Mobile DNA

Recent advances in *in vivo* applications of intein-mediated protein splicing

Natalya I Topilina[1] and Kenneth V Mills[2*]

Abstract

Intein-mediated protein splicing has become an essential tool in modern biotechnology. Fundamental progress in the structure and catalytic strategies of *cis*- and *trans*-splicing inteins has led to the development of modified inteins that promote efficient protein purification, ligation, modification and cyclization. Recent work has extended these *in vitro* applications to the cell or to whole organisms. We review recent advances in intein-mediated protein expression and modification, post-translational processing and labeling, protein regulation by conditional protein splicing, biosensors, and expression of *trans*-genes.

Keywords: Intein, Protein splicing, Post-translational modification, Protein purification, *Trans*-splicing, Conditional protein splicing, *Trans*-genes, Biosensor

Introduction

Protein splicing is a post-translational process by which an intervening polypeptide, called an intein, catalyzes its own excision from the flanking polypeptides, or exteins, as well as ligation of the exteins (Figure 1A).

Many inteins are interrupted by homing endonuclease domains similar to those found in mobile introns. However, the homing endonuclease domain can be deleted from the intein without complete loss of splicing activity and is absent in a class of inteins called mini-inteins [1].

The mechanism of splicing for canonical inteins is a four-step process (Figure 2) [1,2]. First, the peptide bond linking the N-extein and intein is converted to a thioester or ester via nucleophilic attack by the N-terminal Cys or Ser of the intein (Step 1). Second, the N-extein is transferred from the side chain of the first intein residue to the side chain of the first C–extein residue (Cys, Ser or Thr) by transesterification, resulting in a branched ester intermediate (Step 2). Third, the branched ester is resolved by Asn cyclization coupled to peptide bond cleavage (Step 3). This leaves the ligated exteins separated from the intein and linked by an ester bond, while the intein has a C-terminal aminosuccinimide. Finally, the ester bond connecting the ligated exteins is rapidly

converted to the amide bond, and the C-terminal aminosuccinimide of the intein may be hydrolyzed (Step 4).

Two side reactions can occur if the splicing process is disrupted (Figure 1 B and C). The ester or thioester formed in Steps 1 and/or 2 may be cleaved by hydrolysis or thiolysis, uncoupled from Asn cyclization. This process is called N-terminal cleavage, and results in cleavage of the N-extein from the precursor (Figure 1B). Alternatively, Asn cyclization may occur uncoupled from Steps 1 and 2, freeing the C-extein in a process called C-terminal cleavage (Figure 1C).

Some inteins are expressed as two separate fragments, either natively or by protein engineering, and facilitate protein splicing in *trans* (Figure 1D). In protein *trans*-splicing, or PTS, reassociation of the intein fragments is required prior to splicing.
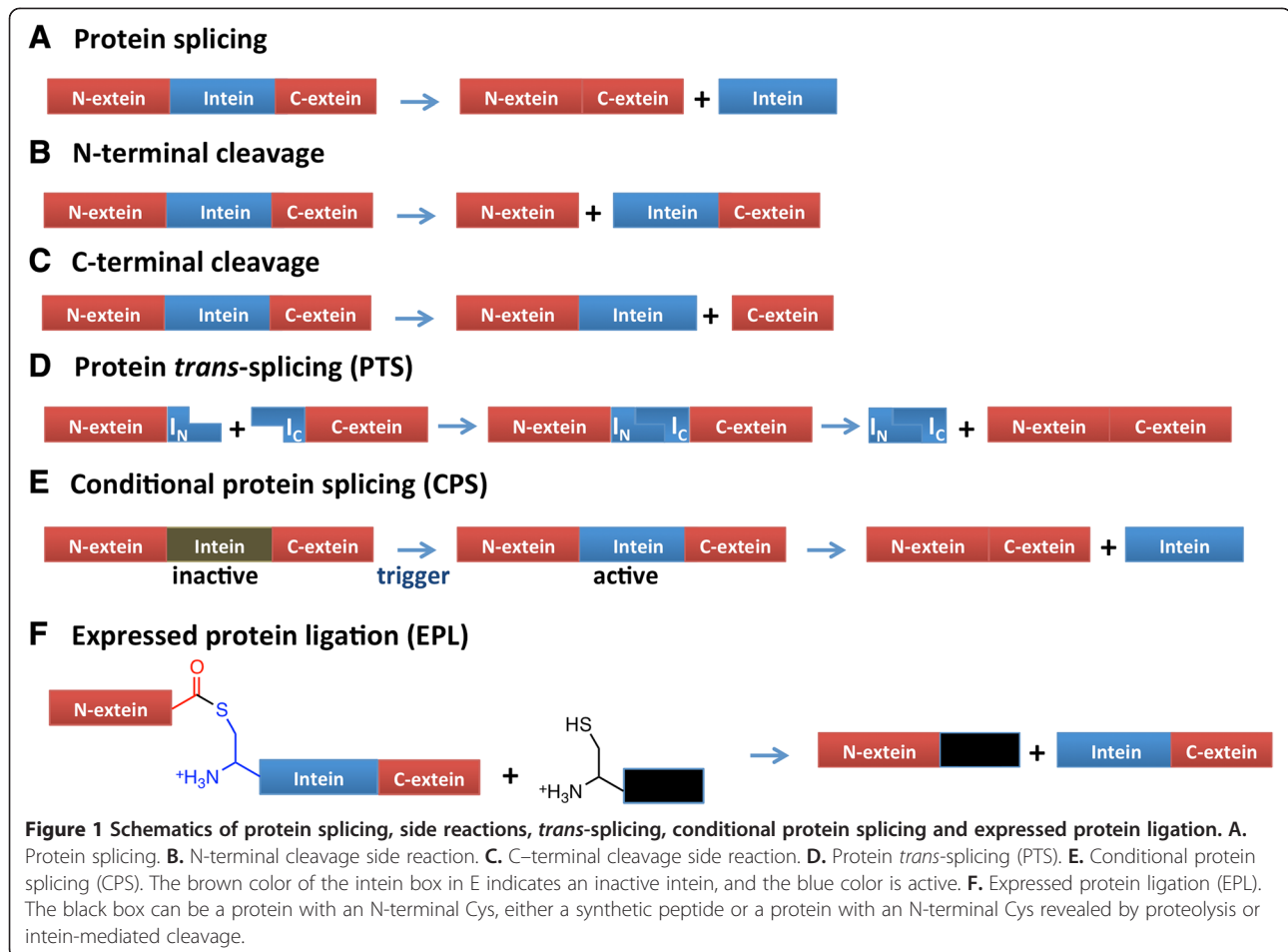
Both *cis*- and *trans*-splicing inteins have been engineered to undergo conditional protein splicing, or CPS (Figure 1E). CPS requires the addition of a trigger to initiate splicing of a precursor fusion protein. Such triggers include light, changes in pH or temperature, change in redox state, or the addition of a small molecule [2-6]. For CPS of *trans*-splicing inteins, split dimerization domains have been fused to the intein fragments to make reassociation conditional on addition of a small molecule or on the affinity of the domains [2-6].

Intein-based methods have been developed to facilitate purification and post-translational modifications of

* Correspondence: kmills@holycross.edu
[2]Department of Chemistry, College of the Holy Cross, 1 College Street, Worcester, MA 01610, USA
Full list of author information is available at the end of the article

Figure 1 Schematics of protein splicing, side reactions, *trans*-splicing, conditional protein splicing and expressed protein ligation. A. Protein splicing. **B.** N-terminal cleavage side reaction. **C.** C–terminal cleavage side reaction. **D.** Protein *trans*-splicing (PTS). **E.** Conditional protein splicing (CPS). The brown color of the intein box in E indicates an inactive intein, and the blue color is active. **F.** Expressed protein ligation (EPL). The black box can be a protein with an N-terminal Cys, either a synthetic peptide or a protein with an N-terminal Cys revealed by proteolysis or intein-mediated cleavage.

recombinant proteins. Expressed protein ligation (EPL) and protein *trans*-splicing (PTS) [7-9] can produce proteins with site-specific incorporation of a diverse set of chemical modifications [6,7,10-13]. We introduce these methods here and will describe more recent applications below.

EPL is a method to modify the C terminus of a recombinant protein [14,15] (Figure 1F). The protein is fused at its C terminus to an intein, which promotes formation of a thioester between the protein and intein. The protein is then transferred to the side chain of a synthetic peptide with an N-terminal Cys (or a protein with N-terminal Cys revealed by proteolysis or intein-mediated cleavage). The peptide may contain non-native amino acids or other chemical probes that can be incorporated by solid phase peptide synthesis. EPL is similar to native chemical ligation [16], which facilitates the ligation of a small synthetic peptide with a C-terminal α-thioester to a peptide with an N-terminal Cys. In most EPL strategies, the ligated segments have no natural affinity to each other and there is an entropic barrier to ligation. However, this entropic barrier can be overcome if the segments to be ligated have affinity for one another [17].
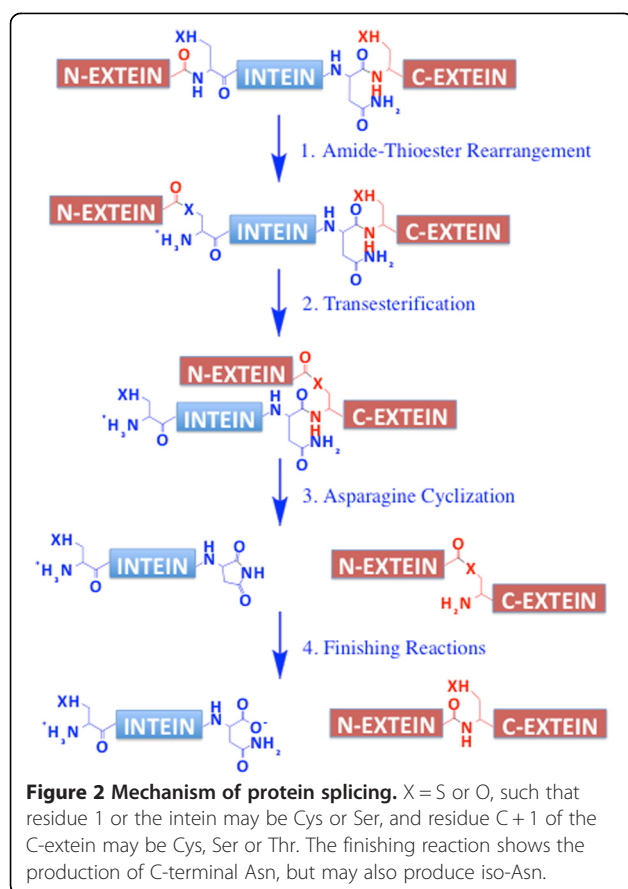
In PTS, natively or artificially split inteins ligate the exteins via a peptide bond [18,19]. The production of semi-synthetic proteins via PTS takes advantage of the affinity between the intein fragments. However, PTS relies on efficient splicing rather than on intein side reactions. Thereby it has additional challenges in that efficient splicing may depend more heavily on the presence of short native exteins and on where the target protein is split. As a result, one may need to incorporate short native extein sequence into the final ligation product or substantially optimize the split site.

