

Protein *Trans*-Splicing of an Atypical Split Intein Showing Structural Flexibility and Cross-Reactivity

Huiling Song^{1,2}, Qing Meng^{1*}, Xiang-Qin Liu^{2*}

1 Institute of Biological Sciences and Biotechnology, Donghua University, Shanghai, People's Republic of China, **2** Department of Biochemistry and Molecular Biology, Dalhousie University, Halifax, Nova Scotia, Canada

Abstract

Inteins catalyze a protein splicing reaction to excise the intein from a precursor protein and join the flanking sequences (exteins) with a peptide bond. In a split intein, the intein fragments (I_N and I_C) can reassemble non-covalently to catalyze a *trans*-splicing reaction that joins the exteins from separate polypeptides. An atypical split intein having a very small I_N and a large I_C is particularly useful for joining synthetic peptides with recombinant proteins, which can be a generally useful method of introducing site-specific chemical labeling or modifications into proteins. However, a large I_C derived from an *Ssp* DnaX intein was found recently to undergo spontaneous C-cleavage, which raised questions regarding its structure-function and ability to *trans*-splice. Here, we show that this I_C could undergo *trans*-splicing in the presence of I_N , and the *trans*-splicing activity completely suppressed the C-cleavage activity. We also found that this I_C could *trans*-splice with small I_N sequences derived from two other inteins, showing a cross-reactivity of this atypical split intein. Furthermore, we found that this I_C could *trans*-splice even when the I_N sequence was embedded in a nearly complete intein sequence, suggesting that the small I_N could project out of the central pocket of the intein to become accessible to the I_C . Overall, these findings uncovered a new atypical split intein that can be valuable for peptide-protein *trans*-splicing, and they also revealed an interesting structural flexibility and cross-reactivity at the active site of this intein.

Citation: Song H, Meng Q, Liu X-Q (2012) Protein *Trans*-Splicing of an Atypical Split Intein Showing Structural Flexibility and Cross-Reactivity. PLoS ONE 7(9): e45355. doi:10.1371/journal.pone.0045355

Editor: Bin Xue, University of South Florida, United States of America

Received: May 15, 2012; **Accepted:** August 20, 2012; **Published:** September 14, 2012

Copyright: © 2012 Song et al. This is an open-access article distributed under the terms of the Creative Commons Attribution License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Funding: This work was supported by a grant to XQL from the National Sciences and Engineering Research Council of Canada and by research grants to QM from the Chinese National High Technology Research and Development Program 863 (NO 2006AA03Z451), the Chinese National Natural Science Foundation of China (NO 30800186 and 31070698), and the Shanghai key projects of basic research (NO 10JC1400300). The funders had no role in study design, data collection and analysis, decision to publish, or preparation of the manuscript.

Competing Interests: The authors have declared that no competing interests exist.

* E-mail: mengqing@dhu.edu.cn (QM); paul.liu@dal.ca (XQL)

Introduction

Inteins are internal protein sequences that can catalyze a protein splicing reaction to excise themselves from a precursor protein and at the same time join the flanking sequences (N- and C-exteins) with a peptide bond [1,2]. The catalytic mechanism of protein splicing typically consists of four steps [2,3]: 1) an N-S or N-O acyl rearrangement at the upstream splicing junction replaces the peptide bond with an ester bond; 2) a transesterification reaction transfers the N-extein from the upstream splice junction to the downstream splice junction; 3) an asparagine cyclization at the C-terminus of the intein breaks the peptide bond, which separates the intein from the exteins; and 4) a S-N or O-N acyl rearrangement forms a peptide bond joining the N- and C-exteins. Inteins can sometimes catalyze protein cleavage at their N- or C-termini when the splicing mechanism is disrupted [3]. For example, when step 1 of the splicing mechanism is blocked or missing, step 3 may still occur and break the peptide bond at the C-terminus of the intein, in what is termed C-cleavage. Whereas different inteins show low levels of similarity at the amino acid sequence level, the crystal structures of different inteins (the protein splicing domain) appear very similar [4,5,6,7]. The splicing domains of inteins are typically shaped like a flattened disk consisting of ~12 β -strands, with the N- and C-terminal parts of the intein folded into a centrally located catalytic pocket.

In a split intein, the intein sequence is broken into two separate fragments, and these intein fragments (I_N and I_C) can reassemble non-covalently to catalyze a protein *trans*-splicing reaction [8,9]. This reaction takes place between two separate polypeptides, with one polypeptide consisting of an N-extein fused to the I_N and the other polypeptide consisting of the I_C fused to a C-extein. During *trans*-splicing, the I_N and I_C are excised, while the N- and C-exteins are concomitantly joined with a peptide bond. Protein-protein *trans*-splicing using split inteins has proven to be a useful technology with a wide range of applications. Examples include production of cytotoxic proteins that cannot be expressed in a single piece [10,11], segmental isotope labeling of proteins for NMR studies [12], and a gene therapy procedure using split genes [13]. For peptide-protein *trans*-splicing, a non-canonical split intein (*Ssp* DnaB S1) consisting of a small I_N and a large I_C has been engineered [14]. This atypical split intein is particularly suitable for splicing synthetic peptides onto the N-terminus of recombinant proteins, because the extremely small I_N (11 aa long) can be more readily produced together with a small N-extein through chemical synthesis. Because the chemically synthesized N-extein can potentially contain any desired chemical group, this atypical split intein has been used successfully in adding fluorescent labels to the N-terminus of recombinant proteins *in vitro* [15] and on the surface of live mammalian cells [16].

