12

Protein Ligation by HINT Domains

Hideo Iwaï and A. Sesilja Aranko

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, Helsinki, Finland

12.1 Introduction

Protein splicing was first reported in the early 1990s [1,2]. The intervening protein sequence responsible for self-excision and concomitant ligation of the flanking peptide sequences (i.e., protein splicing) is termed intein (*in*ternal pro*tein*) [3]. Inteins are analogous to self-splicing group I introns in RNA but self-catalytically splice on the protein level. It is generally accepted that protein splicing catalyzed by a canonical intein involves the four concerted reaction steps of (i) N—S(O) acyl shift, (ii) trans-(thio)esterification, (iii) Asn cyclization, and (iv) S(O)–N acyl shift (Figure 12.1a) [4]. The four-step reaction couples the breakage of the peptide bonds connecting an intein and the flanking polypeptides (N- and C-exteins) with the formation of a covalent peptide bond between the cleaved N- and C-exteins. Protein splicing is solely mediated by the intervening sequence (intein) without any additional cofactors or accessory proteins. Biological significance of inteins is still obscure. Elimination of inteins has no effect on the fitness of their host organisms [5,6]. Therefore, inteins have been considered to be mere parasitic elements although regulatory functions have been suggested for specific inteins [7,8]. The unique chemical reactions involved in protein splicing have been appreciated and utilized for various biotechnological applications such as protein purification, protein modification, and protein ligation [9–11]. Hundreds of inteins have been identified to date from different unicellular host organisms [12]. Recently, more variations in the protein-splicing mechanism have been discovered additionally to the canonical four-step splicing reaction [13,14]. Structural studies identified other closely related domains such as Hedgehog C-terminal processing domains

Chemical Ligation: Tools for Biomolecule Synthesis and Modification, First Edition.
Edited by Luca D. D'Andrea and Alessandra Romanelli.

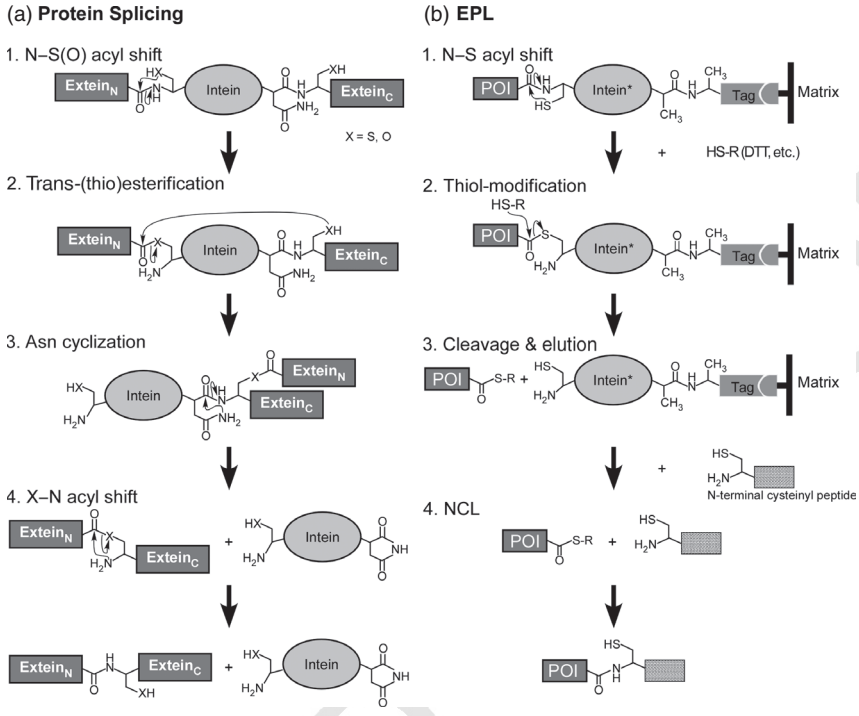

Figure 12.1 Schematic presentation of the chemical mechanism of protein splicing and expressed protein ligation (EPL). (a) The four reaction steps involved in protein splicing by canonical inteins. Step 1, N—S(O) acyl shift: thiol or hydroxyl group of the first residue of the intein attacks the preceding peptide bond forming a (thio)ester bond; step 2, trans-(thio) esterification: thiol or hydroxyl group of the nucleophilic +1 residue replaces the (thio)ester bond formed by the first residue of the intein resulting in a "branched intermediate"; step 3, Asn cyclization: the intein is cleaved off from the exteins; step 4, S(O)—N acyl shift: the released branched (thio)ester intermediate undergoes spontaneous rearrangement to form a more energetically stable peptide bond between N- and C-exteins. (b) Chemical reaction steps of EPL. The precursor protein containing a protein of interest (POI) fused to the inactivated intein is immobilized on the column by an affinity tag for purification. Step 1, N—S acyl shift: the first step of EPL is N—S acyl shift induced by the partially inactivated intein; step 2, thiol modification and cleavage: thiol agent cleaves the thioester bond in the precursor to release POI with the C-terminal thioester; step 3, elution: cleaved POI with the C-terminal thioester is eluted from the column; step 4, NCL step: the released POI with the C-terminal thioester reacts with an N-terminal cysteinyl peptide to form a peptide bond by NCL. Extein$_N$ and Extein$_C$ stand for N- and C-exteins, respectively. Asterisks indicate the mutated intein.

(Hh-C) and bacterial intein-like (BIL) domains sharing the same Hedgehog/INTein (HINT) fold [15,16].

Structural investigations of the HINT superfamily provided the structural and mechanistic basis for further applications of protein-splicing domains. In

this chapter, we discuss the potential use of various HINT domains as protein ligation tools for connecting polypeptide chains.

12.2 Protein Ligation by Protein Splicing

There are two very different approaches for intein-mediated protein ligation. Both approaches utilize an intein or a protein-splicing domain and are thus often considered to be the same approach involving protein splicing. The first approach makes use of inteins in order to create a C-terminally thioester-modified precursor for the subsequent native chemical ligation (NCL) reaction (Figure 12.1b). To produce a C-terminally thioester-modified target protein required for chemoselective ligation by NCL, a target protein is fused upstream of a modified intein for bacterial overexpression. The modified intein is partially inactive and catalyzes only the first N—S acyl migration step (Figure 12.1b). The intermediate thioester bond can be cleaved by a thiol agent to produce an α-thioester bond at the C-terminus of the fused target protein. The thioester-modified protein can be subsequently ligated with an N-terminal cysteinyl peptide via NCL (Figure 12.1b). This intein-mediated chemoselective ligation is often called expressed protein ligation (EPL) or intein-mediated protein ligation (IPL) [17,18]. The ligation step in EPL is identical to NCL. Hence, the limitations and required conditions of NCL are also relevant to EPL [19]. In addition, the stability of the thioester intermediate induced by the modified intein is very sensitive to the amino-acid type at the splicing junction [20,21]. Due to the instability of the thioester intermediate, cleavage of the fused protein could already take place in bacterial cells, thereby reducing the yield of thioester-modified target proteins [20,22]. This premature cleavage might hamper general applications of EPL to various proteins [20,22]. EPL also requires a Cys residue at the N-terminus of the C-terminal peptide fragment, although the use of auxiliary groups could circumvent this limitation [23]. The second approach for protein ligation is to make use of protein trans-splicing (PTS) by split inteins, fully exploiting the complete reaction steps of protein splicing (Figure 12.2). PTS is based on the fact that an intein can be split into two fragments, the function (splicing) of which can be reconstituted by reassembling the two split fragments into a protein-splicing active form (Figure 12.2a) [24–26]. Therefore, the two polypeptides fused with the split intein fragments can be ligated with a peptide bond by protein splicing in trans. To date, more than 100 split inteins have been found in nature or engineered from cis-splicing inteins [27]. The location of the split site and the lengths of the split fragments previously reported vary considerably, although split inteins at the conserved homing endonuclease (HEN) insertion site seem to be the most active ones (Figure 12.2b) [27]. These two approaches of EPL and PTS appear to be very similar because of the involvement of an intein but

