Trans Protein Splicing of Cyanobacterial Split Inteins in Endogenous and Exogenous Combinations

Bareket Dassa,[‡] Gil Amitai,[‡] Jonathan Caspi,[‡] Ora Schueler-Furman,[§] and Shmuel Pietrokovski*[,‡]

Department of Molecular Genetics, The Weizmann Institute of Science, Rehovot 76100, Israel, and Department of Molecular Genetics and Biotechnology, Hadassah Medical School, Hebrew University, Jerusalem 91120, Israel

Received June 13, 2006; Revised Manuscript Received November 6, 2006

ABSTRACT: Inteins are autocatalytic protein domains that post-translationally excise from protein precursors and ligate their flanking regions with a peptide bond, in a process called protein splicing. Intein-containing DNA polymerases of cyanobacteria and nanoarchaea are naturally split into two separate genes at their intein domain. Such naturally occurring split inteins rapidly self-associate and reconstitute protein-splicing activity in trans. Here, we analyze the *in vitro* protein-splicing activity of three naturally split inteins from diverse cyanobacteria: *Oscillatoria limnetica*, *Thermosynechococcus vulcanus*, and *Nostoc* sp. PCC7120. N- and C-terminal halves of these split inteins were mixed in nine combinations, resulting in three endogenous (wild-type) and six exogenous combinations. Protein splicing was detected in all split-intein combinations, despite a 30−50% sequence variation between the homologous proteins. Splicing activity proceeded under a variety of conditions, including the presence of denaturants and reductants and high temperature, ionic strength, and viscosity. Still, in a high concentration of salt (2 M) or urea (6 M), specific combinations spliced significantly better than others. Additionally, copper ions were found to inhibit trans splicing in a reversible double-lock reaction. Our comparative analysis of naturally split inteins in endogenous and exogenous combinations demonstrates the modularity of trans protein-splicing elements and their robust activity. It suggests tight interactions between split-intein halves and conditions for modifying the specificity of intein parts. These results promote the biotechnological use of split inteins for controlled assembly of protein fragments either *in vivo* or *in vitro* and under moderate or extreme conditions.

An intein is an autocatalytic protein domain that post-translationally excises from its protein precursor and ligates its flanking regions with a peptide bond, in a process called protein splicing (*1*). An intein-containing gene can be split in the intein domain into two separate functional genes. This rare genetic event occurred in the ancestor of the DNA polymerase III α subunit (DnaE) of diverse cyanobacteria (*2*, *3*) and in the B-type DNA polymerase (Pol) gene of the archaeon *Nanoarchaeum equitans* (*4*, *5*). These cyanobacterial and nanoarchaeal split inteins are distinct from each other and arose in independent events. The inteins are split into separately inactive, short halves. The N-intein (98−123 amino acids) and C-intein (30−36 amino acids) parts have a high affinity for each other and can rapidly self-associate into a compact intein fold (*6*−*10*). The split-intein complex has a protein-splicing activity in trans, autocatalytically ligating its flanks into a functional protein (Figure 1A).

The best studied naturally occurring split intein is that of S*ynechocystis* sp. PCC6803 cyanobacterium. Co-incubation of its split-intein halves generates both splicing and cleavage (N- or C-terminal) products, together with stable complexes of noncovalently, tightly interacting intein intermediates (*7*). Kinetic studies of self-association and splicing rates of split inteins suggest that trans protein splicing is rapid [$(6.6 \pm 1.3) \times 10^{-5}$ s^{-1}] and occurs at a low nanomolar affinity (*6*−*8*). Split inteins were found to interact with a very fast association rate (2.8×10^7 M^{-1} s^{-1}) and a slow dissociation rate (1.2 s^{-1}) (*9*). The crystal structure of artificially joined split-intein halves shows a high structural similarity to the compact intein fold and its active site (*10*). Recently, trans-splicing activity was shown *in vivo* for split inteins from the cyanobacteria *Nostoc* sp. PCC7120 (Nsp)[1] (*11*) and *Nostoc punctiforme* (*12*) and from the archaeon *N. equitans* (*5*).

