

Recent progress in intein research: from mechanism to directed evolution and applications

Gerrit Volkmann · Henning D. Mootz

Received: 21 December 2011 / Revised: 23 July 2012 / Accepted: 6 August 2012 / Published online: 28 August 2012
© Springer Basel AG 2012

Abstract Inteins catalyze a post-translational modification known as protein splicing, where the intein removes itself from a precursor protein and concomitantly ligates the flanking protein sequences with a peptide bond. Over the past two decades, inteins have risen from a peculiarity to a rich source of applications in biotechnology, biomedicine, and protein chemistry. In this review, we focus on developments of intein-related research spanning the last 5 years, including the three different splicing mechanisms and their molecular underpinnings, the directed evolution of inteins towards improved splicing in exogenous protein contexts, as well as novel applications of inteins for cell biology and protein engineering, which were made possible by a clearer understanding of the protein splicing mechanism.

Keywords Post-translational modification · Protein chemistry · Protein splicing · Enzyme mechanism · Protein engineering · Protein semisynthesis

Introduction

Protein splicing was discovered as a new twist in the synthesis of proteins about 20 years ago. This post-translational equivalent to RNA splicing, first described in 1990 for the catalytic subunit of the vacuolar H⁺-translocating ATPase (TFP1/VMA1) of *Saccharomyces cerevisiae* [1, 2], converts a precursor protein into its mature form by self-removal of

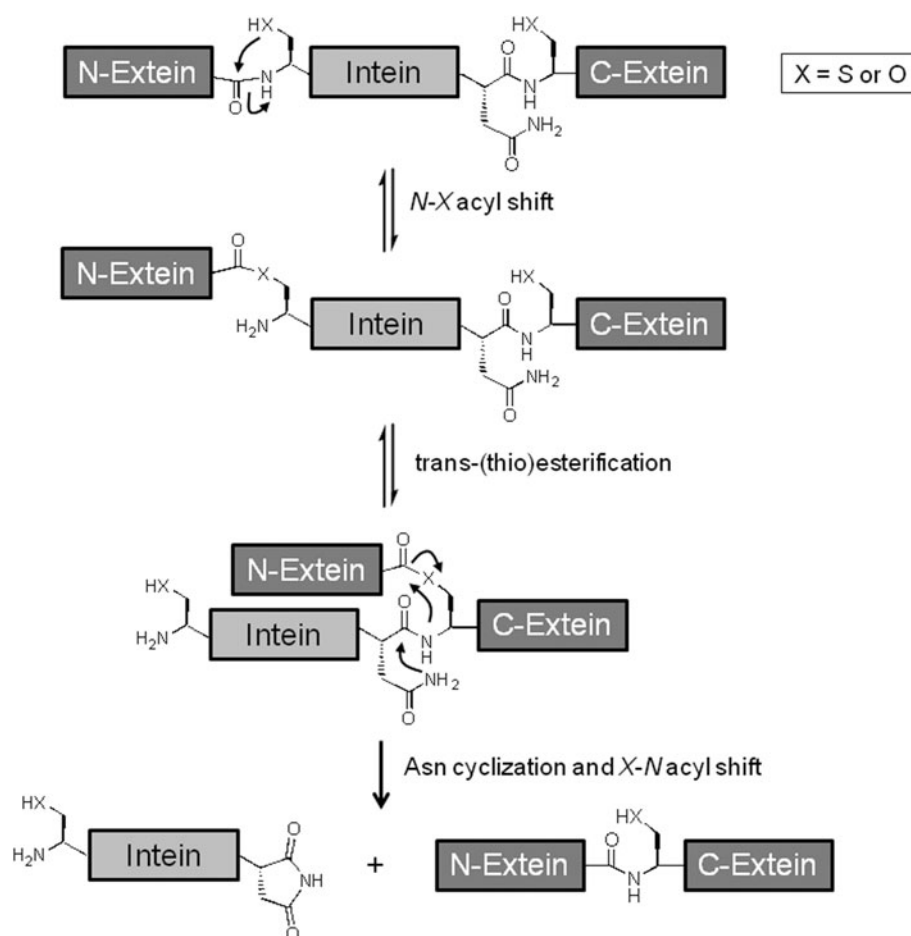
an internal protein, the intein, whereby a peptide bond is concomitantly formed between the flanking sequences (referred to as N- and C-exteins, respectively) (Fig. 1). Inteins exist as either one type of three distinct domain organizations (Fig. 2a): (1) as maxi or bifunctional inteins, where a sequence-specific DNA homing endonuclease is embedded into the protein splicing domain, resulting in an N-terminal (I^N) and C-terminal (I^C) intein fragment, (2) as mini-inteins, which lack the endonuclease and thus contain a contiguous protein splicing domain, or (3) as split inteins, where the protein splicing domain is discontinuous and the I^N and I^C fragments fused to their respective extein sequences are encoded by two independent genes.

We now know that inteins are common to all three kingdoms of life, and they also occur in virus and phage genomes, although they are sporadically distributed and confined to unicellular organisms. The reason for this distribution of inteins has spurred the proposal for numerous evolutionary scenarios, including roles of inteins in genetic mobility and selfish DNA [3–6], but we are still far from a complete understanding as to why inteins exist. It remains stunning why inteins have persisted over millions of years despite the fact that they do not seem to provide an obvious benefit to their hosts. It is especially puzzling why some archaeobacteria harbor more than ten inteins in their proteome, the current record holder being *Methanococcus jannaschii* with 19 inteins [7].

Although higher eukaryotes do not harbor inteins in their genomes, the hedgehog (Hh) protein, involved in eukaryotic developmental processes, is similar to inteins both in structural and functional terms (Fig. 2) [8]. The C-terminal hedgehog domain (Hh-C) resembles a mini-intein, whereas the N-terminal Hh domain (Hh-N) represents the N-extein. In a chemical reaction catalyzed by Hh-C, the Hh-N is modified with a cholesterol molecule

G. Volkmann · H. D. Mootz (✉)
Institute of Biochemistry, University of Münster,
Wilhelm-Klemm-Str. 2, 48149 Münster, Germany
e-mail: henning.mootz@uni-muenster.de

Fig. 1 Intein-mediated protein splicing. Schematic representation of the standard protein splicing mechanism, which proceeds in four steps: (1) N–X acyl shift, (2) *trans*-(thio)esterification, (3) asparagine cyclization, and (4) X–N acyl shift. The first residue of standard (class 1) inteins can either be Cys or Ser, whereas the first residue of the C-extein can be Cys or Ser, as well as Thr



(Fig. 2B), which can thus be viewed as the C-extein and leads to targeting of Hh-N to the plasma membrane. The striking homology between inteins and hedgehog suggested that they were derived from a common ancestor by gene duplication [5, 8]. Another group of intein homologs are intein-like autoprocessing domains found in bacteria and ciliates [9, 10]. The function of these domains remains unclear, but they have been shown to undergo cleavage and atypical splicing reactions.

Inteins can be regarded as single turnover enzymes, and, as such, the orchestration of the catalytic mechanism relies on numerous specific amino acid side chains. The amino acid sequences of inteins are usually numbered outside of their host protein context, i.e., if an intein starts with a cysteine, it is labeled Cys1. The residues flanking the intein N-terminally are counted backwards with negative numerals (the residue upstream of intein residue 1 is the “–1 residue”), whereas those downstream of the last intein residue receive a plus sign (the first C-extein residue is the “+1 residue”) (see Fig. 3a). The mechanistically and/or structurally crucial residues of inteins are grouped into four sequence blocks or motifs (A, B, F, and G; Fig. 3a) [11]. Although inteins found in different host

proteins generally share low sequence identity, some residues are extremely conserved, such as a cysteine or serine at position 1 (residue A:1), a threonine and histidine in block B (residues B:7 and B:10, respectively), as well as a histidine-asparagine pair at the end of the intein (residues G:6 and G:7) followed by a nucleophilic residue (Cys, Ser or Thr) at the beginning of the C-extein (residue G:8; see Fig. 3a for nomenclature). The discovery of certain deviations from these standard intein features has led to a new classification of inteins with respect to their splicing mechanisms [12], and will be discussed in the following section.

The potential to exploit inteins for practical purposes has inspired the development of a diverse array of applications in protein engineering and chemical biology. A number of reviews have focused on recent advances of intein-based applications [13–18], and the reader is referred to these articles for a more in-depth study of the subject, as this review will highlight mostly those applications that were stimulated by biochemical and mechanistic insights into protein splicing. Another emerging field is the use of inteins for pharmaceutical purposes [19, 20], which is also beyond the scope of this review.

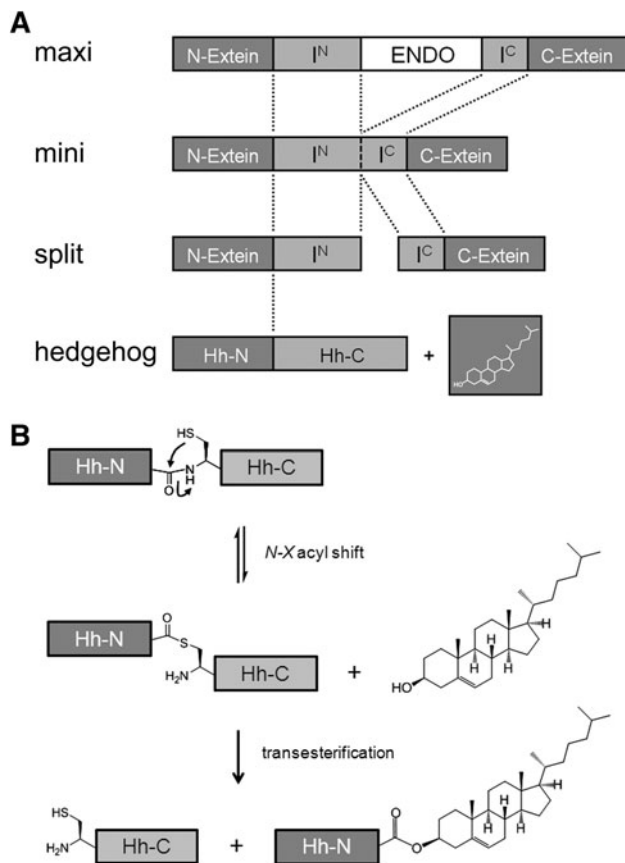


Fig. 2 Intein types found in nature and their similarity to hedgehog. **a** The three different intein types are shown schematically in the context of an N- and C-terminal extein, where I^N and I^C represent the N- and C-terminal intein splicing domains, respectively. ENDO corresponds to the homing endonuclease domain found exclusively in maxi inteins. The C-terminal domain of the hedgehog protein (Hh-C) is homologous to the splicing domain of inteins; in this case, the N-terminal domain (Hh-N) and a cholesterol molecule function as “exteins”. **b** The processing mechanism of hedgehog is initiated by an N–S acyl shift between the first residue of Hh-C (Cys) and the preceding peptide bond, as is the case in many inteins (see Fig. 1). The thioester bond is then attacked by the hydroxyl group of a cholesterol molecule, which results in the esterification of cholesterol to the C-terminus of Hh-N

The protein splicing mechanism: variations on a common theme

Inteins displaying all sequence features mentioned above are grouped in class 1, which represents the majority of inteins (>90 %) currently listed in the intein database InBase [7]. The currently accepted protein splicing mechanism of class 1 inteins was elucidated in a step-by-step fashion and was fully presented in 1996 [21]. The reaction proceeds through a series of nucleophile-driven bond rearrangement and cleavage reactions, and ends with the spontaneous formation of the peptide bond between the extein sequences (Fig. 1, 3b). In the initial N–X acyl shift (X denoting either a sulfur or oxygen atom), the thiol or hydroxyl side chain of the A:1

residue attacks the carbonyl carbon of the preceding peptide bond, so that the N-extein is linked to the side chain of A:1 by a (thio)ester bond (“linear ester intermediate”). In the second step, the N-extein is cleaved from the intein and is transferred onto the side chain of the first C-extein residue, G:8 or +1, in a *trans*-esterification reaction. The intein is completely removed from the resulting “block G branched intermediate (BI)” in the third step by cyclization of the last intein residue G:7 (usually Asn), cleaving the peptide bond between intein and C-extein and leaving a succinimide ring on the intein’s C-terminal end. Finally, the peptide bond between the esterified exteins is formed by an enthalpy-driven X–N acyl shift. Off-pathway reactions include N- and C-terminal cleavage, liberating the N- and C-exteins from the ester intermediates and/or the precursor protein, respectively.

Class 2 inteins lack a nucleophilic residue at position A:1, and attack the N-terminal scissile peptide bond directly with the side chain of the first C-extein residue (Cys), thereby omitting both the N–X acyl shift and the *trans*-esterification steps. Once the block G BI is formed, Asn cyclization and the X–N acyl shift proceed as in the standard splicing mechanism (Fig. 3B). The *Mja* K1bA intein represents the prototype of class 2 inteins and has been studied in much detail [22, 23]. Although the protein structure was recently solved using solution NMR, the precise molecular basis for the alternative splicing mechanism was not obviously apparent from the arrangement of amino acids in the catalytic core [24]. Instead, the structure suggests that the *Mja* K1bA intein is able to directly attack the N-terminal scissile peptide bond due to a gain in available space in the active site, which is ~2.4 Å wider than in other inteins. More members of class 2 need to be investigated structurally to determine whether this is a general feature of class 2 inteins.

