **6**, 4861–4864 (2004).
39. Ohta, Y. *et al.* Cysteine-derived S-protected oxazolidinones: potential chemical devices for the preparation of peptide thioesters. *Org. Lett.* **8**, 467–470 (2006).
40. Muir, T.W., Sondhi, D. & Cole, P.A. Expressed protein ligation: a general method for protein engineering. *Proc. Natl. Acad. Sci. USA* **95**, 6705–6710 (1998).
41. Camarero, J.A. *et al.* Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR. *Proc. Natl. Acad. Sci. USA* **99**, 8536–8541 (2002).
42. Erlanson, D.A., Chytil, M. & Verdine, G.L. The leucine zipper domain controls the orientation of AP-1 in the NFAT center dot AP-1 center dot DNA complex. *Chem. Biol.* **3**, 981–991 (1996).
43. Tolbert, T.J. & Wong, C.H. New methods for proteomic research: preparation of proteins with N-terminal cysteines for labeling and conjugation. *Angew. Chem.*

Int. Edn. Engl. **41**, 2171–2174 (2002).
44. Derbyshire, V. *et al.* Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. *Proc. Natl. Acad. Sci. USA* **94**, 11466–11471 (1997).
45. Mathys, S. *et al.* Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* **231**, 1–13 (1999).
46. Wood, D.W., Wu, W., Belfort, G., Derbyshire, V. & Belfort, M. A genetic system yields self-cleaving inteins for bioseparations. *Nat. Biotechnol.* **17**, 889–892 (1999).
47. Romanelli, A., Shekhtman, A., Cowburn, D. & Muir, T.W. Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proc. Natl. Acad. Sci. USA* **101**, 6397–6402 (2004).
48. Muralidharan, V. *et al.* Domain-specific incorporation of noninvasive optical probes into recombinant proteins. *J. Am. Chem. Soc.* **126**, 14004–14012 (2004).
49. Beligere, G.S. & Dawson, P.E. Conformationally assisted protein ligation using C-terminal thioester peptides. *J. Am. Chem. Soc.* **121**, 6332–6333 (1999).
50. Gentle, I.E., De Souza, D.P. & Baca, M. Direct production of proteins with N-terminal cysteine for site-specific conjugation. *Bioconjug. Chem.* **15**, 658–663 (2004).
51. Shi, J.X. & Muir, T.W. Development of a tandem protein trans-splicing system based on native and engineered split inteins. *J. Am. Chem. Soc.* **127**, 6198–6206 (2005).
52. Southworth, M.W. *et al.* Control of protein splicing by intein fragment reassembly. *EMBO J.* **17**, 918–926 (1998).
53. Mills, K.V., Lew, B.M., Jiang, S.-q. & Paulus, H. Protein splicing in *trans* by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein. *Proc. Natl. Acad. Sci. USA* **95**, 3543–3548 (1998).
54. Sun, W., Yang, J. & Liu, X-Q. Synthetic two-piece and three-piece split inteins for protein *trans*-splicing. *J. Biol. Chem.* **279**, 35281–35286 (2004).
55. Wu, H., Hu, Z. & Liu, X-Q. Protein *trans*-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc. Natl. Acad. Sci. USA* **95**, 9226–9231 (1998).
56. Zuger, S. & Iwai, H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat. Biotechnol.* **23**, 736–740 (2005).
57. Mootz, H.D. & Muir, T.W. Protein splicing triggered by a small molecule. *J. Am. Chem. Soc.* **124**, 9044–9045 (2002).
58. Ozawa, T., Kaihara, A., Sato, M., Tachihara, K. & Umezawa, Y. Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing. *Anal. Chem.* **73**, 2516–2521 (2001).
59. Giriat, I. & Muir, T.W. Protein semisynthesis in living cells. *J. Am. Chem. Soc.* **125**, 7180–7181 (2003).
60. Chin, H.G. *et al.* Protein *trans*-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes. *Proc. Natl. Acad. Sci. USA* **100**, 4510–4515 (2003).
61. Yang, J., Fox, G.C., Jr & Henry-Smith, T.V. Intein-mediated assembly of a functional β-glucuronidase in transgenic plants. *Proc. Natl. Acad. Sci. USA* **100**, 3513–3518 (2003).