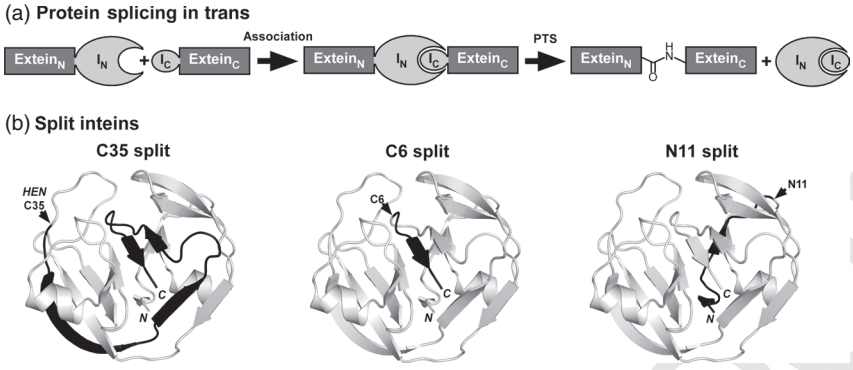

Figure 12.2 Protein trans-splicing (PTS) and the split sites on the intein structure. (a) Schematic presentation of PTS. An intein is split into interacting halves, termed N- and C-inteins (I_N and I_C). (b) Cartoon presentation of the three-dimensional structure of a HINT domain (*Npu*DnaE intein, PDB ID: 2KEQ) showing the natural split site ("C35") and two artificially engineered split sites ("C6" and "N11"). C35 site is located at the conserved HEN insertion site before 35 residues from the C-terminus. The C6 and N11 split inteins were created by shifting the split site toward either C- or N-termini to have 6-residue C-intein and 11-residue N-intein, respectively. The shorter fragments of the split inteins are shown in black. Black arrows indicate the split sites and the conserved HEN insertion site. N and C show the N- and C-termini, respectively.

significantly differ when it comes to the practical use. For example, PTS is not limited to Cys residue at the ligation site, but other nucleophilic residues of Ser and Thr can replace Cys, permitting the application to a wider range of proteins [4].

12.3 Naturally Occurring and Artificially Split Inteins for Protein Ligation

Protein splicing occurs spontaneously and self-catalytically after translation and protein folding, because only the correctly folded three-dimensional structure of an intein is required for the splicing activity. One way to control protein splicing is to use PTS by splitting an intein into two pieces and activating protein-splicing reaction by reassembling the two split fragments upon mixing the two precursors as previously demonstrated [24–26]. This was first demonstrated by reassembling two separately prepared precursor fragments *in vitro* using artificially split inteins [24–26]. This approach has been particularly useful for preparing segmentally isotope-labeled proteins for reducing the complexity of nuclear magnetic resonance (NMR) spectra, because a part of the protein of interest (POI) can be differentially isotope-labeled by preparing one precursor in the labeled medium that could be ligated to the other unlabeled

precursor by PTS (Figure 12.3a) [26,28,29]. Because of the poor solubility of these artificially split intein precursors, the two precursor fragments were usually refolded from denaturing conditions. This refolding process could be labor-intensive and time-consuming, requiring optimization and multiple-step purifications [28,29]. Therefore, the practical application has been restricted to proteins that can be refolded *in vitro*. However, nature also exploits the same approach presumably in order to control the host protein function. Naturally occurring split DnaE intein was first discovered in *Synechocystis* sp. PCC6803 (*Ssp*DnaE intein) [30]. Unlike artificially split inteins, naturally split DnaE inteins seem to be more soluble and capable of trans-splicing *in vivo* without denaturation/renaturation, although trans-splicing of *Ssp*DnaE intein was not very efficient for many applications [30,31]. Our group discovered that the naturally split DnaE intein from *Nostoc punctiforme* (*Npu*DnaE intein) has much more robust trans-splicing efficiency and higher tolerance of any changes at the C-terminal splicing junction compared to *Ssp*DnaE intein [32]. The discovery of *Npu*DnaE intein has significantly improved yields of protein ligation by PTS and was followed by other identifications of similar robust split inteins [32–35]. Soluble and robust split inteins opened new *in vivo* applications of PTS including *in vivo* segmental isotopic labeling, genomic editing, and possible gene therapy [31,36,37].

Naturally occurring split inteins commonly lack HEN, which has been found in many inteins. The split sites of almost all naturally occurring split inteins correspond to the conserved HEN insertion site of canonical inteins (Figure 12.2b). Therefore, N- and C-terminal split intein fragments typically consist of about 90–120 residues and 30–50 residues, respectively [38–40]. Even though these intein fragments seem to be more soluble compared to artificially split inteins, the split fragments are relatively too long to be chemically synthesized. This is the major limiting factor for protein semisynthesis using PTS. PTS can be used for protein semisynthesis, in which one of the split intein precursors is chemically synthetised and the other precursor is recombinantly prepared prior to protein ligation by PTS (Figure 12.3b). Semisynthesis using PTS could take advantage of both recombinant protein expression and chemical synthesis. PTS-mediated protein semisynthesis has been thus used in numerous applications including fluorescent labeling, biotinylation, site-specific introduction of a glycosylation, and protein immobilization [41–43]. Furthermore, PTS permits highly specific labeling *in vivo* if synthetised peptides can penetrate into cells, for example, using cell-penetrating peptides [44]. The relatively long split fragments (typically longer than 30 residues) for chemical synthesis, however, diminish advantages of protein semisynthesis by PTS. Therefore, there have been many attempts to create split inteins with shorter fragments that could be easily prepared chemically [45–51]. Artificial split sites created from several inteins based on the three-dimensional structures of inteins were systematically tested for protein-splicing activity and revealed that the C35 site

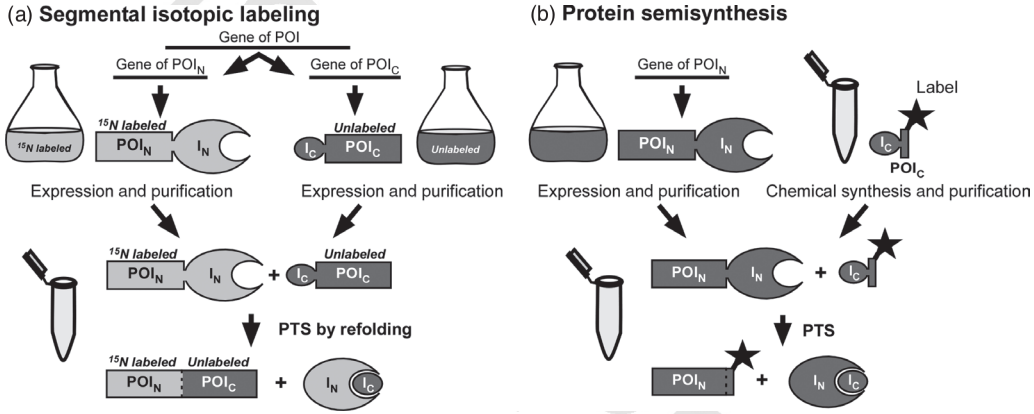

Figure 12.3 Segmental isotope-labeling and protein semisynthesis by PTS. (a) Segmental isotopic labeling *in vitro*. The N- and C-terminal fragments of POI are fused to N- and C-inteins, respectively. One of the precursors is expressed in isotopic (^{15}N)-labeled culture medium and the other precursor in unlabeled culture medium. The two precursors are purified separately and mixed together for PTS by refolding. (b) Protein semisynthesis. An N-terminal fragment of POI is fused with N-intein and recombinantly expressed. The C-terminal precursor containing C-intein and a chemical label is chemically prepared. The two precursor fragments are mixed to initiate PTS, thereby producing the semisynthetically ligated product. I_N and I_C stand for N- and C-inteins, respectively. POI_N and POI_C stand for the N- and C-terminal fragments, respectively, of POI.

(35 residues from the C-terminus of the *Npu*DnaE intein and at the HEN insertion site observed in many canonical inteins) is the best split site to create a functional split intein and corresponds to the native split site of naturally split inteins [27,47]. Albeit lower efficiency, the shortest functional intein fragments discovered consist of 11 residues for N-terminal split intein fragment (N11 split) and 6 residues for the C-terminal intein fragment (C6 split) (Figure 12.2b) [45,47,50,51]. These split sites can be of practical importance because the shorter length allows to chemically synthesize the precursor fragment with possibilities to incorporate nonnatural or modified amino acids. Unfortunately, protein ligation yields seem to be much lower when these shorter fragments are used, particularly for *in vitro* protein ligation. This is probably due to the fact that inteins are presumably not optimized for these split sites, thereby predominantly inducing side reactions of the N- and/or C-cleavages instead of the productive splicing reaction and/or remaining in the inactive conformation. Intein sequences bearing a split site at these positions would be highly desirable for semisynthesis but probably need to be optimized for better splicing activity. Directed evolution of inteins might be able to provide such optimized sequences [52].

