Synthetic Two-piece and Three-piece Split Inteins for Protein trans-Splicing*

Wenchang Sun, Jing Yang, and Xiang-Qin Liu‡

From the Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia B3H 4H7, Canada

An intein is a protein-intervening sequence that catalyzes a protein-splicing reaction in which the intein sequence is precisely excised and its flanking sequences (N- and C-exteins) join with a peptide bond to produce the mature host protein (spliced protein) (1). The mechanism of protein splicing typically has four steps: two acyl rearrangements at the two splicing junctions, a trans-esterification between the two junctions, and a cyclization of the Asn residue at the C-terminal junction (2–4). Crystal structures of inteins revealed a splicing domain consisting of 11–12 β-strands and forming a compact horseshoe shape with the splicing junctions located in the central cleft (5–11). A majority of inteins also have a homing endonuclease domain inserted in the splicing domain sequence (12). These bifunctional inteins are ~350–550 amino acids (aa)1 long, although some extra large inteins are up to 1650 aa long and also contain tandem repeats (13, 14). Nearly 200 intein and intein-like sequences have been found in a wide variety of host proteins and in microorganisms belonging to bacteria, Archaea, and eukaryotes (12, 15). Their sporadic phylogenetic distributions suggest lateral gene transfer through intein homing (16, 17). Inteins generally share only low levels of sequence similarity, but they share striking similarities in structure, reaction mechanism, and evolution (4, 18, 19, 21). It is thought that inteins first originated with just the splicing domain and then acquired the endonuclease domain, with the latter conferring genetic mobility to the intein. During intein evolution, however, some inteins lost their endonuclease domain to become mini-inteins consisting of just the ~130-aa protein-splicing domain plus a linker sequence of various lengths in place of the endonuclease domain (12, 22).

An interesting event of intein evolution is the loss of sequence continuity in some inteins, which apparently produced the DnaE split intein that exists in two fragments and is capable of protein trans-splicing (23). A pair of split DnaE genes produces two precursor polypeptides, with one consisting of the N-terminal part of DnaE (N-extein) followed by the N-terminal part of intein (N-intein) and another consisting of the C-terminal part of DnaE (C-extein) preceded by the C-terminal part of the intein (C-intein). The N- and C-inteins, through their structural complementation, can reassemble and catalyze a protein trans-splicing reaction to produce a mature DnaE. This two-piece split intein has since been found in many cyanobacterial species (14, 24). These naturally occurring split inteins most likely originated from a contiguous intein sequence as a result of genomic rearrangement(s) that broke the intein coding sequence. Interestingly, all these DnaE split inteins share the same split site, which may be explained by a single origin for all these split inteins. However, it may also suggest that splitting at any other site is incompatible with protein trans-splicing and therefore not tolerated, which can be examined by splitting inteins at other sites followed by testing for possible trans-splicing. Synthetic two-piece split inteins have been engineered before in laboratories by splitting the coding sequences of contiguous inteins, but their split sites corresponded with those of the naturally occurring split inteins (25–27). Both naturally occurring and synthetic split inteins have found many practical uses, which include producing trans-spliced recombinant proteins and circularized proteins or peptides for various purposes (28–30). Finding new split sites for functional split inteins can be useful for the many emerging applications involving protein trans-splicing techniques.

Inteins have been viewed as protein equivalents of introns because of some superficial similarities between inteins and self-splicing introns. Both are intervening sequences, can excise themselves through self-splicing, and are genetically mobile through a similar homing mechanism. Furthermore, the natural occurrence of two-piece split inteins parallels the natural occurrence of two-piece split group II introns. A group II intron can also exist in three pieces, and further fragmentation is believed to have led to the origin of nuclear spliceosomal introns (31, 32). Can inteins also function when split at different sites and into more than two pieces? This is not as predict-

---

*This work was supported by a research grant from the Canadian Institutes of Health Research. The costs of publication of this article were defrayed in part by the payment of page charges. This article must therefore be hereby marked "advertisement" in accordance with 18 U.S.C. Section 1734 solely to indicate this fact.