Inteins and naturally split inteins are utilized for various applications of protein ligation, protein cyclization, and protein−protein interactions *in vivo* and *in vitro* (*13*). This is due to the robustness, rapidness, and autocatalytic processing of protein splicing. *In vitro* protein splicing can be inhibited by zinc ions and, to lesser extent, by cadmium ions (*14*, *15*). These are the only known protein-splicing inhibitors, and their effect can be reversed by metal chelators, such as ethylenediaminetetraacetic acid (EDTA). Another advantage in using inteins is their modular activity, which is maintained even when inteins are integrated in foreign

* To whom correspondence should be addressed: Weizmann Institute of Science, Rehovot 76100, Israel. Telephone: 972-8-9342747. Fax: 972-8-9344108. E-mail: shmuel.pietrokovski@weizmann.ac.il.

[‡] The Weizmann Institute of Science.

[§] Hebrew University.

[1] Abbreviations: Tvu, *Thermosynechococcus vulcanus*; Oli, *Oscillatoria limnetica*; Nsp, *Nostoc* species PCC7120; MBP, maltose-binding protein; CBD, chitin-binding domain; EDTA, ethylenediaminetetraacetic acid; DTT, dithiothreitol; TCEP, tris(2-carboxyethyl)phosphine hydrochloride.

10.1021/bi0611762 CCC: $37.00
Published on Web 01/02/2007

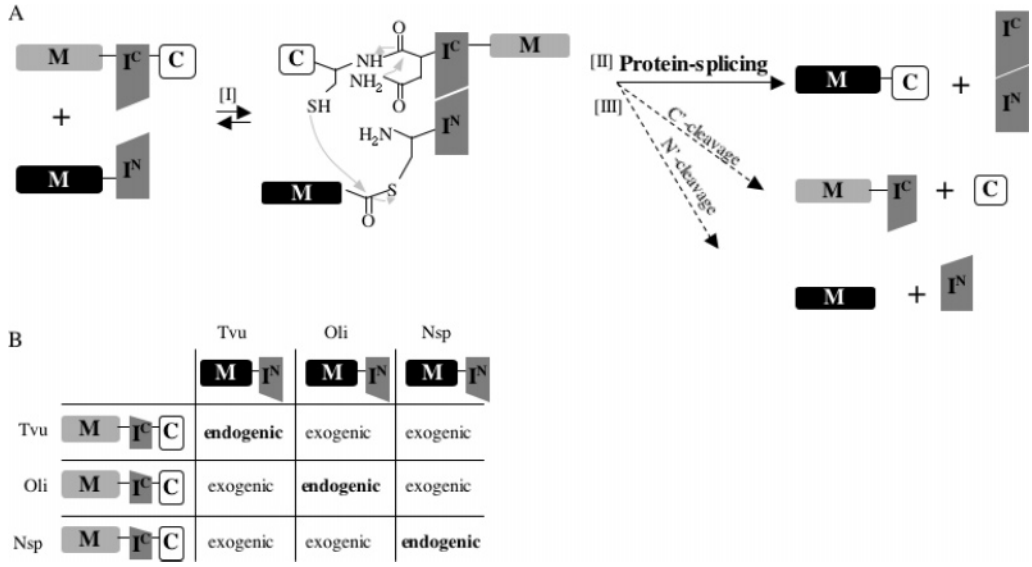


FIGURE 1: (A) Schematic presentation of split-intein constructs used in this work and their trans-splicing products. (I) The two precursors of intein halves (MI^N and MI^CC) associate into a complex and undergo bond rearrangements. (II) During the protein-splicing reaction, the N−S acyl shift at the N-terminal intein yields a reactive thioester bond, which is attacked by a conserved cysteine at the C-terminal intein, followed by cyclization of asparagine. (III) C-Terminal cleavage can result from the cyclization of asparagine at the I^C intein half, and N-terminal cleavage can result from thioester hydrolysis at the I^N intein half. (B) Matrix of all tested combinations of cyanobacterial split-intein halves in endogenous or exogenous mixtures. I^N is the N-terminal intein half; I^C is the C-terminal intein half; M is the maltose-binding protein; and C is the chitin-binding domain. Cyanobacterial species used in this work are Oli, Tvu, and Nsp.

