BioMol Concepts, Vol. 2 (2011), pp. 183–198 • Copyright © by Walter de Gruyter • Berlin • Boston. DOI 10.1515/BMC.2011.014

Review

Protein *trans*-splicing as a protein ligation tool to study protein structure and function

A. Sesilja Aranko and Gerrit Volkmann[a,*]

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, P.O. Box 65, FIN-00014 Helsinki, Finland

*Corresponding author
e-mail: gerrit.volkmann@uni-muenster.de

Abstract

Protein *trans*-splicing (PTS) exerted by split inteins is a protein ligation reaction which enables overcoming the barriers of conventional heterologous protein production. We provide an overview of the current state-of-the-art in split intein engineering, as well as the achievements of PTS technology in the realm of protein structure-function analyses, including incorporation of natural and artificial protein modifications, controllable protein reconstitution, segmental isotope labeling and protein cyclization. We further discuss factors crucial for the successful implementation of PTS in these protein engineering approaches, and speculate on necessary future endeavours to make PTS a universally applicable protein ligation tool.

Keywords: junction sequence effect; protein ligation; protein structure-function relationships; protein semisynthesis; protein *trans*-splicing; split intein.

Introduction

Understanding the structure-function relationship of proteins is crucial for both basic research and applications in biotechnology and biomedicine. Recombinant DNA technology has advanced the expression of proteins in heterologous systems like *Escherichia coli* (*E. coli*), allowing us to investigate by site-directed mutagenesis the contributions of individual amino acid residues to protein activity or stability. Over-expression of foreign genes in bacteria is a convenient and inexpensive way to obtain recombinant proteins but may in some cases suffer from poor yields, misfolding, and lack of post-translational modifications. Furthermore, recombinant protein production is usually limited to the common 20 amino acids. Therefore, considerable efforts have been made for expanding the functionalities of proteins by adding artificial chemical groups through site-specific modification or biosynthetic incorporation.

Protein ligation allows us to combine proteins or peptides from versatile sources. It has the potential to become a powerful technology to overcome the limitations of recombinant protein production, and could open new avenues to engineer proteins in site- and region-specific manners. However, despite the wide range of potential applications, no method for protein ligation that would be universally applicable has been described so far. Protein *trans*-splicing (PTS), a ligation reaction performed by split inteins, is a convenient and easy way to perform protein ligation, thus making it a promising candidate for a wide range of protein ligation applications. Here, we review the PTS ligation approach and its use for structural and functional studies of proteins. The basic four-step mechanism of PTS is currently well understood and will be described, along with the most interesting applications of PTS, including protein semi-synthesis, fluorescent labeling, and segmental isotope labeling of proteins. Furthermore, practical issues including limitations and advantages of the use of PTS are discussed. Finally, we speculate on potential PTS-mediated applications in the future.

Protein *trans*-splicing for protein ligation

Inteins

Protein splicing is a post-translational modification catalyzed by an intervening polypeptide sequence termed 'intein' (from *int*ernal prot*ein*), in which the intein catalyzes self-excision from the host protein and concomitantly ligates the two flanking polypeptide chains termed 'exteins' (from *ext*ernal prot*ein*) with a peptide bond (Figure 1A) (1). Because of the analogy to RNA splicing, inteins can be regarded as protein introns. Some inteins are bifunctional, i.e. they contain a homing endonuclease function as an insertion within the protein splicing domain (2). Endonuclease homing and the protein splicing activity of inteins are structurally and functionally independent (3–7), and so-called mini-inteins can be engineered by removing the endonuclease domain without negatively affecting protein splicing (8–10). Mini-inteins also exist in nature [first described in (11)], and are the most relevant in intein-related applications due to their small size ($\sim$130–160 residues compared to the $>$200 residues of bifunctional inteins).

Up to now, more than 550 predicted inteins have been compiled in InBase, the intein database (12), based on sequence homology, and almost 100 of them have been examined for their protein splicing activity. Inteins have been found in all three domains of life, albeit only in unicellular

[a]Present address: Institut für Biochemie, Westfälische Wilhelms-Universität, Wilhelm-Klemm-Str. 2, D-48149 Münster, Germany.

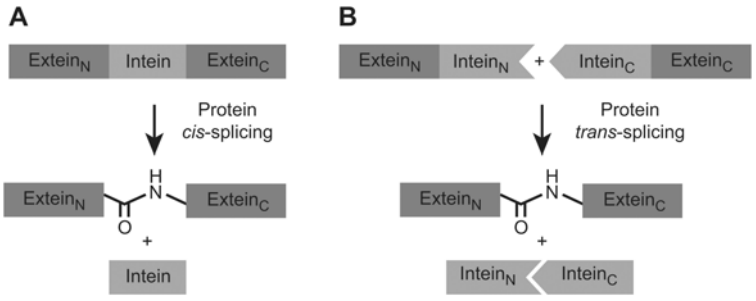

Figure 1 Schematic representation of (A) protein *cis*-splicing as exerted by bifunctional and mini-inteins, and (B) protein *trans*-splicing as exerted by split inteins.

organisms. They most often intervene highly conserved proteins that are essential for survival of the host organism, and are believed to be of ancient origin (13). Until now, apart from the single-turnover protein ligation reaction, no other biological function has been described for the protein splicing domain of mini-inteins, and hence they are often considered to be mere parasitic genetic elements.

Protein splicing is an autocatalytic reaction induced by protein folding, and therefore requires neither cofactors nor energy equivalents. Inteins present several blocks of consensus sequence, four of which participate in protein splicing (blocks A, B, F, G) (14, 15). The homing endonuclease of bifunctional inteins also has conserved sequence motifs that are specific for a certain type of endonuclease (not discussed here). The most conserved residues of the intein protein splicing domain are the first residue, a nucleophilic Ser or Cys (block A), a ThrXXHis motif (block B) located usually between residues 70–105, an Asp and His residue in block F, a His-Asn pair at the C-terminus of the intein as well as a nucleophilic residue (Cys, Ser or Thr) at the beginning of the C-terminal extein (both in block G) (Figure 2). Exceptions to these rules have been identified (16, 17), most remarkable of which are the inteins lacking the conserved Ser or Cys as the first residue. These inteins were grouped as class 2 or 3 inteins (depending on the type of amino acid at intein position 1 and the concurrent deviations from the

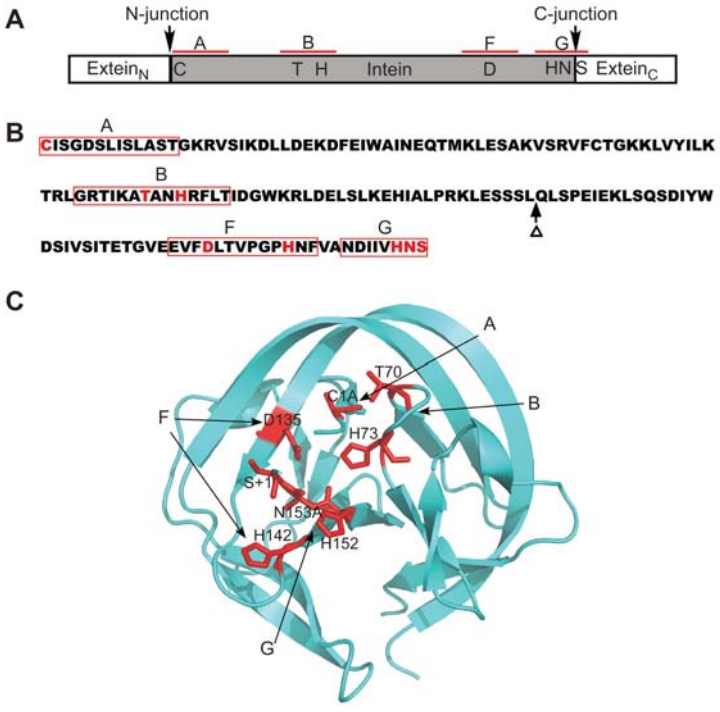

Figure 2 Conserved blocks in intein sequences.
(A) Schematic view of an intein, the relative locations and lengths of sequence blocks/motifs A, B, F and G; highly conserved residues are indicated underneath the red motif bars. Variations in these residues are possible. (B) Amino acid sequence of the *Ssp* DnaB mini-intein with sequence blocks and conserved residues highlighted. The Δ symbol indicates the deletion point of the *Ssp* DnaB homing endonuclease domain. (C) Crystal structure of the *Ssp* DnaB mini-intein [PDB code: 1MI8; (117)] with the location of sequence blocks indicated by arrows and position of the conserved residues shown in B highlighted by red, ball-and-stick representations.

protein splicing mechanism, see below), while inteins which do begin with a nucleophilic residue represent class 1 inteins (18, 19).

Split inteins

Intriguingly, protein splicing can also occur in *trans* by inteins, which are split into two halves and spontaneously reassemble into an active form (Figure 1B). Split inteins are of special interest as they are capable of joining two separate polypeptide chains with a covalent peptide bond, and could thus serve as protein ligases. Protein *trans*-splicing (PTS) was first demonstrated with artificial split inteins derived from bifunctional inteins found in the DNA polymerase of *Pyrococcus species* GB-D (*Psp* Pol-1) (20) and the *Mycobacterium tuberculosis* RecA recombinase (*Mtu* RecA) (21) that were split either with the endonuclease domain present (*Psp* Pol-1) or subsequent to its removal (*Mtu* RecA) (22). The first natural split intein (DnaE intein of *Synechocystis* sp. PCC6803; *Ssp* DnaE) was discovered in the same year (23), and proven to be active in protein *trans*-splicing (24). Most natural split inteins described thus far are present in the *dnaE* gene of cyanobacteria with a common domain organization (N-intein: ~125 aa; C-intein: ~35 aa) and high sequence homology (25, 26), however, one archaeal split intein has been identified (27) and several novel split inteins have been suggested from a recent metagenomic study (28).

In addition to the early artificial split inteins described above, some unusual split inteins have been created with N- and C-terminal fragments that are considerably shorter than the ones found, for example in the DnaE split inteins. For instance, the DnaB intein of *Synechocystis* sp. PCC6803 (*Ssp* DnaB) lacking its endonuclease domain can be split just after the 11th residue (referred to in this review as split site 'N11') with reasonable activity and solubility remaining (29). On the other hand, the DnaE inteins from *Nostoc punctiforme* (*Npu* DnaE) and *Synechocystis* sp. PCC6803 (*Ssp* DnaE) have been split 14 or 15 residues upstream of their C-terminal end (sites 'C14' and 'C15'), respectively, resulting in moderate (*Ssp* DnaE) and good (*Npu* DnaE) ligation yields (30). Even split inteins with a C-intein as short as six residues (site 'C6') have been shown to efficiently catalyze protein *trans*-splicing, derived from an endonuclease-free version of the GyrB intein from *Synechocystis* sp. PCC6803 (*Ssp* GyrB) (31) and a *cis*-splicing version of *Npu* DnaE (32).

