



Protein Splicing of Inteins: A Powerful Tool in Synthetic Biology

Hao Wang¹, Lin Wang¹, Baihua Zhong² and Zhuojun Dai^{1*}

¹Materials Synthetic Biology Center, CAS Key Laboratory of Quantitative Engineering Biology, Guangdong Provincial Key Laboratory of Synthetic Genomics, Shenzhen Institute of Synthetic Biology, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China, ²Materials Interfaces Center, Institute of Advanced Materials Science and Engineering, Shenzhen Institutes of Advanced Technology, Chinese Academy of Sciences, Shenzhen, China

Inteins are protein segments that are capable of enabling the ligation of flanking extein into a new protein, a process known as protein splicing. Since its discovery, inteins have become powerful biotechnological tools for applications such as protein engineering. In the last 10 years, the development in synthetic biology has further endowed inteins with enhanced functions and diverse utilizations. Here we review these efforts and discuss the future directions.

Keywords: inteins, synthetic biology, living therapeutics, protein engineering, split inteins

INTRODUCTION

Inteins are protein segments that are capable of ligating the flanking exteins (external proteins) into a new protein, a process known as protein splicing (Belfort et al., 2006). They are found in many natural organisms, such as bacteria, fungi and lower plants, and are usually embedded within essential proteins (Belfort et al., 2006). For example, Hirata et al. discovered the *Sce* VMA intein (vacuolar membrane ATPase subunit of *Saccharomyces cerevisiae*) in his study of ATPase from *Saccharomyces cerevisiae* by sequence alignment analysis in 1988 (Hirata et al., 1990). Naturally occurred inteins exist in several forms including full-length inteins, mini-inteins and naturally split inteins. The full-length inteins and mini-inteins are both *cis*-splicing inteins, with or without an endonuclease domain. Split inteins are *trans*-splicing inteins, with two fragments transcribed and translated by two independent genes. The *trans*-splicing requires the co-expression of both split intein fragments, namely N-intein (I_N, fused with C-termini of an N-extein) and C-intein (I_C, fused with N-termini of a C-extein). The split intein fragments subsequently associate to recover its activity and catalyze the ligation of N-extein and C-extein (Figure 1) (Stevens et al., 2016). In both *cis* or *trans*-splicing, the intein-mediated activities do not require assistance from any enzyme or co-factor, but only a proper folded structure of the expressed protein (Perler, 1998; Kwong et al., 2016). In the previous literature, the mechanisms and applications of inteins have been extensively reviewed (Wood et al., 1999; Mills et al., 2014). Nevertheless, with the rapid developments in synthetic biology research and technology, we have noticed that new and powerful tools are emerging to discover or evolve inteins with higher splicing efficacy. In the meantime, diverse and complex functions are achieved through engineered inteins mediated protein splicing. Therefore, in this paper, we first introduce the composition and function of naturally occurring inteins, and further review the recent developments and applications of inteins in synthetic biology.

NATURALLY OCCURRED INTEIN: COMPOSITION AND SPLICING MECHANISMS

In nature, protein splicing synthesizes two separate proteins (the inteins and exteins) under the control of a single gene by the precise excision of an internal protein segment and concomitantly

OPEN ACCESS

Edited by:

Jiaofang Huang,
East China University of Science and
Technology, China

Reviewed by:

Baojun Wang,
University of Edinburgh,
United Kingdom
Sesilja Aranko,
Aalto University, Finland

*Correspondence:

Zhuojun Dai
zj.dai@siaat.ac.cn

Specialty section:

This article was submitted to
Synthetic Biology,
a section of the journal
Frontiers in Bioengineering and
Biotechnology

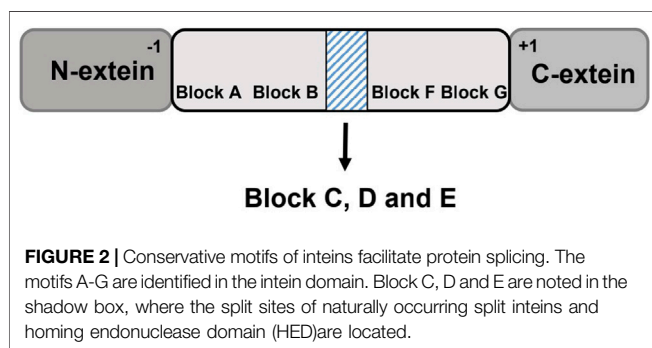
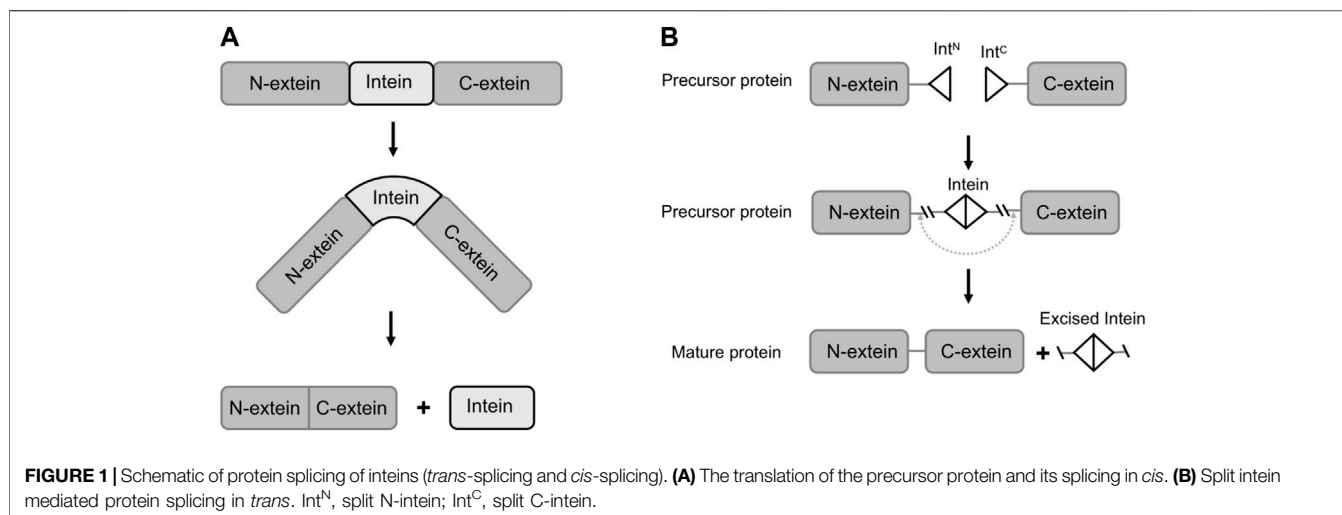
Received: 06 November 2021

Accepted: 25 January 2022

Published: 21 February 2022

Citation:

Wang H, Wang L, Zhong B and Dai Z
(2022) Protein Splicing of Inteins: A
Powerful Tool in Synthetic Biology.
Front. Bioeng. Biotechnol. 10:810180.
doi: 10.3389/fbioe.2022.810180



ligation of the flanking regions (Topilina and Mills, 2014). Oftentimes these two separate proteins are both functional: the excised inteins contain homing endonucleases (HED) that can catalyze the lateral transfer of their DNA coding sequences by an intein homing mechanism, while the ligated exteins are mostly enzymes with specific functions (Chong and Xu, 1997; Perler, 2005; Cheriyan and Perler, 2009; Kwong et al., 2016). Previous research has categorized the inteins into three classes: class 1, class 2 and class 3 based on the mechanism of splicing from the extein. Here we mainly elaborate on the composition and splicing process of the class 1 inteins.