12.4 Conditional Protein Splicing

Controlled protein splicing is often called conditional protein splicing (CPS) and has been of great interest because it allows posttranslational control of protein functions. Intein-mediated posttranslational regulation of protein functions might be more effective compared to transcriptional-level control. Most of two split intein fragments of naturally or artificially split inteins can interact with each other to form an active intein fold within minutes to hours. In some cases, however, the interaction between the split intein halves is slow and/or weak [25,53,54]. One of these weakly interacting split inteins was exploited to create a small-molecule inducible split intein [55]. Two binding partners (FK506 binding protein (FKBP) and FKBP-rapamycin binding (FRB) domains) were fused with the split intein fragments to increase their affinity in the presence of rapamycin because FKBP and FRB domains interact only in the presence of rapamycin with high affinity [55]. Therefore, a small molecule rapamycin could initiate the association of the fused split intein halves, thereby triggering protein splicing (Figure 12.4a) [55]. Despite potential usefulness of rapamycin-controlled CPS, there are a few limitations such as the toxicity of rapamycin. Other strategies for CPS have also been proposed [56,57]. For example, photolabile protective groups stalling the splicing reaction have been used for light-inducible PTS (Figure 12.4b) [56,57]. Not only trans-splicing but also cis-splicing by inteins has been controlled by small molecules, redox status, pH, and temperature [58–62]. However, turning on/off sharply without

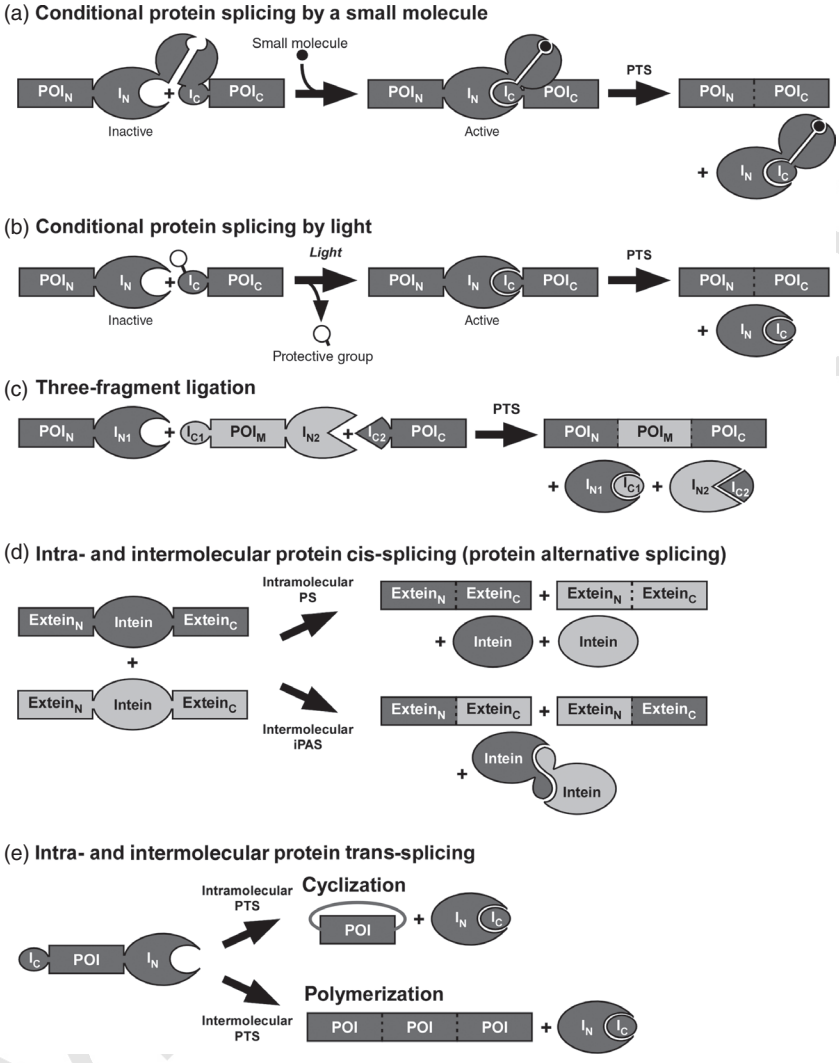

Figure 12.4 Intra- and intermolecular applications with protein cis-splicing (PS) and trans-splicing (PTS). (a) Small-molecule induced CPS: Split intein halves with low affinity are fused with two interacting domains in the presence of a small molecule. Addition of a small molecule induces PTS upon association of the two precursors. (b) Light-induced CPS: light removes a protective group stalling PTS. (c) Three-fragment ligation. POI is split into three pieces (POI_N, POI_M, and POI_C) and fused with two orthogonal split inteins (I_{N1}/I_{C1} and I_{N2}/I_{C2}). (d) Intein-mediated protein alternative splicing (iPAS). Intermolecular protein splicing by protein 3D-domain swapping produces alternatively ligated products instead of cis-spliced products. (e) Cyclization and polymerization. Circularly permutated precursor bearing a split intein ligates the N- and C-termini of a protein by intramolecular protein splicing, thereby resulting in the backbone cyclization. The identical precursor protein could result in polymerization when intermolecular interactions dominate intramolecular interactions. I_N and I_C stand for N- and C-inteins, respectively.

residual activity is a challenge for conditional splicing, and the applications might be limited.

12.5 Inter- and Intramolecular Protein Splicing

Protein splicing is not limited to ligation of two flanking sequences of an intein but can be extended to other intra- and intermolecular reactions such as three-fragment ligation and intramolecular backbone cyclization. Three-fragment ligation by two PTS reactions is particularly useful because it could theoretically allow labeling or modifying a region of a protein at arbitral positions. However, three-fragment ligation requires two split inteins that are not cross-reactive (orthogonal). When two orthogonal split inteins are inserted into a POI, the intervened host protein is thus split into three fragments, which can be religated into one protein by two PTS reactions (Figure 12.4c). To obtain a fully ligated product by two PTS reactions, high efficiencies of individual splicing reactions are desirable. However, naturally split DnaE inteins with highly efficient splicing activity such as *Npu*DnaE intein are cross-reactive to each other due to their high sequence homology [32,34,35]. When two cross-reactive split inteins were used for three-fragment ligation, backbone cyclization of the central precursor protein as well as undesirable intermolecular reaction skipping the central precursor protein could occur due to their cross-activity (Figure 12.4c). Two orthogonal split inteins with robust ligation efficiency are required to obtain sufficient amount of the fully ligated product by three-fragment ligation. Alternatively, this specificity problem has been circumvented by semiorthogonal split inteins with different split fragment lengths or by utilizing differences in ligation kinetics [63,64].

Protein splicing is originally found to be an intramolecular reaction ligating the two flanking sequences of an intein in cis. It has been believed that protein splicing does not happen between different molecules as an intermolecular reaction except for split inteins [65]. However, we demonstrated that not only split inteins catalyze intermolecular reaction but also cis-splicing inteins could be involved in an intermolecular reaction, which was termed intein-mediated protein alternative splicing (iPAS) (Figure 12.4d) [66]. iPAS could occur not only between a cis-splicing intein and a split intein that are cross-reactive but also between two cis-splicing inteins with high homology, although this reaction is very inefficient [66]. It is still unclear whether this alternative splicing on the protein level is a naturally occurring phenomenon because iPAS does not leave any marks on the spliced product, making it difficult to identify naturally occurring iPAS [66]. iPAS could potentially increase the molecular diversity of the ligated proteins with an identical genetic background. For example, two cis-splicing precursors can produce up to four different molecular species by iPAS in a combinatorial manner (Figure 12.4d). iPAS could also modulate protein functions [66]. Thus, intermolecular protein splicing such

as iPAS might be of practical use for regulating protein functions as a biotechnological tool.