Further development of the intein-based method of protein site-specific modification or labeling is of great interest for protein research and engineering. For example, fluorescent or isotope labels can be useful for studying cellular location and trafficking of proteins, and chemical modifications (e.g. unnatural amino acids) can aid studies of protein's structure-function relationship. Standard chemical methods often produce mixed populations of the modified protein, because such methods usually target certain amino acid side chains (thiols, carboxyls, amines) that may exist at multiple locations in the protein [17]. Other methods have been developed for site-specific protein modifications with limited success. Certain polypeptide tags have been used in recombinant proteins to attach a chemical group through an enzymatic reaction (e.g. [18]), but the tag remains incorporated in the modified protein and may interfere with protein function. Specially engineered tRNA charging systems have been used to add unnatural amino acids to proteins during translation [19], but it is difficult or impossible to engineer special tRNA charging systems for every desired unnatural amino acid or chemical modification. Intein-based protein-peptide *trans*-splicing is a newer and potentially more useful method for site-specific protein modifications [20,21,22], because it does not leave a large tag in the modified protein and may be used generally with any chemical moieties on the chemically synthesized extein peptide. To further develop this intein-based method, it is important to find new and atypical split inteins for the peptide-protein *trans*-splicing, because different inteins may exhibit different splicing efficiencies with different exteins [23,24,25,26]. In a recent such study, however, a large C-intein (I_C) derived from the *Ssp* DnaX intein (a natural intein in DnaX protein of *Synechocystis* sp. PCC6803) was found to undergo spontaneous C-cleavage [27], which is unlike the similarly constructed I_C from the *Ssp* DnaB intein (a natural intein in DnaB protein of *Synechocystis* sp. PCC6803). This C-cleavage activity was unexpected, because the large I_C without I_N was thought to have a structural hole in the catalytic pocket of the intein, based on the predicted structure of the intein. This surprise finding raised interesting questions regarding the structure and function of this intein, namely whether it can still catalyze protein *trans*-splicing in the presence of the I_N and, if so, whether its catalytic pocket has unusual structural flexibilities that are not apparent from intein crystal structures.

In this study, we initially found that the unusual I_C of the *Ssp* DnaX intein could undergo protein *trans*-splicing with the small I_N on another protein, and this *trans*-splicing activity completely predominated over the C-cleavage activity. The I_C could also *trans*-splice with small I_N sequences derived from other different inteins, which revealed for the first time a cross-reactivity of atypical split inteins. We also found that the small I_N could be replaced functionally by a nearly complete intein containing the I_N . These findings not only generated a new and second atypical split intein suitable for *trans*-splicing peptides onto the N-terminus of proteins, they also have interesting implications for the structure-function of this atypical intein and perhaps also other inteins. We suggest that the centrally located catalytic pocket of the intein might undergo reversible transitions between an open state for the *trans*-splicing function and a closed state for the C-cleavage function, and this structural flexibility might permit the I_N part of the intein to swing out of the central pocket of the intein.

Results

First we determined whether the 139-aa C-intein (I_C) of the *Ssp* DnaX intein could undergo protein *trans*-splicing when the missing 11-aa N-terminal part (N-intein or I_N) was provided *in trans*,

because the I_C alone had been found to undergo spontaneous C-cleavage [27]. As illustrated in Figure 1A, a maltose binding protein (M) and a thioredoxin protein (T) were used as the N-extein and C-extein, respectively, so that a *trans*-splicing reaction would join these two exteins to form the splicing product MT. As seen in Figure 1B, the splicing product MT was produced both *in vivo* when the two precursor proteins (MI_N and $I_C T$) were co-expressed in *E. coli* and *in vitro* when the purified precursor proteins were incubated together in a test tube.

Interestingly, no C-cleavage activity was detected under these conditions, as indicated by an absence of the cleavage product T. The precursor and product proteins were identified by their predicted sizes and specific recognition of an anti-thioredoxin (anti-T) antibody through Western blotting. For the *in vivo* analysis in *E. coli* cells, only the splicing product MT protein was detected using anti-T antibody, indicating that the precursor protein $I_C T$ had *trans*-spliced completely to form the MT protein. For the *in vitro* analysis, the purified precursor protein MI_N was added in excess to the precursor protein $I_C T$ to drive the *trans*-splicing reaction to greater completion. In producing the $I_C T$ protein alone in *E. coli*, a significant amount of spontaneous C-cleavage occurred as reported previously [27]. In the purification of $I_C T$ using an affinity tag (hexahistidine) contained in the I_C , the cleavage product I_C was co-purified with the remaining $I_C T$ in the purified sample, while the cleavage product T lacked the hexahistidine tag and was absent in the purified sample. The purified $I_C T$ protein did not show new C-cleavage during subsequent co-incubation with the MI_N protein for *trans*-splicing, as indicated by the absence of any new formation of the C-cleavage products I_C and T. Under the *in vitro* conditions used, approximately 85% of the $I_C T$ protein was *trans*-spliced to form the MT protein after 20 hours of incubation at room temperature (Figure 1B). We also tested shorter reaction times and different temperatures (Figure 1C). The efficiency of *trans*-splicing was nearly identical at four tested temperatures (4, 25, and 37°C, and on ice) after 24 hours of reaction. With a shorter reaction time of 15 minutes, the efficiency of *trans*-splicing was a little lower at 4°C and significantly lower on ice.