Finally, inteins belonging to class 3 also lack a nucleophilic A:1 residue but in addition carry the so-called Trp–Cys–Thr (WCT) triplet, of which the Cys residue attacks the N-terminal splice junction (Fig. 3b). The class 3 inteins are exemplified by the *Mycobacteriophage Bethlehem* (MP-Be) DnaB [12] and the *Deinococcus radiodurans* Snf2 intein [25], which both revealed covarying residues scattered over three of the four common intein sequence motifs: a Trp in block B (B:12), a Cys in block F (F:4), and a Thr in block G (G:5, see Fig. 3a). Whereas the Trp and Thr residues of this WCT triplet could be mutated to similar residues without losing protein splicing ability, the block F Cys was found to be absolutely essential for a complete protein splicing reaction, as it could not be replaced with other nucleophilic residues (Ser or Thr). Investigation of the MP-Be DnaB intein revealed the presence of a novel thiol-labile branched intermediate (BI) during protein splicing, which could not have been formed at the +1 residue (Thr) because oxygen esters are generally

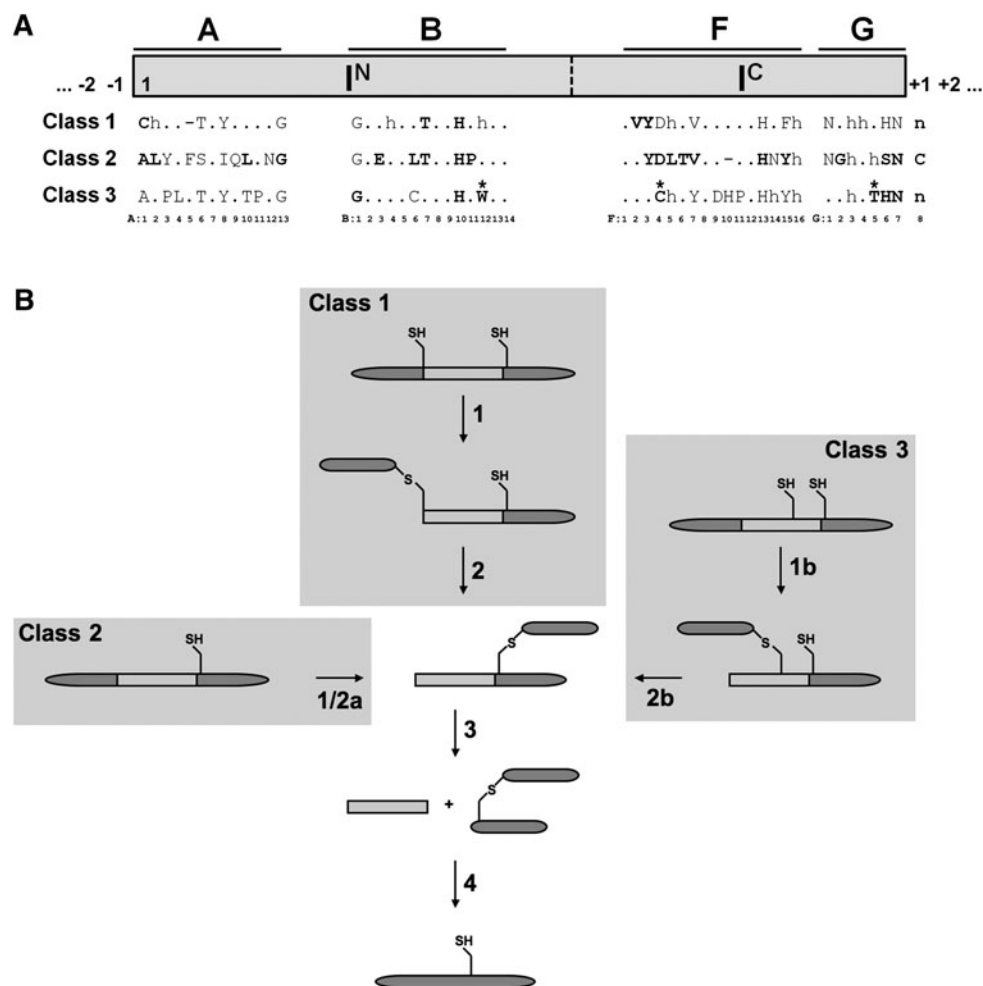


Fig. 3 Inteín sequence motifs and splicing mechanisms. **a** Schematic illustration of a mini-intein with the relative locations of the conserved sequence motifs (A, B, F, and G) are shown on *top*. Nomenclature of amino acid numbering in inteins is indicated (e.g., -1 is the last residue of the N-extein, and +1 is the first residue of the C-extein). The consensus sequences of all motifs are shown for the three inteín classes [12] with *bold-faced, uppercase letters* representing highly conserved residues, and *regular uppercase letters* indicating moderately conserved residues. A “-” symbol is given where acidic residues are conserved, “h” stands for hydrophobic residues, “n” for nucleophilic residues, and “.” if there is no significant conservation at a position. The *asterisks* above residues in class 3 highlight the conserved WCT triplet. The *numbers below the consensus sequences* indicate the nomenclature for numbering of

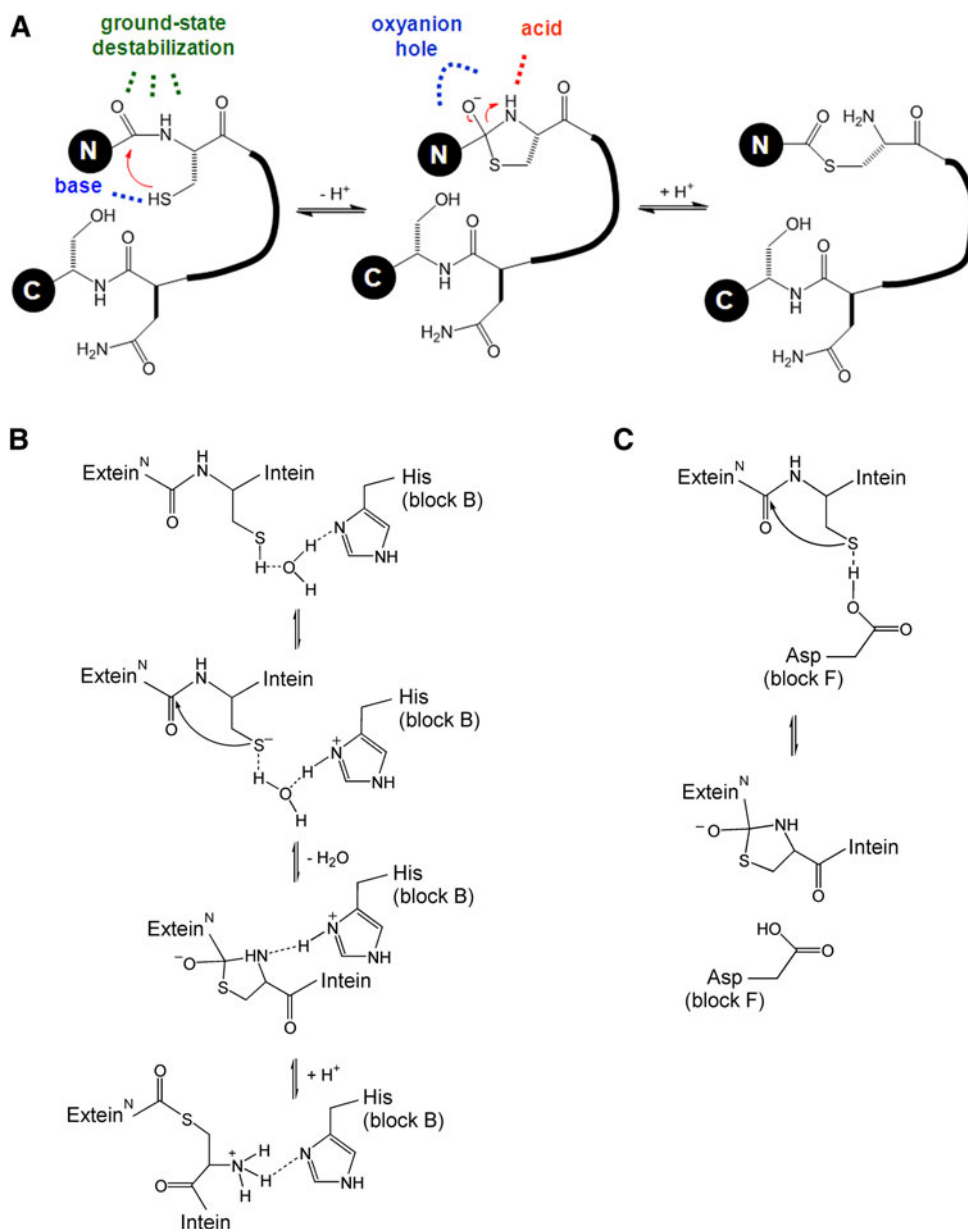
residues within the motifs according to Tori et al. [12], which is also used throughout the text. **b** Splicing mechanisms of the three inteín classes. Splicing of class 1 inteins begins with an N-S acyl shift (if first residue is a Cys) (step 1), followed by *trans*-esterification of the N-extein to the first residue of the C-extein (step 2). Cyclization of the last inteín residue cleaves the inteín off of the esterified exteins (step 3), and a spontaneous S-N acyl shift forms the peptide bond between the exteins (step 4). Class 2 inteins forego step 1 of class 1 inteins and directly attack the N-terminal scissile peptide bond with the first C-extein residue (step 1/2a). Class 3 inteins use an internal Cys residue to form a branched intermediate unique to class 3 inteins (step 1b), which is subsequently attacked by the first residue of the C-extein (step 2b), resulting in the canonical branched ester intermediate

alkaline- and not thiol-labile. Site-directed mutagenesis strongly indicated that this unusual BI was formed by the F:4 Cys residue of the WCT triplet. Homology modeling of the MP-Be DnaB inteín indicated that the F:4 Cys residue is well positioned to interact with the N-terminal scissile peptide bond, and is also in close proximity to the G:8 residue (Thr+1) for a subsequent *trans*-esterification step to form the conventional block G BI [12]. The essential nature of the block F Cys was also confirmed for the *Dra*

Snf2 inteín by site-directed mutagenesis. Importantly, in this study the block F BI could be trapped and purified from a triply mutated *Dra Snf2* inteín (Cys[F:4]Ser, Asn[G:7]Ala, Thr[G:8]Ala), which provided definitive proof for the existence of this unusual branched intermediate [25].

Due to the high degree of sequence divergence in inteins, it is conceivable that more catalytic varieties of protein splicing will be discovered. Interestingly, some

Fig. 4 The initial N–X acyl shift in protein splicing of class 1 inteins. **a** Schematic representation of the peptide bond rearrangement to a (thio)ester bond at the N-terminal splice junction. The intein is represented by a *thick line*, whereas the N- and C-exteins are indicated by *black circles*. The scissile peptide bond with the A:1 residue (Cys in this case) as well as the G:7 and G:8 residues (Asn and Ser, respectively) are shown, along with possible mechanisms that allow the N–S acyl shift to occur. **b** The pK_a shift mechanism during the initial N–S acyl shift as proposed by NMR and quantum mechanical/molecular mechanics calculations, where the B:10 His performs acid–base catalysis with the help of a water molecule [40]. **c** Proposed deprotonation of the A:1 N-nucleophile (Cys) by the F:4 Asp residue during N–S acyl shift [46]



inteins combine sequence features of both class 1 and 3 inteins. For example, the T_{fus} Tfu2914 and Nsp-JS614 TOPRIM inteins present a Cys residue at position A:1, indicative of class 1, as well as the WCT triplet, indicative of class 3. Mutagenetic studies revealed that the A:1 Cys was essential for the first step of the splicing reaction, whereas the F:4 Cys of the WCT triplet was not [26, 27]. These inteins therefore belong, from a mechanistic point of view, to class 1. A different intein (MP-Catera Gp206), on the other hand, which also has a nucleophilic side chain at position A:1 (Ser) together with the WCT triplet, belongs to class 3 because in this case, the F:4 Cys was essential for splicing [27]. These studies show that in some instances,

the primary intein sequence alone cannot accurately predict its classification.

Novel insights into individual steps of the protein splicing mechanism and their coordination

Although the principle pathway of protein splicing is understood, the exact molecular mechanisms of the individual protein splicing steps and how the steps are coordinated within the confined space of the intein's catalytic core have only recently begun to be unraveled at the atomic level.

Step 1: the N–X acyl shift

Given the formal uphill thermodynamic nature of this step, it may be regarded as the most intriguing in the protein splicing pathway. However, such rearrangements of a peptide bond into a thio or oxoester can be found in several other auto-processing proteins, such as the hedgehog proteins (see Fig. 2), as well as glycosyltransferases, pyruvoyl enzymes, SEA domains, and proteasome subunits [28–31]. Yet, in these latter cases, different protein folds and mechanisms are employed and also the subsequent biochemical fate of the generated active esters is quite different. Furthermore, the first reaction of cysteine and serine proteases is reminiscent of the N–X acyl shift in protein splicing, although here the reaction occurs intermolecularly rather than intramolecularly. Nevertheless, despite these differences, the mechanistic challenge to catalyze the reaction is quite similar, and common strategies have been suggested. These include activation of the attacking nucleophilic side chain by deprotonation or polarization, stabilization of the tetrahedral intermediate by an oxyanion hole, as well as ground-state destabilization of the scissile peptide bond through conformational strain (Fig. 4a), and may be used to differing extents by each intein.

The histidine residue in block B (B:10) is the most conserved residue among all inteins, and its crucial character for successful protein splicing has been tracked down to its essential role in the N–X acyl shift by mutagenesis studies [32, 33]. Consistently, structural investigations of numerous inteins unifyingly place this histidine in close proximity to the N-terminal splice junction, where it is in hydrogen bonding distance to the amide nitrogen of the scissile peptide bond [34–39]. In the early structures, it was proposed that the B:10 His residue acts as an acid to facilitate breakdown of the tetrahedral intermediate formed at the N-terminal splice junction [34, 37], however, experimental validation of the theory is difficult and was not forthcoming. Du et al. [40] recently shed new light on the possible precise role of the B:10 His by determining its pK_a before and after the splicing reaction in an engineered version of the *Mtu* RecA intein. Using NMR-based titration experiments, the authors found the pK_a of the B:10 His (His73) to be 7.3 before and <3.5 after protein splicing, respectively. This large pK_a shift was accommodated into a comprehensive reaction scheme for the initial step of protein splicing (Fig. 4b), where the B:10 His residue first acts as a base to deprotonate the side chain of the A:1 N-nucleophile (Cys in *Mtu* RecA), possibly with the help of a water molecule. Following formation of the tetrahedral intermediate by attack of the A:1 Cys side chain on the carbonyl carbon, B:10 His then donates the afore-acquired proton to the leaving group, the amide nitrogen of the former peptide bond. Combined quantum mechanics/

molecular mechanics (QM/MM) modeling backed up the experimentally derived “ pK_a shift mechanism”. The study thus nicely confirmed the prior crystallographic evidence for the B:10 His in breakdown of the tetrahedral intermediate at the N-terminal splice junction, and further provided a solution for activation of the N-nucleophile.