Several amino acid residues and peptide sequences at certain positions of class 1 inteins are highly conservative (conservative motifs), and these conservative motifs are tightly related with the splicing reaction and splicing efficacy (Figure 2) (Chong and Xu, 1997). Shaorong *et al.* identified seven motifs (motif A-G) composed of a series of conserved amino acid residues. For example, motif A is pointed to as the first amino acid residue on the N-terminus of the intein for protein splicing, consisting of a hydroxyl- or thiol-containing residue (Ser, Thr, or Cys), while motif G is often found to be an Asn residue at the other junction site on the C-terminus with a His residue at the penultimate site beside Asn (Mills et al., 2014). Motif C and E are dodecapeptide regions acting as the homing endonucleases (Chong and Xu, 1997). These dodecapeptide motifs could recognize DNA and

catalyze DNA cleavage. By doing so, the *Sce* VMA (Vacuolar Membrane ATPase in *Saccharomyces cerevisiae*) intein initiates a gene recombination process to disperse the intein genes to other strains (Perler, 1998).

The protein splicing of class 1 intein is achieved through structural conformational change and chemical bonds shifting on junction sites between intein and exteins, which can be summarized in four main steps (Figure 3) (Gorbalenya, 1998; Reitter and Mills, 2011). First, a nucleophilic attack by the N-terminal Cys or Ser of the intein converts the peptide bonding between the N-extein and intein to an ester or thioester group. Then, the transesterification transfers the N-extein from the side chain of the N-terminus of intein to the first residue of C-extein at the C-terminus of the intein, forming a branched intermediate. The Asn cyclization of the last amino acid residue on the C-terminus of intein frees the branched ester with a peptide bond cleavage, resulting in ligated exteins with an ester bond linkage. Finally, rapid conversion from the ester bond to the amide bond occurs to form the final ligated peptide (Southworth et al., 2000).

Some of the naturally occurring inteins (mini-intein) lack the endonuclease region and only conduct the protein splicing function (Chong and Xu, 1997; Telenti et al., 1997; Perler, 1998; Southworth et al., 2000). Telenti *et al.* reported the discovery of GyrA inteins retrieved from 7 mycobacterial species and subspecies. Especially, the *Mycobacterium xenopi* GyrA intein (*Mxe* GyrA) consisted of only 198 amino acid residues (AAs), as compared with the inteins with 420 AAs as analyzed from the rest six species. Sequence analysis confirmed the missing of the endonuclease region in *Mxe* GyrA (Telenti et al., 1997). Split inteins were also found in nature that can facilitate protein splicing in *trans* (Figure 1) (Pinto et al., 2020). For instance, Evans *et al.* identified a naturally occurring split intein in the *dnaE* gene which encoded the catalytic subunit of DNA polymerase III in *Synechocystis sp.* PCC6803 (Evans et al., 2000). This DnaE split intein pairs contained the N-terminal half (123 AAs) and C-terminal half (36 AAs), encoded by two separate open reading frames in the genome of the original species (Evans

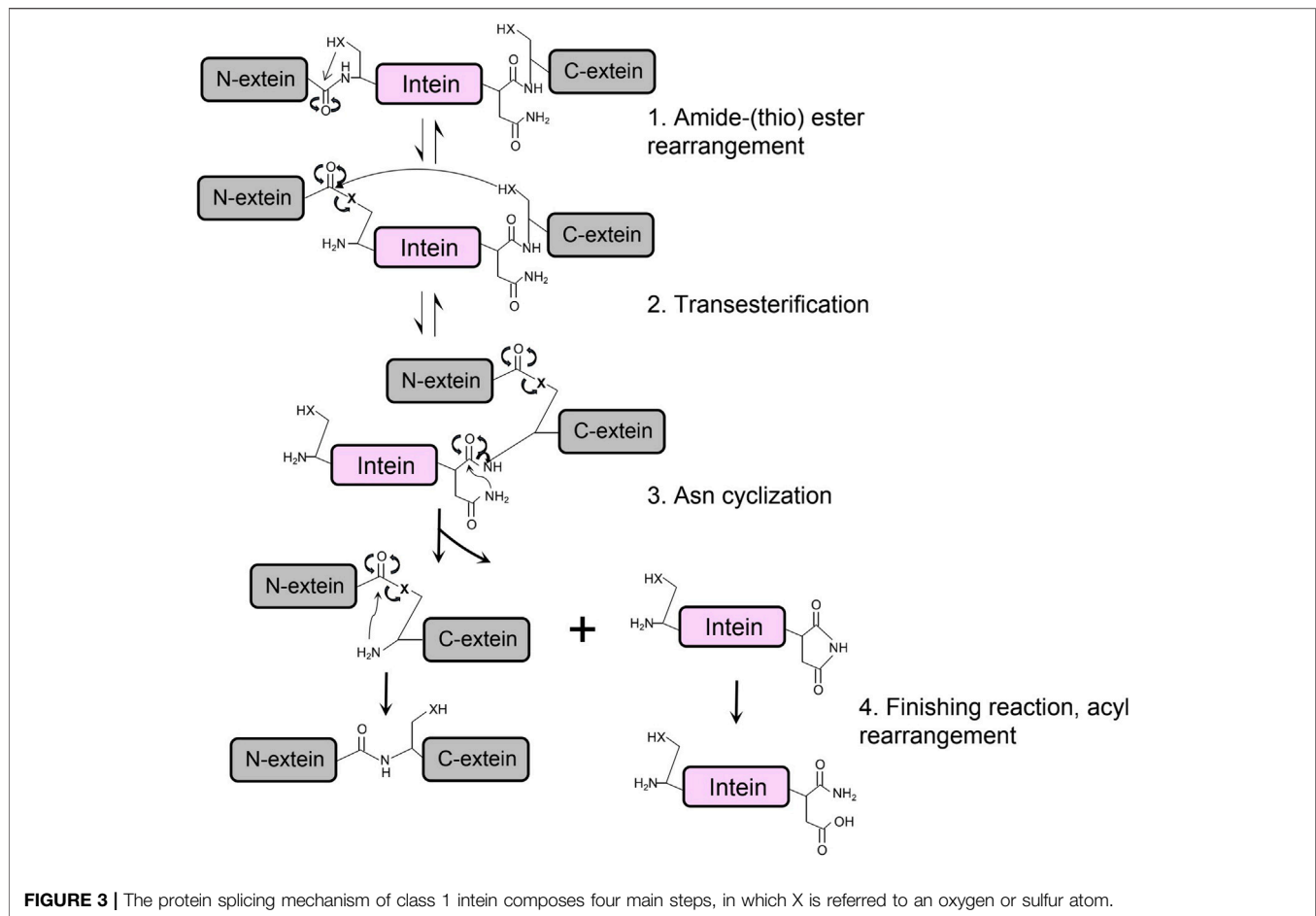


FIGURE 3 | The protein splicing mechanism of class 1 intein composes four main steps, in which X is referred to an oxygen or sulfur atom.

et al., 2000). The two DnaE-intein fragments could be co-expressed in *Escherichia coli* (*E. coli*) and exhibit protein *trans*-splicing activity (Evans et al., 2000).