Several strategies exist for creating artificial split inteins. Since all inteins share a conserved 'horseshoe'-shaped structure common with hedgehog proteins (therefore referred to as the HINT (Hedgehog, Intein) fold) (33), functional split sites can often be selected based on structural homology using known split sites as a reference [splitting *Npu* DnaE based on high homology with *Ssp* DnaE; (30)]. Split inteins have also been engineered by systematically introducing split sites in loop regions of the HINT fold, which provided an opportunity to test different lengths of split intein fragments (30). The experimental data revealed that while all functional split sites were located in loop regions, not all loops were suitable for splitting (Figure 3). Furthermore, although all known intein structures are very alike, not every intein can

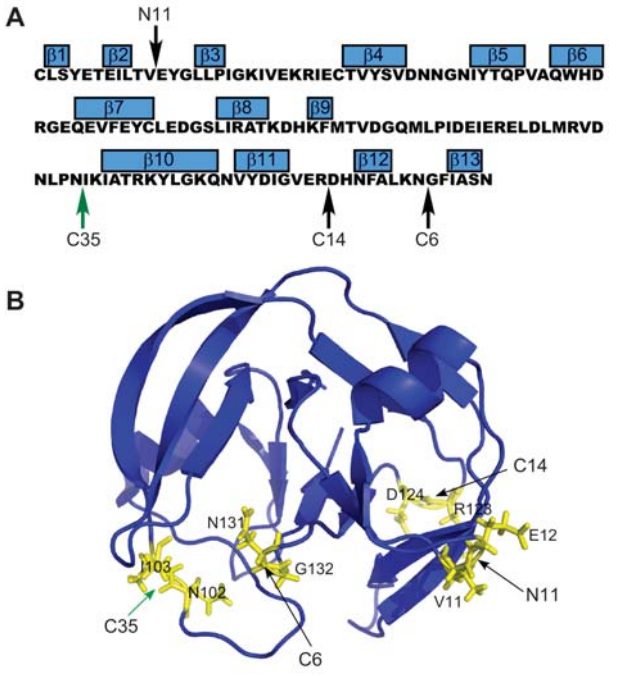

Figure 3 Split sites in the Npu DnaE intein.
(A) Amino acid sequence of a *cis*-splicing version of *Npu* DnaE with β-strands from the nuclear magnetic resonance (NMR) structure and split sites indicated above the sequence. Split site 'C35' (highlighted in green) refers to the natural split intein. (B) NMR structure of *Npu* DnaE [PDB code: 2KEQ; (32)] with the split site positions indicated with arrows. All functional split sites are located in loops. The native and artificial split sites are indicated with green and black arrows, respectively.

be functionally split at a site that works for others; even the highly homologous *Npu* and *Ssp* DnaE inteins behave differently when split at artificial sites (30, 32). The main problems that arise when creating artificial split inteins are either the instability of the split intein fragments and/or their poor association. To assist the difficulty in defining appropriate split sites, it was recently demonstrated that nuclear magnetic resonance (NMR) spectroscopy relaxation data can be used to predict functional split sites (32). The NMR structure of the *Npu* DnaE intein in its *cis*-splicing form shows the common HINT fold consisting of 13 β-strands (Figure 3B). ^{15}N nuclear spin relaxation measurements revealed high internal backbone flexibility in several loop regions that connect the β-strands, and led to the design of the C6 split intein based on the conformational exchange observed at residue Gly132 (located between β-strands 12 and 13, Figure 3).

Despite their lower efficiency compared to natural split inteins, in some cases, artificially split inteins are of considerable interest. Split inteins with extremely short N- or C-terminal halves are much more practical for approaches such as peptide-based protein semi-synthesis (see below). Furthermore, artificially split inteins may provide a benefit for protein ligation of specific protein sequences, as was shown for the Src homology 3 (SH3) domain, which was successfully ligated only with the *Npu* DnaE intein artificially split

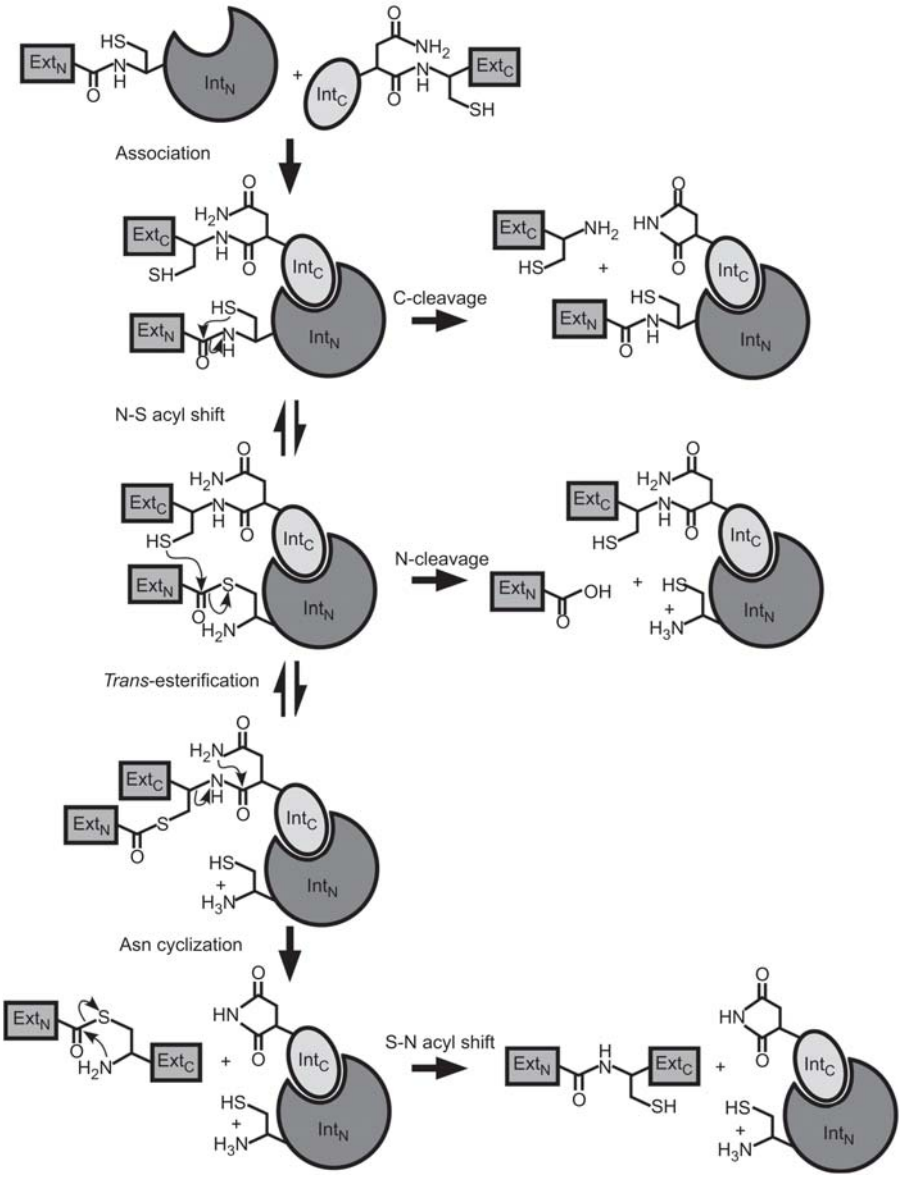

Figure 4 Schematic illustration of the mechanism of protein *trans*-splicing and cleavage reactions.
Ext_N and Ext_C stand for N- and C-exteins, Int_N and Int_C stand for N- and C-terminal fragments of a split intein. Protein *cis*-splicing of class 1 inteins proceeds similarly but no association step is required. See text for further details.

at site C14 (*Npu* $DnaE_{\Delta C14}$) but not when using the naturally split intein (*Npu* $DnaE_{\Delta C35}$), probably due to the smaller size of the C-terminal fragment of *Npu* $DnaE_{\Delta C14}$ (30). Lastly, artificial split inteins can be a solution to intein cross-reactivity: because the well characterized and most widely used naturally split DnaE inteins from cyanobacteria share high sequence homology, their N- and C-terminal split intein fragments cross-splice with each other (26), thus preventing their utility in multi-fragment ligation reactions. Use of a natural split intein in combination with an artificial derivative thereof, which has the split site moved up- or downstream of the natural site, can make multi-fragment ligation possible, as was proven for the combination *Npu* $DnaE_{\Delta C35}$/*Npu* $DnaE_{\Delta C14}$ (34).

Mechanism of protein *trans*-splicing

The mechanism of protein *trans*-splicing has been extensively studied, and the principle chemical reactions are somewhat understood (35). PTS proceeds in four concerted steps (Figure 4), prior to which the two halves of the split intein spontaneously associate (36). In the initial step, the first residue of the N-intein (Ser/Cys) undergoes an N-X acyl-shift with the preceding peptide bond (X denoting either oxygen or sulfur). In the second step, the first residue of the C-extein, the '+1 residue' (either Ser, Cys or Thr, depending on the intein), attacks the new C–X-bond at the N-terminal splice junction in a *trans*-esterification reaction. The third step is cyclization of the last intein residue (most often Asn), which

cleaves the intein from the ligation product. Finally, a peptide bond is spontaneously formed by an X-N acyl shift at the ligation junction.

In addition to the conventional splicing pathway, two alternative mechanisms have been identified for *cis*-splicing inteins that present an Ala at their first position instead of the much more conserved Cys or Ser. Inteins belonging to class 2 'skip' the first step of the conventional splicing reaction (N-X acyl shift) and directly attack the N-terminal scissile peptide bond with the nucleophilic +1 residue (19), whereas in the class 3 pathway, a Cys residue present in block F activates the N-terminal splice junction by nucleophilic displacement (18). The remaining steps in class 2 and 3 splicing proceed according to the conventional protein-splicing pathway.

Both protein *cis*- and *trans*-splicing can be accompanied by two off-pathway reactions, N- and C-terminal cleavage (Figure 4). These cleavage reactions occur when the catalytic center of the intein is slightly perturbed due to, for example foreign extein sequences. The result is the separation of either the N- or C-extein from the precursor protein prior to completion of the splicing reaction. Both cleavage reactions can also selectively be achieved through site-specific mutations. Mutation of the N-nucleophile in class 1 inteins to Ala blocks the initial N-X acyl shift (and subsequently, *trans*-esterification), but still leaves room for Asn cyclization (C-cleavage) (37, 38). Conversely, mutating the C-terminal Asn to Ala stalls the protein splicing reaction before step 3, which can result in N-cleavage by hydrolysis of the labile ester bond formed in steps 1 or 2 (8, 39, 40). Fortuitously, several inteins bearing such mutations can be induced for either cleavage mechanism at will, for example by adding a thiol reagent for N-cleavage or with a change in pH or temperature for C-cleavage [for a review, see (41)]. These observations formed the basis for commercially available protein purification systems using self-cleavable intein tags (42). Furthermore, the thiol-inducible N-cleavage reaction can also be harnessed for site-specific C-terminal protein modification (43, 44).