It is, however, most probably the case that the molecular underpinnings of the N–X acyl shift differ between inteins, as the highly conserved residues in block B (B:7 Thr and B:10 His) are of different importance for this step in different inteins [32, 33, 41–43]. It is especially noteworthy that an interaction between the B:10 His and the side chain of the A:1 N-nucleophile was not emphasized in any of the structural studies on inteins. Instead, in three structures, the nucleophilic side chain of the A:1 residue is in close proximity to a hydrogen-bonding side chain 22 residues upstream of the B:10 His residue (Ser53 [34], Thr51 [36], Arg50 [38]). Interestingly, these residues lie outside of any of the conserved intein sequence motifs [11, 44] but appear to be a common structural feature of the Hedgehog/INTEIN (HINT) fold because the *Drosophila* hedgehog autoprocessing domain shows a similar interaction [8]. Along similar lines, a very recent study by Perler and coworkers investigated a small subgroup of inteins that surprisingly lack the conserved B:10 His residue [45]. The *Tko* CDC21-1 intein was shown to be active in protein splicing, and a positively charged residue 23 amino acids upstream of the B:10 His (Lys58) was found to be essential for the N–S acyl shift. The role of this residue was attributed to a stabilizing effect of the tetrahedral intermediate underling once more the idea that the catalytic strategies are utilized to varying degrees by different inteins. It will be interesting to determine whether the residues present in other inteins at this position, as mentioned above (Ser53 in *Mxe* GyrA, Thr51 in *Ssp* DnaB, and Arg50 in *Ssp* DnaE), play a similar role for the N–X acyl shift in these inteins.

Another residue that appears to be critical for the N–X acyl shift of class 1 inteins is an Asp present in motif F (F:4 residue). Several intein structures have shown this F:4 Asp to be in hydrogen-bonding distance to residues at the N-terminal catalytic center [24, 38, 39], and it was experimentally demonstrated that this residue is a strict requirement for the initial N–S acyl shift in an engineered *Mtu* RecA mini-intein [39]. NMR-based pK_a measurements of the F:4 Asp in a splicing-enhanced engineered *Mtu* RecA mini-intein (Asp422) have now helped to more clearly define this residue’s role during the N–X acyl shift [46]: with a value of ~ 6 , the pK_a of Asp422 is about two units higher than the usual pK_a of Asp residues in proteins. This elevation in pK_a is likely mediated by the A:1 N-nucleophile (Cys) because Asp422 showed a normal pK_a of ~ 4 when the A:1 Cys residue was mutated to Ala. Furthermore, the A:1 Cys had a depressed pK_a (7.5 vs. the

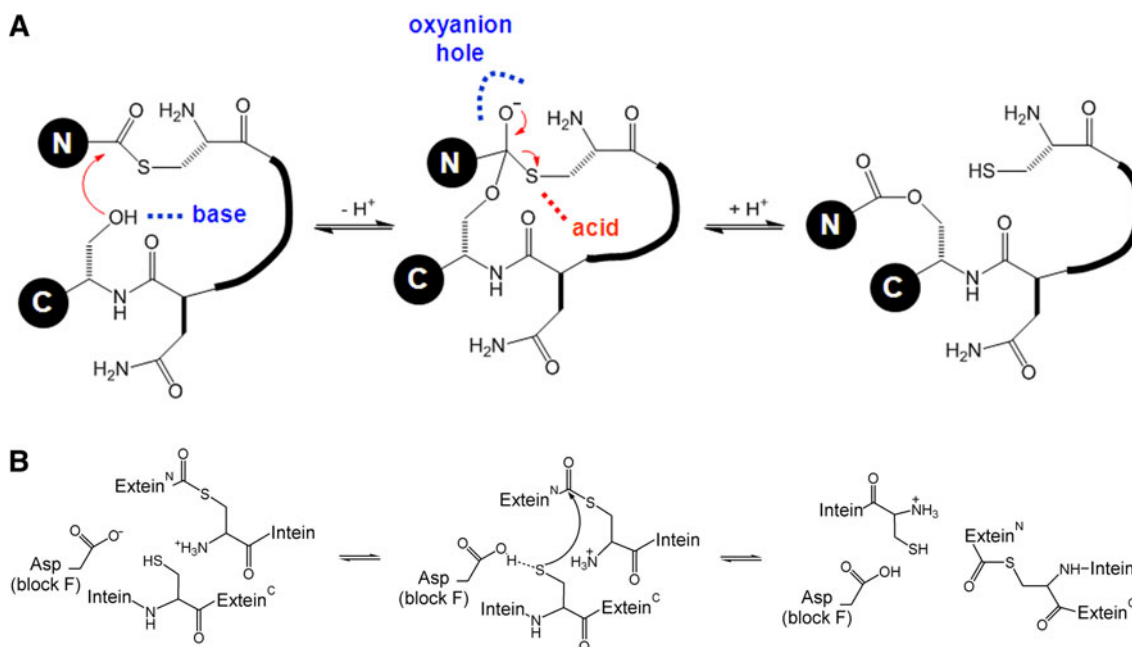


Fig. 5 The *trans*(thio)esterification step as catalyzed by class 1 inteins. **a** Schematic representation of the intein active site after the N–S acyl shift (*left*), where a base might potentially activate the G:8 C-nucleophile for attack of the thioester bond at the N-terminal splice junction. This attack yields a tetrahedral intermediate, which is likely

stabilized by an oxyanion hole. Acid-catalysis then protonates the A:1 side chain, resulting in the branched ester intermediate. **b** Possible role of the F:4 Asp residue during the *trans*-thioesterification reaction by activating the G:8 C-nucleophile [54]

usual 8.5), which depended on the presence of the F:4 Asp residue. Together, these results provided strong evidence for a hydrogen bond between F:4 Asp and A:1 Cys, with the protonated Asp residue stabilizing the thiolate side chain of the Cys residue (Fig. 4c). Remarkably, replacing Asp422 in *Mtu* RecA with residues that could in principle serve as hydrogen bond donors due to high side chain pK_a values substantially decreased or abrogated splicing. These results indicate that the F:4 Asp is uniquely qualified to lower the activation energy for the N–X acyl shift by stabilizing the negative charge of the A:1 Cys thiolate, thereby facilitating the nucleophilic attack of the thiolate on the carbonyl carbon of the N-terminal scissile peptide bond. Although not all inteins have an Asp residue at position F:4, it is important to note that this is the exact position of the catalytic Cys residue in class 3 inteins, which directly attacks the N-terminal scissile peptide bond. Moreover, the F:4 Trp residue present in the class 1 *CneA* Prp8 intein is also crucial for the N–S acyl shift in this intein [42]. It thus appears that during evolution, this position was specifically selected for an involvement at the N-terminal splice junction, although the chemistry of the residue's side chain and the mechanisms involved diverged substantially.

The idea that a base is required for deprotonation of the N-nucleophile in class 1 inteins, however, is challenged by a recent study that used unnatural amino acid substitutions

to probe the formation of thioester bonds at the N-terminal splice junction of a semi-synthetic *Ssp* DnaB split intein [47]. Specifically, the introduction of homocysteine (Hcy) in place of the natural cysteine residue at position A:1 did not significantly decrease the rate of thioester hydrolysis, indicating that the N–S acyl shift was largely unaffected. This observation is striking because the additional methylene group in Hcy must lead to a rearrangement of the active site where the thiol side chain of Hcy would be misaligned for either abstraction of its proton by the B:10 His and/or stabilization of the thiolate by the F:4 Asp residue. This assumption is in agreement with the finding that the B:10 His of the *Mxe* GyrA intein facilitates the N–X acyl shift by destabilizing the ground-state of the N-terminal scissile peptide bond by polarization, rather than by deprotonation of the N-nucleophile [41], which lends further support to the originally conceived role for this residue.

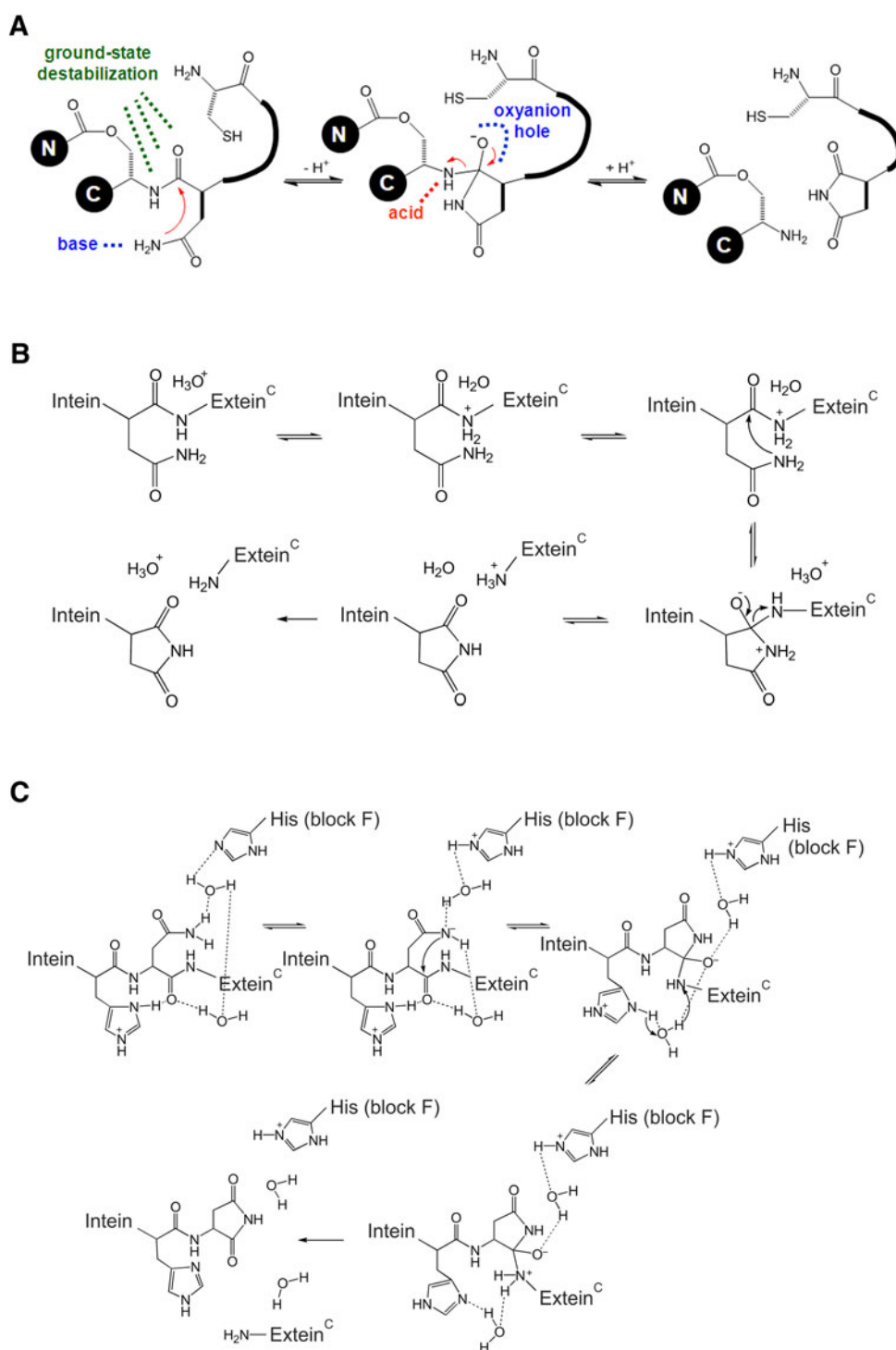
Most recently, a peptide complementation study using a non-canonical *Ssp* GyrB split intein [48] further suggested the involvement of residues located in motif G during the N–X acyl shift [49]. Although the entire motif G was not essential for formation of the linear ester intermediate, the presence of the motif, especially the +1 residue, resulted in a more than tenfold increase in the reaction rate constant. The mostly hydrophobic residues of motif G could be crucial for forming the active site through van der Waals

forces, whereas the side chain of the +1 residue might be involved in polarization of the N-terminal scissile peptide bond. High-resolution structural data will likely be required to ascertain how exactly the motif G residues and the C-nucleophile exert their beneficial effects on the N–X acyl shift.

Another hallmark of the initial N–X acyl shift is the formation of a tetrahedral intermediate resulting from the attack of the N-nucleophilic side chain on the carbonyl carbon of the scissile peptide bond (Fig. 4a). This oxothiazolidine (in case of A:1 Cys) or oxyxazolidine (in case of A:1 Ser) ring structure has been proposed by careful

Fig. 6 The final step in protein splicing catalyzed by all intein classes, Asn cyclization.

a Schematic representation of the active site, starting with the branched ester intermediate, where base-catalysis in combination with ground-state destabilization of the C-scissile peptide bond possibly assist the nucleophilic attack of the G:7 Asn side chain on the peptidyl carbon. The resulting tetrahedral intermediate is likely stabilized by an oxyanion hole, and acid-assisted protonation of the peptidic nitrogen leads to cleavage of the C-scissile peptide bond. **b** Proposed scheme for Asn cyclization in a low-pH environment as deduced by QM/MM calculations [62]. **c** Proposed scheme for Asn cyclization derived by a computational model as based on the *Ssp* DnaB mini-intein crystal structure [66]



investigation of the *Sce* VMA1 N-catalytic center [37], and was also accounted for in the above pK_a shift mechanism, where the QM/MM calculations clearly indicated that such an intermediate was energetically possible. Experimental proof for the oxothiazolidine was obtained through serendipity, when a semi-synthetic split derivative of the *Ssp* DnaB intein was analyzed for kinetic parameters and complex formation [50]. Introduction of three mutations (Gly(-1)Ala in an I^N -containing peptide, and Asn[G:7]Ala/Ser[G:8]Ala in an I^C -containing protein) resulted in the unexpected loss of 18 Da in the I^N peptide, indicative of loss of water. Tandem mass spectrometry confirmed that a thiazoline ring had been formed between the A:1 Cys side chain and the carbonyl carbon of the preceding peptide bond. Because the observed thiazoline could only have been generated by elimination of a water molecule from the proposed oxothiazolidine intermediate, these results thus corroborated the existence of the latter intermediate during the N-S acyl shift.