We then investigated whether the *trans*-splicing reaction could still occur when the small I_N is embedded in a near complete intein. As illustrated in Figure 2A, the intein fragment I_{NL} was designed to contain the N-terminal 144-aa sequence of the 150-aa *Ssp* DnaX intein that lacked the C-terminal 6 aa of the intein, to prevent possible self-cleavage or *cis*-splicing. This I_{NL} was found to *trans*-splice efficiently with I_C *in vitro* at three different temperatures (4, 25, and 37°C), where M and T were the exteins (Figure 2B). Because the 144-aa I_{NL} consists of the small (11-aa) I_N plus other parts of the intein, we asked whether the other parts of the intein also participated in the *trans*-splicing reaction. To answer this question, a double mutation (TXXH to AXXA) was introduced in the Block B motif of the intein, because this conserved intein sequence motif is outside the 11-aa I_N and known to be functionally important in inteins [25]. As seen in Figure 2C, mutating the Block B motif of I_C (resulting in I_{Cm}) destroyed its ability to *trans*-splice with I_N , as expected. In contrast, mutating the Block B motif of I_{NL} (resulting in I_{NLm}) did not affect its ability to *trans*-splice with I_C , indicating that the Block B motif in I_{NL} did not participate in the reaction. We also tested mutated C-intein (I_{Cm}) in a combination with the non-mutated version of I_{NL} , and found that latter could compensate for the mutated Block B motif in the former for the *trans*-splicing reaction. When the Block B motif was mutated in both I_{NL} (resulting in I_{NLm}) and I_C (resulting in I_{Cm}), the *trans*-splicing reaction was abolished, as expected.

To further explore the structural flexibility and versatility of this atypical split intein, we asked whether the large I_C could *trans*-splice