contexts (nevertheless, this can require the addition of a few native flanking residues, e.g., see ref *16*).

Trans protein splicing can be achieved by artificially splitting full-length inteins into two trans-splicing fragments (*17*−*24*). While naturally split inteins can protein-splice in relatively moderate conditions, many artificial split inteins function better under denaturing conditions (*17*, *23*) or with additional heterodimerization domains (*22*).

Naturally split-intein halves of each species are under a selective pressure to co-evolve for an efficient interaction with each other and for splicing their flanking host protein. Split inteins in diverse cyanobacterial groups accumulated natural sequence variations, which are hypothesized to be either (i) neutral changes not affecting their binding and splicing or (ii) changes possibly reflecting a functional conservation. The latter may optimize the activity of split-intein halves to best fit with each other under the specific living conditions of their host species (e.g., high temperature) or under a common sequence drift of both halves. The autocatalytic and bipartite nature of trans splicing can be used to test the hypotheses on the divergence of natural split inteins. Trans-splicing activity can be tested in different combinations of wild-type N- and C-intein halves. This approach is distinguished from previous studies, which used point mutations (*6*), deletions (*15*), or fusion (*9*, *10*) of split inteins.

In this work, we examined the trans-splicing activity of naturally occurring split inteins from three diverse cyanobacteria. Nine possible pairs of N- and C-intein halves were mixed *in vitro*, forming three endogenous (wild-type) combinations and six exogenous combinations of split inteins. The protein-splicing activity of all split inteins was monitored under diverse conditions, including high salt concentrations, the presence of denaturing agents, different temperatures, and the presence of divalent metal ions. Under the latter condition, we identified copper ions as reversible inhibitors of the trans-splicing reaction. Our comparative analysis of splicing activity among all split-intein combinations under different conditions found their activity to be modular and thus potentially useful for engineering and manipulating proteins.

EXPERIMENTAL PROCEDURES

Bacterial Strains and Constructs. Oscillatoria limnetica (Oli) strain "Solar Lake" genomic DNA was provided by E. Padan (Hebrew University, Israel). *Thermosynechococcus vulcanus* (Tvu) genomic DNA was provided by N. Adir (Technion, Israel). Nsp was provided by S. Boussiba (Ben-Gurion University, Israel). Each N-terminal intein domain (I^N) was cloned downstream to the maltose-binding domain (MBP) to enhance its soluble expression and flanked by five native residues KFAEY (except for I^N-Tvu, which was flanked by a serine), and each C-terminal intein domain (I^C) was cloned downstream to the MBP and upstream to the chitin-binding domain (CBD), together with five or six C-terminal flanking residues (CFNKS or CFNKSH), which enhance its activity (*16*). Both I^N and I^C domains were cloned with linkers separating them from the protein tags, to allow for their structural flexibility.

Protein Expression and Purification. Escherichia coli TB1 bacteria were transformed with either I^N or I^C clones and grown at 30−37 °C as described in ref *25*. Overexpressed proteins containing the MBP tag were extracted and affinity-purified using amylose beads as described in ref *25*. The protein extraction buffer contained 20 mM Tris-HCl at pH 7.4, 200 mM NaCl, 10 mM EDTA, and 1 mM Na−azide. In some duplicate experiments, Tris-HCl was replaced with 20 mM Na−phosphate without any significant effect on the results. The protein concentration was measured by the Bradford method, following visualization of calibrated samples on sodium dodecyl sulfate−polyacrylamide gel electrophoresis (SDS−PAGE).