A so-far-underestimated facet of the N-X acyl shift may be the documented prevalence of aminoacyl-cysteinyl peptide bonds to spontaneously rearrange into the thioester [47], in particular under acidic conditions, when protonation of the α -amino group favors the equilibrium to the thioester. Efforts in the field of synthetic peptide chemistry to achieve new synthetic routes to peptide thioesters have exploited this rearrangement under acidic conditions and even at neutral pH. The formed thioester is either cleaved with excess free thiol or stabilized by trapping the free α -amino group into a diketopiperazine through a cysteinyl-prolyl ester switch element [51–53]. When applied to protein splicing, these insights may suggest that the catalytic role of the intein for the N-X acyl shift is less so to drive the forward reaction but rather to prevent the back reaction to the peptide bond in order to effectively remove the thio(ester) from the equilibrium by the next step in the pathway.

Step 2: *trans*-esterification

This step in the protein splicing pathway is probably the most difficult to investigate experimentally. Even though the reaction starts from an energetically activated (thio)ester, one would postulate similar mechanistic requirements for its catalysis as for the first step, at least activation of the G:8 nucleophilic side chain by deprotonation, correct positioning to mediate the nucleophilic attack, and stabilization of the tetrahedral intermediate by an oxyanion hole (Fig. 5a). In fact, if and how the latter two points are brought about remains enigmatic at this point. Most intein structures show distances of 8–9 Å between the G:8 side chain and the N-terminal scissile peptide bond [35, 36, 38] and it is not clear how a defined conformational change is

triggered or if the reaction is enabled by increased local dynamics, for example.

Some recent studies addressed the deprotonation of the G:8 side chain. Mutational analysis of the class 2 *Mja* KlbA intein revealed that the F:4 Asp residue, mentioned above to be important for the N-X acyl shift in class 1 inteins, might also serve a role in the formation of the branched ester intermediate [24]. The observation that replacement of the F:4 Asp in *Mja* KlbA with either Glu or Ala abrogated protein splicing was highly significant because the class 2 inteins start their splicing reaction by directly forming the branched ester intermediate through an attack of the N-terminal scissile peptide bond with the G:8 C-nucleophile [22].

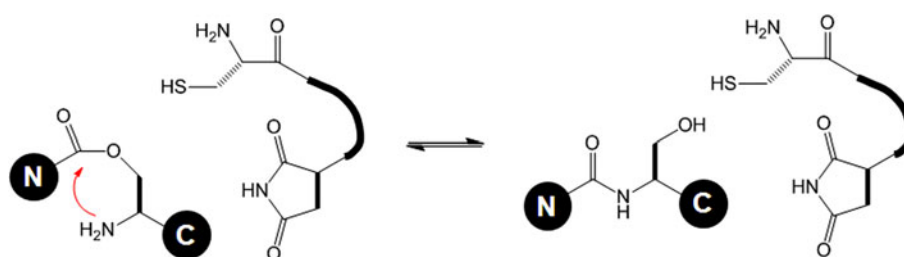
Most recently, the F:4 Asp was also shown to be pivotal for *trans*-esterification in class 1 inteins [54]. Using a sensitive FRET assay for N-terminal cleavage, it became apparent that an engineered *Mtu* RecA intein carrying the native F:4 Asp residue in combination with a mutation of the G:8 C-nucleophile to Ala showed no significant N-cleavage activity in comparison to the intein where the G:8 C-nucleophile was not replaced with Ala. Similar behavior was observed when F:4 Asp was replaced with residues unable to act as hydrogen bond donors, indicating that the F:4 Asp exerts its role on *trans*-esterification through the formation of hydrogen bonds. QM/MM simulations performed on a minimal system including the linear ester intermediate, the F:4 Asp and the G:8 C-nucleophile (Cys), suggested that the side chain of F:4 Asp spontaneously abstracts a proton from the G:8 Cys thiol group. The thiolate thus achieved was then able to attack the carbonyl carbon of the linear ester intermediate, forming the branched ester intermediate (Fig. 5b).

The role of F:4 Asp to act as a hydrogen bond acceptor is supported by the NMR-based pK_a measurements performed by Du et al. [46]. After serving as a hydrogen bond donor during the N-X acyl shift due to its elevated pK_a , the F:4 Asp transfers the proton to the free N-terminus of the intein, and can then accept a proton from the G:8 C-nucleophile. Together, these studies manifest the idea that the first two steps of class 1 protein splicing are interconnected by a complex hydrogen bonding network based on locally depressed and elevated pK_a values of at least three highly conserved intein residues, the A:1 N-nucleophile, the B:10 His, and the F:4 Asp residues (Figs. 4, 5).

Step 3: Asn cyclization

This reaction in the protein splicing pathway represents the first irreversible step because the bond between the intein and the C-extein is cleaved. Therefore, one would expect that it underlies a specific control mechanism to prevent premature cleavage that would result in off-pathway by-products. Indeed, several studies showed that this reaction

Fig. 7 The final, uncatalyzed reaction in protein splicing (X–N acyl shift). The formation of the peptide bond between the exteins from the ester bond is not catalyzed by the intein but represents a spontaneous rearrangement to the more stable amide linkage



is tightly coupled to the first or the second step in the protein splicing pathway [55–58].

Mechanistically, the side chain nitrogen of the last intein residue (G:7, in most cases an asparagine) performs a nucleophilic attack on the carbonyl carbon of the downstream peptide bond, effectively cleaving the intein from the branched ester intermediate with formation of a succinimide ring at the C-terminus of the intein [59, 60] (Fig. 6a). This early observation is intriguing because in a protein context, asparagines usually undergo a deamidation reaction where the amide nitrogen of the downstream peptide bond attacks the carbonyl carbon of the Asn side chain [61]. Two computational studies have recently made advances at providing molecular explanations for how inteins drive Asn cleavage rather than deamidation.

The first study used QM/MM calculations to examine Asn-mediated cleavage in a low-pH environment [62], based on the observation that engineered mini-inteins from both *Mtu* RecA and *Ssp* DnaB show a preponderance for C-terminal cleavage by Asn cyclization at pH < 7 [63, 64]. The computational system consisted of the G:7 Asn side chain preceded by an amine group (acting as a surrogate for the amide nitrogen of the peptide bond linking the penultimate residue to Asn) and followed by the scissile peptide bond. The low-pH environment was mimicked by including a hydronium ion in the computational system. In the derived model (Fig. 6b), the hydronium ion initiates C-cleavage by protonating the amide nitrogen of the scissile peptide bond, making the latter a good leaving group. In turn, this protonation event facilitates the nucleophilic attack of the G:7 Asn side chain on the carbonyl carbon of the scissile peptide bond, whereby one of the amide hydrogens is transferred onto a water molecule, which, in the final step, results in protonation of the former leaving group. The authors scrutinized by semiempirical and high-level quantum calculations the possibility that the G:7 Asn side chain amide is deprotonated prior to cyclization. This scenario was excluded due to relative energies of >30 kcal/mol (depending on the dielectric constant of the medium), which were higher than those calculated for the rate-determining Asn cyclization step (25 kcal/mol in water), and, importantly, higher than the energy obtained from laboratory experiments (~21 kcal/mol [65]).

In the second line of investigation [66], the authors initially studied Asn cyclization with an Asn–Thr dipeptide, which suggested that (1) deprotonation of the Asn side chain amide is required to lower the energy barrier for formation of the tetrahedral intermediate, (2) protonation of the carbonyl oxygen of the scissile peptide bond is required to lower the energy barrier for cleavage of the C–N-bond during the collapse of the tetrahedral intermediate, and (3) intramolecular hydrogen transfer proceeds with high-energy barriers. Inteins must have therefore evolved to fulfill these requirements by providing suitable amino acids in the spatial vicinity of the terminal Asn. Several inteins appear to have such opportune molecular arrangements in their active sites [34–36, 38], which inspired the authors to evaluate whether the mechanism for Asn cyclization and C-terminal cleavage proposed from the crystal structures makes thermodynamic sense. Ding et al. predicted from the *Ssp* DnaB crystal structure [36] that a charge relay system, initiated by the F:13 His residue and further involving a water molecule, is essential for the formation of the tetrahedral intermediate, in which the G:7 Asn side chain is transitionally linked to the carbonyl carbon of the scissile peptide bond. A further prediction was that this tetrahedral intermediate is stabilized by an oxyanion-binding site provided by the side chains of the penultimate G:6 His and the F:4 Asp residue, the latter bridged by a water molecule. Mujika et al. [66] thus used the X-ray structure of the *Ssp* DnaB mini-intein as the starting point for their calculations on a more sophisticated model for Asn cyclization and C-cleavage, which consisted of all the above components except the F:4 Asp side chain (Fig. 6c). The computations largely supported the proposal for the *Ssp* DnaB mini-intein, and comparison with the initial Asn–Thr dipeptide studies clearly showed that the F:13 His and G:6 His side chains and water molecules are well positioned to lower the energy barriers for the crucial transition states. The only deviation from the proposal by Ding et al. was the breakdown of the oxyanion-stabilized tetrahedral intermediate, where the calculations showed the penultimate G:6 His to be in a much more favorable distance and orientation for protonating the peptidic nitrogen of the intermediate than the F:13 His (2.9 vs. 6.0 Å), as had been suggested from the crystal structure of the *Mxe* GyrA mini-intein [34]. The model is further backed up by the pK_a

measurements performed by Du et al. [40], which revealed pK_a values of 6.3 for the penultimate G:6 His and 8.9 for the F:13 His, thus experimentally evoking their roles as hydrogen donors and acceptors, respectively.

The role of the aforementioned F:4 Asp residue, suggested to be pivotal in linking the first two steps of class 1 protein splicing, has also been shown to be somewhat involved in Asn cyclization. Structural investigations of a C-cleavage enhanced *Mtu* RecA mini-intein indicated that the F:4 Asp side chain was in close contact to the C-terminal scissile peptide bond [39]. Furthermore, introduction of Gly at this position yielded an *Mtu* RecA intein that exhibited predominantly C-cleavage activity [63], which could be pinpointed by the crystal structure of this cleavage mutant to the presence of two water molecules. The F:4 Asp may thus be vital for a complete protein splicing reaction by decreasing the access of water to the C-terminal active site.

Asn cyclization is intimately linked to the breakdown of the branched splicing intermediate. Fundamental insights into this link have now been provided through a series of experiments using protein semisynthesis [58]. Here, expressed protein ligation (EPL) was used to prepare a variety of semisynthetic mimics of the *Mxe* GyrA intein branched intermediate, which were unable to revert to the linear intermediate or precursor and, importantly, could be induced for Asn cyclization by a simple temperature shift. This allowed for an unprecedented dissection of kinetic parameters associated with the formation of the intein-succinimide by Asn cyclization. Initially, the authors determined that Asn cyclization truly is the rate-limiting step for protein splicing by the *Mxe* GyrA intein, as suggested earlier [67], because this step proceeded with a rate constant indistinguishable from the overall rate constant of protein splicing, in contrast to the speed of the N–S acyl shift for this intein, which occurs at a rate 100-fold faster than complete splicing [41]. They also found that succinimide formation was ten times slower for intein variants that could not provide the branched intermediate (due to a Cys[A:1]Ala mutation or lack of an N-extein sequence), indicating that the branched intermediate directly stimulates Asn cyclization, thereby ensuring that cleavage of the C-terminal scissile peptide bond is favored only when the time is right. NMR spectroscopy performed on a semisynthetic branched intermediate carrying a unique isotopic handle at the C-terminal scissile peptide bond, the authors could pinpoint this stimulating effect to a markedly different local environment at the peptidic amide nitrogen [58].

X–N acyl shift

After cleavage of the intein from the branched intermediate, the exteins are linked with an oxygen/thioester bond,

which rearranges through X–N acyl migration to a peptide bond (Fig. 7). Early studies with peptides mimicking the esterified exteins revealed that this acyl shift occurs at a much faster rate than the overall protein splicing reaction [59, 68], indicating that peptide bond formation is uncoupled from the rest of the protein splicing mechanism. The process was largely temperature-independent with low activation energies (4–5 kcal/mol [68]), in line with the observed spontaneity of the X–N acyl shift. The experimental data obtained with the purified synthetic peptides thus strongly suggested that the excised intein does not contribute to the final step in the protein splicing reaction, although unequivocal proof was never provided.

To finally address this popular assumption experimentally, Frutos et al. [58] determined the rate of the X–N acyl shift for a model depsipeptide in the absence and presence of an intein-succinimide. The peptide consisted of five N- and four C-extein residues native to the *Mxe* GyrA intein, which were esterified through the G:8 Thr side chain. A photocleavable protective group at the α -amino group of the G:8 Thr allowed for precise control of the start of the O–N acyl migration by light irradiation. The deprotected peptide was then incubated in physiological buffer with or without addition of the *Mxe* GyrA intein-succinimide. Maybe not surprisingly, the rates of peptide bond formation were found to be indistinguishable whether the intein-succinimide was present or not, thereby firmly establishing that the excised intein is not responsible for the final step in protein splicing.

Directed evolution of inteins

Most inteins characterized to date are comparably slow single-turn over enzymes, and often show some level of sequence dependence at the splice junctions, which is difficult to predict and so far only very poorly understood [17, 69, 70]. Many biotechnological applications of inteins, however, require a traceless removal of the intein or at least a change of the flanking amino acid sequence as minimal as possible to preserve the primary sequence of the protein of interest. Inteins that are independent of the residues flanking the splice junctions are therefore highly desirable. Rational engineering of existing intein sequences towards faster splicing and less sequence context dependency has not been forthcoming until this very year, where the efficiency and speed of the *trans*-splicing reaction of several DnaE split inteins could be improved [71], and the *Pho* RadA intein could be rendered more promiscuous towards the amino acid at position –1 [72]. However successful, a generalized way to improve intein function remains a formidable, if not impossible, task, given the variety of mechanistic strategies to catalyze protein splicing, as

outlined above, and the generally low sequence similarity between inteins. Moreover, some applications might require inteins that are tailored to specific reaction parameters, e.g., a certain temperature or denaturing buffer conditions, or the conditional association of the fragment pairs of a split intein, which is just as difficult to achieve by rational design. A conceptually different approach than rational design in order to generate improved or customized inteins is the use of directed evolution based on random mutagenesis and selection, which over the past few years has shown great promise in yielding inteins with superior properties.