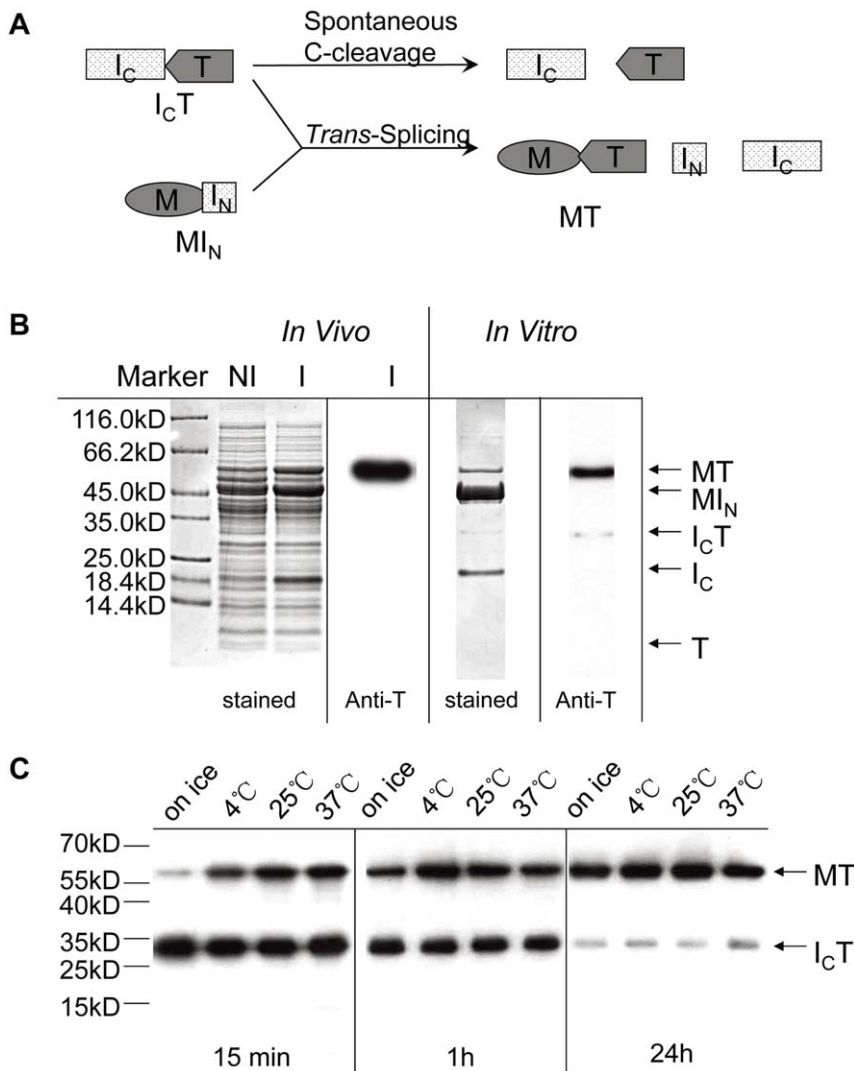


Figure 1. Detection of *trans*-splicing and C-cleavage activities. A. Schematic illustration of the C-cleavage and *trans*-splicing reactions. Precursor protein $I_C T$ is a fusion protein consisting of the 139-aa C-intein (I_C) of the *Ssp* DnaX split intein fused to a thioredoxin protein (T). Precursor protein $M I_N$ is a fusion protein consisting of a maltose binding protein (M) and the 11-aa N-intein (I_N) of the *Ssp* DnaX split intein. B. Experimental analysis of the reactions. For analysis *in vivo*, $M I_N$ and $I_C T$ proteins were co-expressed in *E. coli* cells. Total cellular proteins before (NI) and after (I) the IPTG-induced expression were resolved by SDS-PAGE, and protein bands were visualized either by staining (Coomassie stained) or by Western blotting using an anti-thioredoxin (Anti-T) antibody. For analysis *in vitro*, the $M I_N$ and $I_C T$ proteins were separately produced and purified. These two proteins were then mixed and incubated at room temperature for 20 hours. Reaction products were analyzed and visualized by staining or Western blotting as above. Positions are indicated for the precursor proteins ($M I_N$ and $I_C T$), the splicing product (MT), and the C-cleavage products (I_C and T). Size markers (Marker) are shown on the left. C. Effects of reaction times and temperatures. The *in vitro* reactions were carried out for the specified length of times and at the specified temperatures. Reaction products were analyzed by Western blotting as above.
doi:10.1371/journal.pone.0045355.g001

with small I_N sequences derived from other inteins. The 12-aa I_{NRB} was derived from the N-terminus of the *Rma* DnaB intein (a natural intein in DnaB protein of *Rhodothermus marinus*) [28] that is highly similar to the *Ssp* DnaB intein from which the first atypical split intein was derived [13]. The 12-aa I_{NSG} was derived from the N-terminus of the *Ssp* GyrB intein (a natural intein in GyrB protein of *Synechocystis* sp. PCC6803) [29]. As shown in Figure 3A, the I_{NRB} sequence is 41% identical (58% similar) to the I_N sequence, and the I_{NSG} sequence is 50% identical (75% similar) to the I_N sequence. Under *in vivo* conditions in *E. coli*, both I_{NRB} and I_{NSG} *trans*-spliced efficiently with I_C , as indicated by the accumulation of splicing product MT but not precursor protein $I_C T$ (Figure 3B). Under *in vitro* conditions using purified precursor proteins, I_{NRB} *trans*-spliced with I_C , but I_{NSG} did not. For I_{NRB} , the *in vitro* *trans*-splicing reaction did not go to

completion, with ~60% of the precursor protein $I_C T$ remaining. This was not due to a lesser amount of I_{NRB} , because Coomassie-stained gel pictures showed an excess amount of the precursor protein $M I_{NRB}$ in the reaction (Figure 3B). This may indicate an inefficient use of I_{NRB} under the *in vitro* conditions used, although an efficient use of I_{NRB} was seen in *E. coli* cells. With both I_{NRB} and I_{NSG} , the precursor protein $I_C T$ underwent a small amount of C-cleavage, as indicated by the accumulation of a small amount of the C-cleavage product T.

Discussion

The atypical *Ssp* DnaX intein is found, for the first time, to be capable of protein *trans*-splicing, despite the fact that its large C-intein (I_C) had been known to undergo spontaneous C-cleavage.