Besides the proximal amino acid residues on both sides of intein terminuses, splicing is also dependent on the reaction condition. Both *cis*- and *trans*-splicing could be optimized in certain conditions such as optimized pH or reducing environments (Perler and Adam, 2000; Shi et al., 2015; Zhang et al., 2015; Yu et al., 2016; Wang et al., 2018; Pinto et al., 2020; Kimura et al., 2021). For example, Zhang *et al.* found that a weakly acidic environment (pH ~ 6.0–6.5) could facilitate the *Ssp* DnaB intein C-terminal cleavage activity (*cis*-splicing) (Zhang et al., 2015). Wang *et al.* reported that a basic environment (pH ~ 8.5–9.5) benefitted the *Mxe* GyrA intein cleavage reactions (*cis*-splicing) (Wang et al., 2018). Recently, Pinto *et al.* pointed out that a weak basic condition (pH~ 9.0) plus the minimal concentration of DTT (4 mM) promoted the splicing of the split inteins (*trans*-splicing) (Pinto et al., 2020).

ENGINEERING INTEIN BY SYNTHETIC BIOLOGY

In the last 20 years, synthetic biology has been bridging multiple disciplines to design and build novel biomolecular components,

networks and pathways, and using these elements and knowledge to rewire and reprogram organisms (Khalil and Collins, 2010; Cameron et al., 2014; Singh, 2014; Mao et al., 2021; Tang et al., 2021). As nature's escape artists, inteins can seamlessly stitch two proteins, leading to beneficial consequences such as the large protein assembly and the activation of certain factors without requiring assistance from any enzymes or co-factors (Ramsden et al., 2011; López-Igual et al., 2019). These features of intein hold great promise in addressing needs in biomanufacturing, sensing and diagnoses. However, the deficiency in the splicing efficacy and the lack of enough intein tools largely restrain the applications. Inteins can proficiently excise itself from its natural host protein largely depending on the certain fixed junction sequences between exteins and inteins. Based on the current research, the splicing efficiency is often influenced by the proximal amino acid residues on both sides of intein terminuses, especially when recombinantly expressed in a non-native host (Wood et al., 1999; Pinto et al., 2020). These limitations hinder the applications and popularization of inteins since the successful splicing reaction experienced with one extein may not work on another one (Belfort et al., 2006; Pinto et al., 2020). In the past several years, researchers have successfully utilized strategies and tools from synthetic biology to evolve the inteins with enhanced splicing efficacy and expanded the current intein library.

Directed evolution has been proven to be a robust and reliable method to design and alter proteins towards the desirable biological functions, by generating random mutations in the target gene and imposing stringent selection conditions to identify proteins with optimized functionality. Multiple research groups have successfully used the directed evolution method to create genetic diversity of inteins and identified the ones with enhanced splicing efficiency (Wood et al., 1999; Marshall et al., 2015; Stevens et al., 2017). For example, Wood et al. conducted random mutagenesis on a mini-intein and coupled the intein activity to a selectable growth phenotype for screening. Specifically, they used *E. coli* deficient in cellular thymidylate synthase (TS) thus the resultant strain was unable to grow without thymine. They generated a pool of mini-inteins (containing the first 110 and the last 58 amino acids of the 441-amino acid *Mtu* RecA intein) by mutagenic PCR. To couple the splicing efficiency to the TS reporter system, intein-TS fusions were constructed in such a way that the active TS would be produced by intein-mediated splicing to rescue the host cells (Wood et al., 1999).

In 2017, Stevens et al. integrated both the rational design and directed evolution to engineer a naturally occurred split intein, and the improved version is capable of efficient splicing with tolerance on the local extein contexts. They chose one of the most commonly used split inteins *Npu* DnaE (*Nostoc punctiforme*, embedded within the catalytic subunit of DNA polymerase III), which already showed negligible sensitivity to the N-extein residues. However, *Npu* DnaE split intein has sequence preference to the catalytic cysteine (+1 position) and large hydrophobic residues (+2 position) of the C-extein. A previous study showed that these large hydrophobic residues were essential in maintaining the splicing rate due to a stabilizing effect between Phe₊₂ and His₁₂₅ (a key catalytic residue in the last step of protein splicing, involving the cyclization of Asn₁₃₇) (Sun et al., 2005; Shah et al., 2013). Less bulky +2 residues lead to a more dynamic His₁₂₅ side chain with additional conformations that cannot catalyze the splicing. Therefore, they hypothesized that engineering the loop around His₁₂₅ (residues 122–124) could potentially adjust the His₁₂₅ conformational dynamics and therefore restrain the effect of the +2 residue on splicing kinetics. To implement this loop engineering, they conducted saturation mutagenesis on the His₁₂₅ loop (residues 122–124) of *Npu* DnaE intein and coupled the splicing activity to antibiotic resistance in *E. coli* by reconstituting a split version of the aminoglycoside phosphotransferase (resistance to kanamycin) protein. Especially, this splicing was done in the presence of the unfavorable Gly₊₂. By this strategy, they selected mutants that showed remarkable kanamycin resistance with unfavorable C-extein +2 residue, indicating that the evolved intein could splice with minimal extein residues dependency (Stevens et al., 2017).

Besides evolving specific inteins for better performance, people also built intein bank to further explore the potential useful inteins. As mentioned previously, the splicing efficiency of inteins is largely dictated by the junction sequence between the inteins and exteins, and the deviation from the preferred junction sequence may lead to reduced splicing activity (Kwong

and Wong, 2013). Therefore, an expanded library of sufficiently characterized inteins would enhance the probability of matching the target protein with a certain intein bearing a compatible junction sequence. Additionally, selecting split intein pairs that could work simultaneously and orthogonally with no cross-reactivity holds great promise to increase the splicing throughput and expand the application fields.

In 2020, Pinto et al. assessed 34 split inteins libraries and established a library of mutually orthogonal split inteins for both *in vivo* and *in vitro* applications. This study offered fully characterized and versatile toolboxes for scientists to choose based on the desired applications. They first chose 11 pairs of thoroughly characterized split inteins with high splicing reaction rates and efficiencies, and 3 inteins that were reported previously but not fully characterized (Perler, 2002; Dassa et al., 2009). To further expand the library, they extended their search to the intein database of viral and viral-like inteins assuming that these inteins would have faster splicing rates due to the short life cycle of the virus. They shortlisted the 50 inteins at the initials and screened down to 20 phylogenetically distant inteins, based on the assumption that homology is negatively correlated with orthogonality. With these stringent selection criteria, the total of 34 selected candidate inteins were unlikely to share a common ancestor intein since each of them had unique native exteins and shared low sequence homology. To assess the intein functionality and orthogonality, they developed a fast and indirect measurement of splicing activity reporter system, based on the intein-mediated reconstitution of fluorescent protein mCherry. They evaluated the *in vivo* or *in vitro* orthogonality of the split inteins and identified 15 mutually orthogonal pairs that could be used in diverse applications (Pinto et al., 2020).