Perler and coworkers, for example, have used molecular evolution in combination with positive selection systems to turn both the *Mxe* GyrA mini-intein as well as the *Tli* Pol-2 intein into temperature-sensitive forms as a potential system for growth-based screenings of protein splicing inhibitors [73, 74]. Perrimon and colleagues have evolved the *Sce* VMA1 intein into temperature-sensitive derivatives to enable control of protein activity by temperature-regulated protein splicing in eukaryotes [75]. The Tan group has now evolved five variants of *Sce* VMA1 with different optimal splicing temperatures [76].

Belfort and coworkers used an in vitro intein evolution system based on phage display technology [77] in combination with error-prone PCR (ePCR) to select for improved protein splicing under different temperature and pH conditions. Several mutant *Mtu* RecA inteins were isolated that showed improved splicing efficiency over the parent $\Delta\Delta I_{hh}$ intein, a previously minimized *Mtu* RecA intein containing a hedgehog-derived linker between the I^N and I^C sequences [78]. Surprisingly, the mutations were located at a fair distance from the catalytic centers of the intein, with the hedgehog-linker representing a particular “hot spot” for beneficial mutations (Fig. 8). The phenomenon of enhanced protein splicing due to mutations remote from the intein active site was termed the “ripple effect”, and NMR spectroscopic data of one particular mutant indeed showed that even a single mutation can cause global chemical shift perturbations that relay into the intein active site. Although overall the selected inteins were more active than the parent intein, the highest splicing efficiency achieved with any of the mutants still was only 50 %, and thus only marginally better compared to a single enhancing mutation isolated in an earlier study [63]. It thus remains to be demonstrated whether this phage selection system is powerful enough to yield a robust intein with quantitative splicing efficiency.

Adapting inteins to heterologous flanking sequences in which the wild-type intein shows no or only reduced splicing activity is of particular importance for the generality of intein-based protein technologies. In an effort to directly evolve an intein with a more relaxed junction

sequence dependency, Lockless and Muir [79] developed an in vivo evolution approach for *trans*-splicing split inteins using a genetic selection. The kanamycin resistance protein (KanR), which had already been successfully used in intein evolution [80], was split within a loop region, and a split intein cassette containing the *Npu* DnaE I^N and *Ssp* DnaE I^C split intein fragments were inserted on the DNA level. The C-terminal splice junction was chosen to be Ser–Gly–Val (SGV), a sequence contained within the linker region of the multimodular adaptor protein Crk-II, which was spliced by the chimeric *Npu/Ssp* DnaE split intein only with low efficiency and yield. After three rounds of selection, a mutant split intein was isolated that spliced the SGV C-terminal junction fivefold better than the parent split intein in terms of reaction rate and product yield. Unfortunately, the evolved split intein appeared to be adapted specifically to the SGV sequence because at other C-terminal junctions the *trans*-splicing activity was much lower. Thus, specialization to a particular sequence context, as known for most native inteins, was obtained in this selection scheme rather than the evolution of an intein with relaxed junction sequence dependency.

An obvious way to overcome a specific junction sequence context is to evolve the intein at multiple insertion sites within a host protein. Liu and coworkers subjected the *Ssp* DnaB mini-intein [81] to sequential rounds of directed evolution within the KanR protein [82]. The first insertion site contained two native N-extein and three native C-extein residues but resulted in an inactive intein, thus conferring kanamycin sensitivity to *E. coli* cells harboring this construct. Two rounds of directed evolution yielded a mutant intein with two mutations that was able to splice a functional KanR protein with ~35 % efficiency. This primary mutant intein (1°) was subsequently inserted into a second site in the KanR protein without any native extein residues except the catalytic G:8 Ser residue, which inactivated the 1° intein. Evolution at this site led to a secondary mutant intein (2°) with four additional mutations that spliced the new extein context with >50 % efficiency, while retaining splicing activity at the first insertion site. The 2° mutant intein was next inserted at a third site within KanR, where it showed only low or no splicing activity. A final round of directed evolution yielded tertiary mutant inteins (3°) with improved activity, each containing one or two additional mutations. Significantly, these 3° mutant inteins were also able to splice at several other sites within KanR for which they were not selected for and at which the wild-type, 1 and 2° mutant inteins showed no activity, all the while retaining quantitative splicing in the native extein context of the wild-type *Ssp* DnaB intein. This study shows that by performing sequential rounds of directed evolution at different junction sequences it is possible to generate inteins with a broad tolerance towards the amino acids

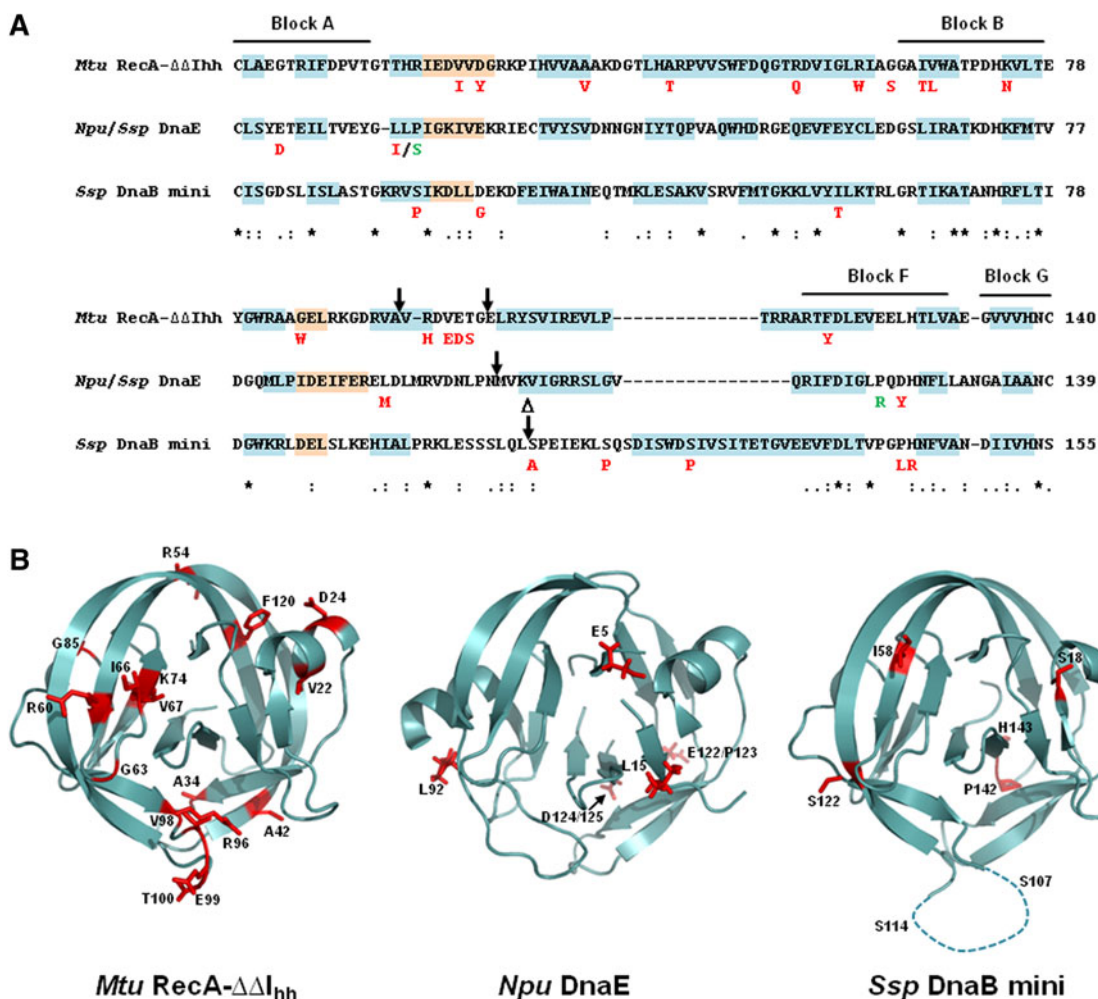


Fig. 8 Molecular evolution of inteins. **a** ClustalW alignment of amino acid sequences from inteins that have been evolved by error-prone PCR and selected for more efficient splicing under various selection pressures. The conserved intein motifs are indicated above the sequences, and secondary structural elements are highlighted (blue boxes β -strands, orange boxes α -helices). In the sequence of *Mtu RecA- $\Delta\Delta I_{hh}$* , the black arrows indicate the position of the hedgehog (hh) protein-derived β -turn sequence (VRDVEVG), which replaces residues 95–402 of the wild-type *Mtu RecA* intein including the endonuclease domain [78]. In the chimeric *Npu/Ssp DnaE* intein sequence, the black arrow indicates the boundary of the split intein fragments (residues 1–102: N-intein from *Npu*, residues 103–138: C-intein from *Ssp*). In the *Ssp DnaB mini*-intein sequence, the black arrow indicates the deletion point (Δ) of the endonuclease domain

[81]. Residues below the sequences indicate mutations introduced by molecular evolution. For the *Mtu RecA- $\Delta\Delta I_{hh}$* intein, the following mutations were also found but omitted for simplicity: D24E, D24N, K74M, R96C, R96P. The F120 residue is often referred to as F421 in the literature (in accordance with the numbering of the wild-type maxi *Mtu RecA* intein). In the case of the *Npu/Ssp DnaE* intein, red residues correspond to mutations isolated during selection for a different splice junction context, whereas green residues indicate mutations discovered when the selection pressure was temperature (37°C). Leu15 was either mutated to Ile or Ser, depending on the experiment. **b** The locations of the mutations shown in **a** were mapped onto the structures of the parent inteins using PyMol. PDB codes: *Mtu RecA- $\Delta\Delta I_{hh}$* , 2IN9 [39]; *Npu DnaE*, 2KEQ [108]; *Ssp DnaB mini*, 1MI8 [36]

adjacent to the intein. Remarkably, although selected as *cis*-splicing inteins, one of the 3° mutant inteins (M86 mutant) had substantially improved characteristics over the wild-type intein in a *trans*-splicing system: the overall rate constant for splicing had increased 60-fold with formation of only small amounts of C-terminal cleavage product, and the M86 mutant was significantly more active in *trans*-splicing when the native Gly-1 was mutated to Ala-1. Moreover, the K_d value between the intein fragments had decreased by an order of magnitude.

The beneficial mutations in the evolved inteins from the latter two studies [79, 82] were again scattered over the entire structure, as seen with the *Mtu RecA* inteins evolved by phage display [77] (Fig. 8). Inteins have subtle dynamic fluctuations in their polypeptide backbone, and even a single splicing-enhancing mutation has been shown to shift the structure to a more stable conformation [83]. It is thus likely that the presence of several mutations in the evolved inteins dramatically affected the dynamic behavior of the individual structures causing changes in internal motions of

catalytically important residues. Moreover, statistical coupling analysis identified several coevolving residues in the intein family, which were speculated to form an interaction network in order to transmit allosteric effects from distant sites to the catalytic center. Indeed, many residues that were found to be mutated in intein evolution experiments are part of or juxtaposed to such coevolving network residues [79].

Altogether, directed evolution promises to be an attractive means to generate inteins with desired characteristics. Obviously, a single or a few mutations can have significant beneficial effects on the intein's activity. This observation is further corroborated by the dramatic differences in activity seen for the DnaE split intein alleles, although they are highly homologous in sequence [56, 57, 84–86]. It suggests that further and even better evolved inteins will be generated by these approaches. In the ideal case, a single intein would combine all such traits including, but not limited to, (1) exceptionally fast reaction kinetics at various temperatures, (2) negligible extent of off-pathway cleavage reactions, (3) good solubility and stability, (4) the ability to splice under denaturing conditions, and (5) quantitative splicing independent of the nature of the extein residues flanking the N- and C-terminal splice junctions. Split intein mutants could also be selected for a higher affinity between the N- and C-terminal fragments. Temperature-dependent (see above) and ligand-dependent [80, 87, 88] inteins have already been evolved, but it remains a formidable challenge to combine these traits with the most robustly splicing inteins, like the split *Npu* DnaE intein, because no or very low activity under a certain condition appears to be in principle mutually exclusive with very high activity under another condition. This has so far only been achieved with the incorporation of covalent chemical modifications (see below). A so-far-unexplored field is the laboratory evolution of inteins with improved properties for applications in expressed protein ligation (EPL), i.e., for the synthesis of protein thioesters [89–91]. Inteins that show no tendency to premature in vivo N- or C-cleavage or that work under strongly denaturing conditions would be highly desirable. The selection of mutants fulfilling these criteria might actually be more straight-forward, as only the step of the N–S acyl shift would require optimization. In contrast, for the selection of mutants with better splicing properties, the coordination between the individual steps of the protein splicing pathway must be maintained, which is likely to be met only by rarer combinations of amino acid substitutions.

So far, the mutations observed in evolved inteins are difficult to rationalize. It can be expected, however, with more such data accumulating, that these will also help in understanding the structure–function relationship for efficient protein splicing.

New applications from the intein tool box

Since their discovery, inteins have been exploited for a myriad of clever biotechnological applications owing to their general promiscuity towards foreign extein sequences, even though their activity may depend upon the specific protein sequence context. Not only the peptide bond forming reaction of inteins is attractive in many applications but also the peptide bond cleavage at one of the splice junctions of partially inactivated intein mutants have enabled various technologies. Well-established applications of the latter kind are the preparation of protein thioesters as reagents for the expressed protein ligation (EPL) approach [90, 91] and the self-cleavage for tag-free affinity purification of proteins [92]. Split inteins are especially attractive tools to join foreign polypeptide sequences prepared by recombinant protein expression and/or chemical synthesis and have received much attention recently [13, 18]. In this section, we highlight a few recent examples of how progress in the biochemical characterization and mechanistic investigation of inteins could be turned into new developments of intein-based applications and their far-reaching potential for research areas such as protein engineering, chemical biology, and cell biology. For a more detailed discussion on these topics, the reader is referred to the more specialized review articles [13, 15, 17–20, 93].