Another unique feature of protein splicing is intramolecular PTS, which can be used to produce proteins and peptides with circular backbones by connecting N- and C-termini with a peptide bond [67,68]. Circular permutation of the precursor protein containing a cis-splicing intein by introducing new termini within an intein and by connecting the N- and C-termini of the host protein will result in backbone cyclization of the host protein (Figure 12.4e). Backbone cyclization has been used for stabilizing proteins by simply connecting the two ends with a peptide bond, because of the reduced entropy in the unfolded status [67–70]. There are many naturally occurring backbone-cyclized peptides identified in nature. These peptides are usually more stable compared to the linear forms and tend to have diverse pharmaceutically interesting biological functions such as antimicrobial and immunosuppressive activities. Natural cyclic peptides are usually synthetised by nonribosomal peptide synthetases, and thus, their recombinant productions are not straightforward. Up to date, no naturally occurring cyclic peptide/protein produced by inteins has been reported. However, peptide cyclization by inteins has been a promising method for producing a library of cyclic peptides *in vivo* for screening bioactive peptides [67,71]. Protein/peptide cyclization by circular permutation of an intein-containing precursor relies on the favorable intramolecular reaction (Figure 12.4e). It is noteworthy that intermolecular reaction with the identical precursor would result in polymerization of the target protein rather than cyclization when favorable conditions for intermolecular reaction were used such as high concentration of the precursor protein (Figure 12.4e) [72].

12.6 Protein Ligation by Other HINT Domains

Inteins share the same three-dimensional architecture with the C-terminal domain of Hedgehog protein (Hh-C), which is termed HINT fold (Figure 12.5a) [15]. Hh-C catalyzes cholesterol modification of the N-terminal domain of the Hedgehog protein (Hh-N). Hh-C induces N—S acyl shift at the N-terminal junction as inteins do, which is followed by the nucleophilic attack of a hydroxyl group in cholesterol that binds to the C-terminal cholesterol binding domain of Hh-C. The first step of protein splicing can be thus catalyzed by HINT fold of Hh-C. Cholesterol modification of Hh-N by Hh-C plays an essential role in the embryonic development unlike obscure biological functions by many inteins. There are also a number of intein-like sequences in bacteria that are similar to mini-inteins because of the absence of obvious HEN domain and the similar sequence motifs found in inteins [73,74]. These BIL domains also share the same HINT fold common to inteins and Hh-C (Figure 12.5a) [16]. BIL domains are characteristic in their distribution, biological functions, and differences in

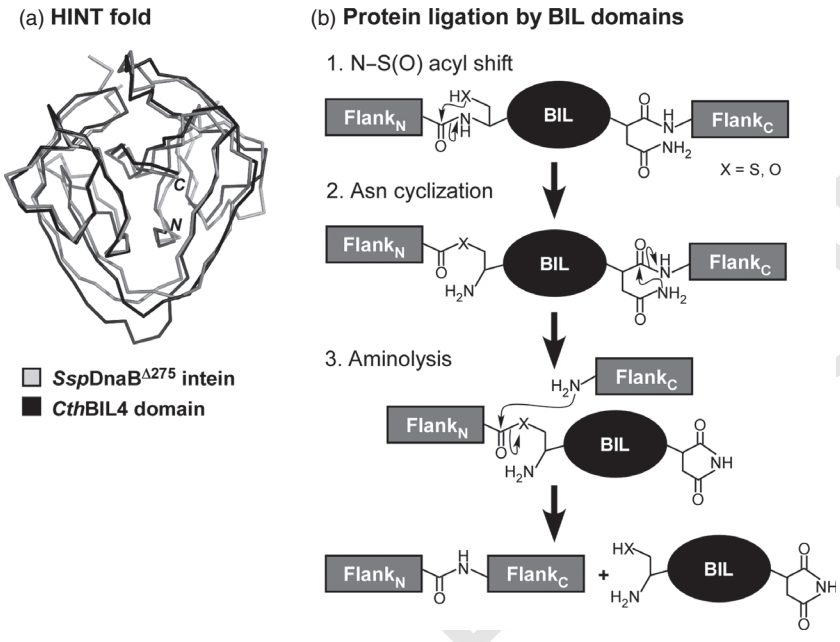

Figure 12.5 HINT domains and protein splicing by BIL domain. (a) A superposition of the structures of DnaE intein from *Synechocystis* sp. PCC6803 (*Ssp*DnaBΔ^{275} intein, PDB ID: 1MI8) colored in light gray and BIL4 domain from *Clostridium thermocellum* (*Cth*BIL4, PDB ID: 2LWY) colored in black. N- and C indicate the N- and C-termini. (b) The mechanism of protein splicing by BIL domains involves the following steps: step 1, N—S(O) acyl shift forming a (thio)ester bond; step 2, Asn cyclization cleaving off the C-flank from the BIL domain; step 3, aminolysis by the N-terminal amine group of the cleaved C-flank. $Flank_N$ and $Flank_C$ stand for the N- and C-terminal flanking sequences, respectively.

the conserved sequence motifs. Importantly, many BIL domains lack the +1 nucleophilic residue that catalyzes the crucial trans-(thio)esterification step in protein splicing [75]. BIL domains are capable of N—S(O) acyl shift like inteins and Hh-C, but no trans-(thio)esterification will occur because they lack the +1 nucleophilic residue. Their biological functions are still unclear. It is likely that BIL domains function as internal proteases cleaving the flanking sequences instead of ligation [73,75]. However, minute amount of protein ligation has been reported by BIL domains due to intermolecular aminolysis reaction after the cleavage (Figure 12.5b) [75]. Ligation by BILs domains is too inefficient to be of practical use. However, we demonstrated that a BIL domain could be converted to an efficient protein-splicing domain simply by introducing the C-terminal nucleophilic residue (+1Cys). This also suggests that BIL domains have probably evolved from mini-inteins [16]. It was also possible to catalyze protein trans-splicing by splitting BIL domains into halves as split inteins [16]. Albeit very low efficiency, BIL domains are capable of protein ligation by the

split BIL domains without the C-terminal +1 nucleophilic residue required for protein splicing, that is, Cys, Thr, or Ser [16]. This indicates a great potential of BIL domains for protein ligation. It might be possible to improve the ligation efficiency by applying further protein-engineering such as directed evolution. Efficient protein ligation without any nucleophilic residue at the +1 position could substantially widen the applications of PTS by split HINT domains because it could overcome the limitation imposed by the sequence requirement of Cys, Ser, or Thr at the splicing junction.

12.7 Bottleneck of Protein Ligation by PTS

One of the important practical aspects in protein ligation using PTS is the junction sequence dependency near the splicing junctions. The +1 position following the intein is essential for trans-esterification step as a nucleophile and is always one of Cys, Ser, or Thr in inteins (Figures 12.1a and 12.6a) [76]. The neighboring residues of the splicing junction (the –1 and +2 positions) are not directly involved in the protein splicing reaction but can significantly influence the protein-splicing efficiency (Figure 12.6) [32,77–79]. The structure–function relationship of this splicing-junction dependency is not well understood. Therefore, at least one to three native flanking residues are often kept in the precursor protein, resulting in an extra foreign sequence insertion when PTS is used for the ligation of foreign proteins. These linker residues at the splicing junction will remain as a "scar" in the ligated product and might disturb the function of the ligated target protein (Figure 12.6a). Unfortunately, the junction sequence dependency has not been investigated in detail except for a few inteins. It is still unpredictable how amino-acid residue types near the ligation junctions influence the splicing efficiency [32,77–79]. It is thus necessary to check the ligation efficiency by PTS in a trial-and-error manner by testing several junction sequences for individual cases. Interestingly, the amino-acid type preference is strongly specific to individual inteins, and some inteins are more tolerant of changes at the splicing junctions than others (Figures 12.6b and c) [32]. For example, *Npu*DnaE and *Npu*DnaB inteins have completely different preferences of amino-acid types (Figures 12.6b and c) [32]. In contrast to protein trans-splicing by split inteins, NCL does not have this requirement. However, branched amino-acid types at the C-terminus are known to be unfavorable in NCL [80].