Another challenge lied in identifying insertion sites of inteins. Besides sharing the similar difficulties in searching a general split site, the introduction of inteins brought an extra layer of complexity since the splicing efficiency largely depends on the extein junction sequences (Ho et al., 2021). To address this issue, Ho et al. used a mini-Mu transposon-based screening approach (intein-assisted bisection mapping (IBM) method) to reveal the split sites for a given protein. They first inserted a transposon randomly into a staging vector, which hosted a coding DNA sequence (CDS) of interest by an *in vitro* transposition reaction. The CDS with successful insertion was isolated by size selection and further ligated into a vector (Nadler et al., 2016; Zeng et al., 2018). The transposon was then substituted with a DNA fragment containing a split intein, the transcription and translation initiation elements for carboxyl-lobe expression and a selection marker. In-frame insertions with the correct orientation will thus split a CDS into two, in which the split intein fragments were fused with the amino-lobes (N-lobes) or carboxyl-lobe (C-lobes) under the separate control of two inducible promoters. The generated library was screened based on the selection rule that the clones displayed the function when both promoters were induced. The clones fulfilling the standard were then sequenced to reveal the split sites at the fusion joints. Using this method, they discovered clusters of split sites on five proteins (Calles and Lorenzo, 2013; Ho et al., 2021). Their work established a generalizable methodology to create split protein-intein fusions for synthetic biology.

APPLYING INTEINS IN SYNTHETIC BIOLOGY

As protein engineering tools, inteins have been widely used in protein purification, protein labeling, and protein cyclization. For example, inteins have been developed as self-cleavable linker during the purification to generate untagged protein (Zhang et al., 2015; Wang et al., 2018). Inteins were also applied in making recombinant C-terminal polypeptide α -thioesters, which were the crucial components for the semi-synthesis of chemically modified proteins using expressed protein ligation (EPL) (Flavell and Muir, 2009; Fong and Wood, 2010; Li, 2015; Sarmiento and Camarero, 2019). Protein *trans*-splicing mediated by split inteins (naturally-occurring or artificially designed) has been used in proteins modification, such as introducing the site-specific modification including phosphorylation, biotinylation, ubiquitination, glycosylation, and segmental isotopic labeling both *in vitro* and in cells (Qin et al., 2002; Rak et al., 2003; Shogren-Knaak et al., 2003; Chacko et al., 2004; Durek et al., 2004; Flavell and Muir, 2009; Borra and Camarero, 2017; Debelouchina and Muir, 2017). Inteins have also been utilized in generating cyclic proteins, in which the split intein fragments were fused to both sides of a target protein, and the N- and C-terminus of a target protein were joined through the association and splicing of the pairs (Topilina and Mills, 2014; Nanda et al., 2020). These applications have been well established and extensively reviewed. Apart from these works, we have noticed that the inteins have been playing important roles in multiple areas of synthetic biology ranging from biocomputing, living therapeutics to material assembly. Here, we further discuss these new efforts.

Biocomputing

Split inteins are ideal tools for implementing digital logic. A protein can be divided into two and fused with a pair of split intein, such that the bipartite fragments remain individually inactive, and protein function is not restored until protein splicing occurs. In the work of Ho *et al.*, after establishing a general workflow (IBM, as mentioned above) to identify the split sites for given proteins, they continued to demonstrate the universality of the method in engineering protein-based logic gates. They discovered multiple split sites within a repressor (tetracycline repressor, TetR) and an activator (the extracytoplasmic sigma factor 20, ECF20). TetR or ECF20 were reconstituted only when both the N-lobes and the C-lobes were present under the induction (induced by arabinose and DAPG (2,4-Diacetylphloroglucinol) respectively). The full protein either repressed or activated the expression of mScarlet (output), and effectively generated a NAND (TetR) or AND (ECF 20) logic (Ho et al., 2021).

In another work, after establishing an extended library of orthogonal split inteins, Pinto *et al.* have further coupled the orthogonal split inteins with orthogonal split extracytoplasmic function (ECF) sigma factors to build modular logic AND gates that can be wired to build complex logic circuits. ECF sigma factors are the smallest and simplest alternative sigma factors. They split three orthogonal ECFs and fused each one with a pair of split intein (three pairs of split inteins are orthogonal to each

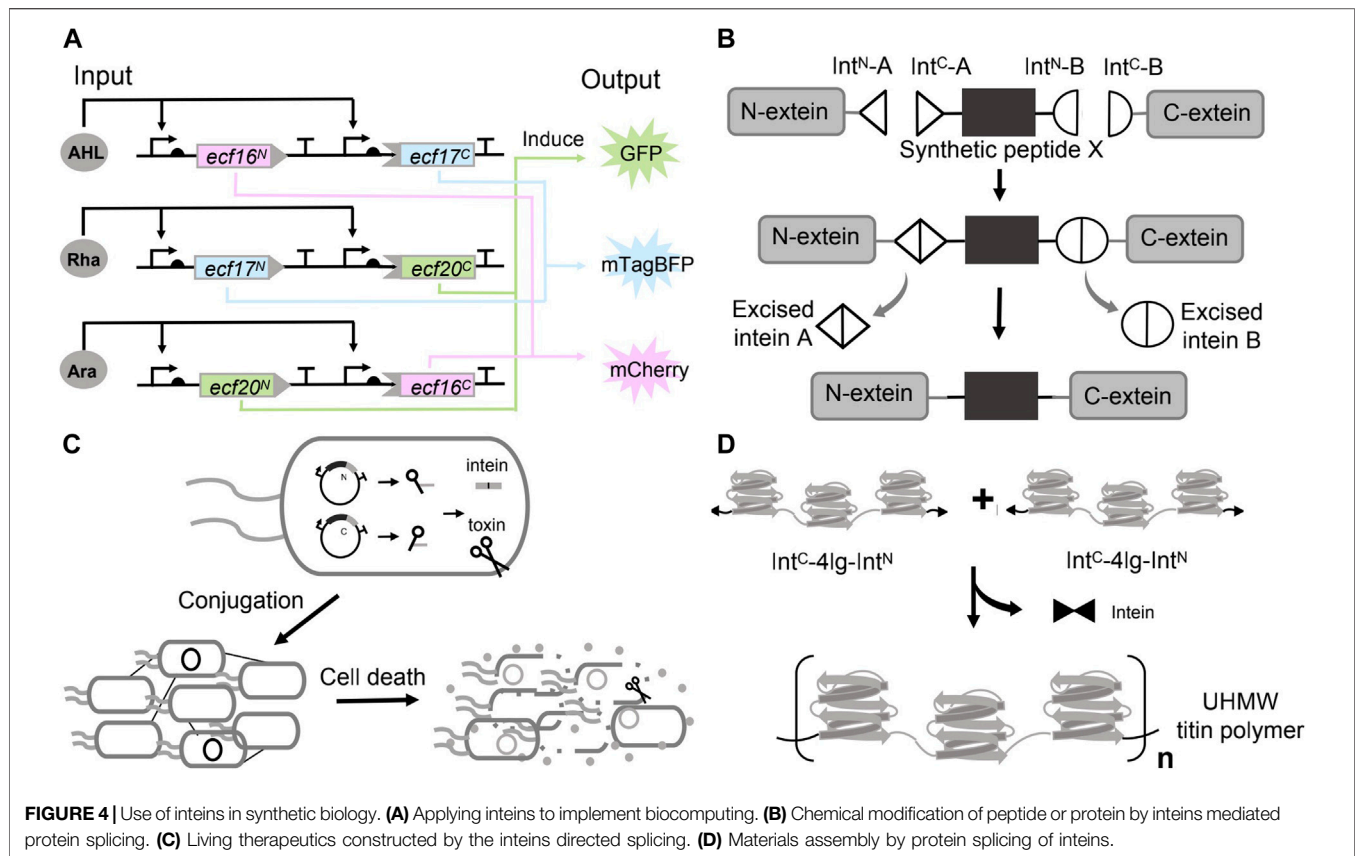
other). The ECF proteins were reconstituted by the intein mediated *trans*-splicing. The fluorescence was detected only when both ECF halves were expressed, confirming the behavior of logic AND gates (Pinto et al., 2020). They have further connected the three AND gates to build a three-input three-output integrated logic circuit (Figure 4A). In each circuit, the promoter (induced by one inducer, and each inducer is defined as one input) drove the expression of a pair of unrelated intein split ECF halves. Therefore, the fluorescent output is only activated by having at least two inputs. The experimental results indicated that the design exhibited the expected logic behavior. These applications suggested the potentials of using orthogonal split inteins and split transcription factors to design complex cellular logic circuits.