Intein-based protein engineering and covalent manipulation

Split inteins can be regarded as nature's protein ligases [13, 18]. A protein or a short peptide tag that may include non-proteinogenic chemical groups can be site-specifically installed either at the N-terminus or at the C-terminus of recombinant proteins by means of protein *trans*-splicing (Fig. 9). Also, entire protein domains or fragments can be linked in this way to reconstitute proteins from smaller segments. This reaction thus represents the intein-mediated equivalent of the powerful chemical ligation reactions

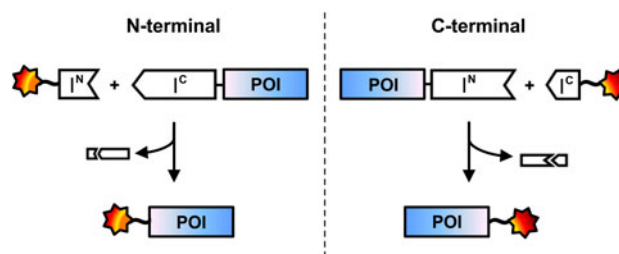


Fig. 9 Split intein-based labeling of recombinant proteins. Recombinant proteins can be modified with chemical groups (represented by a star) either at the N- or C-terminus using semisynthetic protein *trans*-splicing

between peptides and/or proteins, namely native chemical ligation (NCL) or expressed protein ligation (EPL). The most important advantages over the latter reactions include the absence of special functional groups, i.e., no thioester is required, and the inherent affinity in the low nanomolar to micromolar range [50, 94, 95] based on the selective recognition of the split intein fragments. These features bring about practical benefits for the semi-synthesis of proteins: (1) the split intein fusions with the desired peptide or protein sequence can either be synthesized in a straightforward way or be fully genetically encoded, (2) low reactant concentrations are sufficient in the splicing reaction, and (3) the reactions can be performed even in complex mixtures, namely on or inside living cells. Remaining challenges to develop split inteins into robust, generally useful tools are to overcome the sequence dependence around the splice junction, improve the solubility of the split intein fusion proteins, or to isolate inteins that work under strongly denaturing conditions, as already discussed in the previous section.

The labeling of proteins by protein *trans*-splicing is achieved by incorporating the synthetic moiety into the polypeptide sequence that is spliced to the protein of interest. This can be done either through metabolic feeding (i.e., for NMR isotopes), chemical modification of amino acid side chains [67, 96, 97] or even by total chemical synthesis of the peptide with a short intein fragment. Modifications that have been attached to proteins in this way span from small fluorescent probes [67, 96–101] and affinity tags [48, 99] to unusual and even high molecular weight compounds such as crown ethers [97], quantum dots [102], polyethylene glycol [67], and glycosylphosphatidylinositol (GPI)-anchor mimics [103, 104]. Even the immobilization of proteins to a solid support has been achieved by protein *trans*-splicing [105]. The short C-terminal intein fragment of the naturally split *Spy* DnaE (36 aa) and of the artificially split *Mtu* RecA intein (38 aa) [106, 107] were used in the first total synthesis examples of the intein piece by the solid-phase methodology. In recent years, several artificially split intein systems with shorter fragments of 6–15 aa were generated to circumvent the challenging synthesis of peptides ≥ 30 –40 aa [48, 98, 108, 109].

Cellular applications are especially exciting for split inteins because of the unique properties they offer in the modification of a protein's covalent structure. Protein semisynthesis has been demonstrated inside living mammalian cells [110, 111] and *Xenopus* embryos [102], as well as on the surface of eukaryotic cells [99, 101]. While these studies mostly have a proof-of-principle character, it is to be expected that with improved inteins and improved methods for cellular delivery of exogenously added proteins or peptides exciting progress will be feasible. Given their self-removal in the splicing reaction, split inteins are

probably the prime candidates to effect more complex and subtle chemical modifications of cellular proteins, of the kind that cannot be incorporated by the tRNA suppression technology [112], for example to install or mimic post-translational modifications. Other applications apart from protein semisynthesis take advantage of the co-expression of complementary intein fragments to reconstitute intact proteins from two pieces *in vivo*. For example, split inteins can be used in a variety of ways to construct biosensors for applications in cell biology (reviewed in [17]), with recent advancements made towards tracing MAP kinase signaling [113], apoptosis [114], internalization of G-protein-coupled receptors [115], and calcium signalling [116]. Another classical example for an application that exploits the posttranslational character of the protein *trans*-splicing reaction is the reconstitution of foreign gene products incorporated in transgenic plants. After fragmentation of the encoding gene into two pieces, each fused with a split intein gene, the resulting two DNA constructs are separated, e.g., placed into the nuclear and the chloroplast genome, to dramatically reduce the risk of spreading the entire transgene [117–119].

Controlling split intein interaction for *in vivo* manipulation of protein activity

Inteins are also recognized as potentially general tools to control the activity of a protein in the living cell through the protein *trans*-splicing or intein-mediated cleavage reactions. Such an artificial switch of fully genetically encoded intein–protein fusions is ideally triggered by a small molecule or light to allow for a precise and dosable manipulation. It acts on the posttranslational level and therefore can provide a degree of temporal control that cannot be achieved with purely genetic approaches. If the activity of the intein can be switched at will, then this regulatory element could in principle be used to control the function of any protein. *Cis*-splicing inteins can be engineered into such protein switches by inserting a ligand-binding domain, so that protein splicing or the inhibition thereof becomes dependent on the addition of a small-molecule ligand [80, 87, 88, 120]. For *trans*-splicing split inteins the control of the association of the intein fragments was in the focus to generate artificial switches. Referred to as conditional protein splicing (CPS) systems, the rapamycin-binding domains FKBP and FRB were used to bring about proximity and thereby high local concentrations of the split intein fusion constructs to trigger the reconstitution of intein activity [17, 121–124]. Alternatively, the phytochrome B (PhyB) and the PIF3 phytochrome binding domain (PIF3-APB) served to control intein fragment association with light [125]. A premise for this switch design is a low inherent affinity of the intein fragments to

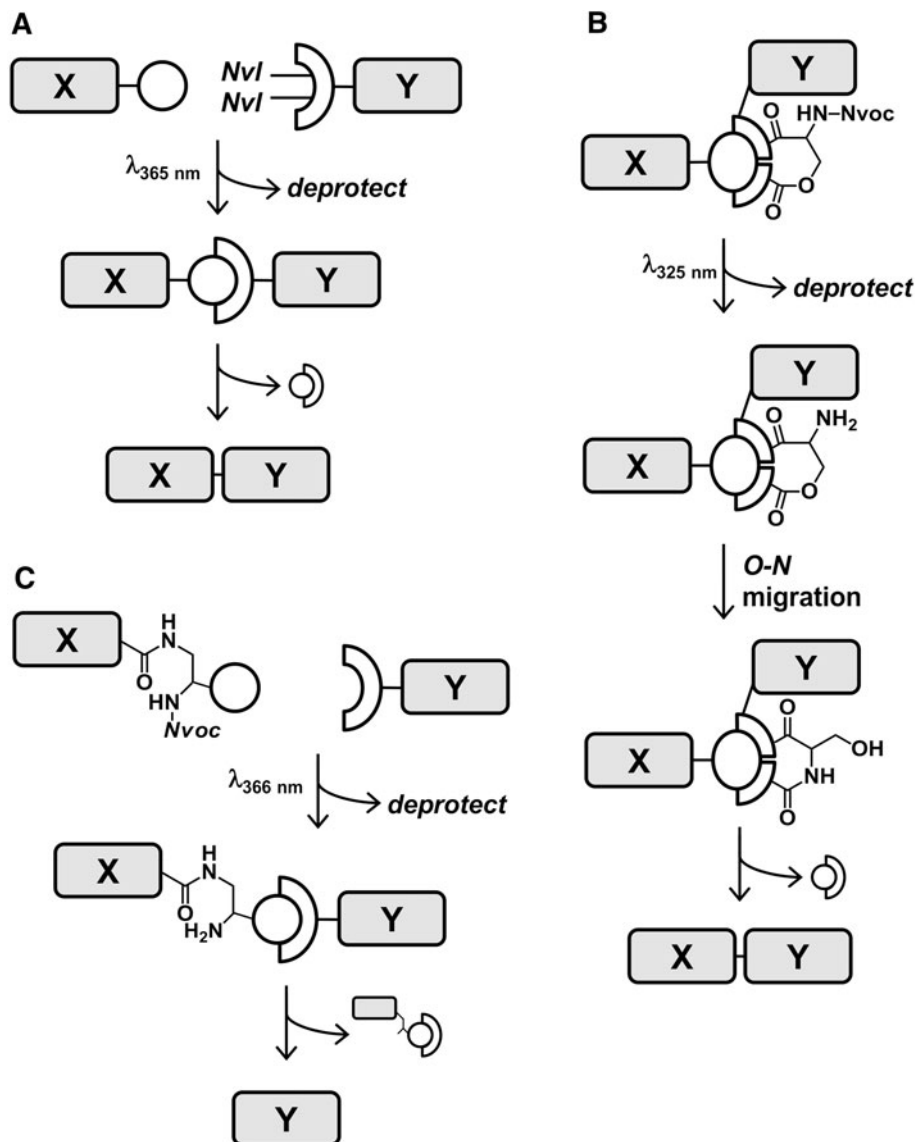


Fig. 10 Controlling intein function with photoremovable groups. **a** Installing two 6-nitroveratryl (Nvl) groups at positions Gly19 and Gly31 in the *Ssp* DnaE I^C split intein fragment disturbs the interaction with the cognate I^N fragment. The high affinity association of the intein fragments and subsequent protein *trans*-splicing is restored only upon irradiation with light ($\lambda_{365\text{ nm}}$). **b** Modification of the penultimate residue in the *Ssp* DnaE I^C fragment (here: Ala[G:7]Ser) with an *o*-nitroveratryloxycarbonyl (Nvoc) group still allows for split intein association but significantly perturbs the active site. Protein *trans*-splicing commences after treatment of the complex with light

($\lambda_{325\text{ nm}}$), which results in deprotection and O–N acyl migration to properly rebuild the intein active site. **c** To engineer a semisynthetic version of the *Ssp* DnaB intein into a light-responsive switch for C-terminal cleavage, the 11-aa I^N fragment was rationally redesigned to contain (1) the cysteine isoster diamino propionic acid at position A:1 with the N-extein sequence X linked to the side chain, thereby resembling the linear thioester intermediate after the N–S acyl shift, and (2) an Nvoc-protected N-terminus. Liberating a free N-terminus by light-irradiation ($\lambda_{365\text{ nm}}$) restores split intein association and C-cleavage activity

prevent constitutive levels of protein *trans*-splicing. This was indeed met by the artificially split *Sce* VMA1 intein that was employed in the mentioned studies. However, it also has become clear that the low affinity was accompanied with overall poor activity of the VMA1 intein in most insertion points [126], thereby limiting the scope of the system. To develop more robust intein switches, several efforts have been undertaken to artificially control the

naturally split DnaE inteins, like the *Npu* DnaE intein, which have proven less context-dependent [85]. The challenge to produce a DnaE intein in an inhibited but activatable form to prevent spontaneous *trans*-splicing could so far not be solved by genetic manipulation. However, two successful routes were reported that employed chemical modification of the intein fragments. By introducing the photocleavable 6-nitroveratryl group (Nvl) at

the backbone amides of Gly19 and Gly31 in the *Ssp* DnaE I^C intein fragment the affinity to the I^N fragment was reduced by ~50 times and splicing was effectively inhibited. Splicing with purified proteins could then be restored by irradiation with light ($\lambda_{365\text{nm}}$) to wild-type levels with respect to both yield and kinetics [127] (Fig. 10a). In the other study, the steric bulk of the photocleavable *o*-nitroveratryloxycarbonyl (Nvoc) group was combined with the introduction of an O-acyl linkage close to C-terminus of the *Ssp* DnaE I^C fragment. Interestingly, this modification abrogated splicing, although assembly of the split intein fragments was unaffected. Obviously, the structural changes did not prevent overall association and folding but caused sufficient local distortion to prevent the correct folding of the active site. Light irradiation restored the splicing function of the intein through liberation of the α -amino group at the O-acyl linkage and the subsequent spontaneous reversion to the peptide bond by O–N acyl migration [128] (Fig. 10b). In a comparable way to these examples for light-induced protein *trans*-splicing, the introduction of the Nvoc group was also applied to control the C-terminal *trans*-cleavage reaction. For this purpose, the I^N fragment (11 aa) of an artificially split *Ssp* DnaB intein was modified with the photocleavable group to block correct folding of the assembled intein complex. Administration of UV light released the protein fused to the I^C fragment by the C-terminal cleavage reaction [129] (Fig. 10c). All of these light-inducible approaches are so far severely limited for in vivo applications because the chemically prepared intein component would need to be introduced into the cells across the cell membrane. However, they show the importance of identifying the critical spots in the intein structure to achieve the desired manipulation of split intein assembly and function. Improved future versions of these intein switches might also be obtained with precise in vivo chemical modification of an intein, e.g., by the tRNA suppressor technology, or may be combined with efficient ways of protein delivery into cells.

Understanding and controlling split intein fragment recognition for in vitro applications

As already noted in the previous section, our understanding of the specific determinants for specific and efficient intein fragment association is still limited. Despite the highly conserved 3D intein structure, split inteins only splice when complementary N- and C-terminal pairs are combined. Given the overall low level of sequence conservation between inteins, this finding may not be surprising. Exceptions from this rule were found in some combinations of homologs of the DnaE intein from different strains. These inteins share the same insertion point in the same host

protein and also a higher level of sequence similarity [84, 86]. For example, the *Ssp* DnaE I^C fragment readily reacted with the *Npu* DnaE I^N fragment, even better than with its cognate *Ssp* DnaE I^N partner [85].