12.8 Comparison with Other Enzymatic Ligation Methods

Ligation of two polypeptide chains by an enzyme has been a dream for chemists, which could be theoretically achievable by reversing the proteolytic

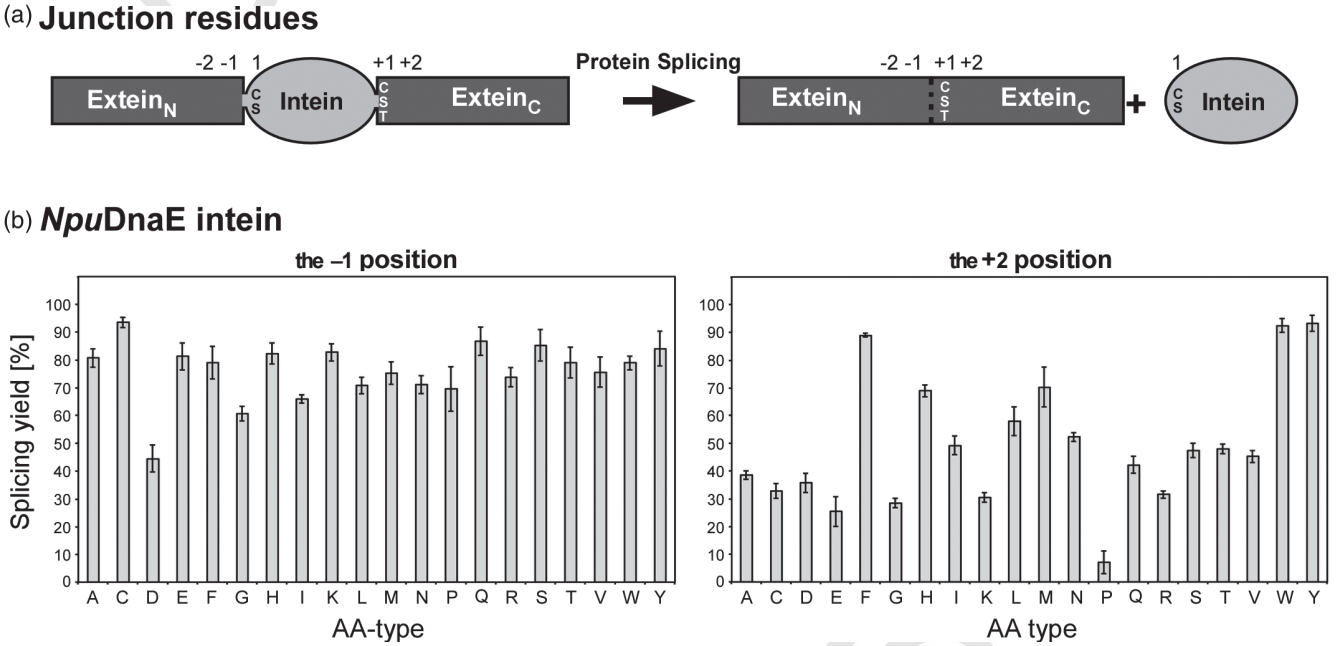

Figure 12.6 Junction sequence dependencies of different inteins. (a) Numbering of junction residues of an intein-containing precursor. The –1 and –2 positions are the first and second residues preceding the intein sequence. The +1 and +2 positions are the first and second residues following the intein sequence. The first residue of an intein is indicated as "1." The +1 residue is the nucleophilic residue and either Cys or Ser or Thr in inteins. (b) Cis-splicing efficiencies of DnaE intein from *Nostoc punctiforme* (*Npu*DnaE intein) with 20 different amino-acid types at the N- and C-terminal junctions (the –1 and +2 positions).

(c) ***Npu*DnaB$^{\Delta 290}$ intein**

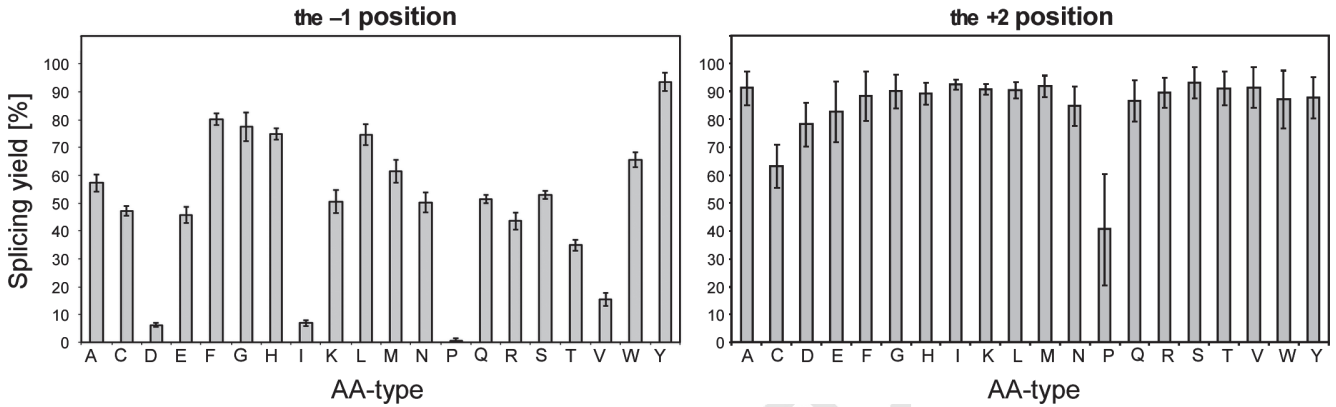

Figure 12.6 (Cont'd) (c) Cis-splicing efficiencies of minimized DnaB intein from *Nostoc punctiforme* (*Npu*DnaBΔ^{290} intein) with 20 different amino-acid types. The left and right panels show the data for the −1 position and the +2 position, respectively. Amino-acid types are indicated by one-letter codes at the bottom. The standard deviations of the mean ($n = 3$) are shown as error bars.

reaction of a protease. This was first demonstrated by an engineered variant of the protease subtilisin BPN', termed subtiligase [81]. Subtiligase catalyzes aminolysis of a peptide ester efficiently and was used for total synthesis of ribonuclease A analogue [82]. The availability of peptide esters as the substrates could restrict its general application. However, subtiligase has recently been employed for N-terminal labeling in proteomic researches where defined peptide esters were used as the N-terminal label [83]. Sortases from *Staphylococcus aureus* are natural peptide ligases catalyzing *trans*-peptidase reactions to covalently link bacterial surface proteins to the cell wall [84]. By mimicking the naturally occurring *trans*-peptidase reaction of sortase A, sortase-mediated protein ligation has been developed using a conserved recognition sequence of "LPXTG" from the bacterial surface proteins in the N-terminal substrate and the N-terminal three-glycine residues of the cell-wall peptide glycan in the C-terminal substrate [84,85]. Sortase forms a thioester bond with the substrate when "LPXTG" sequence is cleaved after the Thr [86]. This thioester intermediate is then released by nucleophilic attack of the N-terminal amine of the polyglycine-containing protein/peptide. This step is very similar to protein splicing by BIL domains [75]. It is noteworthy that the ligated product still contains an "LPXTGGG" sequence as a linker after ligation, which is also the substrate of sortase. This additional linker requirement for sortase-mediated ligation could be potentially an obstacle for some applications. Compared with typical proteases, the turnover by sortase seems to be very slow [85]. Therefore, an extremely high molar ratio of sortase was often added to substrate proteins, even at 1:1 (substrate:enzyme) molar ratio [85]. Moreover, productive ligation reaction by sortase needs to compete with both hydrolysis of thioester intermediate and aminolysis by the released glycine from the "LPXTG" sequence. Therefore, the N-terminal triglycine-containing substrate is added in excess at the concentration of sub mM [85,86]. High concentrations (mM range) are also generally used in the ligation reactions by subtiligase and EPL (Table 12.1) [21,87]. PTS is typically performed in μM concentrations or even less due to high affinity between the intein halves [35,47]. Table 12.1 summarizes the conditions and requirements for various protein ligation methods including PTS. All of the protein ligation methods listed require some linker sequences (Table 12.1). In the case of PTS, substrate is covalently attached to the enzyme (intein), thereby having only a single turnover. Because of the covalently attached substrates, inteins are generally promiscuous accepting various sequences as substrates. For practical applications, native splicing junction sequences are often used in PTS by split inteins because of the aforementioned junction sequence dependency. For example, a junction sequence of "AEY/CFN" or a longer one was often used for *Ssp*DnaE intein [34,72,79]. In the case of *Npu*DnaE intein, "GS/CFNGT" including the sequences from the cloning sites (*Bam*HI and *Kpn*I) is often

Table 12.1 Comparison of the different protein ligation methods.