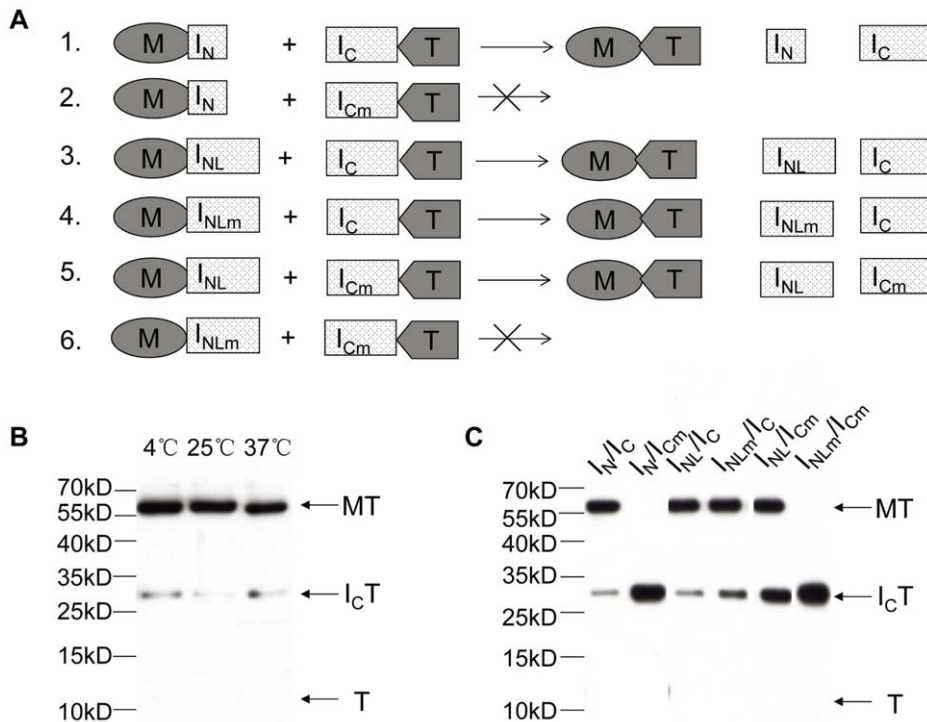


Figure 2. Effects of I_N size and Block B mutation on trans-splicing. A. Schematic illustration of experimental designs. I_{NL} is the N-terminal 144-aa sequence of the *Ssp* DnaX intein. I_{NLm} and I_{Cm} are the same as I_{NL} and I_C , respectively, except that the conserved Block B sequence of the intein was mutated from TXXH to AXXA. Others are the same as in Figure 1A. B. Trans-splicing of I_C T with $M I_{NL}$. A mixture of the two precursor proteins was incubated at the specified temperatures for 20 hours to allow reaction, with the protein bands visualized by Western blotting using anti-T antibody. C. Trans-splicing of different precursor protein pairs listed in panel A. Each pair of precursor proteins was incubated together at room temperature for 20 hours, with the protein bands visualized by Western blotting using anti-T antibody. Names of only intein parts of the precursor proteins are specified on the top of the panel.
doi:10.1371/journal.pone.0045355.g002

This finding has interesting implications on the structure-function of inteins' active site. Previously the I_C part of this intein was found to undergo spontaneous C-cleavage in the absence of I_N [27], which was quite unexpected and unlike other inteins. The highly conserved crystal structures of inteins predict that the N- and C-terminal parts of an intein are located in a central catalytic pocket [5,6,7], as illustrated by a computer modeling of the *Ssp* DnaX intein shown in Figure 4A. The 11-aa I_N sequence forms two small

β -strands named $\beta 1$ and $\beta 2$, with $\beta 1$ being buried deep inside the intein structure. Without I_N , the I_C structure has been predicted to have a structural void (hole) in its catalytic pocket [27], as illustrated in Figure 4C. A similar prediction has also been made for I_C of the *Ssp* DnaB intein [30], where the hole was thought to be a docking place for the I_N to trigger a C-cleavage reaction. To explain why the I_C of *Ssp* DnaX intein (but not of *Ssp* DnaB intein) could undergo spontaneous C-cleavage in the absence of I_N , it was

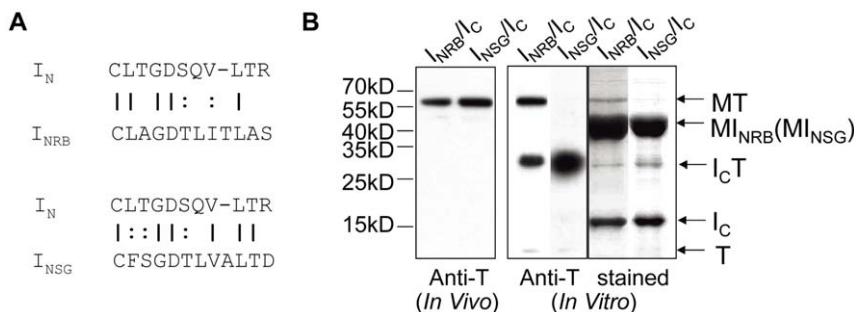


Figure 3. Cross-reactivity of I_C with I_N from different inteins. A. Comparison of the amino acid sequences of I_N from *Ssp* DnaX intein (I_N), *Rma* DnaB intein (I_{NRB}), and *Ssp* GyrB intein (I_{NSG}). Identical and similar residues are marked with a | and a :, respectively. A gap (represented with a -) is introduced in the I_N sequence to maximize the sequence alignment. B. Analysis of trans-splicing reactions between I_C and I_{NRB} (or I_{NSG}) as specified on top. For *in vivo* analysis, $M I_{NRB}$ (or $M I_{NSG}$) and I_C T proteins were co-expressed in *E. coli* cells, and total cellular proteins were analyzed by Western blotting using an anti-thioredoxin (Anti-T) antibody. For *in vitro* analysis, purified $M I_{NRB}$ (or $M I_{NSG}$) and I_C T proteins were co-incubated at room temperature for 20 hours, and the reaction products were analyzed by Western blotting as above. Positions are marked for the precursor I_C T, splicing product MT, and C-cleavage product T. Size markers are shown on the left.
doi:10.1371/journal.pone.0045355.g003

suggested that the predicted hole of this I_C might be sealed or compensated to allow the spontaneous C-cleavage [27]. Our findings in this study show that this I_C can also perform protein *trans*-splicing with I_N , indicating that the predicted hole of I_C is open at least some of the time, in order for the I_N to dock for *trans*-splicing. We suggest that the hole of I_C may exist in two equilibrium states: an open state (hole is open) allowing the docking of I_N for *trans*-splicing, and a closed state (hole is closed) allowing spontaneous C-cleavage without I_N . Our findings also indicate that the open state predominates, because the *trans*-splicing reaction completely suppressed the C-cleavage reaction when I_N was present. Our suggestion is consistent with an earlier study of the *Ssp* DnaB intein, where the horseshoe-like structure of the large I_C was suggested to open up and clamp onto the small I_N [30].