62. Algire, M.A., Maag, D. & Lorsch, J.R. Pi release from eIF2, not GTP hydrolysis, is the step controlled by start-site selection during eukaryotic translation initiation. *Mol. Cell* **20**, 251–262 (2005).
63. Maag, D., Fekete, C.A., Gryczynski, Z. & Lorsch, J.R. A conformational change in the eukaryotic translation preinitiation complex and release of eIF1 signal recognition of the start codon. *Mol. Cell* **17**, 265–275 (2005).
64. Anderson, L.L., Marshall, G.R., Crocker, E., Smith, S.O. & Baranski, T.J. Motion of carboxyl terminus of Gα is restricted upon G protein activation. *J. Biol. Chem.* **280**, 31019–31026 (2005).
65. Yagi, H., Tsujimoto, T., Yamazaki, T., Yoshida, M. & Akutsu, H. Conformational change of H+-ATPase β-monomer revealed on segmental isotope labeling NMR spectroscopy. *J. Am. Chem. Soc.* **126**, 16632–16638 (2004).
66. Tatulian, S.A., Qin, S., Pande, A.H. & He, X.M. Positioning membrane proteins by novel protein engineering and biophysical approaches. *J. Mol. Biol.* **351**, 939–947 (2005).
67. Wu, J-W. *et al.* Crystal Structure of a Phosphorylated Smad2: Recognition of phosphoserine by the MH2 domain and insights on Smad function in TGF-β signaling. *Mol. Cell* **8**, 1277–1289 (2001).
68. Qin, B.Y., Lam, S.S., Correia, J.J. & Lin, K. Smad3 allostery links TGF-β receptor kinase activation to transcriptional control. *Genes Dev.* **16**, 1950–1963 (2002).
69. Chacko, B.M. *et al.* Structural basis of heteromeric Smad protein assembly in TGF-β signaling. *Mol. Cell* **15**, 813–823 (2004).
70. Rak, A. *et al.* Structure of Rab GDP-dissociation inhibitor in complex with prenylated YPT1 GTPase. *Science* **302**, 646–650 (2003).
71. Clayton, D. *et al.* Total chemical synthesis and electrophysiological characterization of mechanosensitive channels from *Escherichia coli* and *Mycobacterium tuberculosis*. *Proc. Natl. Acad. Sci. USA* **101**, 4764–4769 (2004).
72. Valiyaveetil, F.I., Sekedat, M., MacKinnon, R. & Muir, T.W. Glycine as a D-amino acid surrogate in the K+-selectivity filter. *Proc. Natl. Acad. Sci. USA* **101**, 17045–17049 (2004).
73. Doyle, D.A. *et al.* The structure of the potassium channel: molecular basis of K+ conduction and selectivity. *Science* **280**, 69–77 (1998).
74. Burbulis, I., Yamaguchi, K., Gordon, A., Carlson, R. & Brent, R. Using protein-DNA chimeras to detect and count small numbers of molecules. *Nat. Methods* **2**, 31–37 (2005).
75. Lovrinovic, M. *et al.* Synthesis of protein-nucleic acid conjugates by expressed protein ligation. *Chem. Commun.*822–823 (2003).
76. Chong, S. *et al.* Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* **192**, 271–281 (1997).
77. Shingledecker, K., Jiang, S.Q. & Paulus, H. Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: design of a minimal intein and of a *trans*-splicing system involving two intein fragments. *Gene* **207**, 187–195 (1998).
78. Evans, T.C., Jr, Benner, J. & Xu, M-Q. The *in vitro* ligation of bacterially expressed proteins using an intein from *Methanobacterium thermoautotrophicum*. *J. Biol. Chem.* **274**, 3923–3926 (1999).
79. Camarero, J.A., Cotton, G.J., Adeva, A. & Muir, T.W. Chemical ligation of unprotected peptides directly from a solid support. *J. Pept. Res.* **51**, 303–316 (1998).
80. Shin, Y. *et al.* Fmoc-based synthesis of peptide-(α)thioesters: Application to the total chemical synthesis of a glycoprotein by native chemical ligation. *J. Am. Chem. Soc.* **121**, 11684–11689 (1999).