Generation of Semisynthetic Proteins

Chemical modification of proteins holds great potential for therapeutics engineering since it can help to understand the pharmacology and improve the property and effect of the drugs (Wold, 1981; Borra and Camarero, 2017). Incorporating non-canonical amino acids (ncAAs) in a site-specific manner can partially address the issue (Liu and Schultz, 2010). However, limitations exist such as the toxicity of ncAAs to the cells, and the difficulties in incorporating multiple ncAAs using the native transcription and translation machinery (Khoo et al., 2020). These problems could be potentially addressed by dividing the target proteins (with post-translational modification (PTM)) into multiple fragments, in which the fragments with the PTM were chemically synthesized, and stitching these parts back by inteins mediated splicing (Figure 4B) (Sarmiento and Camarero, 2019; Khoo et al., 2020). For example, Khoo *et al.* managed to synthesize the protein with PTM by *trans*-splicing in living eukaryotic cells, a method defined as tandem protein *trans*-splicing (tPTS) (Khoo et al., 2020). They divided the protein of interest (POI) into three fragments, namely the N-terminal, C-terminal parts and a central fragment (peptide X) containing the required modification (Nilsson et al., 2005). These three parts were ligated with two orthogonal split intein pairs (*Cfa* DnaE intein and *Ssp* DnaB). The protein N- and C-fragments were expressed by HEK cells, while the peptide X containing the ncAAs was generated by chemical synthesis and injected into the cells (Khoo et al., 2020). Their approach successfully inserted the synthetic peptide containing the homolysine or ornithine (ncAAs) at K71 into the P2X2 receptors (a trimeric ATP-gated ion channel that can be activated by ATP released during synaptic transmission), as validated by the protein function (Khoo et al., 2020).

Living Therapeutics

Synthetic biology is paying attention to the rising field of living therapeutics. Instead of small molecules and protein drugs, researchers are developing genetically engineered cells as the basis for novel therapeutics. Inteins could endow the engineered cells with diverse and sophisticated functions, such as maintaining a dormant state (un-spliced) in the delivering host and activating to generate the therapeutics through the splicing at the target environment. In 2019, López-Igua *et al.* engineered *E. coli* carrying the split toxins



constructs. The split toxins constructs were delivered through conjugation between the *E. coli* and the pathogen. The toxin–intein antimicrobial reagent was only activated in the pathogen that harbors specific transcription factors (Figure 4C). In their experimental design, they chose the toxic protein CcdB as the antimicrobial reagent since it locked up DNA gyrase with broken double-stranded DNA and ultimately caused cell death (Guéroul et al., 2013). However, even the basal expression of a full-length toxin gene (*ccdB*, driven by the P_{BAD}) was sufficient to kill the *E. coli* host. Consequently, they split the toxin gene by an intein (DnaE). The expression of the split toxin gene and reconstitution of the toxin could only be activated by ToxR, the essential transcription activator controlling the expression of cholera toxin, colonization factor and outer membrane protein of *V. cholerae*. By design, the intein-assisted method enabled targeted killing of pathogenic bacteria without harming beneficial members of host-microbiota (Van Melder et al., 1996).

As another example, inteins could assemble the therapeutics which were too large to deliver in the gene therapy. CRISPR/Cas9 provides a possible solution to target and edit virtually any gene to remove the diseases. A major obstacle, however, is that the size of Cas9 (>4 kb) impedes its efficient delivery. Truong *et al.* reported the use of a split intein mediated Cas9 system in a dual-vector recombinant adeno-associated virus (rAAV) system, in which rAAV invaded cells, delivered the split-Cas9 elements and reconstituted Cas9 via protein splicing of the *Npu* DnaE split intein (Truong et al., 2015).

As another example, the adeno-associated viral (AAV) vector-based gene therapy biologics can cure an inherited form of blindness (Tornabene et al., 2019). Again, the limited cargo capacity of the AAV vector inhibits its potential use (Tornabene et al., 2019). Tornabene *et al.* proved that this problem in retinal gene therapy could be ameliorated by a split intein assisted assembly strategy. In their experiments, they successfully delivered multiple AAV vectors, each encoding one of the fragments of target proteins flanked by the split intein, and reconstituted the large ATP binding cassette subfamily A member 4 (ABCA4) as well as centrosomal protein 290 (CEP290) in the retina of mice, pigs and in human retinal organoids to cure the inherited retinal diseases (Tornabene et al., 2019).

Materials Assembly

Artificial high-performance polymer materials bear unique properties such as high-strength and have broad application prospects. However, these materials are mostly derived from petroleum and are non-degradable and non-sustainable (Wegst et al., 2015; Yang et al., 2017). Synthetic biology has engineered microorganisms to produce a wide range of degradable biomaterials (Chen et al., 2014; Gilbert et al., 2021; Tang et al., 2021). Many natural materials with high mechanical properties are hierarchically assembled ultra-high molecular weight (UHMW) proteins that have highly repetitive amino acid sequences. Nevertheless, these UHMW repetitive proteins

were extremely difficult to produce in microorganisms due to the genetic instability, low transformation efficiency and metabolic burden (Tang and Chilkoti, 2016; Rugbjerg et al., 2018). In 2019, Bowen *et al.* showed that the problem could be addressed through *in vivo* protein polymerization catalyzed by split inteins. They designed the monomer construct that contained the 10 repeats of a *Nephila clavipes* MaSp1 dragline spidroin consensus sequence and flanked by a pair of complementary, fast reacting split inteins in the form of Int^C-monomer-Int^N, where Int^C and Int^N represent the C- and N-half of the split inteins, respectively. Expression of this monomer alone, however, caused the protein cyclization instead of polymerization, probably due to the structural flexibility of the monomer permitted the joint of the N- and C-termini. To prevent intramolecular cyclization, they have devised a seed chain polymerization (SCP) method by first inducing a “seed protein”, which contained only one reactive Int^N domain fused at the C-terminus of the seed. After a certain time, the Int^C-monomer-Int^N cassette was subsequently expressed. Considering both the ligation kinetics and the protein synthesis rate, the Int^C domain at the N-terminus of the monomer should react with a seed or linear chain before its C-terminal Int^N domain can be translated, resulting in linear intermolecular ligation instead of cyclization. Using this method, they have successfully synthesized a spider silk protein with a molecular weight of 300 kDa in *E. coli* (Bowen et al., 2019).