‡ To whom correspondence should be addressed. Tel.: 902-494-1208; Fax: 902-494-1355; E-mail: pxqliu@dal.ca.

1 The abbreviations used are: aa, amino acid; IPTG, isopropyl-1-thio-β-D-galactopyranoside; N- and C-intein, N- and C-terminal part of the intein, respectively; M-intein, middle part of the intein.
able as for group II introns, which have clear base-pairings between the split pieces. The naturally occurring split inteins do not answer this question, because they all have the same split site corresponding to the insertion site of the endonuclease domain in contiguous intein sequences. Previously reported synthetic split inteins do not answer this question either, because they all have a split site similar to that of naturally occurring split inteins that either corresponded to or was inside the endonuclease domain. This question has become more important with the recent discovery of apparent protein-splicing products in human (33), because in this case the reaction mechanism was not known and an intein-like sequence not found. If an intein could be split into multiple pieces, as a group II intron could become a spliceosomal intron, the resulting residual intein sequence could have lost most or all of the known intein signatures.

In this study, we tested a series of potential split sites in the Ssp DnaB mini-intein. We chose the previously engineered trans-splicing DnaB mini-intein that was proficient in protein trans-splicing in loop regions between β-strands through a ribosome binding sequence and a start codon. The resulting split intein coding sequence was isolated as a XhoI-HindIII DNA fragment and placed back into the pMST expression plasmid to replace the original intein coding sequence. The high fidelity Ffu Turbo DNA polymerase (Stratagene) was used in the PCR reactions, and correct DNA sequences were confirmed through DNA sequencing.

**Protein Splicing in Escherichia coli Cells—Production of plasmid-encoded recombinant proteins in E. coli cells, gel electrophoresis, and Western blotting was carried out as previously described (27, 34). Briefly, cells containing individual expression plasmid were grown in liquid Luria broth medium at 37 °C to late log phase (OD600, 0.5). IPTG was added to a final concentration of 0.8 mM to induce production of the recombinant proteins, and the induction was continued for 3 h. Cells were then harvested and lysed in SDS- and dithiothreitol-containing gel loading buffer in a boiling water bath before electrophoresis in SDS-polyacrylamide gel. Western blotting was carried out with either anti-maltose binding protein monoclonal antibody (New England Biolabs) or anti-thioredoxin monoclonal antibody (Invitrogen), using the Enhanced Chemi-Luminescence detection kit. Intensity of the protein band was estimated using a gel documentation system (Gel Doc 1000 coupled with Molecular Analyst software; Bio-Rad).

**RESULTS**

**Two-piece Split Inteins and Protein trans-Splicing—**To find new split sites capable of supporting protein trans-splicing, we chose the previously engineered Ssp DnaB mini-intein that was proficient in protein cis- and trans-splicing (27). In choosing new split sites in the intein sequence, we considered the likely three-dimensional structure of the intein, reasoning that splitting in loop regions between β-strands would more likely produce protein trans-splicing activity. Because a crystal structure of the Ssp DnaB mini-intein was not available at the beginning of this study, we predicted likely β-strands through sequence comparisons with inteins of known crystal structures (data not shown). Our predictions turned out to be largely correct when the crystal structure of this mini-intein became available recently (6). As shown in Fig. 1 and Table I, fourteen

---

**EXPERIMENTAL PROCEDURES**

**Constructing Split Intein Genes—**The coding sequence of Ssp DnaB mini-intein was isolated from the previously described pMST plasmid (27) as an 842-bp XhoI-HindIII DNA fragment. This was placed in a pBluescript plasmid for easier subsequent manipulations. To insert a spacer sequence to split the intein coding sequence, a restriction enzyme cutting site (KasI or MfeI) was introduced at the intended splitting site. This was achieved by inverse PCR on the circular plasmid, using a synthetic three-piece split intein was constructed and shown to be capable of protein trans-splicing various levels of protein trans-splicing activity. In addition, a synthetic three-piece split intein was constructed and shown to be capable of protein trans-splicing. We discuss their implications on intein structure-function, evolution, and potential uses in biotechnology.