Intein biotechnology applications have been extensively and expertly reviewed in the recent literature [2-6]. We aim to describe in detail the most recent advances in this area, including protein expression and modification, post-translational processing and labeling, protein regulation by conditional protein splicing, biosensors, and the expression of *trans*-genes.

Review

Protein expression and modification

Intein-based methods can be used to modify the sequence or structure of recombinant proteins, including

Figure 2 Mechanism of protein splicing. X = S or O, such that residue 1 or the intein may be Cys or Ser, and residue C + 1 of the C-extein may be Cys, Ser or Thr. The finishing reaction shows the production of C-terminal Asn, but may also produce iso-Asn.

protein cyclization or polymerization, expression of proteins with native N-terminal residues, and site-specific proteolysis. Inteins can facilitate the expression of toxic proteins and large proteins from within the same reading frame, can allow for post-translational generation of small peptides, and can serve as selectable genetic markers.

Cyclization

Protein and peptide cyclization can be facilitated by inteins by two methods (reviewed in refs [3,4,20,21].) In the first method (Figure 3A), cyclization can be achieved by having the protein of interest (POI) fused to different inteins at its N and C termini [22,23]. C-terminal cleavage at the N-intein/target junction results in an N-terminal Cys residue. Alternatively, an N-terminal Cys residue may be generated by proteolysis. This Cys reacts with an activated thioester generated at the target/C-intein junction via EPL to produce cyclized protein. In the second method (Figure 3B), split inteins are used to produce cyclized peptides or proteins. The target protein or peptide is expressed as a fusion between C- and N-split intein fragments [24,25]. The inversion of the placement of the N- and C-intein fragments in the precursor

(I_C-protein-I_N) ensures that PTS results in ligation of the internal polypeptide fragment. The cyclization of the target proteins results in enhanced stability and bioactivity [26-31]. One of the most exciting applications of intein-mediated cyclization is the *in vivo* generation of large libraries of genetically-encoded cyclic peptides for high throughput screens [3]. In addition to cyclization, intein splicing from the I_C-protein-I_N precursor may result in polymerization of the target protein [32] (Figure 3C).

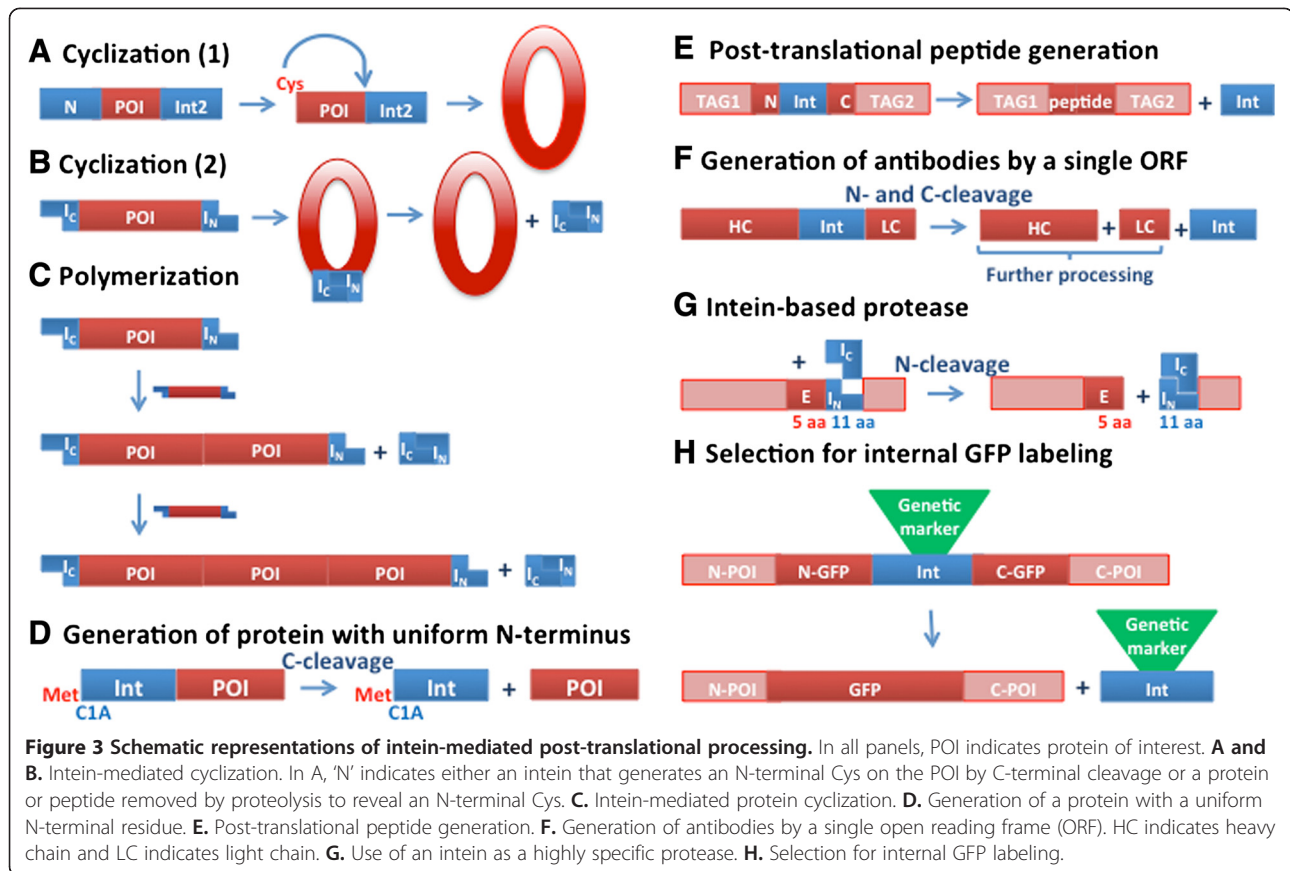### Proteins with uniform N-terminal residues

Inteins have been used to express proteins in *E. coli* with uniform N-terminal residues [33,34]. For instance, proteins expressed in their host organism that have signal sequences often have that signal sequence cleaved by an aminopeptidase, resulting in a protein with an N-terminal residue other that Met. When these proteins are over-expressed in *E. coli* without their native signal sequences, they can be subject to unwanted N-terminal processing by aminopeptidases. To avoid this cleavage and have their native N-terminal residue, a target protein was fused at its N terminus to the *Ssp* DnaB mini-intein, and expressed in an aminopeptidase-deficient strain (Figure 3D). Upon C-terminal cleavage of the intein, the target protein with the desired N-terminal amino acid was generated [33,34].

Expression of peptides, toxic proteins, and proteins from a single reading frame

Intein technology can be used to express challenging targets: small peptides can be expressed as part of properly folded proteins with affinity domains, toxic proteins can be expressed in an inactivated format, and proteins that function in an essential stoichiometry can be expressed from a single open reading frame.

Intein catalysis can facilitate post-translational generation of peptides by protein splicing [35] (Figure 3E). Intein-mediated intracellular peptide production was used to distinguish between the behavior of peptides generated by cleavage of disordered, defective ribosomal products and those generated from well-folded proteins. For example, it was thought that major histocompatibility complex (MHC) class I peptides were mostly derived from cleavage of misfolded protein fragments. However, peptides produced via splicing of either the *Mtu* RecA or *Pch* PRP8 mini-inteins also generated MHC class 1 epitopes. Because the intein precursor protein must be stably folded to facilitate splicing, this suggests that MHC peptides can be produced from stable, well-folded proteins.

Inteins can facilitate the overexpression of toxic proteins. For instance, the *Sce* VMA intein has been used to create building blocks for the semi-synthesis by EPL of active cytotoxic enzymes from inactive fragments,

Figure 3 Schematic representations of intein-mediated post-translational processing. In all panels, POI indicates protein of interest. **A and B.** Intein-mediated cyclization. In A, 'N' indicates either an intein that generates an N-terminal Cys on the POI by C-terminal cleavage or a protein or peptide removed by proteolysis to reveal an N-terminal Cys. **C.** Intein-mediated protein cyclization. **D.** Generation of a protein with a uniform N-terminal residue. **E.** Post-translational peptide generation. **F.** Generation of antibodies by a single open reading frame (ORF). HC indicates heavy chain and LC indicates light chain. **G.** Use of an intein as a highly specific protease. **H.** Selection for internal GFP labeling.

including bovine pancreatic RNase A and a restriction endonuclease from *Haemophilus parainfluenzae* (*Hpa*I) [14]. Another approach is to produce non-toxic protein precursor by inserting an intein in the toxic protein, with activity of the target protein dependent on CPS of the intein. For example, I-TevI endonuclease was expressed by inserting a modified *Mtu* RecA intein that is active only under specific pH conditions [36,37].

Expression of antibodies using a single open reading frame was achieved by fusing the genes for antibody heavy and light chains with an intein [38] (Figure 3F). This fusion protein was successfully expressed and processed in mammalian cells, with intein-directed N- and C- terminal cleavage reactions resulting in antibodies with the correct sequences for both heavy and light chains.

Intein proteases

Split inteins can be used to facilitate *in vivo*, site-specific protein cleavage [39] (Figure 3G). The 11-residue N-terminal fragment of the *Ssp* DnaB S1 split intein was inserted between two target sequences and used as a cleavage site that is recognized by the C-terminal intein fragment. This C-terminal fragment is called an intein-derived protease (IP), because on co-expression of IP and complementation with the N-extein fragment, the target protein is cleaved via N-terminal cleavage of the reconstituted split intein. Site-specific protein cleavage by the IP was demonstrated in bacterial and eukaryotic cells. In contrast to the relatively low substrate specificity of other commonly used proteases, this intein-derived protease has very limited unintended proteolysis of endogenous proteins resulting in minimal cellular toxicity. The authors suggested utilization of the IP as a molecular tool to provide control of protein cleavage inside living cells.

Inteins as genetic markers

Inteins can facilitate *in vivo* gene modification by serving as genetic markers [40] (Figure 3H). Muller and co-workers interrupted the *Pch* PRP8 intein with selectable markers, including aminoglycoside phosphotransferase and imidazoleglycerol-phosphate dehydratase. The interrupted inteins are able to splice, and could serve as selectable markers for expression of the spliced extein, GFP. This split GFP (or GFP:int) construct was used for a single step internal labeling of calmodulin with GFP in yeast.

Intein-mediated protein processing and labeling

EPL and PTS have been successfully used to produce diverse N- and C-terminal modifications of target proteins that are expertly reviewed in the recent literature [7,8,10,12], including protein phosphorylation, lipidation, glycosylation, biotinylation, ubiquitination, and segmental isotope labeling. Here we focus on *in vivo* protein modifications, including protein semi-synthesis on cell surfaces, segmental isotope labeling inside cells, and selective protein labeling inside of living cells.