Trans Protein-Splicing Assay. In the standard assay, amylose-purified split inteins (I^N or I^C) were mixed in the same molar concentrations (0.4 mg/mL) in the protein extraction buffer and incubated at room temperature for the tested time periods (0, 0.16, 0.5, 1, 3, 12, and 24 h). The splicing activity of endogenous combinations was examined in increasing ionic strength (0.2, 1, 2, and 3 M NaCl), at different temperatures (4, 25, 55, and 90 °C), in the presence of glycerol (40%), urea (3, 4, and 6 M), GuHCl (1 and 4 M), dithiothreitol (DTT) (0.2 M), 2-(*N*-morpholino)ethanesulfonic acid (MES) buffer (pH 5.5), or Tween-20 (1%). In the metal-ion inhibition experiments, EDTA was not added to the extraction buffer but either $CuSO_4$, $CuCl_2$, $ZnSO_4$, or $ZnCl_2$ (5 mM) were added and the reaction proceeded at room temperature. For chelation of metal ions, 10 mM EDTA was added and reduction was performed with 2.5 mM tris-(2-carboxyethyl)phosphine hydrochloride (TCEP) or 0.25 M DTT. The extent of the reactions was measured in different time periods, ranging from 1 min to 24 h. The minimal time at room temperature to clearly observe splicing products on SDS−PAGE was 30 min. We thus chose these temperatures and incubation times to test other reaction conditions.

Protein-Splicing Measurement. Amounts of protein-splicing products were quantified by band densitometry (using NIH image and ImageJ softwares) of Coomassie-stained gels. Splicing efficiency was calculated as the fraction of the splicing product (MBP−CBD) divided by half of the precursors and partial cleavage products.

Creation of Structural Models for the Different Intein Combinations. Backbone coordinates and side-chain coordinates wherever the amino acid was conserved, from the structure of the Ssp6803 split intein, Protein Data Bank (PDB) code 1ZD7 (*10*), were used as a template, upon which the split-intein sequences of the different species combinations were modeled, using a simplified Rosetta protocol (*36*, *37*). In brief, only side-chain atoms were remodeled, with no attempt to model backbone conformational changes. The C termini of the I^N parts, which are not conserved and of different lengths, were not included. This simple modeling procedure created for each of the combinations a well-packed structure that does not accumulate any steric clashes, indicating that those sequences fit onto the fold. The models were then analyzed for energy contributions across the interface.