Alternative methods for protein ligation

A practical alternative to recombinant protein production is chemical synthesis. Modern automated solid-phase peptide synthesizer can routinely produce peptides of up to 30–50 residues in length, depending on their sequences. In theory, chemical synthesis allows to incorporate any chemical group into proteins at any position, although it could require considerable workforces to prepare the peptide if the desired unnatural amino acid is not commercially available. The limitation of size in solid-phase peptide synthesis (SPPS) was alleviated by the invention of native chemical ligation (NCL) (45). The chemo-selective reaction of NCL allows ligation of two unprotected peptides bearing a C-terminal α-thioester on the N-terminal reactant and an N-terminal cysteine on the C-terminal reactant. To date, more than 100 proteins have been synthesized by the combination of SPPS with NCL, and even the production of several proteins with up to 200 amino acid residues has been demonstrated (46). NCL in combination with other peptide ligation techniques, such as ΨPro-ligation (47) can further be used to connect three (or more) fragments together (48).

The potential of NCL was significantly expanded when it was realized that recombinant proteins from bacterial expression systems could be combined with NCL. Production of recombinant proteins is much more cost-effective and less labor intensive than SPPS, and offers the great advantage to synthesize polypeptides with sizes of well over 1000 aa. The combination of NCL and recombinant proteins is often referred to as expressed protein ligation (EPL), when at least one of the components is chemically synthesized, or intein-mediated protein ligation (IPL), when both components are recombinantly produced. Recombinant proteins for EPL/IPL can conveniently be afforded using the above mentioned inducible intein cleavage reactions (49). Using the thiol compound 2-mercaptoethanesulfonic acid (MESNA) for N-cleavage appends a highly reactive α-thioester to the C-terminus of the recombinant protein. Providing a Cys residue immediately downstream of the Asn residue at the end of the intein will result in a protein with an N-terminal Cys after C-cleavage. The general disadvantage of EPL/IPL is that successful ligation requires high reactant concentrations (mM range), thus limiting the application if the recombinant reactant is difficult to express and/or purify in high yields, or SPPS of the peptide reactant is otherwise cumbersome. Furthermore, multiple ligation reactions require tedious purification steps that may lower the final yield.

A conceptually simpler approach for peptide ligation is 'reversed proteolysis'. The idea of a protein ligase has been demonstrated with an engineered enzyme named subtiligase, which was derived from the protease subtilisin BPN' (50). Although subtiligase has been successfully used for the total synthesis of ribonuclease A (51) and a few other proteins (52), this ligase has never become widely used. Recently, sortase A from *Staphylococcus aureus*, which covalently attaches proteins to the cell wall, has been used for ligation of two polypeptide chains [for a review, see (53)]. Though promising alternatives, the slow reaction kinetics and necessity to use excess enzyme represent a bottleneck for the use of subtiligase and sortase as protein ligases in a wide range of applications.

Protein *trans*-splicing in studies of protein structure and function

Site-specific modifications

Incorporation of unnatural components such as fluorescent dyes into proteins is an excellent tool to understand their function. Attaching a chemical probe via a traditional residue-specific reaction is a well-established method to post-translationally introduce a modification into a protein. For example, cysteine is often utilized for residue-specific labeling of proteins because of its reactive thiol group, which can easily be modified with haloacetamido- or maleimide-functionalized probes. Although cysteine is one of the least abun-

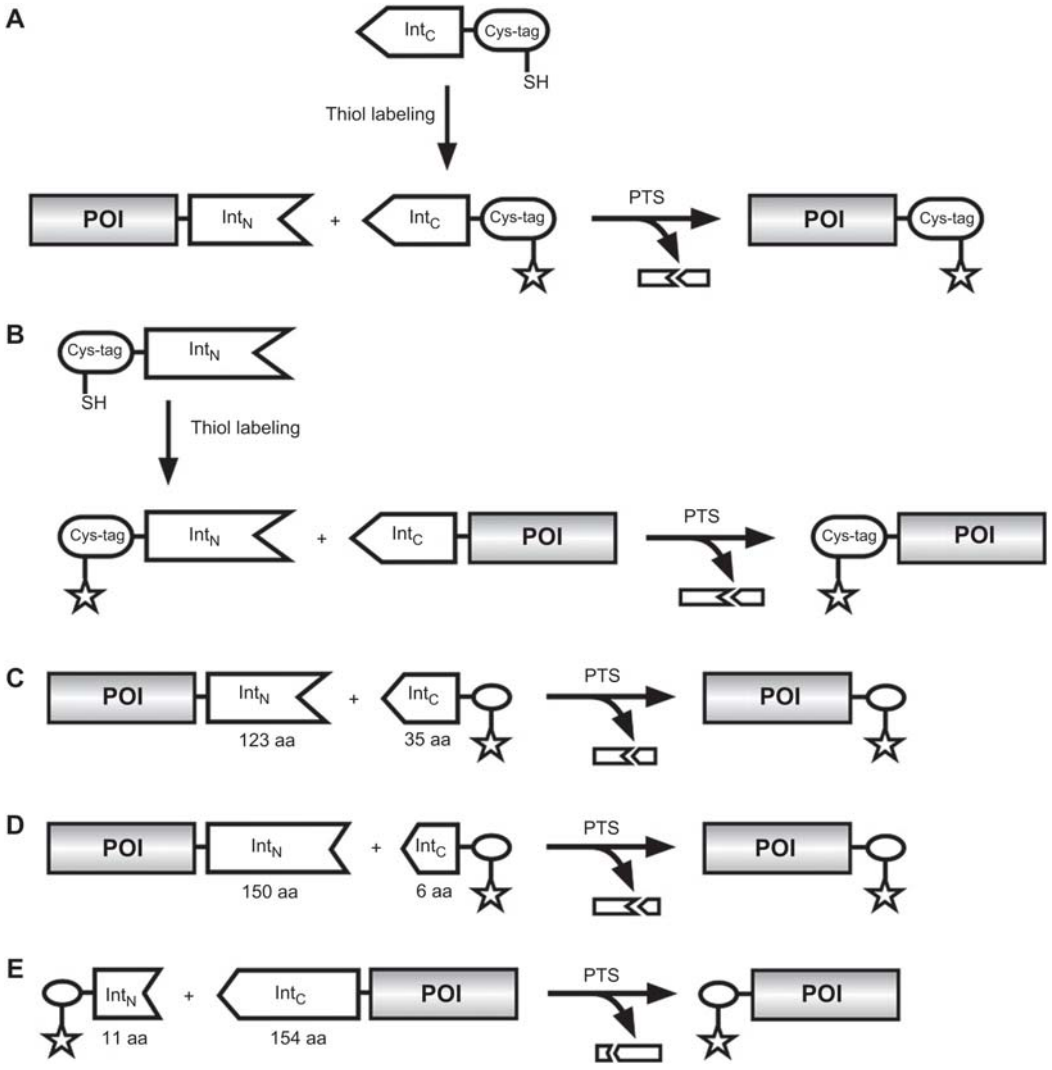

Figure 5 Current strategies for split intein-mediated site-specific protein modification.
Labels/probes are indicated with a star. (A) The original Cys-tag approach designed for C-terminal modification. The Int_C/Cys-tag protein is site-specifically modified on a single Cys residue located in the Cys-tag with a thiol-reactive probe. The split inteins used in the approach were engineered from *Ssp* DnaB (Int_N: 105 aa, Int_C: 48 aa) and *Mxe* GyrA (Int_N: 119 aa, Int_C: 79 aa). The protein of interest (POI) may contain free Cys residues or disulfide bonds (not shown). (B) Deviation of the Cys-tag approach for N-terminal modification. The intein used was the *Psp*-GBD Pol intein (Int_N: 440 aa, Int_C: 97 aa), in which a Cys351Ala mutation was introduced to ensure that thiol-labeling would only occur in the Cys-tag. (C) C-terminal modification catalyzed by the natural *Ssp* DnaE split intein (sizes of Int_N and Int_C are indicated). (D) C-terminal modification catalyzed by the engineered *Ssp* GyrB C6 split intein, which has a much smaller Int_C than the *Ssp* DnaE intein (sizes of Int_N and Int_C are indicated). (E) N-terminal modification catalyzed by the engineered *Ssp* DnaB N11 split intein, which has a very short Int_N (sizes of Int_N and Int_C are indicated). In C-E, the labels were incorporated into Int_C or Int_N during solid-phase peptide synthesis.

dant amino acids in the protein realm, a protein often contains more than one Cys, which complicates the site-specific labeling of only one thiol group. Alternatively, a probe can be introduced during chemical synthesis of the protein, but this approach is hindered by the limited size of the synthesized molecule, and the required equipment and resources (see above).

To overcome this common obstacle in chemical modification, Kurpiers and Mootz have introduced an approach to attach a short cysteine containing peptide to the C-terminus of a protein of interest (POI) by PTS using artificially split derivatives of either the *Ssp* DnaB (54) or the *Mycobacterium xenopi* GyrA intein (*Mxe* GyrA) (55) (Figure 5A). The C-intein fragments were recombinantly produced in fusion with a short extein sequence containing a single Cys residue, and were labeled with probes such as fluorescein-iodoacetamide and polyethylene glycol (PEG)5000-maleimide. Importantly, the inteins used had a Ser (*Ssp* DnaB) or Thr (*Mxe* GyrA) as their C-terminal reactive nucleophilic residue (at position +1), and were thus not disturbed by the labeling reaction, a fact that precludes the use of split inteins with an obligate Cys+1 residue for such an approach. The gentle

reaction conditions of this 'CysTag' approach preserved the function of the modified proteins (β-lactamase, tyrocidine synthetase module TycA), could proceed in a complex protein environment (*E. coli* cell lysate), and the reaction employing the *Mxe* GyrA split intein could even be used in a dialysis procedure to efficiently label proteins that were deposited into inclusion bodies upon expression in *E. coli*.

As an expansion of their CysTag method, which focused on C-terminal protein modification, Mootz and co-workers also devised a scheme for the introduction of modifications into an N-terminal protein region (56) (Figure 5B). This study focused on the *E. coli* porin OmpF, a non-selective, β-barrel outer membrane protein. To modulate OmpF pore function, the authors used the artificial *Psp*-GBD Pol split intein in combination with maleimide chemistry to introduce a bulky crown ether molecule into the pore of the β-barrel, which resulted in altered conductance. Importantly, the *Psp*-GBD Pol intein is able to perform PTS in buffer containing high concentrations of denaturants (e.g. 6 M urea) (20), which was a necessity for the successful modification due to the aggregation and insolubility of the OmpF membrane protein fragments, thus potentially paving the way for the site-specific modification of membrane proteins in general.

An alternative to the CysTag approach, which is based solely on recombinant protein fragments, is the introduction of a desired modification by protein semi-synthesis. As a proof-of-concept, the *Ssp* DnaE split intein was used to introduce a biotin moiety at the C-terminus of a recombinant protein (57) (Figure 5C). The biotin was linked to the Lys+4 residue of the native *Ssp* DnaE C-extein sequence (CFNK) during chemical synthesis of the C-intein containing peptide. Incubation with a recombinant model protein [maltose binding protein (MBP) fused to the N-intein] appended the biotin to the C-terminus of MBP by PTS. This exciting opportunity for C-terminal protein modification formed the basis for a protein microarray platform based on PTS (58). Here, the authors attached the *Ssp* DnaE C-intein to a glass slide via PEGylated thiol-linkers, which were attached to the C-intein during its chemical synthesis. Incubating this C-intein functionalized solid support with model proteins fused to the N-intein led to immobilization of the proteins to the glass slide. Intriguingly, the reaction also proceeded when whole cell lysates containing the N-intein fusion proteins were spotted onto the glass slide, hence potentially enabling high-throughput proteomic studies using this split intein mediated protein microarray approach.