Recently, Bowen *et al.* has further used a similar strategy to fabricate megadalton muscle titin polymers (Figure 4D). They fused the C- and N-terminal halves of a split intein to the N- and C-termini of a short titin subunit containing four Ig domains. The expression of monomers spontaneously initiated the splicing reactions and covalently linked the monomers to form the UHMW proteins (20% of the purified titin polymers have a molecular weight over 5 MDa). In this case, the subunit of four Ig domains was rigid to prevent the cyclization. They next processed these UHMW proteins into macroscale monofilament fibers, and showed that these high-performance fibers exhibited high strength, toughness, and damping energy (Bowen et al., 2021).

SUMMARY AND PERSPECTIVE

Over the last 3 decades, research about inteins is progressively pushing the limits from identification and characterization of

inteins to productive utilization. Recently, the concepts and tools from synthetic biology have been further integrated with inteins to enable new and diverse functions. In this review, we have introduced several inteins mediated applications ranging from living therapeutics engineering to material assembly. Obviously, inteins can empower the system with more flexibility and functionality. Previously research groups have utilized SpyTag/SpyCatcher system in the field of engineered living material. For example, they have decorated the polypeptides with either SpyTags or SpyCatchers, which can react to polymerize to form the protein polymer (Dai et al., 2021). The role of the SpyTag/SpyCatcher could be replaced with the split intein pairs, which can stitch the peptide seamlessly. It is worthy to note that the library of the orthogonal split intein pairs can link monomers with high selectivity. Therefore, different polypeptides could be assembled in a certain sequence to form the “block polymer”. Inteins could further be coupled with tools of optogenetics, such that the living therapeutics could be only activated via light-controlled intein mediated splicing.

Limitations still exist. For example, splitting the target protein often leads to the misfolding of the protein and the formation of inclusion bodies. Although new inteins are consistently discovered, the number of feasible and reliable tools and orthogonal pairs (split inteins) are still deficient. These problems could be possibly improved with protein design by deep learning and large-scale screening by robotic-driven automation. With the advancement and enrichment in the inteins, new genetic parts and networks, we are expecting to see more exciting and creative roles played by inteins in versatile applications.

AUTHOR CONTRIBUTIONS

HW and LW wrote the review. BZ revised the review. ZD conceived and wrote the review.

ACKNOWLEDGMENTS

This review was partially supported by National Key Research and Development Program of China No. 2018YFA0903000, No.2020YFA0908100, Shenzhen Science and Technology Program No. KQTD20180413181837372, National Natural Science Foundation of China 32071427.

REFERENCES

- Belfort, M., Stoddard, B. L., Wood, D. W., and Derbyshire, V. (2006). *Homing Endonucleases and Inteins*. Germany: Springer Science & Business Media.
- Borra, R., and Camarero, J. A. (2017). Protein Chemical Modification inside Living Cells Using Split Inteins. *Methods Mol. Biol.* 1495, 111–130. doi:10.1007/978-1-4939-6451-2_8
- Bowen, C. H., Sargent, C. J., Wang, A., Zhu, Y., Chang, X., Li, J., et al. (2021). Microbial Production of Megadalton Titin Yields Fibers with Advantageous Mechanical Properties. *Nat. Commun.* 12, 5182. doi:10.1038/s41467-021-25360-6
- Bowen, C. H., Reed, T. J., Sargent, C. J., Mpamo, B., Galazka, J. M., and Zhang, F. (2019). Seeded Chain-Growth Polymerization of Proteins in Living Bacterial Cells. *ACS Synth. Biol.* 8, 2651–2658. doi:10.1021/acssynbio.9b00362
- Calles, B., and Lorenzo, V. d. (2013). Expanding the Boolean Logic of the Prokaryotic Transcription Factor XylR by Functionalization of Permissive Sites with a Protease-Target Sequence. *ACS Synth. Biol.* 2, 594–603. doi:10.1021/sb400050k
- Cameron, D. E., Bashor, C. J., and Collins, J. J. (2014). A Brief History of Synthetic Biology. *Nat. Rev. Microbiol.* 12, 381–390. doi:10.1038/nrmicro3239
- Chacko, B. M., Qin, B. Y., Tiwari, A., Shi, G., Lam, S., Hayward, L. J., et al. (2004). Structural Basis of Heteromeric Smad Protein Assembly in TGF- β Signaling. *Mol. Cell.* 15, 813–823. doi:10.1016/j.molcel.2004.07.016