	PTS (intein)	PTS (BIL)	EPL	Sortase	Subtiligase
Typical preparation time	~2 d/~4 d (*In vivo*/*in vitro*) [36]	~2 d/~4 d (*In vivo*/*in vitro*) [16]	>7 d [22]	~2 d/~9 d (*In vivo*/*in vitro*) [88,89]	~3 d
Ligation efficiency	~50–100% [33,36,47]	~90%/<5% (+1C/+1X) [16] (X = A, V)	10–100% [19]	~50–90%[a] [85,89]	~50–90% [87]
Reaction time	20 s–24 h [33,36,47]	~4 h–24 h (+1C/+1X) [16]	~4 h–48 h (for NCL) [19]	~24 h [85]	<1 h [87]
Typical sample concentration	nM–μM [33,36,47,64]	<μM [16]	mM [19]	>μM–mM [84,85]	mM [87]
Linker	+1 residue = C, S, T; sensitive to flanking residues [77–79,90]	+1 residue = A, V, G, S, C, T[b] [16]	+1 residue = C; sensitive to flanking residues [19]	LPXT/GGG[c] (X = D/E/A/N/Q/K) [84,85]	Sensitive to ~7aa residues in the ligation junction [87]
In vivo/*in vitro*	*In vivo*/*in vitro* [36]	*In vivo*/*in vitro* [16]	*In vitro*	*In vivo*/*in vitro* [88,89]	*In vitro*
Multiple-fragment ligation	Yes (stepwise/one-pot) [63,64]	N.R.	Yes (stepwise) [91]	Yes (stepwise) [92]	Yes (stepwise) [82]

N.R., not reported.

a Defined from the fragments that are not in excess.

b Only these have been tested, likely to be sensitive to flanking residues.

c Some variations were reported.

used at the junction in order to obtain higher ligation yields [32,36,93]. Theoretically, only "+1Cys" is minimally required for *Npu*DnaE intein. The mechanism of the splicing junction dependency is still unclear and remains to be elucidated for further development of PTS. PTS is especially a powerful method for *in vivo* applications because it is typically highly specific at a low concentration and self- and autocatalytic. Although *in vivo* ligation using sortase has also been demonstrated, reversible enzymatic reaction could be a major bottleneck (Table 12.1) [88]. Three-fragment ligation is generally even more challenging with EPL, sortase, and subtiligase as compared to one-pot approach using PTS, because they require protective groups and/or stepwise ligation reactions (Table 12.1) [82,91,92]. High yields might be achievable by several different ligation methods under ideally optimized conditions. However, all the aforementioned methods have both advantages and disadvantages. Specific needs for the application are likely to define the best choice of the ligation method.

12.9 Perspective of Protein Ligation by HINT Domains

Needs for protein ligation can be very different for individual applications. For example, most of fluorescent labeling applications might not require very high yields. A few percentages of the ligation efficiency might be sufficient for applications such as visualization of labeled proteins. Many existing artificially split inteins are probably efficient enough for this type of application. BIL domains can also be used without nucleophilic Cys, Ser, and Thr for producing small amounts of ligated products [16]. On the other hand, segmental isotopic labeling or total semisynthesis of a protein would require much higher yields with high efficiency due to the cost of isotopes and chemical synthesis. Especially, multiple-fragment ligation requires higher ligation efficiencies not only because the overall yield is determined by a multiple of the efficiencies for individual ligation steps but also because incomplete ligation complicates the purification of ligated products from the precursors. For practical use, the ligation yield of one ligation step should be >50% considering the following purification steps. Because slight modifications near the splicing junctions could influence the ligation efficiency significantly, it is still important to optimize the junction sequences. Understanding the structure–function relationship of HINT domains is certainly important for rational design of the optimal junction sequences and eliminating laborious case-by-case optimization. Ultimately, highly promiscuous split inteins with robust ligation efficiency would be needed for demanding applications of protein ligation by PTS.

12.10 Conclusions and Future Perspectives

Protein splicing has steadily and increasingly become an important tool for protein ligation despite some drawbacks. Protein ligation by PTS using split inteins is especially versatile because of high specificity and self-catalytic reaction without additional cofactors and accessory proteins. Protein ligation by PTS has opened new possibilities to perform protein engineering *in vivo* and to create proteins that are not even coded in the genes. Robust split inteins with high efficiency are still in demand for widening the applications of PTS even further and for fulfilling our requirements necessary for some of our dream experiments. Orthogonal split inteins with high ligation efficiency are highly desirable particularly for multiple-fragment ligation. Even though they have not been widely exploited, other HINT domains are of potential importance not only for protein ligation but also as proteases or for protein modifications. High promiscuity and prevalence of HINT domains in nature are promising for further innovative use of uncharacterized HINT domains in biotechnological applications, which might provide a new means to overcome current limitations of inteins and other HINT domains. After more than two decades since the discovery of protein splicing, we are probably still at the entrance of this fascinating "protein-splicing" field with great potential in protein science and protein engineering.

Acknowledgment

This work was supported by the Academy of Finland (137995, 1277335), Sigrid Jusélius Foundation, and Biocenter Finland.

References

1 Hirata, R.; Ohsumk, Y.; Nakano, A.; Kawasaki, H.; Suzuki, K.; Anraku, Y.: Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem* 1990, **265**, 6726–6733.

2 Kane, P. M.; Yamashiro, C. T.; Wolczyk, D. F.; Neff, N.; Goebl, M.; Stevens, T. H.: Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 1990, **250**, 651–657.

3 Perler, F. B.; Davis, E. O.; Dean, G. E.; Gimble, F. S.; Jack, W. E.; Neff, N.; Noren, C. J.; Thorner, J.; Belfort, M.: Protein splicing elements: inteins and exteins – a definition of terms and recommended nomenclature. *Nucleic Acids Res* 1994, **22**, 1125–1127.

4 Noren, C.; Wang, J.; Perler, F.: Dissecting the chemistry of protein splicing and its applications. *Angew Chem Int Ed Engl* 2000, **39**, 450–466.

5 Naor, A.; Lazary, R.; Barzel, A.; Papke, R. T.; Gophna, U.: In vivo characterization of the homing endonuclease within the polB gene in the halophilic archaeon *Haloferax volcanii*. *PLoS One* 2011, **6**, e15833.

6 Papavinasasundaram, K. G.; Colston, M. J.; Davis, E. O.: Construction and complementation of a recA deletion mutant of *Mycobacterium smegmatis* reveals that the intein in Mycobacterium tuberculosis recA does not affect RecA function. *Mol Microbiol* 1998, **30**, 525–534.

7 Callahan, B. P.; Topilina, N. I.; Stanger, M. J.; Van Roey, P.; Belfort, M.: Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications. *Nat Struct Mol Biol* 2011, **18**, 630–633.

8 Topilina, N. I.; Green, C. M.; Jayachandran, P.; Kelley, D. S.; Stanger, M. J.; Piazza, C. L.; Nayak, S.; Belfort, M.: SufB intein of Mycobacterium tuberculosis as a sensor for oxidative and nitrosative stresses. *Proc Natl Acad Sci USA* 2015, **112**, 10348–10353.

9 hah, N. H.; Muir, T. W.: Split inteins: nature's protein ligases. *Isr J Chem* 2011, **51**, 854–861.

10 Wood, D. W.; Camarero, J. A.: Intein applications: from protein purification and labeling to metabolic control methods. *J Biol Chem* 2014, **289**, 14512–14519.

11 Volkmann, G.; Iwaï, H.: Protein trans-splicing and its use in structural biology: opportunities and limitations. *Mol Biosyst* 2010, **6**, 2110–2121.

12 Perler, F. B.: InBase: the intein database. *Nucleic Acids Res* 2000, **30**, 383–384.

13 Southworth, M. W.; Benner, J.; Perler, F. B.: An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J* 2000, **19**, 5019–5026.