RESULTS

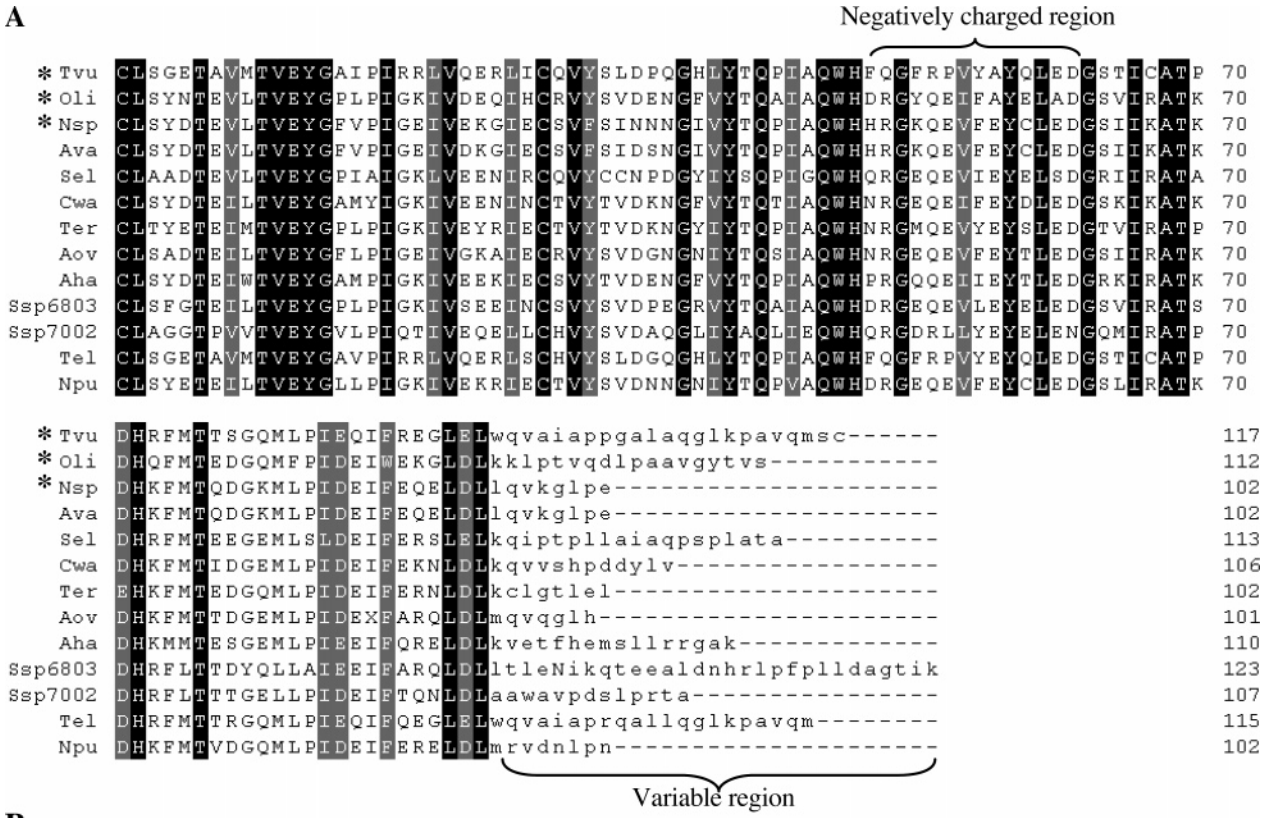
Sequence Features of Cyanobacterial Split Inteins. We experimentally examined three cyanobacterial split-intein pairs, formerly identified by Caspi et al., from Oli, Tvu, and Nsp (*3*). These species belong to distinct cyanobacterial phylogenetic groups (*27*) from different environments (sulfide-rich, hypersaline lake, hot spring, or freshwater). To date, 13 naturally split inteins have been identified in different cyanobacteria species (parts A and B of Figure 2). All known cyanobacterial split inteins are integrated in the same highly conserved position in the DnaE protein and are predicted to be functional, because their catalytic residues are identical to *Synechocystis* sp. PCC6803 and Nsp split inteins, which were previously demonstrated to be functional (*2*, *11*). These inteins also include the six conserved sequence motifs that define the intein fold (*29*−*31*). However, the C-terminal region of each N-terminal intein half is variable in sequence and length and is not conserved among the cyanobacterial split inteins (Figure 2A). Moreover, the selected split inteins in this work share only 52−68% protein sequence identity with each other, thus containing notable natural sequence variations in both intein halves (Figure 2C).

Heterologous Expression and Purification of Three Split Inteins. To demonstrate the protein-splicing activity of the three examined split inteins, their N- or C-terminal halves were each fused to protein tags, overexpressed in *E. coli*, and purified using affinity columns (Figure 1A and the Experimental Procedures). Each N-terminal intein domain (I^N) was cloned downstream to a MBP, and each C-terminal intein domain (I^C) was cloned between an upstream MBP, to enhance its solubility (*32*), and a downstream CBD. It was previously demonstrated that split inteins can splice in non-native protein contexts and self-associate with only five native flanking residues (*16*, *17*). Specifically, the ability of split-intein halves to fold, interact, and splice independently of their foreign flanks was previously studied using MBP or CBD tags and implies that these characteristics are due to the intein fragments and not their extein flanking sequence. Our overexpression of I^C without the MBP tag resulted in an insoluble protein (data not shown), as was previously reported for several other inteins (*17*, *18*). Overexpression of all three I^C domains with MBP resulted in small amounts of this tag, which may result from a premature end of transcription or translation (Figure 3A) (*33*).