- Chen, A. Y., Deng, Z., Billings, A. N., Seker, U. O. S., Lu, M. Y., Citorik, R. J., et al. (2014). Synthesis and Patterning of Tunable Multiscale Materials with Engineered Cells. *Nat. Mater.* 13, 515–523. doi:10.1038/nmat3912
- Cheriyian, M., and Perler, F. B. (2009). Protein Splicing: A Versatile Tool for Drug Discovery. *Adv. Drug Deliv. Rev.* 61, 899–907. doi:10.1016/j.addr.2009.04.021
- Chong, S., and Xu, M.-Q. (1997). Protein Splicing of the *Saccharomyces cerevisiae* VMA Intein without the Endonuclease Motifs. *J. Biol. Chem.* 272, 15587–15590. doi:10.1074/jbc.272.25.15587
- Dai, Z., Yang, X., Wu, F., Wang, L., Xiang, K., Li, P., et al. (2021). Living Fabrication of Functional Semi-interpenetrating Polymeric Materials. *Nat. Commun.* 12, 3422. doi:10.1038/s41467-021-23812-7
- Dassa, B., London, N., Stoddard, B. L., Schueler-Furman, O., and Pietrokovski, S. (2009). Fractured Genes: A Novel Genomic Arrangement Involving New Split Inteins and a New Homing Endonuclease Family. *Nucleic Acids Res.* 37, 2560–2573. doi:10.1093/nar/gkp095
- Debelouchina, G. T., and Muir, T. W. (2017). A Molecular Engineering Toolbox for the Structural Biologist. *Q. Rev. Biophys.* 50, e7. doi:10.1017/S0033583517000051
- Durek, T., Alexandrov, K., Goody, R. S., Hildebrand, A., Heinemann, I., and Waldmann, H. (2004). Synthesis of Fluorescently Labeled Mono- and Diprenylated Rab7 GTPase. *J. Am. Chem. Soc.* 126, 16368–16378. doi:10.1021/ja046164n
- Evans, T. C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., et al. (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. doi:10.1074/jbc.275.13.9091
- Flavell, R. R., and Muir, T. W. (2009). Expressed Protein Ligation (EPL) in the Study of Signal Transduction, Ion Conduction, and Chromatin Biology. *Acc. Chem. Res.* 42, 107–116. doi:10.1021/ar800129c
- Fong, B. A., and Wood, D. W. (2010). Expression and Purification of ELP-Intein-Tagged Target Proteins in High Cell Density *E. coli* Fermentation. *Microb. Cell Fact* 9, 77–11. doi:10.1186/1475-2859-9-77
- Gilbert, C., Tang, T.-C., Ott, W., Dorr, B. A., Shaw, W. M., Sun, G. L., et al. (2021). Living Materials with Programmable Functionalities Grown from Engineered Microbial Co-cultures. *Nat. Mater.* 20, 691–700. doi:10.1038/s41563-020-00857-5
- Gorbalenya, A. E. (1998). Non-canonical Inteins. *Nucleic Acids Res.* 26, 1741–1748. doi:10.1093/nar/26.7.1741
- Guérout, A.-M., Iqbal, N., Mine, N., Ducos-Galand, M., Van Melderren, L., and Mazel, D. (2013). Characterization of the Phd-Doc and Ccd Toxin-Antitoxin Cassettes from *Vibrio* Superintegrons. *J. Bacteriol.* 195, 2270–2283. doi:10.1128/JB.01389-12
- Hirata, R., Ohsumi, Y., Nakano, A., Kawasaki, H., Suzuki, K., and 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. doi:10.1016/s0021-9258(19)39210-5
- Ho, T. Y., Shao, A., Lu, Z., Savilahti, H., Menolascina, F., Wang, L., et al. (2021). A Systematic Approach to Inserting Split Inteins for Boolean Logic Gate Engineering and Basal Activity Reduction. *Nat. Commun.* 12, 1–12. doi:10.1038/s41467-021-22404-9
- Khalil, A. S., and Collins, J. J. (2010). Synthetic Biology: Applications Come of Age. *Nat. Rev. Genet.* 11, 367–379. doi:10.1038/nrg2775
- Khoo, K. K., Galleano, I., Gasparri, F., Wieneke, R., Harms, H., Poulsen, M. H., et al. (2020). Chemical Modification of Proteins by Insertion of Synthetic Peptides Using Tandem Protein Trans-splicing. *Nat. Commun.* 11, 2284. doi:10.1038/s41467-020-16208-6
- Kimura, H., Miura, D., Tsugawa, W., Ikebukuro, K., Sode, K., and Asano, R. (2021). Rapid and Homogeneous Electrochemical Detection by Fabricating a High Affinity Bispecific Antibody-Enzyme Complex Using Two Catcher/Tag Systems. *Biosens. Bioelectron.* 175, 112885. doi:10.1016/j.bios.2020.112885
- Kwong, K. W. Y., Ng, A. K. L., and Wong, W. K. R. (2016). Engineering Versatile Protein Expression Systems Mediated by Inteins in *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 100, 255–262. doi:10.1007/s00253-015-6960-z
- Kwong, K. W. Y., and Wong, W. K. R. (2013). A Revolutionary Approach Facilitating Co-expression of Authentic Human Epidermal Growth Factor and Basic Fibroblast Growth Factor in Both Cytoplasm and Culture Medium of *Escherichia coli*. *Appl. Microbiol. Biotechnol.* 97, 9071–9080. doi:10.1007/s00253-013-5090-8
- Li, Y. (2015). Split-inteins and Their Bioapplications. *Biotechnol. Lett.* 37, 2121–2137. doi:10.1007/s10529-015-1905-2
- Liu, C. C., and Schultz, P. G. (2010). Adding New Chemistries to the Genetic Code. *Annu. Rev. Biochem.* 79, 413–444. doi:10.1146/annurev.biochem.052308.105824
- López-Igual, R., Bernal-Bayard, J., Rodríguez-Paton, A., Ghigo, J.-M., and Mazel, D. (2019). Engineered Toxin-Intein Antimicrobials Can Selectively Target and Kill Antibiotic-Resistant Bacteria in Mixed Populations. *Nat. Biotechnol.* 37, 755–760. doi:10.1038/s41587-019-0105-3
- Mao, N., Aggarwal, N., Poh, C. L., Cho, B. K., Kondo, A., Liu, C., et al. (2021). Future Trends in Synthetic Biology in Asia. *Adv. Genet.* 2, e10038. doi:10.1002/ggn2.10038
- Marshall, C. J., Grosskopf, V. A., Moehling, T. J., Tillotson, B. J., Wiepz, G. J., Abbott, N. L., et al. (2015). An Evolved Mxe GyrA Intein for Enhanced Production of Fusion Proteins. *ACS Chem. Biol.* 10, 527–538. doi:10.1021/cb500689g
- Mills, K. V., Johnson, M. A., and Perler, F. B. (2014). Protein Splicing: How Inteins Escape from Precursor Proteins. *J. Biol. Chem.* 289, 14498–14505. doi:10.1074/jbc.r113.540310
- Nadler, D. C., Morgan, S. A., Flamholz, A., Kortright, K. E., and Savage, D. F. (2016). Rapid Construction of Metabolite Biosensors Using Domain-Insertion Profiling. *Nat. Commun.* 7, 12266. doi:10.1038/ncomms12266
- Nanda, A., Nasker, S. S., Mehra, A., Panda, S., and Nayak, S. (2020). Inteins in Science: Evolution to Application. *Microorganisms* 8, 2004. doi:10.3390/microorganisms8122004
- Nilsson, B. L., Soellner, M. B., and Raines, R. T. (2005). Chemical Synthesis of Proteins. *Annu. Rev. Biophys. Biomol. Struct.* 34, 91–118. doi:10.1146/annurev.biophys.34.040204.144700
- Perler, F. B., and Adam, E. (2000). Protein Splicing and its Applications. *Curr. Opin. Biotechnol.* 11, 377–383. doi:10.1016/s0958-1669(00)00113-0
- Perler, F. B. (2002). InBase: The Intein Database. *Nucleic Acids Res.* 30, 383–384. doi:10.1093/nar/30.1.383
- Perler, F. B. (2005). *Inteins—A Historical Perspective, Homing Endonucleases and Inteins*. Germany: Springer, 193–210.
- Perler, F. B. (1998). Protein Splicing of Inteins and Hedgehog Autoproteolysis: Structure, Function, and Evolution. *Cell* 92, 1–4. doi:10.1016/s0092-8674(00)80892-2
- Pinto, F., Thornton, E. L., and Wang, B. (2020). An Expanded Library of Orthogonal Split Inteins Enables Modular Multi-Peptide Assemblies. *Nat. Commun.* 11, 1529. doi:10.1038/s41467-020-15272-2
- Qin, B. Y., Lam, S. S., Correia, J. J., and Lin, K. (2002). Smad3 Allosteric Links TGF- β Receptor Kinase Activation to Transcriptional Control. *Genes Dev.* 16, 1950–1963. doi:10.1101/gad.1002002
- Rak, A., Pilypenko, O., Durek, T., Watzke, A., Kushnir, S., Brunsveld, L., et al. (2003). Structure of Rab GDP-Dissociation Inhibitor in Complex with Prenylated YPT1 GTPase. *Science* 302, 646–650. doi:10.1126/science.1087761
- Ramsden, R., Arms, L., Davis, T. N., and Muller, E. G. (2011). An Intein with Genetically Selectable Markers Provides a New Approach to Internally Label Proteins with GFP. *BMC Biotechnol.* 11, 71–11. doi:10.1186/1472-6750-11-71
- Reitter, J. N., and Mills, K. V. (2011). Canonical Protein Splicing of a Class 1 Intein that Has a Class 3 Noncanonical Sequence Motif. *J. Bacteriol.* 193, 994–997. doi:10.1128/jb.01287-10
- Rugbjerg, P., Myling-Petersen, N., Porse, A., Sarup-Lytzen, K., and Sommer, M. O. A. (2018). Diverse Genetic Error Modes Constrain Large-Scale Bio-Based Production. *Nat. Commun.* 9, 787. doi:10.1038/s41467-018-03232-w
- Sarmiento, C., and Camarero, J. A. (2019). Biotechnological Applications of Protein Splicing. *Curr. Protein Pept. Sci.* 20, 408–424. doi:10.2174/1389203720666190208110416
- Shah, N. H., Eryilmaz, E., Cowburn, D., and Muir, T. W. (2013). Extein Residues Play an Intimate Role in the Rate-Limiting Step of Protein Trans-splicing. *J. Am. Chem. Soc.* 135, 5839–5847. doi:10.1021/ja401015p
- Shi, Y.-F., Cao, Z.-A., and Shen, Z.-Y. (2015). Impact of Polybasic Alcohols on Biocompatibility and Selectivity of Penicillin G Acylase for Kinetically Controlled Synthesis. *PeerJ Preprints* 3, e1122v1. doi:10.7287/PEERJ.PREPRINTS.1122