14 Tori, K.; Dassa, B.; Johnson, M. A.; Southworth, M. W.; Brace, L. E.; Ishino, Y.; Pietrokovski, S.; Perler, F. B.: 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**, 2515–2526.

15 Hall, T. M.; Porter, J. A.; Young, K. E.; Koonin, E. V.; Beachy, P. A.; Leahy, D. J.: Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* 1997, **91**, 85–97.

16 Aranko, A. S.; Oeemig, J. S.; Iwaï, H.: Structural basis for protein trans-splicing by a bacterial intein-like domain – protein ligation without nucleophilic side chains. *FEBS J* 2013, **280**, 3256–3269.

17 Muir, T. W.; Sondhi, D.; Cole, P. A.: Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* 1998, **95**, 6705–6710.

18 Evans, T. C. Jr.; Benner, J.; Xu, M. Q.: Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 1998, **7**, 2256–2264.

19 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* 1999, **96**, 10068–10073.

20 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* 1998, **273**, 10567–10577.

21 Xu, M. Q.; Evans, T. C. Jr.: Intein-mediated ligation and cyclization of expressed proteins. *Methods* 2001, **24**, 257–277.

22 Minato, Y.; Ueda, T.; Machiyama, A.; Shimada, I.; Iwaï, H.: Segmental isotopic labeling of a 140 kDa dimeric multi-domain protein CheA from Escherichia coli by expressed protein ligation and protein trans-splicing. *J Biomol NMR* 2012, **53**, 191–207.

23 Low, D. W.; Hill, M. G.; Carrasco, M. R.; Kent, S. B.; Botti, P.: Total synthesis of cytochrome b562 by native chemical ligation using a removable auxiliary. *Proc Natl Acad Sci USA* 2001, **98**, 6554–6559.

24 Mills, K. V.; Lew, B. M.; 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**, 3543–3548.

25 Southworth, M. W.; Adam, E.; Panne, D.; Byer, R.; Kautz, R.; Perler, F. B.: Control of protein splicing by intein fragment reassembly. *EMBO J* 1998, **17**, 918–926.

26 Yamazaki, T.; Otomo, T.; Oda, N.; Kyogoku, Y.; Uegaki, K.; Ito, N.; Ishino, Y.; Nakamura, H.: Segmental isotope labeling for protein NMR using peptide splicing. *J Am Chem Soc* 1998, **120**, 5591–5592.

27 Aranko, A. S.; Wlodawer, A.; Iwaï, H.: Nature's recipe for splitting inteins. *Protein Eng Des Sel* 2014, **27**, 263–271.

28 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* 1999, **38**, 16040–16044.

29 Otomo, T.; Teruya, K.; Uegaki, K.; Yamazaki, T.; Kyogoku, Y.: Improved segmental isotope labeling of proteins and application to a larger protein. *J Biomol NMR* 1999, **14**, 105–114.

30 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* 1998, **95**, 9226–9231.

31 Züger, S.; Iwai, H.: Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. *Nat Biotechnol* 2005, **23**, 736–740.

32 Iwai, H.; Züger, S.; Jin, J.; Tam, P. H.: Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett* 2006, **580**, 1853–1858.

33 Carvajal-Vallejos, P.; Pallisse, R.; Mootz, H. D.; Schmidt, S. R.: Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources. *J Biol Chem* 2012, **287**, 28686–28696.

34 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**, 322–330.

35 Shah, N. H.; Dann, G. P.; Vila-Perello, M.; Liu, Z.; Muir, T. W.: Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. *J Am Chem Soc* 2012, **134**, 11338–11341.

36 Muona, M.; Aranko, A. S.; Raulinaitis, V.; Iwaï, H.: Segmental isotopic labeling of multi-domain and fusion proteins by protein trans-splicing in vivo and in vitro. *Nat Protoc* 2010, **5**, 574–587.

37 Li, J.; Sun, W.; Wang, B.; Xiao, X.; Liu, X. Q.: Protein trans-splicing as a means for viral vector-mediated in vivo gene therapy. *Hum Gene Ther* 2008, **19**, 958–964.

38 Paulus, H.: Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 2000, **69**, 447–496.

39 Caspi, J.; Amitai, G.; Belenkiy, O.; Pietrokovski, S.: Distribution of split DnaE inteins in cyanobacteria. *Mol Microbiol* 2003, **50**, 1569–1577.

40 Dassa, B.; London, N.; Stoddard, B. L.; 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**, 2560–2573.

41 Volkmann, G.; Liu, X. Q.: Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein. *PLoS One* 2009, **4**, e8381.

42 Olschewski, D.; Seidel, R.; Miesbauer, M.; Rambold, A. S.; Oesterhelt, D.; Winklhofer, K. F.; Tatzelt, J.; Engelhard, M.; Becker, C. F.: Semisynthetic murine prion protein equipped with a GPI anchor mimic incorporates into cellular membranes. *Chem Biol* 2007, **14**, 994–1006.

43 Kwon, Y.; Coleman, M. A.; Camarero, J. A.: Selective immobilization of proteins onto solid supports through split-intein-mediated protein trans-splicing. *Angew Chem Int Ed Engl* 2006, **45**, 1726–1729.

44 Giriat, I.; Muir, T. W.: Protein semi-synthesis in living cells. *J Am Chem Soc* 2003, **125**, 7180–7181.

45 Appleby, J. H.; Zhou, K.; Volkmann, G.; Liu, X. Q.: Novel split intein for trans-splicing synthetic peptide onto C terminus of protein. *J Biol Chem* 2009, **284**, 6194–6199.

46 Aranko, A. S.; Züger, S.; Buchinger, E.; Iwaï, H.: In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PloS One* 2009, **4**, e5185.

47 Aranko, A. S.; Oeemig, J. S.; Zhou, D.; Kajander, T.; Wlodawer, A.; Iwaï, H.: Structure-based engineering and comparison of novel split inteins for protein ligation. *Mol Biosyst* 2014, **10**, 1023–1034.

48 Elleuche, S.; Poggeler, S.: Trans-splicing of an artificially split fungal mini-intein. *Biochem Biophys Res Commun* 2007, **355**, 830–834.

49 Lee, Y. T.; Su, T. H.; Lo, W. C.; Lyu, P. C.; Sue, S. C.: Circular permutation prediction reveals a viable backbone disconnection for split proteins: an approach in identifying a new functional split intein. *PLoS One* 2012, **7**, e43820.

50 Oeemig, J. S.; Aranko, A. S.; Djupsjöbacka, J.; Heinämäki, K.; Iwaï, 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**, 1451–1456.

51 Sun, W.; Yang, J.; Liu, X. Q.: Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J Biol Chem* 2004, **279**, 35281–35286.

52 Appleby-Tagoe, J. H.; Thiel, I. V.; Wang, Y.; Wang, Y.; Mootz, H. D.; Liu, X. Q.: Highly efficient and more general cis- and trans-splicing inteins through sequential directed evolution. *J Biol Chem* 2011, **286**, 34440–34447.

53 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* 1998, **207**, 187–195.

54 Ozawa, T.; Nogami, S.; Sato, M.; Ohya, Y.; Umezawa, Y.: A fluorescent indicator for detecting protein–protein interactions in vivo based on protein splicing. *Anal Chem* 2000, **72**, 5151–5157.

55 Mootz, H. D.; Muir, T. W.: Protein splicing triggered by a small molecule. *J Am Chem Soc* 2002, **124**, 9044–9045.

56 Berrade, L.; Kwon, Y.; Camarero, J. A.: Photomodulation of protein trans-splicing through backbone photocaging of the DnaE split intein. *Chembiochem* 2010, **11**, 1368–1372.

57 Vila-Perelló, M.; Hori, Y.; Ribó, M.; Muir, T. W.: Activation of protein splicing by protease- or light-triggered O to N acyl migration. *Angew Chem Int Ed Engl* 2008, **47**, 7764–7767.

58 Buskirk, A. R.; Ong, Y. C.; Gartner, Z. J.; Liu, D. R.: Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc Natl Acad Sci USA* 2004, **101**, 10505–10510.