The above suggestion is further supported by our finding that the *trans*-splicing reaction could occur when the small I_N was replaced with I_{NL} , where I_{NL} is a near complete intein containing the I_N . The I_N part of I_{NL} must have participated in the *trans*-splicing reaction, whereas the remaining part of I_{NL} (at least the conserved Block B motif) apparently did not participate in the reaction, because a mutation in the Block B motif of I_{NL} did not prevent *trans*-splicing. To participate in the *trans*-splicing reaction, the I_N part of I_{NL} needs to move out from its buried position in I_{NL} (Figure 4D), before it can dock into the catalytic pocket (open hole) of I_C for *trans*-splicing. This suggests that the I_{NL} structure is also able to open up in order for the I_N part to ‘swing out’ to an exposed position. Consistent with this suggestion, a mutated C-intein (I_{Cm}) *trans*-spliced with I_{NL} , suggesting that the I_{Cm} structure could open up to allow its C-terminal part to “swing out” to an exposed position for participation in the *trans*-splicing reaction. Furthermore, we discovered cross-reactivity between I_C of the atypical *Ssp* DnaX split intein and the small I_N from two other inteins, which is the first finding of cross-reactivity for such atypical split inteins. Considering that the I_N of the other inteins is only 40–50% identical to the native I_N in amino acid sequence, it is interesting that the non-native I_N can correctly dock into the

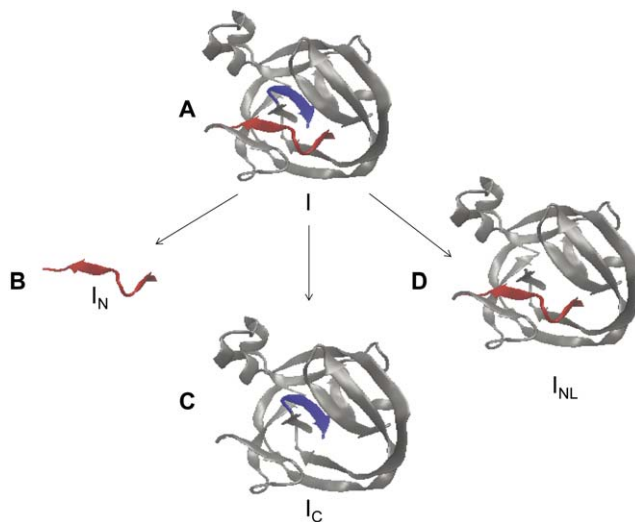


Figure 4. Structural modeling of the intein and its parts. Computer-based modeling of the *Ssp* DnaX intein (A) and its parts (B, C, D) used the VMD program (<http://www.ks.uiuc.edu/Research/vmd/>) and the crystal structure of the *Ssp* DnaB mini-intein [7]. The 11-aa I_N part of the intein is shown in red. The C-terminal 6-aa part of the intein is shown in blue.

doi:10.1371/journal.pone.0045355.g004

structural hole of the I_C and catalyze the *trans*-splicing reaction. Overall, these findings revealed an interesting structural flexibility at or near the catalytic pocket of inteins, which can have significant implications on future engineering of split inteins for peptide-protein or protein-protein *trans*-splicing for various applications.

The *trans*-splicing function of this atypical *Ssp* DnaX split intein also makes a significant addition to the intein-based toolbox for general uses. Previously only the *Ssp* DnaB intein has been engineered into such an atypical split intein, and was named the S1 split intein [14]. Unlike other forms of split inteins, the S1 split intein has an extremely small I_N and is therefore particularly useful for splicing synthetic peptides onto the N-terminus of target proteins. The synthetic peptide can easily accommodate the 11-aa I_N plus a small N-extein to be spliced onto the N-terminus of a target protein, with the target protein being a fusion protein containing the I_C . This peptide-protein *trans*-splicing is useful for site-specific labeling or modifications of proteins, because the synthetic N-extein may be engineered to carry a variety of chemical moieties, including fluorescent groups, modified or unnatural amino acids, and drug molecules, as long as the chemical moiety does not block *trans*-splicing. Finding and understanding new S1 split inteins, as we have done in this study, is important for wide uses of this peptide-protein *trans*-splicing method, because different inteins have been known to splice differently when used on different target proteins [25,31]. It is impressive that this new S1 split intein could perform the *trans*-splicing reaction at temperatures ranging from $\sim 1^\circ\text{C}$ (on ice) to 37°C , although the reaction speed was somewhat lower at $1\text{--}4^\circ\text{C}$ temperatures. This temperature tolerance may be due to the fact that this intein was derived from a natural intein found in a cyanobacterium (*Syemochocystis* sp. PCC6803) that lives under a wide range of environmental temperatures. This robust nature of the S1 split intein can be an advantage in practical applications where one may need to achieve *trans*-splicing under low temperatures. Our finding of cross-reactivity between the I_C of the atypical *Ssp* DnaX split intein and the small I_N from two other inteins also has interesting implications. On the one hand, it permits different choices for the I_N for doing peptide-protein *trans*-splicing and suggests that the I_N sequence may tolerate many sequence changes, which can be useful information for designing and producing synthetic peptides containing I_N . On the other hand, two atypical split inteins may not be used together in a mixed system to achieve labeling or modification of two different target proteins in a protein-specific manner.