Protein semi-synthesis on a cell surface

PTS has been used for protein semi-synthesis on a cell surface (Figure 4A). The C terminus of the human transferrin receptor was labeled with a fluorescent group (5-carboxy-fluorescein) on the surface of Chinese hamster ovary (CHO) cells using the *Ssp* GyrB split intein [41]. Likewise, the N terminus of the monomeric red fluorescent protein (mRFP) was modified with a biotin tag via PTS on the surface of CHO cells [42]. PTS can be employed for ligation of an endogenous polypeptide to a membrane protein on mammalian cells [43]. Mootz and coworkers attached enhanced GFP (eGFP) to transmembrane- and GPI-anchored proteins via a PTS reaction between the *Npu* DnaE Int_C fragment fused with the membrane-localized protein and *E. coli* overexpressed eGFP-Int_N fusion.

To overcome the low binding affinity between split intein pairs that have a short N-terminal fragment, an auxiliary receptor-ligand interaction was integrated, as outlined in Figure 4A [42].

In-cell protein labeling

PTS can facilitate segmental isotopic labeling *in* vivo, as well as the *in vivo* addition of chemical probes to specific target proteins. Cell-based PTS can provide tools for NMR analysis [10,44] (Figure 4B). Labeled and unlabeled precursor fragments can be produced within a single culture. The use of a dual expression system allows for sequential expression of the precursors in media enriched with different isotopes. Incorporation of unlabeled solubilizing tags into isotopically labeled target proteins was demonstrated by sequential overexpression of unlabeled domain B1 of the immunoglobulin binding protein G (GB1) and labeled prion-inducing domain of yeast Sup35p, each fused to *Ssp* DnaE split intein fragments. PTS resulted in production of the protein fusion with improved stability and solubility created by the NMR-invisible tag [44].

Recently, several intein-based methods for selective protein labeling inside of living cells have been developed. All of these methods employ newly developed split inteins with very small N- or C-intein fragments, ranging from 6 to 15 amino acids [45-49]. Because the intein fragments are so short, they are easier to synthesize and more likely to penetrate the cell to allow for *in vivo* labeling. Intein-based *in vivo* labeling has several advantages over methods based on molecular recognition and chemical modifications [50]. As opposed to direct
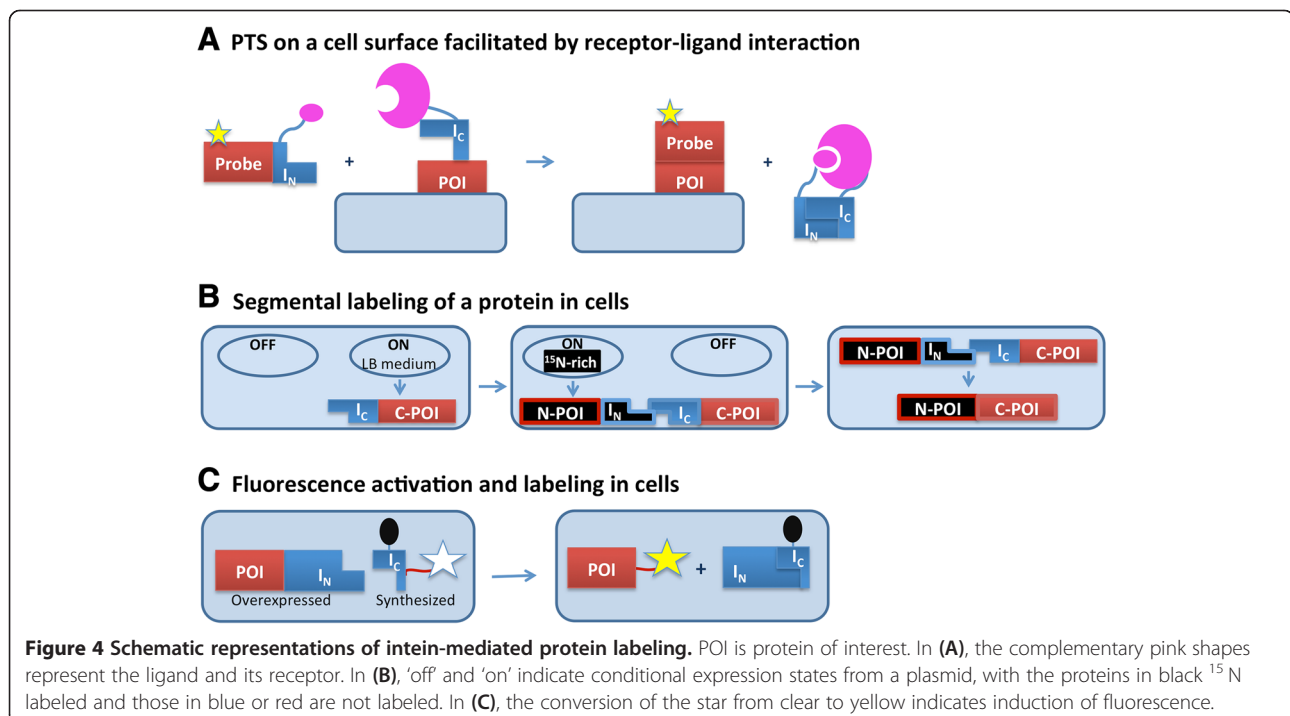

Figure 4 Schematic representations of intein-mediated protein labeling. POI is protein of interest. In **(A)**, the complementary pink shapes represent the ligand and its receptor. In **(B)**, 'off' and 'on' indicate conditional expression states from a plasmid, with the proteins in black ^{15}N labeled and those in blue or red are not labeled. In **(C)**, the conversion of the star from clear to yellow indicates induction of fluorescence.

chemical modifications, intein-based labeling relies on intein-based recognition and minimizes the background from non-reacted reagents.

One such intein-based labeling method uses native chemical ligation, in both bacterial and mammalian systems, to label glutathione-*S*-transferase (GST) and eGFP *in vivo*. The target proteins are expressed as a C-terminal fusion to the *Ssp* DnaB intein, and are designed to have an N-terminal Cys after intein C-terminal cleavage. The target proteins are then labeled by a cell-permeable, thioester-containing small molecule tag such as biotin or a fluorophore [51]. However, this ligation is inefficient because the label and the target have no native affinity for one another, and therefore excess of one reagent needs to be used, resulting in high background signal. Such difficulty was overcome by Camarero and coworkers, who utilized PTS to increase the affinity between target protein and probe, and used a quencher to reduce the signal from unreacted reagent [52] (Figure 4C). The quencher was introduced to the C-terminal *Ssp* or *Npu* DnaE intein fragment, while the fluorophore was part of the C-extein. Therefore, the quencher and fluorophore were part of the same molecule before splicing. On PTS, the fluorophore is ligated to the protein of interest and dissociated from the I_C-linked quencher.

A PTS-based site-specific conjugation of a quantum dot to the C terminus of pleckstrin homology (PH) domain was performed using the *Ssp* DnaE mini-intein inside of *Xenopus* embryos [53]. The authors speculate that their technique allows for covalent conjugation of any nanostructure and/or nanodevice to any protein within the cells of the developing embryo. Later the same group showed that the approach can be extended to accomplish N-terminal protein tagging using the *Ssp* DnaB mini-intein [54]. This work demonstrated the possibility of site-specific conjugation of quantum dots to several proteins simultaneously, allowing for multi-parameter imaging [54].

Regulation of protein function by conditional protein splicing

For protein splicing to regulate the activity of a protein *in vivo*, it must splice conditionally, either in *cis* or in *trans.* CPS is activated by a trigger, such as a small molecule, light, temperature, pH or change in redox state (reviewed in [7]). For CPS to be physiologically relevant, it remains to be shown that native inteins are sensitive to stimulus in their native extein contexts, expressed in their host organism. Such evidence would counteract the belief that inteins persist solely as selfish genetic elements, and are difficult to remove because they interrupt key proteins such as DNA polymerase and recombinase [55]. Rather, some inteins still may play a beneficial role

for their host, which might provide a positive selective pressure to retain the intein.
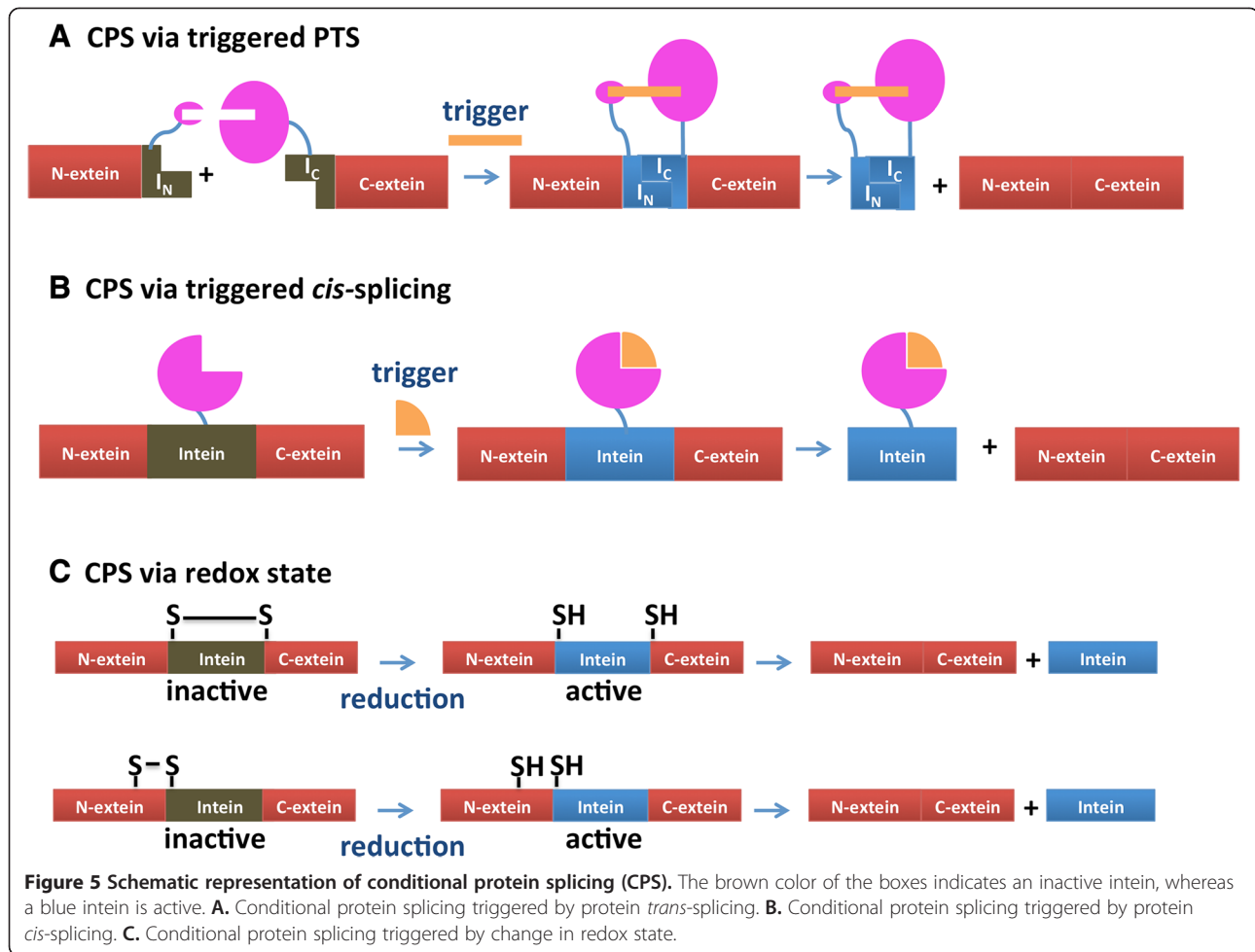
Small molecule induced CPS