However groundbreaking, total chemical synthesis of modified peptides containing the 36 aa *Ssp* DnaE C-intein may not be straightforward and amenable to every laboratory interested in this technology, and obtaining such peptides from commercial sources is often associated with significant costs to the researcher. Moreover, this semi-synthetic approach only allows modification of a protein's C-terminus. Fortunately, two artificial split inteins engineered by Liu and co-workers have shown promising results in site-specific protein modification by semi-synthesis. The C-intein of the *Ssp* GyrB C6 split intein [referred to as *Ssp* GyrB S11 (31)] is six-times smaller than the *Ssp* DnaE C-intein, and thus represents a much simpler and cost-effective molecule for chemical synthesis. Indeed, this non-canonical split intein was shown to be an effective tool to label proteins with fluorescein and biotin at their C-termini (59) (Figure 5D). The gentle conditions of the PTS reaction preserved the native state of the modified proteins, allowed immobilization of a desired protein from a complex mixture to a solid support, and showed promising results in labeling of cell surface receptor proteins on cultured mammalian cells.

For N-terminal protein modification, the *Ssp* DnaB N11 split intein [referred to as *Ssp* DnaB S1 (29)] was used, which has an N-terminal fragment of only 11 aa. Again, synthesis of this sequence is more cost-effective to obtain by chemical synthesis than the *Ssp* DnaE C-intein, and allows for N-terminal modification, which may in some cases be preferred over C-terminal modification. Labels such as fluorescein were efficiently appended to recombinant proteins by PTS without altering their functions (60) (Figure 5E). The gentle PTS conditions also facilitated fluorescent labeling of a cell surface protein expressed on mammalian cells. However, to be effective in this biological scenario, the *Ssp* DnaB N11 split intein had to be re-engineered into a small-molecule integrated version based on the binding of trimethoprim to dihydrofolate reductase (61).

Dual fluorescence labeling is of interest for protein structure-function studies involving fluorescence resonance energy transfer (FRET), and has been demonstrated on a ubiquitin (Ub) dimer using a three-step procedure, where the authors combined residue-specific chemical labeling and split intein-mediated PTS (62). In the first step of the reaction, the C-terminal Ub (Ub_C, with a unique Cys at position 47) fused to the *Npu* DnaE C-intein (with a Cys+1Ser mutation) was labeled with a thiol-reactive Alexa Fluor 647 (AF647) probe. This labeled protein was then reacted with a protein consisting of an N-terminal Ub (Ub_N, with a unique Cys at position 7) fused to the *Npu* DnaE N-intein, resulting in a mono-labeled Ub dimer. Dual-labeling was achieved in the final step by incubating the PTS product with a thiol-reactive Alexa Fluor 594 (AF594) probe. The final $Ub_{N,AF594}$-$Ub_{C,AF647}$ protein was then used to compare protein folding of Ub monomers and dimers using FRET spectroscopy. For this dual labeling approach using the *Npu* DnaE split intein, the Cys+1 residue had to be mutated to Ser+1 in order to prevent its modification during the chemical labeling steps. Although the PTS reaction proceeded about 16-times slower because of this mutation, the final yields of dual-labeled protein were not negatively affected.

The aforementioned split intein-based protein modification schemes (Figure 5) should find great utility for *in vitro* applications or for labeling surface-accessible membrane proteins, but they cannot directly be used for protein modification inside living cells. Advances towards such an *in vivo* PTS-mediated modification have been made by Giriat and Muir (63). In this seminal study, a peptide consisting of the *Ssp* DnaE C-intein, followed by three native C-extein residues (CFN) and a FLAG epitope (I_C-FLAG), was linked to a cell penetrating peptide (CPP, containing residues 43–58 of the *Drosophila* homeotic transcription factor ANTP) with a

disulfide-bridge, involving the *Ssp* DnaE Cys+1 residue. The target protein for PTS was expressed in the cytoplasm of mammalian cells, and consisted of green fluorescent protein (GFP) fused to the *Ssp* DnaE N-intein. When the dipeptide complex was taken up by the cells, the reducing environment of the cytoplasm led to dissociation of I_C-FLAG from the CPP, effectively activating the Cys+1 residue of I_C-FLAG and resulting in C-terminal FLAG-tagging of GFP by PTS. The authors further showed the generality of their *in vivo* semi-synthetic approach by successfully FLAG-tagging four other proteins. Although in this case no chemical modification was introduced into the target proteins, this system promises to be a valuable tool to selectively modify proteins in living cells. The strength of this approach is that it is bio-orthogonal, thus only the protein fused to the N-intein will receive the modification whereas all other proteins inside the cell should be inert to the PTS-based modification. It is conceivable to speculate that by providing CPPs directed towards designated organelles, PTS could also be used to modify proteins in specific cellular compartments.

Introduction of post-translational modifications

A large portion of eukaryotic proteins is post-translationally modified by mechanisms such as glycosylation, lipidation and phosphorylation. These modifications can be crucial for protein stability and function, and their presence is therefore important when the protein is investigated, for example as a potential drug target. Unfortunately, post-translationally modified proteins are challenging to study because their purification from native sources such as mammalian cells requires special facilities, work force and expenses, and still may only result in insufficient yields for further study of the protein. Moreover, some post-translational modifications such as phosphorylation are often dynamic in nature, which can result in a heterogenously modified protein preparation. For successful structural studies of proteins using x-ray crystallography or NMR spectroscopy, however, the protein sample should be as homogenous as possible (64, 65). Bacterial expression hosts such as *E. coli* are usually much easier to handle, cost less and can produce multimilligram amounts of desired protein per liter of culture, yet post-translational modifications in *E. coli* are rare and differ from those in eukaryotes. Although efforts to produce natively modified proteins in bacteria are promising (66), there is still a long way to go.

Chemical synthesis of the entire protein would enable exact control over the modified state of the product but is expensive and limited to relatively small proteins (see above). A semi-synthetic approach appears as an elegant solution to this problem. Producing the main part of the protein in a model organism and adding a chemically synthesized modified fragment *via* protein ligation is an economical and convenient tool to introduce post-translational modifications, as demonstrated by Becker and co-workers for the site-specific introduction of lipid molecules using a combination of NCL and PTS (67). EPL has also been extensively used for similar purposes, and the reader is referred to recent reviews on this subject (68, 69).

In their study, Becker and co-workers were looking for a way to produce a recombinant version of the cellular prion protein (PrP^C) modified with a glycosylphosphatidylinositol (GPI) anchor. This lipid modification targets PrP^C to membranes and is associated with the conversion of PrP^C to the pathological form PrP^{Sc}, an insoluble conformer causing prion disease (70), however, this assumption requires experimental validation. In order to obtain a GPI-anchored PrP^C, the authors first produced by NCL a peptide consisting of the *Ssp* DnaE C-intein followed by a short C-extein sequence modified with a GPI anchor mimick (peptide I_C^{Palm}). This peptide was then reacted with a recombinant form of PrP^C expressed in *E. coli* cells, which was fused to the *Ssp* DnaE N-intein. PTS led to modification of PrP^C with the GPI anchor mimick at its C-terminus (Figure 6). This modified protein, dubbed $rPrP^{Palm}$, could be shown to efficiently integrate into the cell membrane of mammalian cells, thus paving the way to study the conversion of PrP^C to PrP^{Sc} in its natural environment. Using a similar experimental strategy, the same group recently demonstrated that proteins fused to the *Ssp* DnaE N-intein can directly be immobilized to I_C^{Palm}-functionalized liposomes by PTS (71) (Figure 6), which should prove useful in the transfer of custom made proteins into eukaryotic cells.

In vivo protein reconstitution

In nature, split inteins reconstitute a functional protein (e.g. the catalytic DNA polymerase III α-subunit) from two separately expressed fragments by PTS. This potential to reconstitute a protein's function using split inteins has been the basis for numerous cell-based assays to monitor protein-protein interactions and protein targeting, and to control protein function.

Umezawa and co-workers have developed a plethora of such in-cell assays using reconstitution of split versions of enhanced green fluorescent protein (eGFP) or luciferase by PTS as a read-out for protein-protein interactions (72). Later, the sensitivity of the system was enhanced by reconstituting a transcription factor, which led to expression of a reporter (luciferase) only upon protein-protein interactions and PTS (73). PTS-based reconstitution of eGFP or luciferase further proved useful for monitoring protein trafficking into the mitochondrial matrix (74), the nucleoplasm (75) or the endoplasmic reticulum (76) (Figure 7A). Protein reconstitution has also been used to engineer transgenic plants, where herbicide resistance (HR) proteins were reconstituted by PTS (77–80) (Figure 7B). By incorporating the split HR genes into different genomes of the plant, the engineered plants could be considered 'environmentally safe' because only incomplete and inactive HR fragments would be released into the environment during plant reproduction. Lastly, PTS-mediated protein reconstitution has also shown promise in gene therapy (81).

Reconstitution of a protein by PTS is especially useful if the reconstitution reaction, and thus the activation of the protein, can be controlled. This inventive idea, referred to as 'conditional protein splicing', was first realized by engineering an artificially split *Sce* VMA1 intein, that by itself was inactive towards PTS due to non-productive split intein fragment assembly. However, when the split intein fragments