**Trans-splicing Inteins**

![Table 1: List of tested split sites](image)

| Split site | Positiona | N-proteinb | C-proteinc | Efficiency of splicingd |
|-----------|-----------|------------|------------|-------------------------|
| S0        | Gln-105   | 55.8       | 18.3       | 100                     |
| S1        | Ala-11    | 44.9       | 29.1       | 48                      |
| S2        | Glu-25    | 46.4       | 27.5       | 0                       |
| S3        | Glu-55    | 47.7       | 26.3       | 0                       |
| S4        | Lys-53    | 49.7       | 24.3       | 0                       |
| S5        | Ala-71    | 51.7       | 22.3       | 0                       |
| S6        | Asp-79    | 52.7       | 21.3       | 96                      |
| S7        | Ser-88    | 53.8       | 20.2       | 71                      |
| S8        | Glu-100   | 55.2       | 18.8       | 100                     |
| S9        | Glu-132   | 58.7       | 15.2       | 0                       |
| S10       | Pro-140   | 59.6       | 14.3       | 0                       |
| S11       | Asn-148   | 60.4       | 13.5       | 0                       |
| S12       | Lys-44    | 48.7       | 25.3       | 0                       |
| S13       | Gly-64    | 50.9       | 23.0       | 0                       |
| S1 and S0 | Ala-11 and Gln-105 | 44.9 | 18.3 | 87–100 |

---

a Amino acid position preceding the split site.
b Predicted from gene constructs of Figs. 2 and 3.
c Amount of spliced protein divided by the sum of spliced protein and C-protein.
d Calculated from Fig. 1.
different split sites (S0-S13) were selected in the Ssp DnaB mini-intein sequence, with most of them located in a loop region between predicted β-strands. To produce and test the new split inteins in E. coli, we used a previously described pMST plasmid that encodes a fusion protein consisting of the Ssp DnaB mini-intein between a maltose binding protein (M) at the N terminus and a thioredoxin (T) at the C terminus (Fig. 2). Splitting the intein coding sequence was achieved by inserting a spacer sequence at individual selected sites. The spacer sequence contained a stop codon, a ribosome binding (Shine-Dalgarno) sequence, and a start codon. This produced a two-gene operon behind an IPTG-inducible promoter (Ptac). The resulting two-gene operon is behind an IPTG-inducible promoter (Ptac). Bottom, observation of protein splicing on Western blots using Anti-M or Anti-T antibodies. Lane 1 shows a standard of the spliced protein. Positions of the N-protein, C-protein, and spliced protein are marked N, C, and S, respectively. In the Anti-M panel, lanes for S7-S13 are not shown because the S and N bands overlapped.

Western blotting (Fig. 2) and by their predicted sizes (Table I). As seen in Fig. 2, a spliced protein was observed with four split sites (S1, S6, S7, and S8), indicating protein trans-splicing activity. To estimate the efficiency of protein trans-splicing, the intensity of individual protein bands on Western blot was used to measure the amount of that protein, and the efficiency of protein trans-splicing was calculated as the amount of spliced protein divided by the sum of the spliced protein and the C-protein. The efficiency of protein splicing was estimated to be 48, 96, 71, and 85% for split sites S1, S6, S7, and S8, respectively. The other split sites showed no detectable amount of spliced protein (representatives shown in Fig. 2), indicating a lack of protein trans-splicing activity. The N-protein was not used in estimating efficiency because it was from the first gene of the two-gene operon and always accumulated in excessive amounts compared with C-protein from the second gene. This was most likely because of a less than 100% translational coupling of the two genes, as had been observed previously.