- Shogren-Knaak, M. A., Fry, C. J., and Peterson, C. L. (2003). A Native Peptide Ligation Strategy for Deciphering Nucleosomal Histone Modifications. *J. Biol. Chem.* 278, 15744–15748. doi:10.1074/jbc.m301445200
- Singh, V. (2014). Recent Advancements in Synthetic Biology: Current Status and Challenges. *Gene* 535, 1–11. doi:10.1016/j.gene.2013.11.025
- Southworth, M. W., Benner, J., and Perler, F. B. (2000). An Alternative Protein Splicing Mechanism for Inteins Lacking an N-Terminal Nucleophile. *EMBO J.* 19, 5019–5026. doi:10.1093/emboj/19.18.5019
- Stevens, A. J., Brown, Z. Z., Shah, N. H., Sekar, G., Cowburn, D., and Muir, T. W. (2016). Design of a Split Intein with Exceptional Protein Splicing Activity. *J. Am. Chem. Soc.* 138, 2162–2165. doi:10.1021/jacs.5b13528
- Stevens, A. J., Sekar, G., Shah, N. H., Mostafavi, A. Z., Cowburn, D., and Muir, T. W. (2017). A Promiscuous Split Intein with Expanded Protein Engineering Applications. *Proc. Natl. Acad. Sci. USA* 114, 8538–8543. doi:10.1073/pnas.1701083114
- Sun, P., Ye, S., Ferrandon, S., Evans, T. C., Xu, M.-Q., and 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. doi:10.1016/j.jmb.2005.09.039
- Tang, N. C., and Chilkoti, A. (2016). Combinatorial Codon Scrambling Enables Scalable Gene Synthesis and Amplification of Repetitive Proteins. *Nat. Mater.* 15, 419–424. doi:10.1038/nmat4521
- Tang, T.-C., An, B., Huang, Y., Vasikaran, S., Wang, Y., Jiang, X., et al. (2021). Materials Design by Synthetic Biology. *Nat. Rev. Mater.* 6, 332–350. doi:10.1038/s41578-020-00265-w
- Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W. R., Jr, and Perler, F. B. (1997). The Mycobacterium *Xenopi* GyrA Protein Splicing Element: Characterization of a Minimal Intein. *J. Bacteriol.* 179, 6378–6382. doi:10.1128/jb.179.20.6378-6382.1997
- Topilina, N. I., and Mills, K. V. (2014). Recent Advances in *In Vivo* Applications of Intein-Mediated Protein Splicing. *Mob DNA* 5, 5–14. doi:10.1186/1759-8753-5-5
- Tornabene, P., Trapani, I., Minopoli, R., Centrulo, M., Lupo, M., de Simone, S., et al. (2019). Intein-mediated Protein Trans-splicing Expands Adeno-Associated Virus Transfer Capacity in the Retina. *Sci. Transl. Med.* 11. doi:10.1126/scitranslmed.aav4523
- Truong, D.-J. J., Kühner, K., Kühn, R., Werfel, S., Engelhardt, S., Wurst, W., et al. (2015). Development of an Intein-Mediated Split-Cas9 System for Gene Therapy. *Nucleic Acids Res.* 43, 6450–6458. doi:10.1093/nar/gkv601
- Van Melder, L., Thi, M. H. D., Lecchi, P., Gottesman, S., Couturier, M., and Maurizi, M. R. (1996). ATP-dependent Degradation of CcdA by Lon Protease. *J. Biol. Chem.* 271, 27730–27738. doi:10.1074/jbc.271.44.27730
- Wang, H., Hu, X., Thiyagarajan, S., Lai, C. Y., Ng, K. L., Lam, C. C., et al. (2018). A Practical Approach to Unveiling Auto-Catalytic Cleavages Mediated by Mxe GyrA Intein and Improving the Production of Authentic bFGF. *J. Adv. Res. Biotechnol.* 3, 1. doi:10.15226/2475-4714/4/1/00140
- Wegst, U. G. K., Bai, H., Saiz, E., Tomsia, A. P., and Ritchie, R. O. (2015). Bioinspired Structural Materials. *Nat. Mater.* 14, 23–36. doi:10.1038/nmat4089
- Wold, F. (1981). *In Vivo* Chemical Modification of Proteins (Post-Translational Modification). *Annu. Rev. Biochem.* 50, 783–814. doi:10.1146/annurev.bi.50.070181.004031
- Wood, D. W., Wu, W., Belfort, G., Derbyshire, V., and Belfort, M. (1999). A Genetic System Yields Self-Cleaving Inteins for Bioseparations. *Nat. Biotechnol.* 17, 889–892. doi:10.1038/12879
- Yang, Y. J., Holmberg, A. L., and Olsen, B. D. (2017). Artificially Engineered Protein Polymers. *Annu. Rev. Chem. Biomol. Eng.* 8, 549–575. doi:10.1146/annurev-chembioeng-060816-101620
- Yu, F., Yang, Z.-h., Fan, H.-d., and Zuo, Z.-y. (2016). The Construction of Prokaryotic Expression Vector for Human Gene IL-24 and Expression and Purification of its Protein. *Biotechnol. Bull.* 32, 84. doi:10.13560/j.cnki.biotech.bull.1985.2016.02.011
- Zeng, Y., Jones, A. M., Thomas, E. E., Nassif, B., Silberg, J. J., and Segatori, L. (2018). A Split Transcriptional Repressor that Links Protein Solubility to an Orthogonal Genetic Circuit. *ACS Synth. Biol.* 7, 2126–2138. doi:10.1021/acssynbio.8b00129
- Zhang, Y., Zhang, K., Wan, Y., Zi, J., Wang, Y., Wang, J., et al. (2015). A pH-Induced, Intein-Mediated Expression and Purification of Recombinant Human Epidermal Growth Factor in *Escherichia Coli*. *Biotechnol. Prog.* 31, 758–764. doi:10.1002/btpr.2086

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Publisher's Note: All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

Copyright © 2022 Wang, Wang, Zhong and Dai. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.