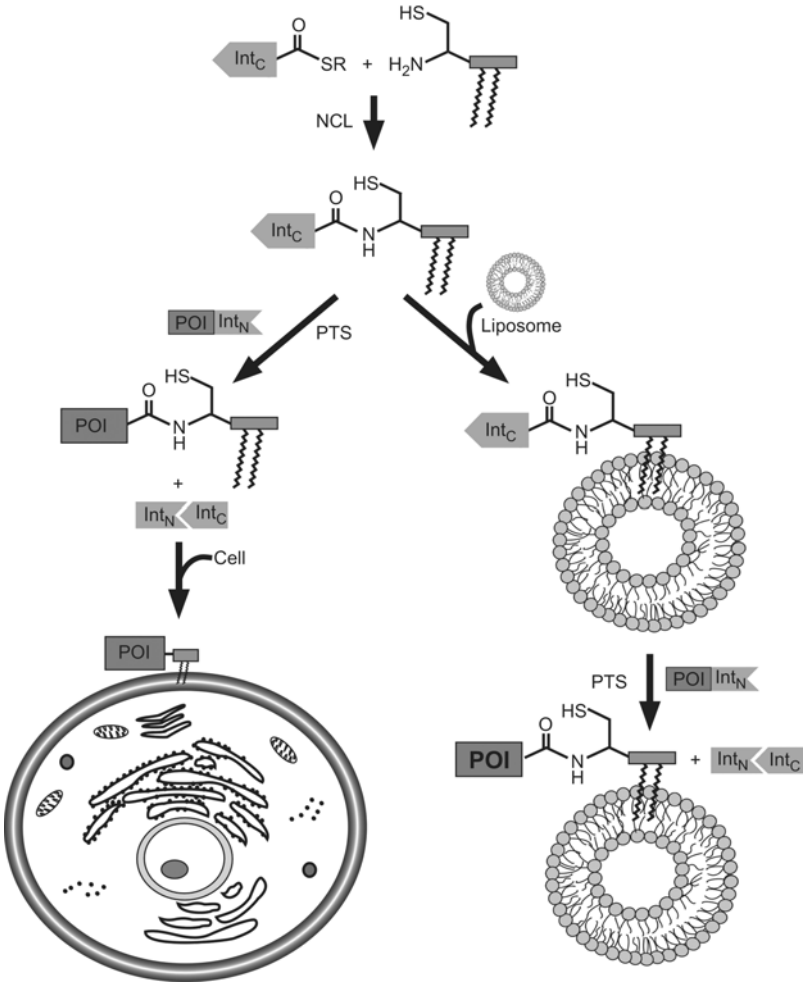

Figure 6 Site-specific lipidation via protein *trans*-splicing (PTS).
Using NCL, the C-terminal fragment of the *Ssp* DnaE split intein was fused to a short, doubly palmitoylated C-extein sequence (yielding peptide I_C^{Palm}). The latter lipidated extein sequence could then be incorporated at the C-terminus of a protein of interest (POI) via PTS, and the lipidated POI attached to the outer membrane of a mammalian cell (left path). In another experimental set-up, the native chemical ligation (NCL) product was directly incorporated into liposomes, and the POI was immobilized via PTS (right path).

were fused with the protein partners FK506 binding protein (FKBP) and the FKBP-rapamycin binding (FRB) domain of FKBP-rapamycin associated protein (FRAP), whose interaction depends on the presence of the small molecule rapamycin, PTS could in turn be achieved by providing rapamycin (82) (Figure 7C). This rapamycin-dependent PTS system has been shown to reconstitute proteins not only in test tubes, but also in live cells (83) and even whole organisms (84), allowing for control of protein function within a defined time period. Recently, light-inducible PTS systems have also been developed (85–87), which could be promising tools to activate protein function *in vivo* in a spatiotemporal manner using highly focused microscopic light sources.

Segmental isotope labeling for NMR studies

Classical NMR spectroscopy using isotopically labeled samples is generally limited to relatively small proteins of 15–20 kDa because larger proteins produce spectra with broadening and overlapping of signals, which has forced scientists to study individual domains of large proteins rather than the entire molecule. Incorporating isotope labels only in a defined segment of a large protein overcomes the problem of signal overlapping, and provides a more accurate structural view of the protein region than when it is removed from the greater protein context and is more straightforward than selective amino acid labelling. Segmental isotope labeling can be achieved by PTS when only one of the split intein-equipped protein reactants has been expressed in isotope-containing culture medium, and experimental schemes for labeling of terminal and medial segments in large proteins have been devised [reviewed in (88)]. Segmental isotope labeling by PTS has facilitated unambiguous spectral assignments of a tri-modular protein, whose three domains shared a high degree of sequence similarity, and would have otherwise been impossible to assign due to signal overlap (34). Segmental isotope labeling also afforded the near complete assignment (up to 90%) of the 52 kDa F_0F_1 ATP-synthase

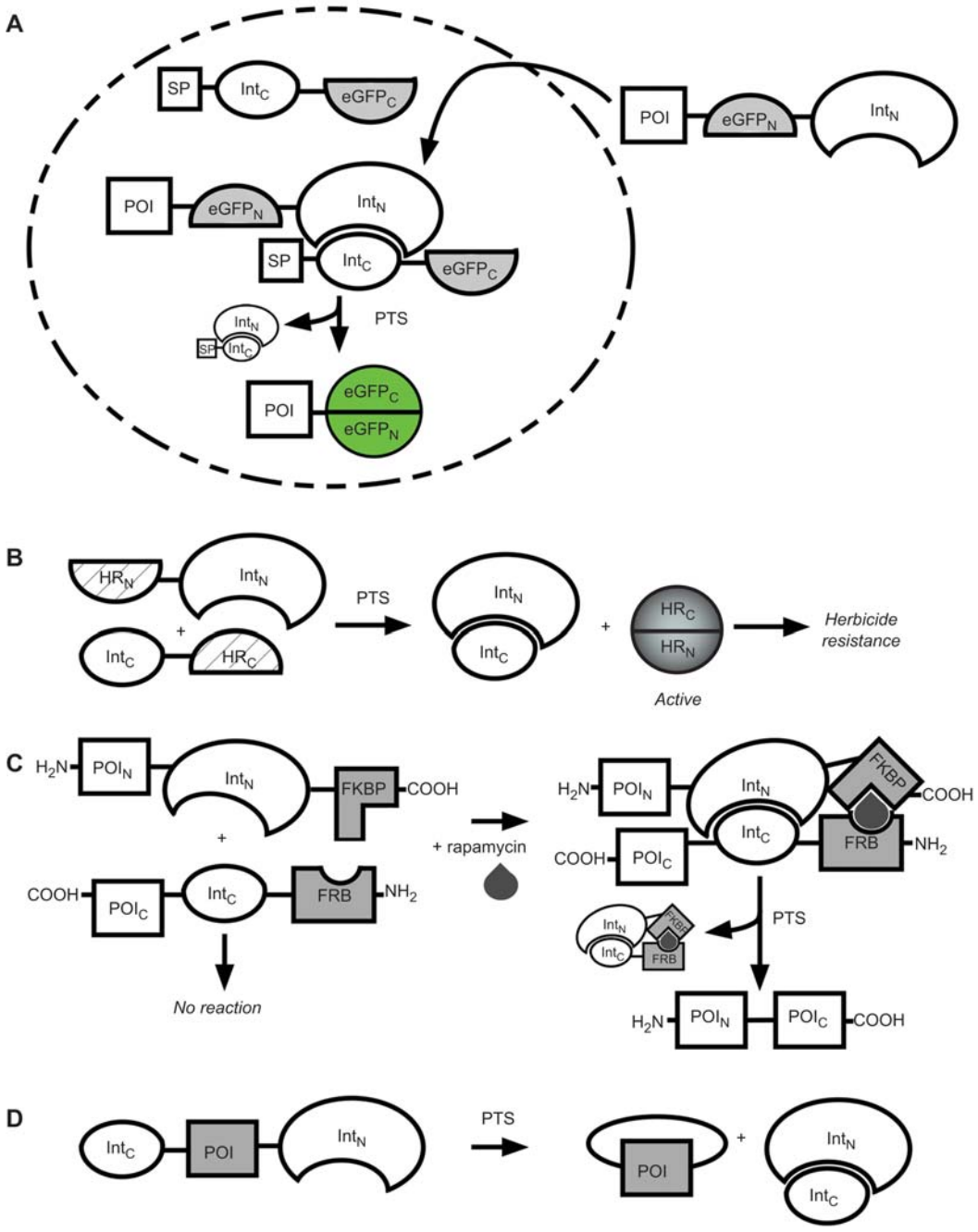

Figure 7 PTS-based protein reconstitution approaches.
(A) *In vivo* protein compartmentalization assay. An Int_C-$eGFP_C$ protein is constitutively localized to a specific organelle via an N-terminal signal peptide (SP). A protein of interest (POI) with unknown signal peptide properties is expressed in fusion with $eGFP_N$ and Int_N. Only upon successful translocation of this protein into the cellular compartment is eGFP reconstituted and able to fluoresce. (B) Reconstitution of a herbicide resistance (HR) protein from split fragments. The genes of the two precursor proteins must be encoded on separate genomes (e.g., in the nucleus and chloroplast) to avoid spreading of the transgene. (C) Induction of protein *trans*-splicing (PTS) and protein reconstitution via addition of rapamycin. PTS and reconstitution of the POI only occurs when rapamycin is present and the split intein fragments are brought in close proximity due to the rapamcyin-induced binding of FKBP to FRB. (D) PTS-mediated protein cyclization.

β-subunit from four differently labeled proteins (89). Furthermore, titration of the segmental-isotope labeled β-subunit proteins with ATP and measurements of residual dipolar couplings provided clear evidence that binding of the nucleotide induced the movement of the C-terminal domain towards the N-terminal domain by an angle of 35°, a conformational shift whose precise nature was poorly defined by prior crystallographic studies.

In addition to assignment of large proteins, PTS-assisted segmental isotope labeling allows the fusion of an unlabeled,

and thus 'invisible', solubility tag to the protein of interest (90). This may enable NMR-based structure determination of proteins that are inherently prone to insolubility, but could also represent a general benefit for protein NMR spectroscopy because the high concentrations required for NMR measurements can have adverse effects on protein solubility.

Protein backbone cyclization

Cyclization is an easy way to stabilize a protein *via* the reduction in entropy and resistance to proteolysis (91). This is especially tempting in the case of small peptides, which have limited possibilities for intramolecular stabilization. The protein ligation reaction of split inteins can be exploited to make custom cyclic peptides and proteins: when the polypeptide is sandwiched between a circularly permuted split intein, intramolecular PTS results in peptide bond formation between N- and C-terminus of the polypeptide (92) (Figure 7D). PTS-catalyzed cyclization has become widely used with numerous protein targets studied (92–96). The cyclic proteins all retained their native biological function, but were resistant to proteolysis by exopeptidases. They further exhibited the expected higher stability towards temperature and denaturating agents compared to their linear counterparts, which allowed for detailed thermodynamic studies of the effects of protein cyclization on protein stability (93, 97). PTS has also recently been used to generate random cyclic peptide libraries (98), which could prove useful in the identification of novel peptide drugs by high-throughput screenings.

Protein cyclization becomes especially interesting when the biological function of the cyclic protein can be assessed *in vivo*. In a recent study, Byers and colleagues used PTS to produce a cyclic version of bacterial acyl carrier protein (ACP) (99), a small acidic protein of a highly flexible nature, which can alternate between an unfolded and a folded conformation, and is essential for fatty acid synthesis in prokaryotes (100). The authors found that in stark contrast to its linear counterpart, the cyclic ACP appeared to be trapped in a permanently folded conformational state, as evidenced by extensive biophysical characterization of the protein. Intriguingly, expression of the cyclic ACP was able to efficiently complement growth of an *E. coli* ACP-knockout strain, thus challenging the belief that the dynamic character of ACP is obligatory for its biological function (99).

Expert opinion: how to make PTS work?

As outlined above, split inteins are valuable protein ligation tools with great potential to aid in deciphering protein structure-function relationships. However, to use split inteins for any of the experiments outlined above or in future techniques, several variables require careful consideration in order to achieve the anticipated results.