PTS facilitated by ligand-induced dimerization domains allows for activation of splicing by small molecules (Figure 5A). The *Sce* VMA intein was split and fused to rapamycin binding domains FKBP12 and FRB, such that the addition of rapamycin induces intein reassociation and PTS [56,57]. This has found *in vivo* applications such as the controllable generation of firefly luciferase in cultured cells and in *Drosophila melanogaster* [58], and PTS of a tobacco etch virus protease in yeast [59]. A mutated form of FKBP12 can be used to induce spontaneous reassociation and PTS of the split intein; in this case, the addition of rapamycin prevents reassocation and inhibits splicing [60]. Recently, Silver and coworkers demonstrated that the FKBP12 and FRB domains could be replaced with complementary coiled coil domains to induce luciferase activity in mammalian cells via specific coiled coil interactions rather than addition of a small molecule, presumably by inducing PTS of the luciferase segments [61].

Engineered inteins also have been created to control *cis*-protein splicing (Figure 5B). The *Mtu* RecA intein was interrupted by the human estrogen receptor ligand binding domain in place of its endonuclease domain and modified by directed evolution to splice only with the addition of 4-hydroxytamoxifen in *S. cerevisiae* [62]. This was extended to mammalian cells, to facilitate splicing of fluorescent reporter proteins and transcription factors that facilitate the hedgehog pathway [63,64], and to create a CPS-activated histone H2A variant in *S. cerevisiae* [65]. The *Mtu* RecA intein was also interrupted by the human thyroid hormone receptor β and shown to control splicing of β-galactosidase and β-lactamase in *E. coli* in response to thyroid hormone [66].

Therefore, it is possible to design inteins that can be turned either on or off with a small molecule. Such small molecule control of splicing may be used to control the active of native, non-engineered inteins. For instance, cisplatin has recently been discovered to inhibit protein splicing both *in vitro* and in *E. coli* and in *M. tuberculosis* [67] and divalent cations can prevent protein splicing *in vitro* [68-73].

CPS activated by temperature, light or pH

Perrimon and coworkers developed a temperature-sensitive version of the *Sce* VMA intein that allows for timed control of protein splicing induced by temperature changes, both in yeast and *D. melanogaster* [74]. The intein introduces control of the activation of the transcription factors Gal4 and Gal80, which in turn allows for temperature-dependent activation or repression of the

Figure 5 Schematic representation of conditional protein splicing (CPS). The brown color of the boxes indicates an inactive intein, whereas a blue intein is active. **A.** Conditional protein splicing triggered by protein *trans*-splicing. **B.** Conditional protein splicing triggered by protein *cis*-splicing. **C.** Conditional protein splicing triggered by change in redox state.

transcription of target genes. This is exciting as it could allow for more general control of protein activity by CPS, given that control is linked to transcriptional activation, which could theoretically control any gene, rather than linked to interrupting a specific target protein by post-translational control. The temperature range was recently broadened by mutagenesis [75], and was used to control the activity of a T7 RNA polymerase in *E. coli,* and hence to control the expression of *lacZ* under control of the T7 promoter [76]. The temperature sensitive mutant was also used to conditionally activate an essential gene in *Dictyostelium discoideum* to identify the function of the gene that is associated with a disorder that predisposes patients to leukemia [77]. Such temperature-dependent CPS activity might have general physiological relevance, as native inteins from extreme thermophiles have been shown to be conditional on elevated temperature for activity [78-85].

PTS also can be regulated by photoactivation, either by intein fusion to a photodimerization domain [86] or by the addition of protecting groups that are photo-

cleavable [87,88], as reviewed in [7]. More recently, Mootz and coworkers have designed a split *Ssp* DnaB intein than can induce C-terminal cleavage on irradiation. They used this system to liberate staphylocoagulase from the I_C segment, which in turn activated native prothrombin, both *in vitro* and in plasma [89]. Protein splicing side reactions may also be enhanced by changes in pH [90].

CPS induced by reduction

CPS can be controlled by the redox state of a disulfide bond that prevents an intein fusion protein from promoting splicing or side reactions (Figure 5C). For instance, the isolation of an unspliced precursor via *in vitro* PTS can be facilitated by reassociation of split intein fragments under oxidizing conditions, with activity induced by the addition of reducing agents [18,91]. Recently, a study in mice showed that a disulfide bond between N- and C-extein residues improved PTS facilitated by the split *Ssp* DnaE intein, as measured by extein activity [92]. Premature *in vivo* cleavage or splicing of

cis-splicing inteins can also be controlled by the introduction of Cys residues in intein or extein positions in order to use inteins in biotechnology applications [93] or to study the mechanism of splicing [94].

Lately, there has been evidence that this disulfide-bond control of splicing activity may have physiological relevance. Belfort, Callahan and coworkers designed a redox trap into the fused, *cis*-splicing version of the *Ssp* DnaE intein, by introducing a Cys (Cys-3) residue in the N-extein. This intein could facilitate N-terminal cleavage only under reducing conditions in *E. coli* and allows for purification of uncleaved precursor and subsequent *in vivo* cleavage after addition of reducing agents [95]. This intein redox trap was subsequently used as a FRET-based biosensor for cellular redox state, showing that inteins can control the response of exteins by their conditional activity [96]. They also discovered the MoaA intein from *Pyrococcus abyssi* has a native disulfide bond also between Cys-3 and Cys1 that can control intein activity [95]. The *Pab* PolII intein has a disulfide bond between Cys1 and Cys + 1 that prevents splicing, and the influence of flanking extein residues on both splicing activity and disulfide bond formation appears to be linked [97]. The activity of the *Mma* PolII intein is dependent on an internal intein disulfide bond, and shows differential splicing activity based on the redox state of the *E. coli* strain or localization to the periplasm or cytoplasm [98].

Intein-based biosensors

CPS permits splicing in response to a specific trigger and lays a foundation for the development of intein-based sensors. Most of these sensors have three functionally and structurally distinct modules: a sensing module, an output module and intein-derived signal transducer (Figure 6A). Signal recognition by a sensor module leads to CPS or conformational changes of the intein connector and activation of the reporter protein. An advantage of intein-based sensors is that the modular design allows for easy exchange of sensor and reporter elements, and, in cases where splicing is involved, the presence of the intein may be trace-less after sensing. Intein-based biosensors have been developed to detect protein-protein interactions, changes in DNA methylation patterns, protein trafficking, small molecules, protease activity and redox state of the cell.

Sensing protein-protein interactions

Intein biosensors for protein-protein interactions utilize PTS facilitated by split intein fragments that have low binding affinity for each other. Design of these biosensors involves creation of two fusion proteins, each containing one protein binding partner, a split intein fragment and a fragment of a reporter protein (Figure 6B). Interaction of the binding partners facilitates split intein reconstitution
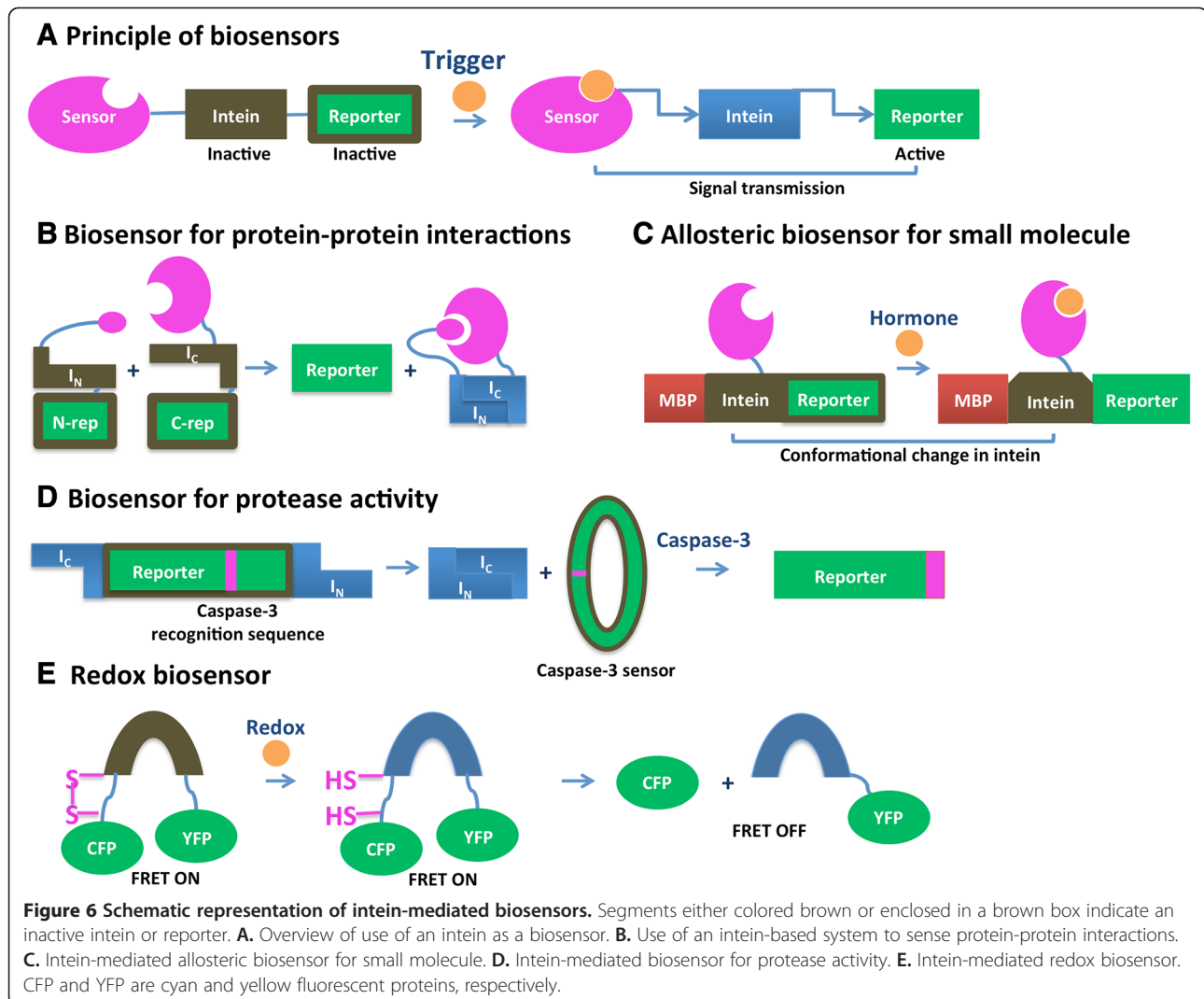
and splicing-induced complementation and activation of a reporter protein. Umezawa and coworkers applied this sensor design to demonstrate protein-protein interactions in various *in vivo* systems ranging from *E. coli* to transgenic animals. In their original work, an *E. coli*-based biosensor was developed to monitor binding between calmodulin and its target peptide M13, using GFP reconstitution as the reporter, mediated by the artificially split *Sce* VMAI intein [99]. Next, an insulin-induced interaction between phosphorylated insulin receptor substrate 1 and its target (the N-terminal SH2 domain of PI 3-kinase) was observed in mammalian cells by luciferase reconstitution by the naturally split *Ssp* DnaE intein [100]. Then, they demonstrated a bioluminescence imaging method to noninvasively and quantitatively image protein-protein interactions in mice by intein-mediated reconstitution of split firefly luciferase proteins driven by the interaction of two strongly interacting proteins, MyoD and Id [101]. To increase the sensitivity of detection, protein splicing was employed to produce a functional transcription factor that modulates a reporter gene, firefly luciferase [102,103]. In this work, the epidermal growth factor (EGF)-induced interactions of an oncogenic product Ras and its target Raf-1 was monitored by bioluminescence signals in mammalian cells. Notably, this interaction was not detected by traditional two-hybrid systems.