Three-piece Split Intein Showing Protein trans-Splicing Activity—We then tested the possibility of producing a three-piece split intein capable of protein trans-splicing, encouraged by the finding of multiple new split sites compatible with protein trans-splicing. We chose to split the Ssp DnaB mini-intein at both the new split site, S1, and the previously known split site,
S0 (27). Because S1 and S0 are the farthest separated sites that showed protein trans-splicing, we reasoned that the resulting larger size (94 aa) of the middle part of the intein (M-intein) would be better for stability in E. coli cells. The resulting N-terminal part (N-intein) and the C-terminal (C-intein) part of the intein were 11 and 49 aa long, respectively. A three-gene operon was constructed and introduced into E. coli cells to produce the N-protein (maltose binding protein plus N-intein), the M-intein, and the C-protein (C-intein plus thioredoxin) in that order (Fig. 3, Plasmid A). Western blotting using specific antibodies revealed a spliced protein (maltose binding protein plus thioredoxin) in addition to the N-protein precursor. As a negative control, no spliced protein was observed when the M-intein gene was absent (Fig. 3, Plasmid B). When the M-intein gene was placed behind the C-protein gene (Fig. 3, Plasmid C), the spliced protein was again observed as it was when the M-intein was in front of the C-protein gene. As expected, the N-protein was produced in excessive amounts, because it was in the first position of the three-gene operon. When the C-protein gene was in the second position and in front of the M-intein gene, a small amount of the C-protein remained unspliced, presumably because of less production of the M-intein from its gene at the third position. However, when the C-protein gene was at the third position and after the M-intein gene, all C-protein was spliced.

**DISCUSSION**

After testing 13 split sites in the Ssp DnaB mini-intein, we have found four functional split sites capable of producing a protein trans-splicing activity. Results of this functional analysis can shed more light on intein structure-function when considered together with the recently reported crystal structure of this intein (6). As illustrated in Fig. 4, the four functional split sites (S1, S6, S7, S8) are all located in loop regions between β-strands, indicating that these loop regions are flexible and can tolerate disruptions. Interestingly, the S6 site corresponds to a location where an extra tongs subdomain was found in the Sce VMA intein crystal structure, and the subdomain appeared to assist the homing endonuclease in DNA binding (8). Therefore, the flexibility and tolerance at the S6 site that were indicated by our findings may have been exploited in nature in the placing of the tongs subdomain. Similarly, the S8 site and the nearby S0 site correspond to a location where the homing endonuclease domain is placed in natural inteins. In contrast to the functional split sites in loop regions, three of the non-functional split sites (S4, S9, and S12) all broke the long β-strands (β5 and β10). This demonstrated the importance of maintaining the integrity of the long β-strands, which structurally form the backbone of the horseshoe-shaped intein. Not all split sites in a loop region produced trans-
splicing activity, indicating that some loop sequences cannot be broken, possibly because they are required to hold the flanking β-strands in correct conformation. Because our intein-splitting methods also added two amino acid residues to each of the split ends, we cannot rule out the possibility that these additional residues might have prevented the trans-splicing function.

For trans-splicing to occur, the two precursor proteins need to recognize and bind each other. The extein sequences, maltose binding protein and thioredoxin, are unlikely contributors to this binding because they are from unrelated proteins not known to interact. Therefore, the two intein fragments must be responsible for bringing about the recognition and binding of the precursor proteins, and for this they need to have sufficient structural interactions. The crystal structure of this Ssp DnaB mini-intein (6), which is very similar to the protein-splicing domain of other inteins (5, 7–9, 11), has several pairs of antiparallel β-strands. The two strands of each pair are expected to interact through various noncovalent bonds, including hydrogen bonds (10). For all of the tested split sites, except S2 and S10, the resulting intein fragments could potentially interact through at least one pair of antiparallel β-strands (Fig. 4). For four of the five functional split sites (S6, S7, S8, and S0), the resulting intein fragments could interact through the ultralong antiparallel β-strands β5/β10. Three of the sites (S6, S8, S0) could interact through one additional antiparallel β-strand (β7/β8 or 9/β10), which may explain the high trans-splicing efficiencies observed with them. The functional split site S1 is perhaps most interesting because it produced an N-intein that is just 11 aa long and could interact with the C-intein through just the short antiparallel β-strands β2/β3. This minimal interaction between the N- and C-inteins may explain the relatively low efficiency of trans-splicing observed with the S1 split site.