Materials and Methods

Plasmid Construction

Plasmid pMSX-S1 for *in vivo* experiments was constructed as described previously [14], in which the two open reading frames expressing the MI_N and I_{CT} proteins were separated by a spacer sequence. A restriction enzyme cutting site *Afl* II was introduced at the split site. To construct plasmid pMSX-S1N expressing the MI_N protein alone, a DNA fragment between *Afl* II and *Hind* III sites was deleted from plasmid pMSX-S1. Plasmid pMSX-S1C expressing the I_{CT} protein alone was from [27]. Plasmid pMSX-S1NL expressing the MI_{NL} protein was constructed by replacing the I_N coding sequence in pMSX-S1N with the first 141 codons of the *Ssp* DnaX intein [32], which used standard recombinant DNA methods including PCR, DNA cutting (*Xho*I-*Afl* II), and ligation. Plasmid pMRB-S1N expressing the MI_{NRB} protein and plasmid pMSG-S1N expressing the MI_{NSG} protein were constructed by replacing the I_N coding sequence in pMSX-S1N with the first 12 codons of the *Rma* DnaB intein [28] and *Ssp* GyrB intein [29],

respectively. To construct pMRBSX-S1, pMSGSX-S1, the I_N coding sequence between XhoI and Afl II in pMSX-S1 was replaced with I_{NRB} or I_{NSG} sequence respectively. To construct plasmid pmMSX-S1C (or pmMSX-S1NL) expressing the $I_{Cm}T$ (or MI_{NLm}) protein, site-directed mutations were introduced into plasmid pMSX-S1C (or pMSX-S1NL), using a standard method of inverse PCR.

Protein Expression, Purification and in vitro Reactions

Plasmids pMSX-S1, pMSX-S1N, pMSX-S1NL, pMRB-S1N, pMSG-S1N and pmMSX-S1NL were each transformed into *Escherichia coli* DH5 α cells, while plasmids pMSX-S1C and pmMSX-S1C were each transformed into *Escherichia coli* BL21(DE3) cells, all using a standard *E. coli* transformation protocol. The transformed *E. coli* cells were grown in 50 mL of Luria Broth (LB) medium at 37°C to mid-logarithmic phase (OD₆₀₀ of ~0.6) and induced by 0.8 mM IPTG to express, at room temperature overnight, the plasmid-encoded protein(s) of interest. To analyze total cellular proteins, cells were harvested by centrifugation and solubilized directly in SDS loading buffer. For protein purification, harvested cells were lysed using a French Press (14,000 PSI), and the cell lysate was centrifuged to remove any insoluble materials. To purify the MI_N , MI_{NL} , and MI_{NLm} proteins containing the maltose binding protein, amylose resin was used according to the manufacturer's instructions (New England Biolabs). To purify the I_{CT} and $I_{Cm}T$ proteins containing a hexahistidine tag, Ni-NTA resin (QIAGEN) was used according to

the manufacturer's instructions. For *in vitro* trans-splicing or cleavage reactions, the specified precursor proteins were mixed and incubated under specified conditions, with 1 mM DTT added to all *in vitro* reactions. Western blotting used an anti-thioredoxin (anti-T) antibody (Invitrogen) and the Enhanced Chem-luminescence detection kit (GE Healthcare), all according to the manufacturer's instructions.

Computational Simulations of Intein Structure

A simulated three-dimensional structure of *Ssp* DanX mini-intein was obtained by using the fully automated homology-modeling pipeline SWISS-MODEL [33,34]. The intein amino acid sequence [27] was uploaded to the Automatic Modeling Workspace (<http://swissmodel.expasy.org/workspace>), the crystal structure of *Ssp* DnaB mini-intein was selected automatically as a closest template by homology, and structural models of *Ssp* DanX mini-intein were generated and presented in NewCartoon style.

Acknowledgments

We thank Dr. Xingmei Qi for providing the plasmid pMSX-S1C and Dr. David Spencer for a critical reading of this manuscript.

Author Contributions

Conceived and designed the experiments: XQL. Performed the experiments: HS. Analyzed the data: HS QM XQL. Contributed reagents/materials/analysis tools: QM XQL. Wrote the paper: HS XQL.