Sensing DNA methylation

Similar biosensor design was utilized in living cells for reporting sequence specific changes in DNA methylation via luminescence [104]. The biosensor design comprises two fusions, each with a polydactyl zinc finger domain fused to a split intein fragment and to a split-luciferase domain. The luciferase reporter can be reconstituted by conditional protein splicing upon binding of two polydactyl zinc finger domains to their DNA targets. This biosensor is capable of detecting the loss of epigenetic silencing and increased accessibility of a DNA sequence near the promoter region of the L1PA2 subfamily of Line-1 retro-elements upon treatment with a demethylating drug.

Sensing protein localization and internalization

Intein biosensors can be used to perform high-throughput screens to identify protein localization in a specific compartment. Design of these biosensors involves creation of two fusion proteins, each containing a split *Ssp* DnaE intein fragment and a fragment of a reporter protein (GFP or luciferase). In addition, one of the fusion proteins has a target protein and the other is designed to localize in a specific compartment, such that PTS can only occur in the appropriate cellular compartment. This sensor design was used to study translocation of protein

Figure 6 Schematic representation of intein-mediated biosensors. Segments either colored brown or enclosed in a brown box indicate an inactive intein or reporter. **A.** Overview of use of an intein as a biosensor. **B.** Use of an intein-based system to sense protein-protein interactions. **C.** Intein-mediated allosteric biosensor for small molecule. **D.** Intein-mediated biosensor for protease activity. **E.** Intein-mediated redox biosensor. CFP and YFP are cyan and yellow fluorescent proteins, respectively.

to the nucleus [105] or mitochondria [106], as well as release of proteins from the mitochondria to the cytosol [107].

In addition to the detection of protein subcellular localization, protein splicing has been used for quantitative analysis of G-protein-coupled receptor (GPCR) internalization. In this case, interaction between activated GPCR and intracellular beta-arrestin2 results in *Npu* DnaE intein-mediated reconstitution of luciferase [108].

A protein localization-dependent sensor also was developed for corticosterone detection in animals [109]. Again, the biosensor has two components. The first is a cytosol-localized fusion of glucocorticoid receptor with C-terminal fragments of the *Ssp* DnaE intein and split luciferase. The second is a nucleus-localized fusion of the N-terminal fragments of the intein and luciferase. Upon corticosterone binding, the glucocorticoid receptor is translocated into the nucleus, facilitating intein

fragment complementation and splicing and therefore activation of luciferase.

Sensing small molecules

Allosteric intein biosensors can be used to detect small molecules. Wood and coworkers have designed sensors for ligand detection by human nuclear hormone receptors utilizing allosteric effects induced by ligand-receptor binding. The sensor does not rely on protein splicing; the intein is used as an allosteric transmitter that allows communication between the hormone receptor and the reporter. The sensors exploit a four-domain fusion protein in which a nuclear receptor of interest is inserted in a loop region of the non-splicing *Mtu* RecA intein. The intein is fused to the *E. coli* maltose binding protein and a T4 bacteriophage thymidylate synthase reporter (Figure 6C). The biosensor is based on a thymidylate-synthase deficient *E. coli* cell strain expressing the fusion

protein, such that cell growth depends on the thymidylate synthase reporter activity modulated by presence of the hormone in a dose-dependent manner. Initially, the human estrogen (ERα) and thyroid hormone (TRβ-1) receptors were used to develop highly sensitive methods for detection of nuclear hormone receptors ligands [66]. Later, an optimized estrogen sensor was created that is capable of identifying diverse estrogenic compounds and differentiating between their agonistic/antagonistic pharmacological effects [110]. Subtype-specific nuclear hormone receptor sensors were developed for estrogen [111] and thyroid hormone [112] receptors using the human estrogen (ERα and ERβ) and the thyroid (TRα-1 and TRβ-1) receptors as sensing domains. The peroxisome proliferator-activated receptor gamma (PPARγ) ligand binding domain was used to create a series of bacterial biosensors for the identification of functional PPARγ ligands [113]. This study showed that the linker region between the intein and the thymidylate synthase reporter influences the quality of the transmission of the allosteric signal induced by ligand binding.

A splicing-dependent allosteric intein biosensor was employed by Liu and coworkers for construction of an *E. coli*-based estrogen detector [114]. The sensing element of this system is the estrogen-sensitive *Sce* VMA(ER) intein that was generated by replacement of the endonuclease region with the human estrogen receptor α. The VMA(ER) gene was inserted in the constitutively expressed chromosomal *lacZ* gene. The principal difference of this sensor from that discussed above is that the detection here relies on estrogen-dependent intein splicing and activation of the reporter protein.

Sensing protease activity

A biosensor for protease activity was developed based on *in vivo* intein-promoted protein cyclization [29] (Figure 6D). Firefly luciferase was fused to a caspase-3 recognition sequence and cyclized by the inverted *Ssp* DnaE split intein. In the absence of caspase activity, the activity of the cyclized luciferase was diminished due to steric hinderance. However, the activity of luciferase is fully restored upon caspase-dependent cleavage, enabling real-time quantitative sensing of caspase-3 activity in mice.

Sensing oxidation state

A bacterial redox sensor was developed utilizing the disulfide-bond control of the *Ssp* DnaE intein splicing activity [96] (Figure 6E). The *Ssp* DnaE intein with an engineered disulfide trap is inactive in the oxidized form and triggered by a reducing environment to produce N-terminal cleavage [95], as described above. This redox sensitive intein was fused with a FRET reporter to detect hyperoxic *E. coli* mutants.

Delivery and control of *trans*-genes
Delivery and control of trans-genes in plantae

Inteins have been used to control *trans*-gene expression; the initial examples were in plants and were reviewed by Evans and coworkers in 2005 [115]. Briefly, plant genes were first split and fused with segments of split inteins, with extein activity demonstrated in *E. coli*. The advantage of transferring genes that confer desired traits as split genes is that it minimizes the chance of transfer of the gene to unwanted hosts, such as transfer of herbicide resistance from crops to weeds, since the weed would need to receive both fragments of the gene separately. For example, the split site for acetolactate synthase (ALS) was determined by rational design, and *trans*-splicing to produce herbicide-resistant ALS was demonstrated in *E. coli* [116]. The split site for *Salmonella typhimurium* 5-enolpyruvylshikimate-3-phosphate synthase (EPSPS) was selected by a library-based approach in *E. coli*, and the split intein segments facilitated EPSPS activity via fragment reassociation, to create an active split EPSPS protein, with PTS not required [117,118]. This subtlety is important to note when evaluating claims that PTS is responsible for trans-gene activity *in vivo*. Full-length EPSPS was later generated by PTS, with expression directed to the chloroplast, in *Nicotiana tabacum* [118]. This was reproduced with a more highly herbicide resistant EPSPS from *Pseudomonas fluorescens* both in *E. coli* and in *N. tabacum* [119].

A split intein was also used to generate functional, transgenic β-glucuronidase (GUS) in *Arabidopsis thaliana*, both by PTS and by intein-mediated reassociation of GUS fragments [120]. The split GUS-intein fusion studies were extended to demonstrate that PTS occurs via plasmid-induced expression in leaf cells of soybean, pea, maize and barley [121].

Functional reconstitution of the barnase from *Bacillus amyloliquifaciens* can be facilitated by the split version of the *Ssp* DnaB intein [122]. The split genes were placed under the control of a promoter for the tapetum, such that reconstitution of the barnase results in male sterile plants, given that the ribonuclease activity of the barnase is toxic to the tapetum cells in the anther [123]. The split barnase can be activated by split intein segments to facilitate cytoxicity via transient agroinfiltration of leaves from *N. benthamiana* or to result in male sterility in transgenic *A. thaliana*. Rather than select for the split barnase genes with separate genetic markers, each fragment was genetically linked to a split gene for an acetolactate synthase (ALS)-intein fragment fusion [19]. Therefore, expression of both the resistance gene for sulfonylurea herbicides (ALS) and expression of the toxic barnase require reconstitution of separate split intein pairs [116]. Although PTS was not definitively shown by a biochemical assay, the split genes did

function as predicted to allow for selection of male sterile *A. thaliana* [123]. This work has been extended to demonstrate PTS in transgenic *Triticum aestivum* [124,125].

Intein-mediated and thermoregulated control of transgenic maize has been recently described with a split xylanase, which when active breaks down plant cell walls [126]. Building on their computational analysis of the characteristics of intein insertion sites [127], Raab and coworkers interrupted a thermostable xylanase from *Dictyoglomus thermophilum* with the DnaE-1 intein from *Thermus thermophilus* [126]. Random mutagenesis via error-prone PCR was used to generate mutants of the xylanase-intein fusion protein that are splicing active only at high temperature but retain wild-type xylanase activity. Transgenic maize expressing uninterrupted and active xylanase produces shriveled seeds with low seed mass, but maize expressing the xylanase-intein fusion produces normal seeds, suggesting that toxic xylanase activity is prevented by the intein insertion. The maize that expresses either xylanase produces more glucose during processing. Therefore, the thermoactive intein allows for the xylanase to be expressed but not active during maize growth, when it would be toxic, and then to be active during bio-mass processing, when it is useful.

A split DnaB intein was used to facilitate polymerization of spider silk flagelliform protein in tobacco plant leaves by both stable and transient transfection [32] (Figure 3C).