We have also demonstrated for the first time that a three-piece split intein could function in protein trans-splicing. For the three-piece split intein to function, the M-intein would need to interact and bind properly with both the small N-intein and the C-intein, which probably involves the short antiparallel β-strands β2/β3 and the long antiparallel β-strands β5/β10, respectively. It is not known whether the M-intein could be reused after trans-splicing one pair of the N- and C-proteins, which would require reversible binding of the M-intein to the N- and C-inteins. Interestingly, the three-piece split intein with both S1 and S0 split sites exhibited more efficient trans-splicing than the two-piece split intein with just the S1 split site. This is not what one would normally expect, because a three-molecule reaction (such as for the three-piece split intein) is in general more difficult than a two-molecule reaction (such as for the two-piece split intein). One possible explanation is that the M-intein folded better as a separate protein compared with when it was a part of the C-intein in the two-piece split intein case.

Our findings have implications for intein evolution. In nature, a split intein could arise from a contiguous intein as a result of genomic rearrangements breaking the intein coding sequence. Considering the large number of naturally occurring split inteins (14, 23, 24), why do they all have a single and common split site? Because we found multiple new split sites compatible with trans-splicing, we can now exclude the possibility that this common split site of natural split inteins is the only site compatible with protein trans-splicing. Our results therefore support the theory that all the known natural split inteins had a single origin from a contiguous intein (24). Our findings have also demonstrated the potential for split inteins of independent origins to have different split sites and for additional intein fragmentation to produce split inteins consisting of three or even more pieces. Highly fragmented inteins could be difficult to recognize in nature. Because intein sequences are poorly conserved, intein recognition has relied on a combination of three criteria: a large intervening sequence in the host protein, multiple and modestly conserved intein sequence motifs, and a few conserved residues at the splice junctions. However, our results showed that the N-intein linked to the N-extein could be as short as 11 aa and that a major portion (M-intein) of the intein could be unlinked from both exteins. In addition, other studies have shown that even the few conserved junction residues can be absent (35–38). Taken together, these findings could suggest the possible existence of highly derived inteins that show no obvious sequence similarity to known self-splicing inteins. This possibility could be reminiscent of the presumed emergence of nuclear spliceosomal introns from self-splicing group II introns, and one may not recognize a spliceosomal intron using criteria for group II introns. The recent report of apparent protein-splicing activity in human (39), which was without a recognizable intein sequence, highlights this possibility, indeed.

The ability to produce artificial split inteins with new and multiple split sites also has implications for intein study and intein use in biotechnology. Previously known split inteins, naturally occurring or artificial, all have a similar split site corresponding to or inside the endonuclease domain. They have shown variability and limitations in solubility, reaction kinetics, and compatibility with certain extein sequences (20, 25–28, 30). New split sites identified in this study that produced intein fragments of various lengths and interacting β-strands can be useful for comparative studies of split intein properties such as fragment binding and reaction kinetics; they can also expand the toolbox of protein trans-splicing in biotechnology. Although these functional split sites were discovered in the Ssp DnaB mini-intein, they most likely can be generalized to other inteins, considering that different inteins showed a similar splicing domain structure consisting of similar β-strands.

REFERENCES
1. Perler, F. B., Davis, E. O., Dean, G. E., Gimble, F. S., Jack, W. E., Noff, N., Noren, C. J., Thorner, J., and Belfort, M. (1994) Nucleic Acids Res. 22, 1125–1127
2. Xu, M. Q., and Perler, F. B. (1996) EMBO J. 15, 5146–5153
3. Perler, F. B., Xu, M. Q., and Paulus, H. (1997) Curr. Opin. Chem. Biol. 1, 292–299
4. Paulus, H. (2000) Annu. Rev. Biochem. 69, 447–496
5. Duan, X., Gimble, F. S., and Quincho, F. A. (1997) Cell 89, 555–564
6. Ding, Y., Xu, M. Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2006) J. Biol. Chem. 278, 36133–36142