Reaction conditions

Since split inteins work in an autocatalytic fashion, they do not require external cofactors or energy equivalents. PTS reactions can thus conveniently be carried out both *in vivo* and *in vitro* with only two requirements: (a) the split intein fragments have to be accessible for the reaction (discussed below), and (b) if the N-nucleophile or +1 residue is a cysteine, a reducing agent should be added in order to prevent the formation of intra- or intermolecular disulfide bridges, which would otherwise inhibit the PTS reaction. A reducing environment *in vitro* is easily achieved by adding a suitable reducing agent, such as tris(2-carboxyethyl)phosphine (TCEP). Reducing agents with a nucleophile, such as dithiothreitol (DTT), should be avoided as they may induce N-terminal cleavage of the intein (101), hence abrogating completion of the PTS reaction. TCEP should not be used with phosphate buffers, as these are known to oxidize the compound. Furthermore, tris(hydroxymethyl)aminomethane (Tris) has been shown to form stable adducts with thioester intermediates during intein-mediated reactions (102), and should therefore also be avoided. Because cations such as zinc, cadmium and copper can inhibit the PTS reaction (103), a chelating agent such as ethylenediamine tetraacetic acid (EDTA) should be present in the reaction buffer.

The pH and temperature of an *in vitro* PTS reaction depends on the natural environment from which the intein has been obtained. Mesophilic inteins, such as the DnaE split inteins from cyanobacteria, perform PTS efficiently at ambient temperature, but some thermophilic inteins may require higher temperatures (up to 80°C) to splice efficiently (27). The choice between a mesophilic vs. a thermophilic split intein for PTS thus depends on the temperature tolerance of the target protein(s). Lastly, since none of the currently described inteins have been obtained from acidophilic or alkaliphilic organisms, the reaction pH should be anywhere between 6 and 8.

Junction sequences

The study of inteins in heterologous protein contexts suggested that they can be transplanted anywhere into any protein, as long as the insertion site provides the crucial native +1 residue of the intein. The nature of the residues that flank the N- and C-terminus of an intein can, however, be crucial for the success of protein splicing. The early efforts to create inducible protein cleavage systems based on inteins showed that the efficiency of both N- and C-cleavage can greatly be influenced by the amino acid flanking the N- and C-catalytic residues [positions -1 and +1; (41, 104)]. A recent study revealed that for N-cleavage this modulation in efficiency can even extend to positions -2 and +3 (105). Surprisingly little is known about this extein junction sequence effect on the complete splicing reaction. The effect of the -1 position on protein splicing of the *Sce* VMA1 intein was systematically examined and showed a broad tolerance (101). The *Npu* and *Ssp* DnaE split inteins were similarly screened at position +2, and even though they are highly homologous, the importance of the +2 residue varied dramatically: a chimeric split intein containing the *Npu* DnaE N-intein and the *Ssp* DnaE C-intein tolerated almost any residue at position +2 (12 mutations from the native Phe+2 resulted in >50% PTS), whereas splicing of the complete *Ssp* DnaE was sensitive to most mutations (only two mutations resulted in >50% PTS)

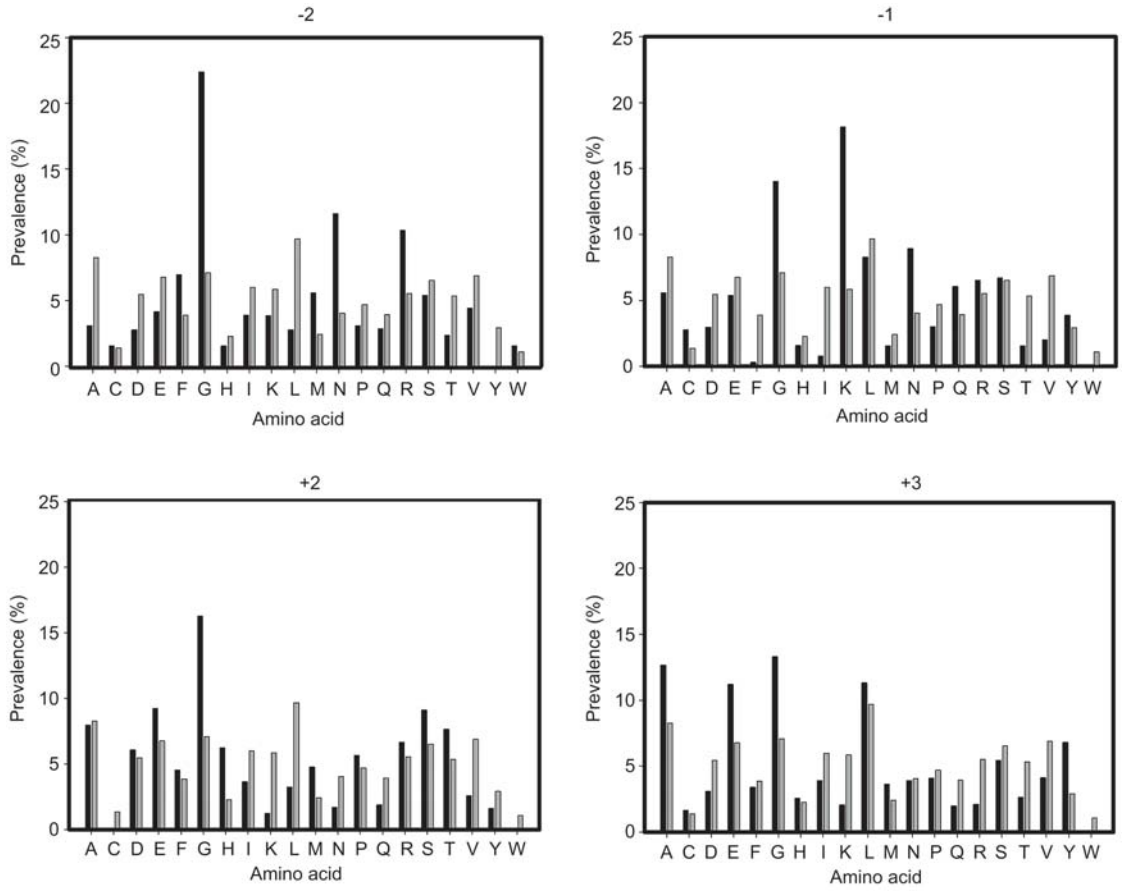

Figure 8 Conservation of intein junction sequences.
Bar graph representation of sequence conservation at extein positions -2, -1, +2, and +3. All complete intein junction sequences (positions -2, -1, +2 and +3) were retrieved from Inbase [http://www.neb.com/neb/inteins.html, (12); 550 entries as of Nov 3, 2010]. To prepare the bar graphs, the prevalence of each amino acid at the positions -2, -1, +2, and +3 was calculated (%), and values were plotted using Sigma Plot. Prevalence values for the average occurrence of each amino acid in proteins were taken from the UniProtKB/Swiss-Prot protein knowledgebase release 2011_02 statistics available online (http://au.expasy.org/sprot/relnotes/relstat.html). Black bars represent the amino acid prevalence of the extein residue indicated above the graph, grey bars represent the average amino acid prevalence in the protein knowledge base.

(106). The extein junction sequence effect is most likely a result of adaptation of inteins to their specific natural insertion site within a host protein, and the presence of non-native residues may lead to distortion of the peptide bonds at the junctions (107–110). This in effect causes a somewhat random pattern of amino acid distributions at positions -2, -1, +2 and +3 (105) (Figure 8), and results in the practical conundrum that not all inteins splice well in every extein residue context.

When deciding in practice where to insert an intein into a target protein, one should consider adding a few residues (two to five) from the native splicing junctions during initial tests for PTS. If introduction of extra residues in the final spliced protein product is undesired, the intein should be inserted at a location that contains residues with properties similar to the native extein residues, as is the case with any site-directed mutagenesis, where protein function is to be retained. When making any changes, it is advisable not to alter the nature of the +1 residue; although Cys has the most nucleophilic amino acid side chain of all three possible +1 residues, one cannot accurately predict how the micro-environment in the intein's catalytic center affects nucleophilicity, hence changing a Ser+1 or Thr+1 to a Cys does not necessarily result in better protein splicing [see, e.g. (35)]. In contrast, mutating a native Cys+1 to Ser or Thr almost always results in a decrease splicing efficiency [e.g. (62)].

Extein protein effect and solubility

Although the extein proteins do not participate in the splicing reaction, they may have a great influence on the efficiency of the protein splicing reaction. The reasons underlying this phenomenon are not fully understood but are considered to be caused by charge repulsion of the extein proteins and the conformational strain thus exerted on the intein catalytic core. Currently, there is no method to predict the possible extein protein effect, except for experimental evaluation. It is known, however, that some inteins are less sensitive to the

extein protein effect than others and are thus considered to be more promiscuous towards the protein ligation targets. Even among the native DnaE split inteins, which share considerable sequence homology, large differences in the reaction yields have been observed (26, 111). Furthermore, it has been demonstrated that changing the split site can enhance the extein tolerance of an intein (30).

In addition to the poorly defined extein related charge-effect, the solubility of the extein proteins is crucial. For PTS to occur, the two halves of a split intein have to associate into a correctly assembled HINT fold, which means in practical terms that at least one of the intein fragments needs to be soluble. For sequential PTS *in vivo* such as in segmental isotope labeling, it is generally advisable that the soluble fragment be expressed prior to the insoluble counterpart, which would still allow split intein association before the insoluble partner is ultimately deposited in non-productive aggregates such as inclusion bodies. When performing *in vitro* ligation, both fragments have to be soluble; alternatively, some split inteins can also perform PTS during removal of denaturing agents by dialysis (55). As several inteins can fold in buffer containing up to 6 M urea (20), refolding of even membrane proteins has been demonstrated (56). Choosing a productive split site within the extein protein will greatly affect the solubility of the intein-extein fusion protein. Target proteins for ligation are usually split either between domains or near the N- or C-terminal ends. Splitting between individually folded domains often ensures the solubility of the split halves, and allows for optimization of the junction sequence by adding an appropriate linker region (34, 36). Splitting at either end of a protein is mainly done when performing semi-synthesis, for example site-specific protein modification (see above) and is not expected to have a great impact on the solubility of the extein protein.

Outlook: the future of protein *trans*-splicing

Protein ligation by PTS enables crossing the limits of traditional recombinant protein expression, and, indeed, many versatile applications of PTS have already been demonstrated. However, PTS is not yet universally applicable and for PTS to gain wider use, more basic research on split inteins and PTS is required in order to achieve protein ligation between practically any pair of proteins or peptides. Systematic mutational studies combined with structural investigations are likely required to help decipher how the splicing junction sequences influence the PTS reaction. Considerable efforts should also be devoted to the as of yet unexplained effect of the extein proteins on PTS. The ultimate goal of such endeavors would be a computational platform that predicts the outcome of split intein-catalyzed PTS. The search for novel split inteins, natural or artificial, with broad tolerance towards the splicing junction and extein proteins, and which perform PTS with high reaction rates and yields of ligated product, are also highly desireable. Of note, such split inteins should not be prone to cross-reactivity with other split inteins to enable PTS-mediated ligation of multiple fragments in a single step.

An alternative approach to the search for novel inteins is to evolve currently available split inteins and, in the process, enhance the desired properties. Directed molecular evolution (DME) of inteins has been shown to be a feasible undertaking, and many *cis*-splicing inteins have been improved and advanced towards specific reaction conditions (112–115); an experimental platform for the *in vitro* DME of split inteins has also been reported recently (116). It remains to be seen whether a unique 'super intein', which combines all favorable characteristics, can be developed through DME, or if the solution will lie in the identification of a set of inteins, each of them suitable for a specific experimental purpose.