Cross-reactivity between non-cognate pairs of split inteins can be undesirable for some in vitro applications, for example when utilizing two split inteins in one pot to build a protein sequence from three individual fragments. Such a multi-fragment ligation is of special interest for segmental isotope labeling of proteins for NMR spectroscopy (reviewed in [16]) to allow a central domain to be selectively investigated by NMR. Earlier reports have used at least one artificial split intein [94, 130, 131] to ensure orthogonality between phylogenetically more distant inteins. Owing to the better solubility of the naturally occurring split DnaE inteins, two recent studies have used rational design to overcome the issue of cross-reactivity when using two of these in one reaction. In one report, a new split site was introduced to generate a variant of the *Npu* DnaE intein that cannot cross-react with the native allele. Guided by protein flexibility observed in the NMR structure of this intein [108], the I^C sequence was shortened from 36 to 16 residues [109], which abolished its reactivity with the I^N of the wild-type split intein due to the missing 20 residues [132]. The 20-aa overlap between the I^N of the engineered split intein and the wild-type I^C, however, did not prevent cross-reactivity, which required the three-piece ligation reaction to be carried out in two steps. In the second study, attempts were made to first understand the molecular basis for the intein fragment recognition before manipulating it. Earlier reports had already suggested that pronounced electrostatic interactions are important for the extremely rapid association [86, 94]. Again using the *Npu* DnaE intein, selected acidic residues in I^N were therefore mutated to basic residues and vice versa for the I^C sequence [133]. These charge-swapped intein fragments displayed remarkably diminished cross-reactivity with the wild-type fragments, while catalyzing protein *trans*-splicing among themselves, albeit with a tenfold slower rate than the wild-type split intein. Three-piece ligations could be performed in one pot in a sequential manner, due to the gain in kinetic control over the two separate protein *trans*-splicing events.

Conclusions and outlook

When protein splicing was first described in the *Saccharomyces cerevisiae* VMA1 protein, it was mind-boggling that such a biochemical reaction existed but remained unnoticed during the over 100 years of protein research. The self-processing reaction represents a unique alteration

of a protein's primary sequence that can be exploited for manifold applications. While much has been learned since then about the general pathway of protein splicing and the structure of inteins, many specific questions remain open. However, maybe the most important general question also still lacks a clear answer—why are there inteins at all? A new appreciation on this might be gained by examining inteins in their natural hosts—instead of studying their behavior in heterologous or in vitro environments—to possibly unravel unprecedented clues for roles of inteins in cellular metabolism. In this respect, a recent study on the *Pyrococcus abyssi* MoaA intein is an intriguing example because this intein splices most efficiently when the cytoplasm in *E. coli* cells mimics a reducing environment [134]. Since the *moaA* gene product is an oxygen-sensitive protein with an Fe–S cluster, the *Pab* MoaA intein might serve as a rheostat in its natural host, ensuring that mature MoaA protein is only generated when oxygen levels are low.

The more detailed molecular underpinnings of peptide bond formation catalyzed by inteins are starting to unravel and have been discussed in this article. Since the formulation of the basic chemical reaction in protein splicing in 1996, one of the recent central conclusions is that various inteins can follow quite distinct molecular strategies for the protein splicing pathway, even within a particular intein class. Obviously, this is also reflected by the low sequence conservation between inteins, which in turn raises the question of whether inteins may exist that have so far eluded the common bioinformatic tools due to an even higher degree in sequence diversification. Other major challenges for a better molecular characterization of inteins include the understanding of the extein sequence dependence, not only on the level of the primary sequence flanking the intein but also in the context of the protein 3-D structure and folding pathway. An intriguing open question is also how conformational changes are brought about within the intein structure during splicing, i.e., to facilitate the *trans*-esterification step, as has been proposed numerous times in the literature [23, 35, 36, 38, 40, 55, 56, 58, 135–137]. In general, we believe that a deeper understanding of the different protein splicing mechanisms, as well as new means to control or circumvent the context-dependency of inteins, for example with the use of evolved super-mutants, will greatly strengthen the applicability of the many intein-related technologies to challenges in protein engineering and protein biotechnology.

Acknowledgments We apologize to those researchers whose work could not be covered in detail due to space limitations and the special focus of this work. We thank all coworkers, past and present, for their contributions to the group's research. Funding in the Mootz lab was provided by the DFG (grant DFG MO 1073/3-1) and the HFSP (Grant RGP0031/2010).

References

- Hirata R, Ohsumi Y, Nakano A, Kawasaki H, Suzuki K, Anraku Y (1990) Molecular structure of a gene, VMA1, encoding the catalytic subunit of H(+)-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. *J Biol Chem* 265:6726–6733
- Kane PM, Yamashiro CT, Wolczyk DF, Neff N, Goebel M, Stevens TH (1990) Protein splicing converts the yeast TFP1 gene product to the 69-kD subunit of the vacuolar H(+)-adenosine triphosphatase. *Science* 250:651–657
- Perler FB (1998) Protein splicing of inteins and hedgehog autoproteolysis: structure, function, and evolution. *Cell* 92:1–4
- Liu XQ (2000) Protein-splicing intein: genetic mobility, origin, and evolution. *Annu Rev Genet* 34:61–76
- Petrokovski S (2001) Intein spread and extinction in evolution. *Trends Genet* 17:465–472
- Gogarten JP, Senejani AG, Zhaxybayeva O, Olenzinski L, Hilario E (2002) Inteins: structure, function, and evolution. *Annu Rev Microbiol* 56:263–287
- Perler FB (2002) InBase: the intein database. *Nucleic Acids Res* 30:383–384
- Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, Leahy DJ (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* 91:85–97
- Dassa B, Haviv H, Amitai G, Petrokovski S (2004) Protein splicing and auto-cleavage of bacterial intein-like domains lacking a C'-flanking nucleophilic residue. *J Biol Chem* 279:32001–32007
- Dassa B, Yanai I, Petrokovski S (2004) New type of poly-ubiquitin-like genes with intein-like autoprocessing domains. *Trends Genet* 20:538–542
- Perler FB, Olsen GJ, Adam E (1997) Compilation and analysis of intein sequences. *Nucleic Acids Res* 25:1087–1093
- Tori K, Dassa B, Johnson MA, Southworth MW, Brace LE, Ishino Y, Petrokovski S, Perler FB (2010) 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* 285:2515–2526
- Mootz HD (2009) Split inteins as versatile tools for protein semisynthesis. *Chem Bio Chem* 10:2579–2589
- Elleuche S, Poggeler S (2010) Inteins, valuable genetic elements in molecular biology and biotechnology. *Appl Microbiol Biotechnol* 87:479–489
- Vila-Perello M, Muir TW (2010) Biological applications of protein splicing. *Cell* 143:191–200
- Volkmann G, Iwai H (2010) Protein *trans*-splicing and its use in structural biology: opportunities and limitations. *Mol Biosys* 6:2110–2121
- Aranko AS, Volkmann G (2011) Protein *trans*-splicing as a protein ligation tool to study protein structure and function. *Biomol Concepts* 2:183–198
- Shah NH, Muir TW (2011) Split inteins: nature's protein ligases. *Isr J Chem* 51:854–861
- Cheriyian M, Perler FB (2009) Protein splicing: a versatile tool for drug discovery. *Adv Drug Deliv Rev* 61:899–907
- Sancheti H, Camarero JA (2009) "Splicing up" drug discovery. Cell-based expression and screening of genetically encoded libraries of backbone-cyclized polypeptides. *Adv Drug Deliv Rev* 61:908–917
- Xu MQ, Perler FB (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J* 15:5146–5153
- Southworth MW, Benner J, Perler FB (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J* 19:5019–5026

23. Saleh L, Southworth MW, Considine N, O'Neill C, Benner J, Bollinger JM Jr, Perler FB (2012) Branched intermediate formation is the slowest step in the protein splicing reaction of the Ala1 KlbA intein from *Methanococcus jannaschii*. *Biochemistry* 50:10576–10589
24. Johnson MA, Southworth MW, Herrmann T, Brace L, Perler FB, Wuthrich K (2007) NMR structure of a KlbA intein precursor from *Methanococcus jannaschii*. *Protein Sci* 16:1316–1328
25. Brace LE, Southworth MW, Tori K, Cushing ML, Perler F (2010) The *Deinococcus radiodurans* Snf2 intein caught in the act: detection of the class 3 intein signature block F branched intermediate. *Protein Sci* 19:1525–1533
26. Reitter JN, Mills KV (2011) Canonical protein splicing of a class one intein that has a class three non-canonical sequence motif. *J Bacteriol* 193:994–997
27. Tori K, Perler FB (2011) Expanding the definition of class 3 inteins and their proposed phage origin. *J Bacteriol* 193:2035–2041
28. Paulus H (2000) Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 69:447–496
29. Johansson DG, Wallin G, Sandberg A, Macao B, Aqvist J, Hard T (2009) Protein autoproteolysis: conformational strain linked to the rate of peptide cleavage by the pH dependence of the N → O acyl shift reaction. *J Am Chem Soc* 131:9475–9477
30. Brannigan JA, Dodson G, Duggleby HJ, Moody PC, Smith JL, Tomchick DR, Murzin AG (1995) A protein catalytic framework with an N-terminal nucleophile is capable of self-activation. *Nature* 378:416–419
31. Ditzel L, Huber R, Mann K, Heinemeyer W, Wolf DH, Groll M (1998) Conformational constraints for protein self-cleavage in the proteasome. *J Mol Biol* 279:1187–1191
32. Kawasaki M, Nogami S, Satow Y, Ohya Y, Anraku Y (1997) 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* 272:15668–15674
33. Ghosh I, Sun L, Xu MQ (2001) Zinc inhibition of protein *trans*-splicing and identification of regions essential for splicing and association of a split intein. *J Biol Chem* 276:24051–24058
34. Klabunde T, Sharma S, Telenti A, Jacobs WR Jr, Sacchettini JC (1998) Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat Struct Biol* 5:31–36
35. Poland BW, Xu MQ, Quirocho FA (2000) Structural insights into the protein splicing mechanism of PI-SceI. *J Biol Chem* 275:16408–16413
36. Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, Lesage G, Rao Z (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J Biol Chem* 278:39133–39142
37. Mizutani R, Nogami S, Kawasaki M, Ohya Y, Anraku Y, Satow Y (2002) Protein-splicing reaction via a thiazolidine intermediate: crystal structure of the VMA1-derived endonuclease bearing the N and C-terminal propeptides. *J Mol Biol* 316:919–929
38. Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z (2005) Crystal structures of an intein from the split dnaE gene of *Synechocystis* sp. PCC6803 reveal the catalytic model without the penultimate histidine and the mechanism of zinc ion inhibition of protein splicing. *J Mol Biol* 353:1093–1105
39. Van Roey P, Pereira B, Li Z, Hiraga K, Belfort M, Derbyshire V (2007) Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *J Mol Biol* 367:162–173
40. Du Z, Shemella PT, Liu Y, McCallum SA, Pereira B, Nayak SK, Belfort G, Belfort M, Wang C (2009) Highly conserved histidine plays a dual catalytic role in protein splicing: a pKa shift mechanism. *J Am Chem Soc* 131:11581–11589
41. Romanelli A, Shekhtman A, Cowburn D, Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein–intein junction. *Proc Natl Acad Sci USA* 101:6397–6402
42. Pearl EJ, Tyndall JD, Poulter RT, Wilbanks SM (2007) Sequence requirements for splicing by the Cne PRP8 intein. *FEBS Lett* 581:3000–3004
43. Du Z, Liu J, Albracht CD, Hsu A, Chen W, Marieni MD, Colelli KM, Williams JE, Reitter JN, Mills KV, Wang C (2011) Structural and mutational studies of a hyperthermophilic intein from DNA polymerase II of *Pyrococcus abyssi*. *J Biol Chem* 286:38638–38648
44. Pietrokovski S (1998) Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci* 7:64–71
45. Tori K, Cheriyan M, Pedamallu CS, Contreras MA, Perler FB (2012) The *Thermococcus kodakaraensis* Tko CDC21-1 intein activates its N-terminal splice junction in the absence of a conserved histidine by a compensatory mechanism. *Biochemistry* 51:2496–2505
46. Du Z, Zheng Y, Patterson M, Liu Y, Wang C (2011) pK(a) coupling at the intein active site: implications for the coordination mechanism of protein splicing with a conserved aspartate. *J Am Chem Soc* 133:10275–10282
47. Schwarzer D, Ludwig C, Thiel IV, Mootz HD (2012) Probing intein-catalyzed thioester formation by unnatural amino acid substitutions in the active site. *Biochemistry* 51:233–242
48. Appleby JH, Zhou K, Volkmann G, Liu XQ (2009) Novel Split Intein for *trans*-splicing synthetic peptide onto C-terminus of protein. *J Biol Chem* 284:6194–6199
49. Volkmann G, Liu XQ (2011) Intein lacking conserved C-terminal motif G retains controllable N-cleavage activity. *FEBS J* 278:3431–3446
50. Ludwig C, Schwarzer D, Mootz HD (2008) Interaction studies and alanine scanning analysis of a semi-synthetic split intein reveal thiazoline ring formation from an intermediate of the protein splicing reaction. *J Biol Chem* 283:25264–25272
51. Kang J, Richardson JP, Macmillan D (2009) 3-Mercaptopropionic acid-mediated synthesis of peptide and protein thioesters. *Chem Commun (Camb)* 407–409
52. Kang J, Macmillan D (2010) Peptide and protein thioester synthesis via N → S acyl transfer. *Org Biomol Chem* 8:1993–2002
53. Kawakami T, Aimoto S (2007) Sequential peptide ligation by using a controlled cysteinyl prolyl ester (CPE) autoactivating unit. *Tetrahedron Lett* 48:1903–1905
54. Pereira B, Shemella PT, Amitai G, Belfort G, Nayak SK, Belfort M (2011) Spontaneous proton transfer to a conserved intein residue determines on-pathway protein splicing. *J Mol Biol* 406:430–442
55. Chong S, Williams KS, Wotkowicz C, Xu MQ (1998) Modulation of protein splicing of the *Saccharomyces cerevisiae* vacuolar membrane ATPase intein. *J Biol Chem* 273:10567–10577
56. Martin DD, Xu MQ, Evans TC Jr (2001) Characterization of a naturally occurring *trans*-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry* 40:1393–1402
57. Zettler J, Schutz V, Mootz HD (2009) The naturally split Npu DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction. *FEBS Lett* 583:909–914