References

- Perler FB, Davis EO, Dean GE, Gimble FS, Jack WE, et al. (1994) Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucleic Acids Res* 22: 1125–1127.
- Perler FB (2005) Protein splicing mechanisms and applications. *IUBMB Life* 57: 469–476.
- Xu MQ, Perler FB (1996) The mechanism of protein splicing and its modulation by mutation. *Embo J* 15: 5146–5153.
- Klabunde T, Sharma S, Telenti A, Jacobs WR, Jr., Sacchettini JC (1998) Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nat Struct Biol* 5: 31–36.
- Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, Morikawa K (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. *J Mol Biol* 300: 889–901.
- Hall TM, Porter JA, Young KE, Koonin EV, Beachy PA, et al. (1997) Crystal structure of a Hedgehog autoprocessing domain: homology between Hedgehog and self-splicing proteins. *Cell* 91: 85–97.
- Ding Y, Xu MQ, Ghosh I, Chen X, Ferrandon S, et al. (2003) Crystal structure of a mini-intein reveals a conserved catalytic module involved in side chain cyclization of asparagine during protein splicing. *J Biol Chem* 278: 39133–39142.
- Wu H, Hu Z, Liu XQ (1998) Protein trans-splicing by a split intein encoded in a split DnaE gene of *Synechocystis* sp. PCC6803. *Proc Natl Acad Sci U S A* 95: 9226–9231.
- Evans TC, Jr., Martin D, Kolly R, Panne D, Sun L, 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.
- Evans TC, Jr., Benner J, Xu MQ (1998) Semisynthesis of cytotoxic proteins using a modified protein splicing element. *Protein Sci* 7: 2256–2264.
- Wu W, Wood DW, Belfort G, Derbyshire V, Belfort M (2002) Intein-mediated purification of cytotoxic endonuclease I-TevI by insertional inactivation and pH-controllable splicing. *Nucleic Acids Res* 30: 4864–4871.
- Toshio Yamazaki TO, Natsuko Oda, Yoshimasa Kyogoku, Koichi Uegaki, Nobutoshi Ito, Yoshizumi Ishino, Haruki Nakamura (1998) Segmental isotope labeling for protein NMR using peptide splicing. *Journal of the American Chemical Society* 120: 2.
- Li J, Sun W, Wang B, Xiao X, Liu XQ (2008) Protein trans-splicing as a means for viral vector-mediated in vivo gene therapy. *Hum Gene Ther* 19: 958–964.
- Sun W, Yang J, Liu XQ (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing. *J Biol Chem* 279: 35281–35286.
- Ludwig C, Pfeiff M, Linne U, Mootz HD (2006) Ligation of a synthetic peptide to the N terminus of a recombinant protein using semisynthetic protein trans-splicing. *Angew Chem Int Ed Engl* 45: 5218–5221.
- Ando T, Tsukiji S, Tanaka T, Nagamune T (2007) Construction of a small-molecule-integrated semisynthetic split intein for in vivo protein ligation. *Chem Commun (Camb)*: 4995–4997.
- Gilmore JM, Scheck RA, Esser-Kahn AP, Joshi NS, Francis MB (2006) N-terminal protein modification through a biomimetic transamination reaction. *Angew Chem Int Ed Engl* 45: 5307–5311.
- Lin CW, Ting AY (2006) Transglutaminase-catalyzed site-specific conjugation of small-molecule probes to proteins in vitro and on the surface of living cells. *J Am Chem Soc* 128: 4542–4543.
- Xie J, Schultz PG (2006) A chemical toolkit for proteins—an expanded genetic code. *Nat Rev Mol Cell Biol* 7: 775–782.
- Muir TW (2003) Semisynthesis of proteins by expressed protein ligation. *Annu Rev Biochem* 72: 249–289.
- Tan LP, Yao SQ (2005) Intein-mediated, in vitro and in vivo protein modifications with small molecules. *Protein Pept Lett* 12: 769–775.
- Kalia J, Raines RT (2006) Reactivity of intein thioesters: appending a functional group to a protein. *ChemBiochem* 7: 1375–1383.
- Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M (2009) Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci U S A* 106: 11005–11010.
- Iwai H, Zuger S, Jin J, Tam PH (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett* 580: 1853–1858.
- Noren CJ, Wang J, Perler FB (2000) Dissecting the Chemistry of Protein Splicing and Its Applications. *Angew Chem Int Ed Engl* 39: 450–466.
- Appleby-Tagoe JH, Thiel IV, Wang Y, Wang Y, Mootz HD, et al. (2011) Highly efficient and more general cis- and trans-splicing inteins through sequential directed evolution. *J Biol Chem* 286: 34440–34447.
- Qi X, Meng Q, Liu XQ (2011) Spontaneous C-cleavage of a mini-intein without its conserved N-terminal motif A. *FEBS Lett* 585: 2513–2518.
- Liu XQ, Hu Z (1997) A DnaB intein in *Rhodothermus marinus*: indication of recent intein homing across remotely related organisms. *Proc Natl Acad Sci U S A* 94: 7851–7856.
- Gorbalenya AE (1998) Non-canonical inteins. *Nucleic Acids Res* 26: 1741–1748.
- Volkman G, Sun W, Liu XQ (2009) Controllable protein cleavages through intein fragment complementation. *Protein Sci* 18: 2393–2402.
- Paulus H (2001) Inteins as enzymes. *Bioorg Chem* 29: 119–129.
- Liu XQ, Hu Z (1997) Identification and characterization of a cyanobacterial DnaX intein. *FEBS Lett* 408: 311–314.
- Kiefer F, Arnold K, Kunzli M, Bordoli L, Schwede T (2009) The SWISS-MODEL Repository and associated resources. *Nucleic Acids Res* 37: D387–392.
- Kopp J, Schwede T (2004) The SWISS-MODEL Repository of annotated three-dimensional protein structure homology models. *Nucleic Acids Res* 32: D230–234.