Highlights

- Protein *trans*-splicing (PTS) by split inteins enables ligation of separately produced polypeptide chains.
- PTS is a convenient and cost-effective way to introduce natural and artificial modifications into proteins of any size.
- PTS-mediated *in vivo* protein reconstitution allows tight control over and high-throughput screening of protein interactions, localization and activity.
- PTS-based segmental isotope labeling is a straightforward method to overcome the size limits of NMR spectroscopy and to analyze single domains in the context of full-length proteins.
- Split inteins can afford cyclic proteins to increase stability and to study protein folding and conformational flexibility.
- Choice of the splicing junction sequence is crucial for the success of any PTS-based approach.
- With more basic research into PTS, this protein ligation tool has the potential to become a universally applicable technique to study protein structure-function relationships.

Acknowledgements

We would like to thank Hideo Iwai for giving us the opportunity to write this review, and to the two anonymous reviewers for their helpful comments.

References

1. Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, Neff N, Noren CJ, Thorner J, Belfort M. Protein splicing elements: inteins and exteins – a definition of terms and recommended nomenclature. Nucleic Acids Res 1994; 22: 1125–7.
2. Belfort M, Roberts RJ. Homing endonucleases: keeping the house in order. Nucleic Acids Res 1997; 25: 3379–88.
3. Duan X, Gimble FS, Quiocho FA. Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity. Cell 1997; 89: 555–64.
4. Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K. Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. J Mol Biol 2000; 300: 889–901.

5. Hodges RA, Perler FB, Noren CJ, Jack WE. Protein splicing removes intervening sequences in an archaea DNA polymerase. Nucleic Acids Res 1992; 20: 6153–7.
6. Fitzsimons Hall M, Noren CJ, Perler FB, Schildkraut I. Creation of an artificial bifunctional intein by grafting a homing endonuclease into a mini-intein. J Mol Biol 2002; 323: 173–9.
7. Kawasaki M, Nogami S, Satow Y, Ohya Y, Anraku Y. Identification of three core regions essential for protein splicing of the yeast Vma1 protozyme. A random mutagenesis study of the entire Vma1-derived endonuclease sequence. J Biol Chem 1997; 272: 15668–74.
8. Chong S, Xu MQ. Protein splicing of the Saccharomyces cerevisiae VMA intein without the endonuclease motifs. J Biol Chem 1997; 272: 15587–90.
9. Derbyshire V, Wood DW, Wu W, Dansereau JT, Dalgaard JZ, Belfort M. Genetic definition of a protein-splicing domain: functional mini-inteins support structure predictions and a model for intein evolution. Proc Natl Acad Sci USA 1997; 94: 11466–71.
10. Wu H, Xu MQ, Liu XQ. Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein. Biochim Biophys Acta 1998; 1387: 422–32.
11. Telenti A, Southworth M, Alcaide F, Daugelat S, Jacobs WR Jr, Perler FB. The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein. J Bacteriol 1997; 179: 6378–82.
12. Perler FB. InBase: the Intein Database. Nucleic Acids Res 2002; 30: 383–4.
13. Liu XQ. Protein-splicing intein: genetic mobility, origin, and evolution. Annu Rev Genet 2000; 34: 61–76.
14. Pietrokovski S. Modular organization of inteins and C-terminal autocatalytic domains. Protein Sci 1998; 7: 64–71.
15. Perler FB, Olsen GJ, Adam E. Compilation and analysis of intein sequences. Nucleic Acids Res 1997; 25: 1087–93.
16. Chen L, Benner J, Perler FB. Protein splicing in the absence of an intein penultimate histidine. J Biol Chem 2000; 275: 20431–5.
17. Mills KV, Manning JS, Garcia AM, Wuerdeman LA. Protein splicing of a Pyrococcus abyssi intein with a C-terminal glutamine. J Biol Chem 2004; 279: 20685–91.
18. Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Pietrokovski S, Perler FB. Splicing of the mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. J Biol Chem 2010; 285: 2515–26.
19. Southworth MW, Benner J, Perler FB. An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. EMBO J 2000; 19: 5019–26.
20. Southworth MW, Adam E, Panne D, Byer R, Kautz R, Perler FB. Control of protein splicing by intein fragment reassembly. EMBO J 1998; 17: 918–26.
21. Mills KV, Lew BM, Jiang S, Paulus H. Protein splicing in trans by purified N- and C-terminal fragments of the Mycobacterium tuberculosis RecA intein. Proc Natl Acad Sci USA 1998; 95: 3543–8.
22. Shingledecker K, Jiang SQ, 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 1998; 207: 187–95.
23. Gorbalenya AE. Non-canonical inteins. Nucleic Acids Res 1998; 26: 1741–8.
24. Wu H, Hu Z, Liu XQ. Protein trans-splicing by a split intein encoded in a split DnaE gene of Synechocystis sp. PCC6803. Proc Natl Acad Sci USA 1998; 95: 9226–31.
25. Caspi J, Amitai G, Belenkiy O, Pietrokovski S. Distribution of split DnaE inteins in cyanobacteria. Mol Microbiol 2003; 50: 1569–77.
26. Dassa B, Amitai G, Caspi J, Schueler-Furman O, Pietrokovski S. Trans protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. Biochemistry 2007; 46: 322–30.
27. Choi JJ, Nam KH, Min B, Kim SJ, Soll D, Kwon ST. Protein trans-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite Nanoarchaeum equitans. J Mol Biol 2006; 356: 1093–106.
28. Dassa B, London N, Stoddard BL, Schueler-Furman O, Pietrokovski S. Fractured genes: a novel genomic arrangement involving new split inteins and a new homing endonuclease family. Nucleic Acids Res 2009; 37: 2560–73.
29. Sun W, Yang J, Liu XQ. Synthetic two-piece and three-piece split inteins for protein trans-splicing. J Biol Chem 2004; 279: 35281–6.
30. Aranko AS, Zuger S, Buchinger E, Iwai H. In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. PLoS One 2009; 4: e5185.
31. Appleby JH, Zhou K, Volkmann G, Liu XQ. Novel split intein for trans-splicing synthetic peptide onto C terminus of protein. J Biol Chem 2009; 284: 6194–9.
32. Oeemig JS, Aranko AS, Djupsjobacka J, Heinamaki K, Iwai H. Solution structure of DnaE intein from Nostoc punctiforme: structural basis for the design of a new split intein suitable for site-specific chemical modification. FEBS Lett 2009; 583: 1451–6.
33. Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ. Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. Cell 1997; 91: 85–97.
34. Busche AE, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dotsch V, Iwai H. Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. Angew Chem Int Ed Engl 2009; 48: 6128–31.
35. Xu MQ, Perler FB. The mechanism of protein splicing and its modulation by mutation. EMBO J 1996; 15: 5146–53.
36. Shi J, Muir TW. Development of a tandem protein trans-splicing system based on native and engineered split inteins. J Am Chem Soc 2005; 127: 6198–206.
37. Brenzel S, Kurpiers T, Mootz HD. Engineering artificially split inteins for applications in protein chemistry: biochemical characterization of the split Ssp DnaB intein and comparison to the split Sce VMA intein. Biochemistry 2006; 45: 1571–8.
38. Mathys S, Evans TC, Chute IC, Wu H, Chong S, Benner J, Liu XQ, Xu MQ. 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 1999; 231: 1–13.
39. Chong S, Mersha FB, Comb DG, Scott ME, Landry D, Vence LM, Perler FB, Benner J, Kucera RB, Hirvonen CA, Pelletier JJ, Paulus H, Xu MQ. Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. Gene 1997; 192: 271–81.
40. Evans TC Jr, Benner J, Xu MQ. Semisynthesis of cytotoxic proteins using a modified protein splicing element. Protein Sci 1998; 7: 2256–64.
41. Xu MQ, Paulus H, Chong S. Fusions to self-splicing inteins for protein purification. Methods Enzymol 2000; 326: 376–418.
42. Xu MQ, Evans TC Jr. Purification of recombinant proteins from E. coli by engineered inteins. Methods Mol Biol 2003; 205: 43–68.