59 Davis, K. M.; Pattanayak, V.; Thompson, D. B.; Zuris, J. A.; Liu, D. R.: Small molecule-triggered Cas9 protein with improved genome-editing specificity. *Nat Chem Biol* 2015, **11**, 316–318.

60 Wood, D. W., Wu, W.; Belfort, G.; Derbyshire, V.; Belfort, M.: A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol* 1999, **17**, 889–892.

61 Nicastri, M. C.; Xega, K.; Li, L.; Xie, J.; Wang, C.; Linhardt, R. J.; Reitter, J. N.; Mills, K. V.: Internal disulfide bond acts as a switch for intein activity. *Biochemistry* 2013, **52**, 5920–5927.

62 Zeidler, M. P.; Tan, C.; Bellaiche, Y.; Cherry, S.; Häder, S.; Gayko, U.; Perrimon, N.: Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nat Biotechnol* 2004, **22**, 871–876.

63 Busche, A. E.; Aranko, A. S.; Talebzadeh-Farooji, M.; Bernhard, F.; Dotsch, V.; Iwaï, 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**, 6128–6131.

64 Shah, N. H.; Vila-Perello, M.; Muir, T. W.: Kinetic control of one-pot trans-splicing reactions by using a wild-type and designed split intein. *Angew Chem Int Ed Engl* 2011, **50**, 6511–6515.

65 Kawasaki, M.; Satow, Y.; Ohya, Y.; Anraku, Y.: Protein splicing in the yeast Vma1 protozyme: evidence for an intramolecular reaction. *FEBS Lett* 1997, **412**, 518–520.

66 Aranko, A. S.; Oeemig, J. S.; Kajander, T.; Iwaï, H.: Intermolecular domain swapping induces intein-mediated protein alternative splicing. *Nat Chem Biol* 2013, **9**, 616–622.

67 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* 1999, **96**, 13638–13643.

68 Iwai, H.; Lingel, A.; Plückthun, A.: Cyclic green fluorescent protein produced in vivo using an artificially split PI-PfuI intein from *Pyrococcus furiosus. J Biol Chem* 2001, **276**, 16548–16554.

69 Iwai, H. and Pluckthun, A.: Circular beta-lactamase: stability enhancement by cyclizing the backbone. *FEBS Lett* 1999, **459**, 166–172.

70 Williams, N. K.; Liepinsh, E.; Watt, S. J.; Prosselkov, P.; Matthews, J. M.; Attard, P.; Beck, J. L.; Dixon, N. E.; Otting, G.: Stabilization of native protein fold by intein-mediated covalent cyclization. *J Mol Biol* 2005, **346**, 1095–1108.

71 Scott, C. P.; Abel-Santos, E.; Jones, A. D.; Benkovic, S. J.: Structural requirements for the biosynthesis of backbone cyclic peptide libraries. *Chem Biol* 2001, **8**, 801–815.

72 Evans, T. C. Jr, Martin, D.; Kolly, R.; Panne, D.; Sun, L.; Ghosh, I.; Chen, L.; Benner, J.; Liu, X. Q.; Xu, M. Q.: Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of Synechocystis species PCC6803. *J Biol Chem* 2000, **275**, 9091–9094.

73 Amitai, G.; Belenkiy, O, Dassa, B.; Shainskaya, A.; Pietrokovski, S.: Distribution and function of new bacterial intein-like protein domains. *Mol Microbiol* 2003, **47**, 61–73.

74 Dori-Bachash, M.; Dassa, B.; Peleg, O.; Pineiro, S. A.; Jurkevitch, E.; Pietrokovski, S.: Bacterial intein-like domains of predatory bacteria: a new domain type characterized in *Bdellovibrio bacteriovorus*. *Funct Integr Genomics* 2009, **9**, 153–166.

75 Dassa, B.; Haviv, H.; Amitai, G.; Pietrokovski, S.: Protein splicing and auto-cleavage of bacterial intein-like domains lacking a C′-flanking nucleophilic residue. *J Biol Chem* 2004, **279**, 32001–32007.

76 Shao, Y.; Xu, M. Q.; Paulus, H.: Protein splicing: evidence for an N–O acyl rearrangement as the initial step in the splicing process. *Biochemistry* 1996, **35**, 3810–3815.

77 Cheriyan, M.; Pedamallu, C. S.; Tori, K.; Perler, F.: Faster protein splicing with then Nostoc punctiforme DnaE intein using non-native extein residues. *J Biol Chem* 2013, **288**, 6202–6211.

78 Oeemig, J. S.; Zhou, D.; Kajander, T.; Wlodawer, A.; Iwaï, H.: NMR and crystal structures of the Pyrococcus horikoshii RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *J Mol Biol* 2012, **421**, 85–99.

79 Shah, N. H.; Eryilmaz, E.; Cowburn, D.; Muir, T. W.: Extein residues play an intimate role in the rate-limiting step of protein trans-splicing. *J Am Chem Soc* 2013, **135**, 5839–5847.

80 Johnson, E. C.; Kent, S. B.: Insights into the mechanism and catalysis of the native chemical ligation reaction. *J Am Chem Soc* 2006, **128**, 6640-6646.

81 Abrahmsén, L.; Tom, J.; Burnier, J.; Butcher, K. A.; Kossiakoff, A.; Wells, J. A.: Engineering subtilisin and its substrates for efficient ligation of peptide bonds in aqueous solution. *Biochemistry* 1991, **30**, 4151–4159.

82 Jackson, D. Y.; Burnier, J.; Quan, C.; Stanley, M.; Tom, J.; Wells, J. A.: A designed peptide ligase for total synthesis of ribonuclease A with unnatural catalytic residues. *Science* 1994, **266**, 243–247.

83 Mahrus, S.; Trinidad, J. C.; Barkan, D. T.; Sali, A.; Burlingame, A. L.; Wells, J. A.: Global sequencing of proteolytic cleavage sites in apoptosis by specific labeling of protein N termini. *Cell* 2008, **134**, 866–876.

84 Mazmanian, S. K.; Ton-That, H.; Schneewind, O.: Sortase-catalysed anchoring of surface proteins to the cell wall of Staphylococcus aureus. *Mol Microbiol* 2001, **40**, 1049–1057.

85 Mao, H.; Hart, S. A.; Schink, A.; Pollok, B. A.: Sortase-mediated protein ligation: a new method for protein engineering. *J Am Chem Soc* 2004, **126**, 2670–2671.

86 Ton-That, H.; Liu, G.; Mazmanian, S. K.; Faull, K. F.; Schneewind, O.: Purification and characterization of sortase, the transpeptidase that cleaves surface proteins of *Staphylococcus aureus* at the LPXTG motif. *Proc Natl Acad Sci USA* 1999, **96**, 12424–12429.

87 Chang, T. K.; Jackson, D. Y.; Burnier, J. P.; Wells, J. A.: Subtiligase: a tool for semisynthesis of proteins. *Proc Natl Acad Sci USA* 1994, **91**, 12544–12548.

88 Strijbis, K.; Spooner, E.; Ploegh, H. L.: Protein ligation in living cells using sortase. *Traffic* 2012, **13**, 780–789.

89 Guimaraes, C. P.; Witte, M. D.; Theile, C. S.; Bozkurt, G.; Kundrat, L.; Blom, A. E., Ploegh, H. L.: Site-specific C-terminal and internal loop labeling of proteins using sortase-mediated reactions. *Nat Protoc* 2013, **8**, 1787–1799.

90 Cooper, A. A.; Chen, Y. J.; Lindorfer, M. A.; Stevens, T. H.: Protein splicing of the yeast TFP1 intervening protein sequence: a model for self-excision. *EMBO J* 1993, **12**, 2575–2583.

91 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* 1999, **121**, 1100–1101.

92 Antos, J. M.; Chew, G. L.; Guimaraes, C. P.; Yoder, N. C.; Grotenbreg, G. M.; Popp, M. W.; Ploegh, H. L.: Site-specific N- and C-terminal labeling of a single polypeptide using sortases of different specificity. *J Am Chem Soc* 2009, **131**, 10800–10801.

93 Muona, M.; Aranko, A. S.; Iwaï, H.: Segmental isotopic labelling of a multidomain protein by protein ligation by protein trans-splicing. *Chembiochem* 2008, **9**, 2958-2961.