Trans Protein Splicing of Endogenous Split Inteins. Pairs of purified split-intein halves (I^N and I^C), each pair half from one of the three examined species, were mixed *in vitro* in equal molar ratios (see the Experimental Procedures). These three endogenous combinations of split inteins had a protein-splicing activity (Figure 3). The expected protein-splicing product (MBP−CBD, 50 kD) and partial N- or C-terminal cleavage products (MBP, 43 kD; MBP−I^C, 47 kD) appeared on SDS−PAGE only when the two intein halves were mixed together. The spliced product MBP−CBD migrated on gel identically to a control fusion protein, which was previously described and verified by mass spectrometry (*25*) (Figure S2 in the Supporting Information). Product identities were also verified by Western blots using antibodies against MBP or CBD and by affinity purification using either amylose or chitin beads (not shown).

Splicing products were observed within a few minutes of incubation, and their amount increased with incubation time, reaching a maximal level after 24 h. Splicing efficiencies were estimated as 45% for Oli split intein, 36% for Tvu, and 15% for Nsp. It is notable that the protein-splicing activity of the Nsp split intein was 2−3-fold lower than those of the Tvu and Oli inteins under all examined conditions (Figure 4). The activity of the endogenous Nsp pair did not increase upon overnight incubation, even when doubling or tripling the relative amount of I^C (not shown).

Protein-Splicing Activity of Endogenous versus Exogenous Split-Intein Combinations. To study the effect of natural sequence variations on the trans protein-splicing activity of different split inteins, each of the three I^N domains (I^N-Nsp, I^N-Oli, and I^N-Tvu) was mixed with each of the I^C domains (I^C-Nsp, I^C-Oli, and I^C-Tvu). This resulted in nine different combinations of split-intein pairs (Figure 1B). Three of these combinations are wild-type (endogenous) pairs, whose activ-



C

Compared sequences	% Sequence identity (similarity)
I^N-Tvu, I^N-Nsp	53 (73)
I^N-Tvu, I^N-Oli	55 (77)
I^N-Nsp, I^N-Oli	68 (85)
I^C-Tvu, I^C-Nsp	52 (72)
I^C-Tvu, I^C-Oli	63 (80)
I^C-Nsp. I^C-Oli	64 (81)

FIGURE 2: Sequence conservation of N-terminal (A) and C-terminal (B) halves of cyanobacterial naturally split inteins. Identical residues are in black, and similar residues are in gray. Proteins examined in this work are marked with an asterisk. The C-terminal regions of the N-terminal intein halves, shown in lower case, are dissimilar and not aligned to each other. (C) Sequence identity and similarity between the split inteins examined in this work, along the intein domains. Tvu, *Thermosynechococcus vulcanus*; Tel, *Thermosynechococcus elongates* BP-1; Ava, *Anabaena variabilis* ATCC29413; Nsp, *Nostoc* species PCC7120; Aov, *Aphanizomenon ovalisporum*; Npu, *Nostoc punctiforme* ATCC29133; Cwa, *Crocosphaera watsonii* WH8501; Aha, *Aphanothece halophytica*; Ter, *Trichodesmium erythraeum* IMS101; Oli, *Oscillatoria limnetica*; Ssp6803, *Synechocystis* species PCC6803; Sel, *Synechococcus elongatus* PCC7942 (*Synechococcus* species PCC6301 is a very similar strain of PCC7942 with almost identical DnaE genes that are thus not shown); Ssp7002, *Synechococcus* species PCC7002.