Trans-genes in other organisms

Recently, PTS in mammalian cells and in mice has been used to test delivery of *trans*-genes by adenovirus delivery vectors. The split fragments of the *Ssp* DnaE intein were fused to heavy and light chain genes for the B-domain deleted factor VIII, and delivered to mammalian cells or mice by separate viral vectors [92,128]. Splicing activity is suggested by increased coagulation activity and the concentrations of functional protein in the plasma, suggesting that PTS could be used for the *in vivo* generation of proteins too large to be delivered by traditional viral vectors. The split *Ssp* DnaE intein also was used to facilitate split Cre reconstitution in mice. In short, fragments of split Cre recombinase can be fused to separate promoters that drive expression under different conditions. Under conditions where both Cre fragments are expressed, Cre is reconstituted and facilitates expression of genes under control of the Cre-LoxP system. The *Ssp* DnaE intein was shown to help improve functional Cre fragment complementation [129].

Conclusions

It is exciting to observe that biotechnology applications of protein splicing have begun to move from proof-of-concept experiments to productive applications in which the intein is the tool rather than the object of study itself. However, the variability in how inteins behave in heterologous contexts can be a limiting factor in the general applicability of intein-based biotechnology. Going forward, recent advances in the understanding of the role of flanking extein residues in the splicing process [130-136] may improve our ability to predict or ameliorate this challenge. The discovery of split inteins that splice even faster than the *Npu* DnaE intein [137] also may increase the efficiency of PTS-based applications. The growing number of examples of *in vivo* protein manipulation using intein catalysis also promises advancement in intein-based tools for systems and functional biology.

Abbreviations

ACP: acyl carrier protein; ALS: acetolactate synthase; CHO: Chinese hamster ovary; CPS: conditional protein splicing; eGFP: enhanced GFP; EPL: expressed protein ligation; EPSPS: 5-enolpyruvylshikimate-3-phosphate synthase; ER: human estrogen receptor; GB1: domain B1 of the immunoglobulin binding protein G; GFP: green fluorescent protein; GPCR: G-protein-coupled receptor; GST: glutathione-*S*-transferase; GUS: β-glucuronidase; HC: heavy chain; *Hay*: *Haemophilus parainfluenzae*; I_C: C-terminal intein fragment; I_N: N-terminal intein fragment; IP: intein-derived protease; LC: light chain; mRFP: the monomeric red fluorescent protein; MHC: major histocompatibility complex; *Mtu*: *Mycobacterium tuberculosis*; *Pch*: *Penicillium chrysogenum*; PH: plextrin homology; POI: protein of interest; PTS: protein *trans*-splicing; *Sce*: *Saccharomyces cerevisiae*; *Ssp*: *Synechocystis* sp. PCC6803; TR-1: human thyroid receptor.

Authors' contributions

NT and KM co-authored the article and read and approved the final manuscript.

Authors' information

NT is a post-doctoral scholar in the lab of Professor Marlene Belfort at the University at Albany; KM is associate professor and chair of the chemistry department at the College of the Holy Cross.

Author details

[1]Department of Biological Sciences, University at Albany, State University of New York, 1400 Washington Avenue, Albany, NY 12222, USA. [2]Department of Chemistry, College of the Holy Cross, 1 College Street, Worcester, MA 01610, USA.

References

1. Paulus H: **Protein splicing and related forms of protein autoprocessing.** *Annu Rev Biochem* 2000, **69**:447–496.
2. Volkmann G, Mootz HD: **Recent progress in intein research: from mechanism to directed evolution and applications.** *Cell Mol Life Sci: CMLS* 2013, **70**(7):1185–1206.
3. Cheriyan M, Perler FB: **Protein splicing: a versatile tool for drug discovery.** *Adv Drug Delivery Rev* 2009, **61**(11):899–907.
4. Elleuche S, Poggeler S: **Inteins, valuable genetic elements in molecular biology and biotechnology.** *Appl Microbiol Biotechnol* 2010, **87**(2):479–489.

5. Shah NH, Muir TW: **Split inteins: nature's protein ligases.** *Israel J Chem* 2011, **51**(8–9):854–861.
6. Vila-Perello M, Muir TW: **Biological applications of protein splicing.** *Cell* 2010, **143**(2):191–200.
7. Mootz HD: **Split inteins as versatile tools for protein semisynthesis.** *Chembiochem* 2009, **10**(16):2579–2589.
8. Flavell RR, Muir TW: **Expressed protein ligation (EPL) in the study of signal transduction, ion conduction, and chromatin biology.** *Acc Chem Res* 2009, **42**(1):107–116.
9. Schwarzer D, Cole PA: **Protein semisynthesis and expressed protein ligation: chasing a protein's tail.** *Curr Opin Chem Biol* 2005, **9**(6):561–569.
10. Volkmann G, Iwai H: **Protein trans-splicing and its use in structural biology: opportunities and limitations.** *Mol BioSyst* 2010, **6**(11):2110–2121.
11. Hackenberger CP, Schwarzer D: **Chemoselective ligation and modification strategies for peptides and proteins.** *Angew Chem* 2008, **47**(52):10030–10074.
12. Muralidharan V, Muir TW: **Protein ligation: an enabling technology for the biophysical analysis of proteins.** *Nat Methods* 2006, **3**(6):429–438.
13. Xu MQ, Evans TC Jr: **Recent advances in protein splicing: manipulating proteins *in vitro* and *in vivo*.** *Curr Opin Biotechnol* 2005, **16**(4):440–446.
14. Evans TC Jr, Benner J, Xu MQ: **Semisynthesis of cytotoxic proteins using a modified protein splicing element.** *Protein science* 1998, **7**(11):2256–2264.
15. Muir TW, Sondhi D, Cole PA: **Expressed protein ligation: a general method for protein engineering.** *Proc Natl Acad Sci USA* 1998, **95**(12):6705–6710.
16. Dawson PE, Muir TW, Clark-Lewis I, Kent SB: **Synthesis of proteins by native chemical ligation.** *Science* 1994, **266**(5186):776–779.
17. Camarero JA, Shekhtman A, Campbell EA, Chlenov M, Gruber TM, Bryant DA, Darst SA, Cowburn D, Muir TW: **Autoregulation of a bacterial sigma factor explored by using segmental isotopic labeling and NMR.** *Proc Natl Acad Sci USA* 2002, **99**(13):8536–8541.
18. Mills KV, Lew BM, Jiang S, Paulus H: **Protein splicing in trans by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein.** *Proc Natl Acad Sci USA* 1998, **95**(7):3543–3548.
19. Wu H, Hu Z, Liu XQ: **Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803.** *Proc Natl Acad Sci USA* 1998, **95**(16):9226–9231.
20. Aboye TL, Camarero JA: **Biological synthesis of circular polypeptides.** *J Biol Chem* 2012, **287**(32):27026–27032.
21. Berrade L, Camarero JA: **Expressed protein ligation: a resourceful tool to study protein structure and function.** *Cell Mol Life Sci* 2009, **66**(24):3909–3922.
22. Evans TC Jr, Benner J, Xu MQ: **The cyclization and polymerization of bacterially expressed proteins using modified self-splicing inteins.** *J Biol Chem* 1999, **274**(26):18359–18363.
23. Camarero JA, Muir TW: **Biosynthesis of a head-to-tail cyclized protein with improved biological activity.** *J Am Chem Soc* 1999, **121**(23):5597–5598.
24. Scott CP, Abel-Santos E, Wall M, Wahnon DC, Benkovic SJ: **Production of cyclic peptides and proteins *in vivo*.** *Proc Natl Acad Sci USA* 1999, **96**(24):13638–13643.
25. Evans TC Jr, Martin D, Kolly R, Panne D, Sun L, Ghosh I, Chen L, Benner J, Liu XQ, Xu MQ: **Protein trans-splicing and cyclization by a naturally split intein from the dnaE gene of *Synechocystis* species PCC6803.** *J Biol Chem* 2000, **275**(13):9091–9094.
26. Jeffries CM, Graham SC, Stokes PH, Collyer CA, Guss JM, Matthews JM: **Stabilization of a binary protein complex by intein-mediated cyclization.** *Protein Sci* 2006, **15**(11):2612–2618.
27. Iwai H, Lingel A, Pluckthun A: **Cyclic green fluorescent protein produced *in vivo* using an artificially split PI-Pful intein from *Pyrococcus furiosus*.** *J Biol Chem* 2001, **276**(19):16548–16554.
28. Iwai H, Pluckthun A: **Circular beta-lactamase: stability enhancement by cyclizing the backbone.** *FEBS letters* 1999, **459**(2):166–172.
29. Kanno A, Yamanaka Y, Hirano H, Umezawa Y, Ozawa T: **Cyclic luciferase for real-time sensing of caspase-3 activities in living mammals.** *Angew Chem* 2007, **46**(40):7595–7599.
30. Abel-Santos E, Scott CP, Benkovic SJ: **Use of inteins for the *in vivo* production of stable cyclic peptide libraries in *E. coli*.** *Methods Mol Biol* 2003, **205**:281–294.
31. Williams NK, Liepinsh E, Watt SJ, Prosselkov P, Matthews JM, Attard P, Beck JL, Dixon NE, Otting G: **Stabilization of native protein fold by intein-mediated covalent cyclization.** *J Mol Biol* 2005, **346**(4):1095–1108.