81. Ingenito, R., Bianchi, E., Fattori, D. & Pessi, A. Solid phase synthesis of peptide C-terminal thioesters by Fmoc/t-Bu chemistry. *J. Am. Chem. Soc.* **121**, 11369–11374 (1999).
82. Alsina, J., Yokum, T.S., Albericio, F. & Barany, G. Backbone amide linker (BAL) strategy for N(α)-9-fluorenylmethoxycarbonyl (Fmoc) solid-phase synthesis of unprotected peptide *p*-nitroanilides and thioesters(1). *J. Org. Chem.* **64**, 8761–8769 (1999).
83. Swinnen, D. & Hilvert, D. Facile, Fmoc-compatible solid-phase synthesis of peptide C-terminal thioesters. *Org. Lett.* **2**, 2439–2442 (2000).
84. Camarero, J.A., Hackel, B.J., deYoreo, J.J. & Mitchell, A.R. Fmoc-based synthesis of peptide α-Thioesters using an aryl hydrazine support. *J. Org. Chem.* **69**, 4145–4151 (2004).
85. Perler, F.B. InBase, the intein database. *Nucleic Acids. Res.* **30**, 383–384 (2002).
86. Xu, M.Q. & Perler, F.B. The mechanism of protein splicing and its modulation by mutation. *EMBO J.* **15**, 5146–5153 (1996).
87. Chong, S. *et al.* Utilizing the C-terminal cleavage activity of a protein splicing element to purify recombinant proteins in a single chromatographic step. *Nucleic Acids Res.* **26**, 5109–5115 (1998).
88. Banki, M.R., Feng, L. & Wood, D.W. Simple bioseparations using self-cleaving elastin-like polypeptide tags. *Nat. Methods* **2**, 659–662 (2005).
89. Ge, X. *et al.* Self-cleavable stimulus responsive tags for protein purification without chromatography. *J. Am. Chem. Soc.* **127**, 11228–11229 (2005).
90. Liu, X-Q. & Yang, J. Split *dnaE* genes encoding multiple novel inteins in *Trichodesmium erythraeum*. *J. Biol. Chem.* **278**, 26315–26318 (2003).
91. Nichols, N.M. & Evans, T.C., Jr. Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split Ssp DnaE intein. *Biochemistry* **43**, 10265–10276 (2004).
92. Schnolzer, M. & Kent, S.B. Constructing proteins by dovetailing unprotected synthetic peptides: backbone-engineered HIV protease. *Science* **256**, 221–225 (1992).
93. Gaertner, H.F. *et al.* Construction of protein analogs by site-specific Condensation of unprotected fragments. *Bioconjug. Chem.* **3**, 262–268 (1992).
94. Hang, H.C. & Bertozzi, C.R. Chemoselective approaches to glycoprotein assembly. *Acc. Chem. Res.* **34**, 727–736 (2001).
95. Kolb, H.C. & Sharpless, K.B. The growing impact of click chemistry on drug discovery. *Drug Discov. Today* **8**, 1128 (2003).
96. Bianchi, E., Ingenito, R., Simon, R.J. & Pessi, A. Engineering and chemical synthesis of a transmembrane protein: the HCV protease cofactor protein NS4A. *J. Am. Chem. Soc.* **121**, 7698–7699 (1999).
97. Hunter, C.L. & Kochendoerfer, G.G. Native chemical ligation of hydropobic peptides in lipid bilayer systems. *Bioconjug. Chem.* **15**, 437–440 (2004).
98. Kochendoerfer, G.G. *et al.* Total chemical synthesis of the integral membrane protein influenza A virus M2: role of its C-terminal domain in tetramer assembly. *Biochemistry* **38**, 11905–11913 (1999).
99. Yeo, D.S. *et al.* Cell-permeable small molecule probes for site-specific labeling of proteins. *Chem. Commun.*2870–2871 (2003).
100. Dawson, P.E., Churchill, M.J., Ghadiri, M.R. & Kent, S.B.H. Modulation of reactivity in native chemical ligation through the use of thiol additives. *J. Am. Chem. Soc.* **119**, 4325–4329 (1997).