58. Frutos S, Goger M, Giovani B, Cowburn D, Muir TW (2010) Branched intermediate formation stimulates peptide bond cleavage in protein splicing. *Nat Chem Biol* 6:527–533
59. Xu MQ, Comb DG, Paulus H, Noren CJ, Shao Y, Perler FB (1994) Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *EMBO J* 13:5517–5522
60. Shao Y, Xu MQ, Paulus H (1995) Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence. *Biochemistry* 34:10844–10850
61. Stephenson RC, Clarke S (1989) Succinimide formation from aspartyl and asparaginyl peptides as a model for the spontaneous degradation of proteins. *J Biol Chem* 264:6164–6170
62. Shemella P, Pereira B, Zhang Y, Van Roey P, Belfort G, Garde S, Nayak SK (2007) Mechanism for intein C-terminal cleavage: a proposal from quantum mechanical calculations. *Biophys J* 92:847–853
63. Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M (1999) A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol* 17:889–892
64. Mathys S, Evans TC, Chute IC, Wu H, Chong S, Benner J, Liu XQ, Xu MQ (1999) Characterization of a self-splicing mini-intein and its conversion into autocatalytic N- and C-terminal cleavage elements: facile production of protein building blocks for protein ligation. *Gene* 231:1–13
65. Wood DW, Derbyshire V, Wu W, Chartrain M, Belfort M, Belfort G (2000) Optimized single-step affinity purification with a self-cleaving intein applied to human acidic fibroblast growth factor. *Biotechnol Prog* 16:1055–1063
66. Mujika JI, Lopez X, Mulholland AJ (2009) Modeling protein splicing: reaction pathway for C-terminal splice and intein scission. *J Phys Chem B* 113:5607–5616
67. Kurpiers T, Mootz HD (2008) Site-specific chemical modification of proteins with a prelabelled cysteine tag using the artificially split Mxe GyrA intein. *ChemBioChem* 9:2317–2325
68. Shao Y, Paulus H (1997) Protein splicing: estimation of the rate of O–N and S–N acyl rearrangements, the last step of the splicing process. *J Pept Res* 50:193–198
69. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M (2009) Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci USA* 106:11005–11010
70. Ellila S, Jurvansuu JM, Iwai H (2011) Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in *Escherichia coli*. *FEBS Lett* 585:3471–3477
71. Shah NH, Dann GP, Vila-Perello M, Liu Z, Muir TW (2012) Ultrafast protein splicing is common among cyanobacterial split inteins: implications for protein engineering. *J Am Chem Soc* 134(28):11338–11341
72. Øemig JS, Zhou D, Kajander T, Wlodawer A, Iwai H (2012) NMR and crystal structures of the *Pyrococcus horikoshii* RadA intein guide a strategy for engineering a highly efficient and promiscuous intein. *J Mol Biol* 421(1):85–99
73. Adam E, Perler FB (2002) Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in *Escherichia coli* DNA gyrase subunit A. *J Mol Microbiol Biotechnol* 4:479–487
74. Cann IK, Amaya KR, Southworth MW, Perler FB (2004) Bacteriophage-based genetic system for selection of nonsplicing inteins. *Appl Environ Microbiol* 70:3158–3162
75. Zeidler MP, Tan C, Bellaiche Y, Cherry S, Hader S, Gayko U, Perrimon N (2004) Temperature-sensitive control of protein activity by conditionally splicing inteins. *Nat Biotechnol* 22:871–876
76. Tan G, Chen M, Foote C, Tan C (2009) Temperature-sensitive mutations made easy: generating conditional mutations by using temperature-sensitive inteins that function within different temperature ranges. *Genetics* 183:13–22
77. Hiraga K, Soga I, Dansereau JT, Pereira B, Derbyshire V, Du Z, Wang C, Van Roey P, Belfort G, Belfort M (2009) Selection and structure of hyperactive inteins: peripheral changes relayed to the catalytic center. *J Mol Biol* 393:1106–1117
78. Hiraga K, Derbyshire V, Dansereau JT, Van Roey P, Belfort M (2005) Minimization and stabilization of the *Mycobacterium tuberculosis* recA intein. *J Mol Biol* 354:916–926
79. Lockless SW, Muir TW (2009) Traceless protein splicing utilizing evolved split inteins. *Proc Natl Acad Sci USA* 106:10999–11004
80. Buskirk AR, Ong YC, Gartner ZJ, Liu DR (2004) Directed evolution of ligand dependence: small-molecule-activated protein splicing. *Proc Natl Acad Sci USA* 101:10505–10510
81. Wu H, Xu MQ, Liu XQ (1998) Protein *trans*-splicing and functional mini-inteins of a cyanobacterial dnaB intein. *Biochim Biophys Acta* 1387:422–432
82. Appleby-Tagoe JH, Thiel IV, Wang Y, Wang Y, Mootz HD, Liu XQ (2011) Highly efficient and more general *cis*- and *trans*-splicing inteins through sequential directed evolution. *J Biol Chem* 286:34440–34447
83. Du Z, Liu Y, Ban D, Lopez MM, Belfort M, Wang C (2010) Backbone dynamics and global effects of an activating mutation in minimized Mtu RecA inteins. *J Mol Biol* 400:755–767
84. Caspi J, Amitai G, Belenkiy O, Pietrovski S (2003) Distribution of split DnaE inteins in cyanobacteria. *Mol Microbiol* 50:1569–1577
85. Iwai H, Züger S, Jin J, Tam PH (2006) Highly efficient protein *trans*-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett* 580:1853–1858
86. Dassa B, Amitai G, Caspi J, Schueler-Furman O, Pietrovski S (2007) *Trans* protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. *Biochemistry* 46:322–330
87. Peck SH, Chen I, Liu DR (2011) Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells. *Chem Biol* 18:619–630
88. Skretas G, Wood DW (2005) Regulation of protein activity with small-molecule-controlled inteins. *Protein Sci* 14:523–532
89. Muir TW, Sondhi D, Cole PA (1998) Expressed protein ligation: a general method for protein engineering. *Proc Natl Acad Sci USA* 95:6705–6710
90. Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* 72:249–289
91. Evans TC Jr, Benner J, Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 7:2256–2264
92. Xu MQ, Evans TC Jr (2003) Purification of recombinant proteins from *E. coli* by engineered inteins. *Methods Mol Biol* 205:43–68
93. Muralidharan V, Muir TW (2006) Protein ligation: an enabling technology for the biophysical analysis of proteins. *Nat Methods* 3:429–438
94. Shi J, Muir TW (2005) Development of a tandem protein *trans*-splicing system based on native and engineered split inteins. *J Am Chem Soc* 127:6198–6206
95. Lu W, Sun Z, Tang Y, Chen J, Tang F, Zhang J, Liu JN (2011) Split intein facilitated tag affinity purification for recombinant proteins with controllable tag removal by inducible auto-cleavage. *J Chromatogr A* 1218:2553–2560
96. Kurpiers T, Mootz HD (2007) Regioselective cysteine bioconjugation by appending a labeled cysteine tag to a protein by using protein splicing in *trans*. *Angew Chem Int Ed Engl* 46:5234–5237

97. Brenzel S, Cebi M, Reiss P, Koert U, Mootz HD (2009) 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* 10:983–986
98. Ludwig C, Pfeiff M, Linne U, Mootz HD (2006) Ligation of a synthetic peptide to the N-terminus of a recombinant protein using semisynthetic protein *trans*-splicing. *Angew Chem Int Ed Engl* 45:5218–5221
99. Volkmann G, Liu XQ (2009) Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein. *PLoS ONE* 4:e8381
100. Yang JY, Yang WY (2009) Site-specific two-color protein labeling for FRET studies using split inteins. *J Am Chem Soc* 131:11644–11645
101. Ando T, Tsukiji S, Tanaka T, Nagamune T (2007) Construction of a small-molecule-integrated semisynthetic split intein for in vivo protein ligation. *Chem Commun (Camb)* 4995–4997
102. Charalambous A, Andreou M, Skourides PA (2009) Intein-mediated site-specific conjugation of quantum dots to proteins in vivo. *J Nanobiotechnol* 7:9
103. Olschewski D, Seidel R, Miesbauer M, Rambold AS, Oesterhelt D, Winkhofer KF, Tatzelt J, Engelhard M, Becker CF (2007) Semisynthetic murine prion protein equipped with a GPI anchor mimic incorporates into cellular membranes. *Chem Biol* 14:994–1006
104. Chu NK, Olschewski D, Seidel R, Winkhofer KF, Tatzelt J, Engelhard M, Becker CF (2010) Protein immobilization on liposomes and lipid-coated nanoparticles by protein *trans*-splicing. *J Pept Sci* 16:582–588
105. Kwon Y, Coleman MA, Camarero JA (2006) Selective immobilization of proteins onto solid supports through split-intein-mediated protein *trans*-splicing. *Angew Chem Int Ed Engl* 45:1726–1729
106. Lew BM, Mills KV, Paulus H (1999) Characteristics of protein splicing in *trans*-mediated by a semisynthetic split intein. *Biopolymers* 51:355–362
107. Evans TC Jr, Martin D, Kolly R, Panne D, Sun L, Ghosh I, Chen L, Benner J, Liu XQ, Xu MQ (2000) Protein *trans*-splicing and cyclization by a naturally split intein from the *dnaE* gene of *Synechocystis* species PCC6803. *J Biol Chem* 275:9091–9094
108. Øemig JS, Aranko AS, Djupsjöbacka J, Heinämäki K, Iwai H (2009) 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* 583:1451–1456
109. Aranko AS, Züger S, Buchinger E, Iwai H (2009) In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. *PLoS ONE* 4:e5185
110. Giriat I, Muir TW (2003) Protein semi-synthesis in living cells. *J Am Chem Soc* 125:7180–7181
111. Borra R, Dong D, Elnagar AY, Woldemariam GA, Camarero JA (2012) In-cell fluorescence activation and labeling of proteins mediated by FRET-quenched split inteins. *J Am Chem Soc* 134:6344–6353
112. Liu CC, Schultz PG (2010) Adding new chemistries to the genetic code. *Annu Rev Biochem* 79:413–444
113. Kanno A, Ozawa T, Umezawa Y (2009) Bioluminescent imaging of MAPK function with intein-mediated reporter gene assay. *Methods Mol Biol* 574:185–192
114. Kanno A, Umezawa Y, Ozawa T (2009) Detection of apoptosis using cyclic luciferase in living mammals. *Methods Mol Biol* 574:105–114
115. Zhang Y, Yang W, Chen L, Shi Y, Li G, Zhou N (2011) Development of a novel DnaE intein-based assay for quantitative analysis of G-protein-coupled receptor internalization. *Anal Biochem* 417:65–72
116. Wong SS, Kotera I, Mills E, Suzuki H, Truong K (2012) Split-intein-mediated re-assembly of genetically encoded Ca(2+) indicators. *Cell Calcium* 51:57–64
117. Gils M, Marillonnet S, Werner S, Grutzner R, Giritich A, Engler C, Schachschneider R, Klimyuk V, Gleba Y (2008) A novel hybrid seed system for plants. *Plant Biotechnol J* 6:226–235
118. Kempe K, Rubtsova M, Gils M (2009) Intein-mediated protein assembly in transgenic wheat: production of active barnase and acetolactate synthase from split genes. *Plant Biotechnol J* 7:283–297
119. Chin HG, Kim GD, Marin I, Mersha F, Evans TC Jr, Chen L, Xu MQ, Pradhan S (2003) Protein *trans*-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes. *Proc Natl Acad Sci USA* 100:4510–4515
120. Yuen CM, Rodda SJ, Vokes SA, McMahon AP, Liu DR (2006) Control of transcription factor activity and osteoblast differentiation in mammalian cells using an evolved small-molecule-dependent intein. *J Am Chem Soc* 128:8939–8946
121. Mootz HD, Muir TW (2002) Protein splicing triggered by a small molecule. *J Am Chem Soc* 124:9044–9045
122. Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW (2003) Conditional protein splicing: a new tool to control protein structure and function in vitro and in vivo. *J Am Chem Soc* 125:10561–10569
123. Mootz HD, Blum ES, Muir TW (2004) Activation of an auto-regulated protein kinase by conditional protein splicing. *Angew Chem Int Ed Engl* 43:5189–5192
124. Schwartz EC, Saez L, Young MW, Muir TW (2007) Post-translational enzyme activation in an animal via optimized conditional protein splicing. *Nat Chem Biol* 3:50–54
125. Tyszkiewicz AB, Muir TW (2008) Activation of protein splicing with light in yeast. *Nat Methods* 5:303–305
126. Sonntag T, Mootz HD (2011) An intein-cassette integration approach used for the generation of a split TEV protease activated by conditional protein splicing. *Mol Biosyst* 7:2031–2039
127. Berrade L, Kwon Y, Camarero JA (2010) Photomodulation of protein *trans*-splicing through backbone photocaging of the DnaE split intein. *Chem Bio Chem* 11:1368–1372
128. Vila-Perello M, Hori Y, Ribo M, Muir TW (2008) Activation of protein splicing by protease- or light-triggered O to N acyl migration. *Angew Chem Int Ed Engl* 47:7764–7767
129. Binschik J, Zettler J, Mootz HD (2011) Photocontrol of protein activity mediated by the cleavage reaction of a split intein. *Angew Chem Int Ed Engl* 50:3249–3252
130. Otomo T, Ito N, Kyogoku Y, Yamazaki T (1999) NMR observation of selected segments in a larger protein: central-segment isotope labeling through intein-mediated ligation. *Biochemistry* 38:16040–16044
131. Brenzel S, Kurpiers T, Mootz HD (2006) 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* 45:1571–1578
132. Busche AE, Aranko AS, Talebzadeh-Farooji M, Bernhard F, Dötsch V, Iwai H (2009) 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* 48:6128–6131
133. Shah NH, Vila-Perello M, Muir TW (2011) Kinetic control of one-pot *trans*-splicing reactions by using a wild-type and designed split intein. *Angew Chem Int Ed Engl* 50:6511–6515

134. Callahan BP, Topilina NI, Stanger MJ, Van Roey P, Belfort M (2011) Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications. *Nat Struct Mol Biol* 18:630–633
135. Perler FB (2005) Protein splicing mechanisms and applications. *IUBMB Life* 57:469–476
136. Mills KV, Dorval DM, Lewandowski KT (2005) Kinetic analysis of the individual steps of protein splicing for the *Pyrococcus abyssi* PolII intein. *J Biol Chem* 280:2714–2720
137. Saleh L, Perler FB (2006) Protein splicing in *cis* and in *trans*. *Chem Rec* 6:183–193