32. Hauptmann V, Weichert N, Menzel M, Knoch D, Paege N, Scheller J, Spohn U, Conrad U, Gils M: **Native-sized spider silk proteins synthesized in planta via intein-based multimerization.** *Transgenic Res* 2013, **22**(2):369–377.
33. Kiraly O, Guan L, Szepessy E, Toth M, Kukor Z, Sahin-Toth M: **Expression of human cationic trypsinogen with an authentic N terminus using intein-mediated splicing in aminopeptidase P deficient *Escherichia coli*.** *Protein Expres Purif* 2006, **48**(1):104–111.
34. Kiraly O, Guan L, Sahin-Toth M: **Expression of recombinant proteins with uniform N-termini.** *Methods Mol Biol* 2011, **705**:175–194.
35. Farfan-Arribas DJ, Stern LJ, Rock KL: **Using intein catalysis to probe the origin of major histocompatibility complex class I-presented peptides.** *Proc Natl Acad Sci USA* 2012, **109**(42):16998–17003.
36. Miao J, Wu W, Spielmann T, Belfort M, Derbyshire V, Belfort G: **Single-step affinity purification of toxic and non-toxic proteins on a fluidics platform.** *Lab Chip* 2005, **5**(3):248–253.
37. Wu W, Wood DW, Belfort G, Derbyshire V, Belfort M: **Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing.** *Nucleic Acids Res* 2002, **30**(22):4864–4871.
38. Kunes YZ, Gion WR, Fung E, Salfeld JG, Zhu RR, Sakorafas P, Carson GR: **Expression of antibodies using single-open reading frame vector design and polyprotein processing from mammalian cells.** *Biotechnol Progr* 2009, **25**(3):735–744.
39. Volkmann G, Volkmann V, Liu XQ: **Site-specific protein cleavage *in vivo* by an intein-derived protease.** *FEBS Lett* 2012, **586**(1):79–84.
40. Ramsden R, Arms L, Davis TN, Muller EG: **An intein with genetically selectable markers provides a new approach to internally label proteins with GFP.** *BMC Biotechnol* 2011, **11**:71.
41. Volkmann G, Liu XQ: **Protein C-terminal labeling and biotinylation using synthetic peptide and split-intein.** *PLoS One* 2009, **4**(12):e8381.
42. Ando T, Tsukiji S, Tanaka T, Nagamune T: **Construction of a small-molecule-integrated semisynthetic split intein for *in vivo* protein ligation.** *Chem Comm* 2007, **47**:4995–4997.
43. Dhar T, Mootz HD: **Modification of transmembrane and GPI-anchored proteins on living cells by efficient protein trans-splicing using the Npu DnaE intein.** *Chem Comm* 2011, **47**(11):3063–3065.
44. Zuger S, Iwai H: **Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies.** *Nat Biotechnol* 2005, **23**(6):736–740.
45. Appleby JH, Zhou K, Volkmann G, Liu XQ: **Novel split intein for trans-splicing synthetic peptide onto C terminus of protein.** *J Biol Chem* 2009, **284**(10):6194–6199.
46. Aranko AS, Zuger S, Buchinger E, Iwai H: **In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins.** *PLoS One* 2009, **4**(4):e5185.
47. Oeemig JS, Aranko AS, Djupsjobacka J, Heinamaki K, Iwai H: **Solution structure of DnaE intein from *Nostoc punctiforme*: structural basis for the design of a new split intein suitable for site-specific chemical modification.** *FEBS Lett* 2009, **583**(9):1451–1456.
48. Muona M, Aranko AS, Iwai H: **Segmental isotopic labeling of a multidomain protein by protein ligation by protein trans-splicing.** *Chembiochem* 2008, **9**(18):2958–2961.
49. Sun W, Yang J, Liu XQ: **Synthetic two-piece and three-piece split inteins for protein trans-splicing.** *J Biol Chem* 2004, **279**(34):35281–35286.
50. Jung D, Min K, Jung J, Jang W, Kwon Y: **Chemical biology-based approaches on fluorescent labeling of proteins in live cells.** *Mol BioSyst* 2013, **9**(5):862–872.
51. Chattopadhaya S, Srinivasan R, Yeo DS, Chen GY, Yao SQ: **Site-specific covalent labeling of proteins inside live cells using small molecule probes.** *Bioorg Med Chem* 2009, **17**(3):981–989.
52. Borra R, Dong D, Elnagar AY, Woldemariam GA, Camarero JA: **In-cell fluorescence activation and labeling of proteins mediated by FRET-quenched split inteins.** *J Am Chem Soc* 2012, **134**(14):6344–6353.
53. Charalambous A, Andreou M, Skourides PA: **Intein-mediated site-specific conjugation of quantum dots to proteins *in vivo*.** *J Nanobiotechnology* 2009, **7**:9.
54. Charalambous A, Antoniades I, Christodoulou N, Skourides PA: **Split-inteins for simultaneous, site-specific conjugation of quantum dots to multiple protein targets *in vivo*.** *J Nanobiotechnology* 2011, **9**:37.
55. Pietrokovski S: **Intein spread and extinction in evolution.** *Trends Genet* 2001, **17**(8):465–472.

56. Mootz HD, Blum ES, Tyszkiewicz AB, Muir TW: **Conditional protein splicing: a new tool to control protein structure and function** *in vitro* **and** *in vivo*. *J Am Chem Soc* 2003, **125**(35):10561–10569.

57. Mootz HD, Muir TW: **Protein splicing triggered by a small molecule**. *J Am Chem Soc* 2002, **124**(31):9044–9045.

58. Schwartz EC, Saez L, Young MW, Muir TW: **Post-translational enzyme activation in an animal via optimized conditional protein splicing**. *Nat Chem Biol* 2007, **3**(1):50–54.

59. Sonntag T, Mootz HD: **An intein-cassette integration approach used for the generation of a split TEV protease activated by conditional protein splicing**. *Mol BioSyst* 2011, **7**(6):2031–2039.

60. Brenzel S, Mootz HD: **Design of an intein that can be inhibited with a small molecule ligand**. *J Am Chem Soc* 2005, **127**(12):4176–4177.

61. Selgrade DF, Lohmueller JJ, Lienert F, Silver PA: **Protein scaffold-activated protein trans-splicing in mammalian cells**. *J Am Chem Soc* 2013, **135**(20):7713–7719.

62. Buskirk AR, Ong YC, Gartner ZJ, Liu DR: **Directed evolution of ligand dependence: small-molecule-activated protein splicing**. *Proc Natl Acad Sci USA* 2004, **101**(29):10505–10510.

63. Peck SH, Chen I, Liu DR: **Directed evolution of a small-molecule-triggered intein with improved splicing properties in mammalian cells**. *Chem Biol* 2011, **18**(5):619–630.

64. Yuen CM, Rodda SJ, Vokes SA, McMahon AP, Liu DR: **Control of transcription factor activity and osteoblast differentiation in mammalian cells using an evolved small-molecule-dependent intein**. *J Am Chem Soc* 2006, **128**(27):8939–8946.

65. Hartley PD, Madhani HD: **Mechanisms that specify promoter nucleosome location and identity**. *Cell* 2009, **137**(3):445–458.

66. Skretas G, Wood DW: **Regulation of protein activity with small-molecule-controlled inteins**. *Protein Sci* 2005, **14**(2):523–532.

67. Zhang L, Zheng Y, Callahan B, Belfort M, Liu Y: **Cisplatin inhibits protein splicing, suggesting inteins as therapeutic targets in** *Mycobacteria*. *J Biol Chem* 2011, **286**(2):1277–1282.

68. Ghosh I, Sun L, Xu MQ: **Zinc inhibition of protein trans-splicing and identification of regions essential for splicing and association of a split intein***. *J Biol Chem* 2001, **276**(26):24051–24058.

69. Mills KV, Paulus H: **Reversible inhibition of protein splicing by zinc ion**. *J Biol Chem* 2001, **276**(14):10832–10838.

70. Nichols NM, Benner JS, Martin DD, Evans TC Jr: **Zinc ion effects on individual Ssp DnaE intein splicing steps: regulating pathway progression**. *Biochemistry* 2003, **42**(18):5301–5311.

71. Sun P, Ye S, Ferrandon S, Evans TC, Xu MQ, Rao Z: **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* 2005, **353**(5):1093–1105.

72. Zhang L, Xiao N, Pan Y, Zheng Y, Pan Z, Luo Z, Xu X, Liu Y: **Binding and inhibition of copper ions to RecA inteins from** *Mycobacterium tuberculosis*. *Chemistry* 2010, **16**(14):4297–4306.

73. Zhang L, Zheng Y, Xi Z, Luo Z, Xu X, Wang C, Liu Y: **Metal ions binding to recA inteins from** *Mycobacterium tuberculosis*. *Mol BioSyst* 2009, **5**(6):644–650.

74. Zeidler MP, Tan C, Bellaiche Y, Cherry S, Hader S, Gayko U, Perrimon N: **Temperature-sensitive control of protein activity by conditionally splicing inteins**. *Nat Biotechnol* 2004, **22**(7):871–876.

75. Tan G, Chen M, Foote C, Tan C: **Temperature-sensitive mutations made easy: generating conditional mutations by using temperature-sensitive inteins that function within different temperature ranges**. *Genetics* 2009, **183**(1):13–22.

76. Liang R, Liu X, Liu J, Ren Q, Liang P, Lin Z, Xie X: **A T7-expression system under temperature control could create temperature-sensitive phenotype of target gene in** *Escherichia coli*. *J Microbiol Methods* 2007, **68**(3):497–506.

77. Wong CC, Traynor D, Basse N, Kay RR, Warren AJ: **Defective ribosome assembly in Shwachman-Diamond syndrome**. *Blood* 2011, **118**(16):4305–4312.

78. Cambon-Bonavita MA, Schmitt P, Zieger M, Flaman JM, Lesongeur F, Raguenes G, Bindel D, Frisch N, Lakkis Z, Dupret D, *et al*: **Cloning, expression, and characterization of DNA polymerase I from the hyperthermophilic archaea** *Thermococcus fumicolans*. *Extremophiles* 2000, **4**(4):215–225.

79. Mills KV, Dorval DM, Lewandowski KT: **Kinetic analysis of the individual steps of protein splicing for the** *Pyrococcus abyssi* **PolII intein**. *J Biol Chem* 2005, **280**(4):2714–2720.

80. Mills KV, Manning JS, Garcia AM, Wuerdeman LA: **Protein splicing of a** *Pyrococcus abyssi* **intein with a C-terminal glutamine**. *J Biol Chem* 2004, **279**(20):20685–20691.

81. Shao Y, Paulus H: **Protein splicing: estimation of the rate of O-N and S-N acyl rearrangements, the last step of the splicing process**. *J Pept Res* 1997, **50**(3):193–198.

82. Shao Y, Xu MQ, Paulus H: **Protein splicing: characterization of the aminosuccinimide residue at the carboxyl terminus of the excised intervening sequence**. *Biochemistry* 1995, **34**(34):10844–10850.

83. Xu MQ, Comb DG, Paulus H, Noren CJ, Shao Y, Perler FB: **Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation**. *EMBO J* 1994, **13**(23):5517–5522.

84. Xu MQ, Perler FB: **The mechanism of protein splicing and its modulation by mutation**. *EMBO J* 1996, **15**(19):5146–5153.

85. Xu MQ, Southworth MW, Mersha FB, Hornstra LJ, Perler FB: **In vitro protein splicing of purified precursor and the identification of a branched intermediate**. *Cell* 1993, **75**(7):1371–1377.

86. Tyszkiewicz AB, Muir TW: **Activation of protein splicing with light in yeast**. *Nat Methods* 2008, **5**(4):303–305.

87. Berrade L, Kwon Y, Camarero JA: **Photomodulation of protein trans-splicing through backbone photocaging of the DnaE split intein**. *Chembiochem* 2010, **11**(10):1368–1372.

88. Vila-Perello M, Pratt MR, Tulin F, Muir TW: **Covalent capture of phospho-dependent protein oligomerization by site-specific incorporation of a diazirine photo-cross-linker**. *J Am Chem Soc* 2007, **129**(26):8068–8069.

89. Binschik J, Zettler J, Mootz HD: **Photocontrol of protein activity mediated by the cleavage reaction of a split intein**. *Angew Chem* 2011, **50**(14):3249–3252.

90. Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M: **A genetic system yields self-cleaving inteins for bioseparations**. *Nat Biotechnol* 1999, **17**(9):889–892.

91. Shi J, Muir TW: **Development of a tandem protein trans-splicing system based on native and engineered split inteins**. *J Am Chem Soc* 2005, **127**(17):6198–6206.