43. Wood RJ, Pascoe DD, Brown ZK, Medlicott EM, Kriek M, Neylon C, Roach PL. Optimized conjugation of a fluorescent label to proteins via intein-mediated activation and ligation. Bioconjug Chem 2004; 15: 366–72.
44. Teruya K, Nishizawa K, Doh-ura K. Semisynthesis of a protein with cholesterol at the C-terminal, targeted to the cell membrane of live cells. Protein J 2010; 29: 493–500.
45. Dawson PE, Muir TW, Clark-Lewis I, Kent SB. Synthesis of proteins by native chemical ligation. Science 1994; 266: 776–9.
46. Kent SB. Total chemical synthesis of proteins. Chem Soc Rev 2009; 38: 338–51.
47. Tam JP, Yu Q, Miao Z. Orthogonal ligation strategies for peptide and protein. Biopolymers 1999; 51: 311–32.
48. Tam JP, Eom KD. Mimicking reverse protein splicing by three-segment tandem peptide ligation. Protein Pept Lett 2005; 12: 743–9.
49. Evans TC Jr, Xu MQ. Intein-mediated protein ligation: harnessing nature's escape artists. Biopolymers 1999; 51: 333–42.
50. Chang TK, Jackson DY, Burnier JP, Wells JA. Subtiligase: a tool for semisynthesis of proteins. Proc Natl Acad Sci USA 1994; 91: 12544–8.
51. Jackson DY, Burnier J, Quan C, Stanley M, Tom J, Wells JA. A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. Science 1994; 266: 243–7.
52. Braisted AC, Judice JK, Wells JA. Synthesis of proteins by subtiligase. Methods Enzymol 1997; 289: 298–313.
53. Proft T. Sortase-mediated protein ligation: an emerging biotechnology tool for protein modification and immobilisation. Biotechnol Lett 2010; 32: 1–10.
54. Kurpiers T, Mootz HD. Regioselective cysteine bioconjugation by appending a labeled cystein tag to a protein by using protein splicing in trans. Angew Chem Int Ed Engl 2007; 46: 5234–7.
55. Kurpiers T, Mootz HD. Site-specific chemical modification of proteins with a prelabelled cysteine tag using the artificially split Mxe GyrA intein. Chembiochem 2008; 9: 2317–25.
56. Brenzel S, Cebi M, Reiss P, Koert U, Mootz HD. Expanding the scope of protein trans-splicing to fragment ligation of an integral membrane protein: towards modulation of porin-based ion channels by chemical modification. Chembiochem 2009; 10: 983–6.
57. Evans TC Jr, Martin D, Kolly R, Panne D, Sun L, Ghosh I, Chen L, Benner J, Liu XQ, Xu MQ. Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803. J Biol Chem 2000; 275: 9091–4.
58. Kwon Y, Coleman MA, Camarero JA. Selective immobilization of proteins onto solid supports through split-intein-mediated protein trans-splicing. Angew Chem Int Ed Engl 2006; 45: 1726–9.
59. Volkmann G, Liu XQ. Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein. PLoS One 2009; 4: e8381.
60. 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. Angew Chem Int Ed Engl 2006; 45: 5218–21.
61. Ando T, Tsukiji S, Tanaka T, Nagamune T. Construction of a small-molecule-integrated semisynthetic split intein for in vivo protein ligation. Chem Commun Camb 2007; 47: 4995–7.
62. Yang JY, Yang WY. Site-specific two-color protein labeling for FRET studies using split inteins. J Am Chem Soc 2009; 131: 11644–5.
63. Giriat I, Muir TW. Protein semi-synthesis in living cells. J Am Chem Soc 2003; 125: 7180–1.
64. Wyss DF, Choi JS, Li J, Knoppers MH, Willis KJ, Arulanandam AR, Smolyar A, Reinherz EL, Wagner G. Conformation and function of the N-linked glycan in the adhesion domain of human CD2. Science 1995; 269: 1273–8.
65. De Beer T, Van Zuylen CW, Leeflang BR, Hard K, Boelens R, Kaptein R, Kamerling JP, Vliegenthart JF. NMR studies of the free alpha subunit of human chorionic gonadotropin. Structural influences of N-glycosylation and the beta subunit on the conformation of the alpha subunit. Eur J Biochem 1996; 241: 229–42.
66. Schwarz F, Huang W, Li C, Schulz BL, Lizak C, Palumbo A, Numao S, Neri D, Aebi M, Wang LX. A combined method for producing homogeneous glycoproteins with eukaryotic N-glycosylation. Nat Chem Biol 2010; 6: 264–6.
67. Olschewski D, Seidel R, Miesbauer M, Rambold AS, Oesterhelt D, Winklhofer KF, Tatzelt J, Engelhard M, Becker CF. Semisynthetic murine prion protein equipped with a GPI anchor mimic incorporates into cellular membranes. Chem Biol 2007; 14: 994–1006.
68. Flavell RR, Muir TW. Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology. Acc Chem Res 2009; 42: 107–16.
69. Berrade L, Camarero JA. Expressed protein ligation: a resourceful tool to study protein structure and function. Cell Mol Life Sci 2009; 66: 3909–22.
70. Taylor DR, Hooper NM. The prion protein and lipid rafts. Mol Membr Biol 2006; 23: 89–99.
71. Chu NK, Olschewski D, Seidel R, Winklhofer KF, Tatzelt J, Engelhard M, Becker CF. Protein immobilization on liposomes and lipid-coated nanoparticles by protein trans-splicing. J Pept Sci 2010; 16: 582–8.
72. Ozawa T, Umezawa Y. Detection of protein-protein interactions in vivo based on protein splicing. Curr Opin Chem Biol 2001; 5: 578–83.
73. Kanno A, Ozawa T, Umezawa Y. Intein-mediated reporter gene assay for detecting protein-protein interactions in living mammalian cells. Anal Chem 2006; 78: 556–60.
74. Ozawa T, Sako Y, Sato M, Kitamura T, Umezawa Y. A genetic approach to identifying mitochondrial proteins. Nat Biotechnol 2007; 21: 287–93.
75. Kim SB, Ozawa T, Watanabe S, Umezawa Y. High-throughput sensing and noninvasive imaging of protein nuclear transport by using reconstitution of split Renilla luciferase. Proc Natl Acad Sci USA 2004; 101: 11542–7.
76. Ozawa T, Nishitani K, Sako Y, Umezawa Y. A high-throughput screening of genes that encode proteins transported into the endoplasmic reticulum in mammalian cells. Nucleic Acids Res 2005; 33: e34.
77. Chin HG, Kim GD, Marin I, Mersha F, Evans TC Jr, Chen L, Xu MQ, Pradhan S. Protein trans-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes. Proc Natl Acad Sci USA 2003; 100: 4510–5.
78. Dun BQ, Wang XJ, Lu W, Zhao ZL, Hou SN, Zhang BM, Li GY, Evans TC Jr, Xu MQ, Lin M. Reconstitution of glyphosate resistance from a split 5-enolpyruvyl shikimate-3-phosphate synthase gene in Escherichia coli and transgenic tobacco. Appl Environ Microbiol 2007; 73: 7997–8000.
79. Gils M, Marillonnet S, Werner S, Grutzner R, Giritch A, Engler C, Schachschneider R, Klimyuk V, Gleba Y. A novel hybrid seed system for plants. Plant Biotechnol J 2008; 6: 226–35.
80. Kempe K, Rubtsova M, Gils M. Intein-mediated protein assembly in transgenic wheat: production of active barnase and acetolactate synthase from split genes. Plant Biotechnol J 2009; 7: 283–97.
81. Li J, Sun W, Wang B, Xiao X, Liu XQ. Protein trans-splicing

as a means for viral vector-mediated in vivo gene therapy. Hum Gene Ther 2008; 19: 958–64.
82. Mootz HD, Muir TW. Protein splicing triggered by a small molecule. J Am Chem Soc 2002; 124: 9044–5.
83. 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. J Am Chem Soc 2003; 125: 10561–9.
84. Schwartz EC, Saez L, Young MW, Muir TW. Post-translational enzyme activation in an animal via optimized conditional protein splicing. Nat Chem Biol 2007; 3: 50–4.
85. Vila-Perello M, Hori Y, Ribo M, Muir TW. Activation of protein splicing by protease- or light-triggered O to N acyl migration. Angew Chem Int Ed Engl 2008; 47: 7764–7.
86. Tyszkiewicz AB, Muir TW. Activation of protein splicing with light in yeast. Nat Methods 2008; 5: 303–5.
87. Berrade L, Kwon Y, Camarero JA. Photomodulation of protein trans-splicing through backbone photocaging of the DnaE split intein. Chembiochem 2010; 11: 1368–72.
88. Volkmann G, Iwai H. Protein trans-splicing and its use in structural biology: opportunities and limitations. Mol Biosyst 2010; 6: 2110–21.
89. Yagi H, Tsujimoto T, Yamazaki T, Yoshida M, Akutsu H. Conformational change of H+-ATPase beta monomer revealed on segmental isotope labeling NMR spectroscopy. J Am Chem Soc 2004; 126: 16632–8.
90. Zuger S, Iwai H. Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat Biotechnol 2005; 23: 736–40.
91. Kee KS, Jois SD. Design of beta-turn based therapeutic agents. Curr Pharm Des 2003; 9: 1209–24.
92. Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ. Production of cyclic peptides and proteins in vivo. Proc Natl Acad Sci USA 1999; 96: 13638–43.
93. 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 2001; 276: 16548–54.
94. Williams NK, Prosselkov P, Liepinsh E, Line I, Sharipo A, Littler DR, Curmi PM, Otting G, Dixon NE. In vivo protein cyclization promoted by a circularly permuted Synechocystis sp. PCC6803 DnaB mini-intein. J Biol Chem 2002; 277: 7790–8.
95. Jeffries CM, Graham SC, Stokes PH, Collyer CA, Guss JM, Matthews JM. Stabilization of a binary protein complex by intein-mediated cyclization. Protein Sci 2006; 15: 2612–8.
96. Siebold C, Erni B. Intein-mediated cyclization of a soluble and a membrane protein in vivo: function and stability. Biophys Chem 2002; 96: 163–71.
97. Williams NK, Liepinsh E, Watt SJ, Prosselkov P, Matthews JM, Attard P, Beck JL, Dixon NE, Otting G. Stabilization of native protein fold by intein-mediated covalent cyclization. J Mol Biol 2005; 346: 1095–108.
98. Deschuyteneer G, Garcia S, Michiels B, Baudoux B, Degand H, Morsomme P, Soumillion P. Intein-mediated cyclization of randomized peptides in the periplasm of Escherichia coli and their extracellular secretion. ACS Chem Biol 2010; 5: 691–700.
99. Volkmann G, Murphy PW, Rowland EE, Cronan JE Jr, Liu XQ, Blouin C, Byers DM. Intein-mediated cyclization of bacterial acyl carrier protein stabilizes its folded conformation but does not abolish function. J Biol Chem 2010; 285: 8605–14.
100. Chan DI, Vogel HJ. Current understanding of fatty acid biosynthesis and the acyl carrier protein. Biochem J 2010; 430: 1–19.
101. Chong S, Williams KS, Wotkowicz C, Xu MQ. Modulation of protein splicing of the Saccharomyces cerevisiae vacuolar membrane ATPase intein. J Biol Chem 1998; 273: 10567–77.
102. Peroza EA, Freisinger E. Tris is a non-innocent buffer during intein-mediated protein cleavage. Protein Expr Purif 2008; 57: 217–25.
103. Mills KV, Paulus H. Reversible inhibition of protein splicing by zinc ion. J Biol Chem 2001; 276: 10832–8.
104. Mills KV, Connor KR, Dorval DM, Lewandowski KT. Protein purification via temperature-dependent, intein-mediated cleavage from an immobilized metal affinity resin. Anal Biochem 2006; 356: 86–93.
105. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M. Modulation of intein activity by its neighboring extein substrates. Proc Natl Acad Sci USA 2009; 106: 11005–10.
106. Iwai H, Zuger S, Jin J, Tam PH. Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. FEBS Lett 2006; 580: 1853–8.
107. Poland BW, Xu MQ, Quiocho FA. Structural insights into the protein splicing mechanism of PI-SceI. J Biol Chem 2000; 275: 16408–13.
108. Mizutani R, Nogami S, Kawasaki M, Ohya Y, Anraku Y, Satow Y. Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides. J Mol Biol 2002; 316: 919–29.
109. Klabunde T, Sharma S, Telenti A, Jacobs WR Jr, Sacchettini JC. Crystal structure of GyrA intein from Mycobacterium xenopi reveals structural basis of protein splicing. Nat Struct Biol 1998; 5: 31–6.
110. Romanelli A, Shekhtman A, Cowburn D, Muir TW. 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 2004; 101: 6397–402.
111. Zettler J, Schutz V, Mootz HD. The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein trans-splicing reaction. FEBS Lett 2009; 583: 909–14.
112. Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M. A genetic system yields self-cleaving inteins for bioseparations. Nat Biotechnol 1999; 17: 889–92.
113. Tan G, Chen M, Foote C, Tan C. Temperature-sensitive mutations made easy: generating conditional mutations by using temperature-sensitive inteins that function within different temperature ranges. Genetics 2009; 183: 13–22.
114. Lockless SW, Muir TW. Traceless protein splicing utilizing evolved split inteins. Proc Natl Acad Sci USA 2009; 106: 10999–1004.
115. Hiraga K, Soga I, Dansereau JT, Pereira B, Derbyshire V, Du Z, Wang C, Van Roey P, Belfort G, Belfort M. Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. J Mol Biol 2009; 393: 1106–17.
116. Garbe D, Thiel IV, Mootz HD. Protein trans-splicing on an M13 bacteriophage: towards directed evolution of a semisynthetic split intein by phage display. J Pept Sci 2010; 16: 575–81.
117. Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, Lesage G, Rao Z. Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. J Biol Chem 2003; 278: 39133–42.

Received February 4, 2011; accepted March 10, 2011