92. Zhu F, Liu Z, Wang X, Miao J, Qu H, Chi X: **Inter-chain disulfide bond improved protein trans-splicing increases plasma coagulation activity in C57BL/6 mice following portal vein FVIII gene delivery by dual vectors**. *Sci China Life Sci* 2013, **56**(3):262–267.

93. Cui C, Zhao W, Chen J, Wang J, Li Q: **Elimination of** *in vivo* **cleavage between target protein and intein in the intein-mediated protein purification systems**. *Protein Expres Purif* 2006, **50**(1):74–81.

94. Saleh L, Southworth MW, Considine N, O'Neill C, Benner J, Bollinger JM Jr, Perler FB: **Branched intermediate formation is the slowest step in the protein splicing reaction of the Ala1 KlbA intein from** *Methanococcus jannaschii*. *Biochemistry* 2011, **50**(49):10576–10589.

95. Callahan BP, Topilina NI, Stanger MJ, Van Roey P, Belfort M: **Structure of catalytically competent intein caught in a redox trap with functional and evolutionary implications**. *Nat Struct Mol Biol* 2011, **18**(5):630–633.

96. Callahan BP, Stanger M, Belfort M: **A redox trap to augment the intein toolbox**. *Biotechnol Bioeng* 2013, **110**(6):1565–1573.

97. Chen W, Li L, Du Z, Liu J, Reitter JN, Mills KV, Linhardt RJ, Wang C: **Intramolecular disulfide bond between catalytic cysteines in an intein precursor**. *J Am Chem Soc* 2012, **134**(5):2500–2503.

98. Nicastri MC, Xega K, Li L, Wang C, Xie J, Linhardt RJ, Reitter JN, Mills KV: **An internal disulfide bond acts as a switch for intein activity**. *Biochemistry* 2013, **52**:5920–5927.

99. Ozawa T, Nogami S, Sato M, Ohya Y, Umezawa Y: **A fluorescent indicator for detecting protein-protein interactions** *in vivo* **based on protein splicing**. *Anal Chem* 2000, **72**(21):5151–5157.

100. Ozawa T, Kaihara A, Sato M, Tachihara K, Umezawa Y: **Split luciferase as an optical probe for detecting protein-protein interactions in mammalian cells based on protein splicing**. *Anal Chem* 2001, **73**(11):2516–2521.

101. Paulmurugan R, Umezawa Y, Gambhir SS: **Noninvasive imaging of protein-protein interactions in living subjects by using reporter protein complementation and reconstitution strategies**. *Proc Natl Acad Sci USA* 2002, **99**(24):15608–15613.

102. Kanno A, Ozawa T, Umezawa Y: **Intein-mediated reporter gene assay for detecting protein-protein interactions in living mammalian cells.** *Anal Chem* 2006, **78**(2):556–560.

103. Kanno A, Ozawa T, Umezawa Y: **Bioluminescent imaging of MAPK function with intein-mediated reporter gene assay.** *Methods Mol Biol* 2009, **574**:185–192.

104. Huang X, Narayanaswamy R, Fenn K, Szpakowski S, Sasaki C, Costa J, Blancafort P, Lizardi PM: **Sequence-specific biosensors report drug-induced changes in epigenetic silencing in living cells.** *DNA Cell Biol* 2012, **31**(Suppl 1):S2–S10.

105. Kim SB, Ozawa T, Watanabe S, Umezawa Y: **High-throughput sensing and noninvasive imaging of protein nuclear transport by using reconstitution of split *Renilla* luciferase.** *Proc Natl Acad Sci USA* 2004, **101**(32):11542–11547.

106. Ozawa T, Sako Y, Sato M, Kitamura T, Umezawa Y: **A genetic approach to identifying mitochondrial proteins.** *Nat Biotechnol* 2003, **21**(3):287–293.

107. Kanno A, Ozawa T, Umezawa Y: **Genetically encoded optical probe for detecting release of proteins from mitochondria toward cytosol in living cells and mammals.** *Anal Chem* 2006, **78**(23):8076–8081.

108. Zhang Y, Yang W, Chen L, Shi Y, Li G, Zhou N: **Development of a novel DnaE intein-based assay for quantitative analysis of G-protein-coupled receptor internalization.** *Anal Biochem* 2011, **417**(1):65–72.

109. Kim SB, Ozawa T, Umezawa Y: **Genetically encoded stress indicator for noninvasively imaging endogenous corticosterone in living mice.** *Anal Chem* 2005, **77**(20):6588–6593.

110. Skretas G, Meligova AK, Villalonga-Barber C, Mitsiou DJ, Alexis MN, Micha-Screttas M, Steele BR, Screttas CG, Wood DW: **Engineered chimeric enzymes as tools for drug discovery: generating reliable bacterial screens for the detection, discovery, and assessment of estrogen receptor modulators.** *J Am Chem Soc* 2007, **129**(27):8443–8457.

111. Skretas G, Wood DW: **A bacterial biosensor of endocrine modulators.** *J Mol Biol* 2005, **349**(3):464–474.

112. Gierach I, Li J, Wu WY, Grover GJ, Wood DW: **Bacterial biosensors for screening isoform-selective ligands for human thyroid receptors alpha-1 and beta-1.** *FEBS Open Bio* 2012, **2**:247–253.

113. Li J, Gierach I, Gillies AR, Warden CD, Wood DW: **Engineering and optimization of an allosteric biosensor protein for peroxisome proliferator-activated receptor gamma ligands.** *Biosens Bioelectron* 2011, **29**(1):132–139.

114. Liang R, Zhou J, Liu J: **Construction of a bacterial assay for estrogen detection based on an estrogen-sensitive intein.** *Appl Environ Microbiol* 2011, **77**(7):2488–2495.

115. Evans TC Jr, Xu MQ, Pradhan S: **Protein splicing elements and plants: from transgene containment to protein purification.** *Annu Rev Plant Biol* 2005, **56**:375–392.

116. Sun L, Ghosh I, Paulus H, Xu MQ: **Protein trans-splicing to produce herbicide-resistant acetolactate synthase.** *Appl Environ Microbiol* 2001, **67**(3):1025–1029.

117. Chen L, Pradhan S, Evans TC Jr: **Herbicide resistance from a divided EPSPS protein: the split *Synechocystis* DnaE intein as an *in vivo* affinity domain.** *Gene* 2001, **263**(1–2):39–48.

118. Chin HG, Kim GD, Marin I, Mersha F, Evans TC Jr, Chen L, Xu MQ, Pradhan S: **Protein trans-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes.** *Proc Natl Acad Sci USA* 2003, **100**(8):4510–4515.

119. Dun BQ, Wang XJ, Lu W, Zhao ZL, Hou SN, Zhang BM, Li GY, Evans TC Jr, Xu MQ, Lin M: **Reconstitution of glyphosate resistance from a split 5-enolpyruvyl shikimate-3-phosphate synthase gene in *Escherichia coli* and transgenic tobacco.** *Appl Environ Microbiol* 2007, **73**(24):7997–8000.

120. Yang J, Fox GC Jr, Henry-Smith TV: **Intein-mediated assembly of a functional beta-glucuronidase in transgenic plants.** *Proc Natl Acad Sci USA* 2003, **100**(6):3513–3518.

121. Yang J, Henry-Smith TV, Qi M: **Functional analysis of the split *Synechocystis* DnaE intein in plant tissues by biolistic particle bombardment.** *Transgenic Res* 2006, **15**(5):583–593.

122. Wu H, Xu MQ, Liu XQ: **Protein trans-splicing and functional mini-inteins of a cyanobacterial dnaB intein.** *Biochim Biophys Acta* 1998, **1387**(1–2):422–432.

123. Sakakibara N, Han M, Rollor CR, Gilson RC, Busch C, Heo G, Kelman Z: **Cloning, purification, and partial characterization of the *Halobacterium* sp. NRC-1 minichromosome maintenance (MCM) helicase.** *Open Microbiol J* 2008, **2**:13–17.

124. Gils M, Rubtsova M, Kempe K: **Split-transgene expression in wheat.** *Methods Mol Biol* 2012, **847**:123–135.

125. Kempe K, Rubtsova M, Gils M: **Intein-mediated protein assembly in transgenic wheat: production of active barnase and acetolactate synthase from split genes.** *Plant Biotechnol J* 2009, **7**(3):283–297.

126. Shen B, Sun X, Zuo X, Shilling T, Apgar J, Ross M, Bougri O, Samoylov V, Parker M, Hancock E, et al: **Engineering a thermoregulated intein-modified xylanase into maize for consolidated lignocellulosic biomass processing.** *Nat Biotechnol* 2012, **30**(11):1131–1136.

127. Apgar J, Ross M, Zuo X, Dohle S, Sturtevant D, Shen B, de la Vega H, Lessard P, Lazar G, Raab RM: **A predictive model of intein insertion site for use in the engineering of molecular switches.** *PLoS One* 2012, **7**(5):e37355.

128. Zhu F, Liu Z, Chi X, Qu H: **Protein trans-splicing based dual-vector delivery of the coagulation factor VIII gene.** *Sci China Life Sci* 2010, **53**(6):683–689.

129. Wang P, Chen T, Sakurai K, Han BX, He Z, Feng G, Wang F: **Intersectional Cre driver lines generated using split-intein mediated split-Cre reconstitution.** *Sci Rep* 2012, **2**:497.

130. Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M: **Modulation of intein activity by its neighboring extein substrates.** *Proc Natl Acad Sci USA* 2009, **106**(27):11005–11010.

131. Cheriyan M, Pedamallu CS, Tori K, Perler F: **Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues.** *J Biol Chem* 2013, **288**(9):6202–6211.

132. Oeemig JS, Zhou D, Kajander T, Wlodawer A, Iwai H: **NMR and crystal structures of the *Pyrococcus horikoshii* RadA intein guide a strategy for engineering a highly efficient and promiscuous intein.** *J Mol Biol* 2012, **421**(1):85–99.

133. Pearl EJ, Bokor AA, Butler MI, Poulter RT, Wilbanks SM: **Preceding hydrophobic and beta-branched amino acids attenuate splicing by the CnePRP8 intein.** *Biochimica et Biophysica Acta* 2007, **1774**(8):995–1001.

134. Romanelli A, Shekhtman A, Cowburn D, Muir TW: **Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction.** *Proc Natl Acad Sci USA* 2004, **101**(17):6397–6402.

135. Shah NH, Eryilmaz E, Cowburn D, Muir TW: **Extein residues play an intimate role in the rate-limiting step of protein trans-splicing.** *J Am Chem Soc* 2013, **135**(15):5839–5847.

136. Nogami S, Satow Y, Ohya Y, Anraku Y: **Probing novel elements for protein splicing in the yeast Vma1 protozyme: a study of replacement mutagenesis and intragenic suppression.** *Genetics* 1997, **147**(1):73–85.

137. Carvajal-Vallejos P, Pallisse R, Mootz HD, Schmidt SR: **Unprecedented rates and efficiencies revealed for new natural split inteins from metagenomic sources.** *J Biol Chem* 2012, **287**(34):28686–28696.