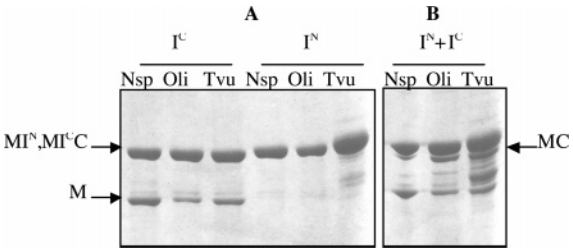


FIGURE 3: Purification and trans protein splicing of split-intein halves. (A) Affinity-purified split-intein halves from Oli, Tvu, and Nsp cyanobacterial species. (B) Trans protein splicing of endogenous split-intein halves. Protein precursors (MI^N and MI^CC) were purified using amylose beads and incubated *in vitro* in equal molar ratios for 30 min at room temperature. Proteins were separated on 10% SDS−PAGE and visualized with Coomassie Blue staining. I^N is the N-terminal intein half; I^C is the C-terminal intein half; M is the maltose-binding protein; and C is the chitin-binding domain.

ity was described above. Six combinations are exogenous split-intein halves, which do not occur naturally together. The protein-splicing activity of each combination was quantified and compared among the nine combinations after 30 min (Figure 4A). In general, all pairs of split inteins had protein-splicing activity, together with cleavage products. Interestingly, the activity of endogenous split-intein combinations was not significantly higher than that of exogenous combinations. Additionally, it is notable that all three combinations containing the I^C-Nsp had the lowest amounts of splicing products under all examined conditions (three right lanes in Figure 4A) and all three combinations with I^C-Tvu resulted with a cleavage product, MI^C, which migrated differently on gel (three left lanes in Figure 4A). MI^C migrated as a doublet that was identified by Western blot using anti-MBP antibodies (data not shown). These effects were not observed in combinations sharing a common I^N intein half.

Protein Splicing in Extreme Conditions. Protein-splicing activity was examined in extreme conditions to compare the efficiency of all nine combinations (Figures 4B and 5). The chosen conditions were derived from the natural environment of the species hosting the tested split inteins and from conditions affecting protein−protein interactions. Tvu is a thermophile, growing optimally at 55 °C; therefore, splicing activity was tested in 4, 25, and 55 °C. Oli is a moderate halophile, growing optimally at 1−2.5 M NaCl, which inspired us to test the splicing activity at high ionic strength (2 M NaCl), although this does not reflect the salt concentration in the cyanobacterial cytoplasm. Additional effectors of protein−protein interactions included the presence of denaturing agents (up to 6 M urea or 1 M GuHCl), which disrupt hydrogen bonds and hydrophobic interactions, elevated viscosity (40% glycerol or 1% Tween-20), and the presence of a reducing agent (0.2 M DTT), which breaks any disulphide bridges and induces N-terminal cleavage in inteins. All of the above conditions were sampled from preliminary experiments with the endogenous combinations, which were tested in different concentrations or time courses (see the Experimental Procedures).

Protein splicing proceeded under these extreme conditions, although it was less efficient compared to splicing under moderate conditions. Protein-splicing and cleavage reactions were totally inhibited only upon incubation in extreme denaturing conditions (4 M GuHCl) or in an acidic environment (pH 5.5) for all combinations. Additionally, incubation at temperatures higher than 60 °C resulted in aggregation of the proteins, possibly because of denaturation of the fusion proteins.

Significant differences in the protein-splicing activity of several split-intein combinations were observed under specific conditions (Figure 5). In the presence of a high salt concentration (2 M NaCl), the splicing activity of two combinations, namely, an endogenous combination I^N-Oli/I^C-Oli and an exogenous combination I^N-Nsp/I^C-Oli, was barely affected by the addition of salt. Their splicing levels were at least 2-fold larger than those of the other combinations (Figure 5A). Differential accumulation of protein-splicing products among the different combinations was optimally observed on SDS−PAGE when limiting the splicing reaction to 30 min.

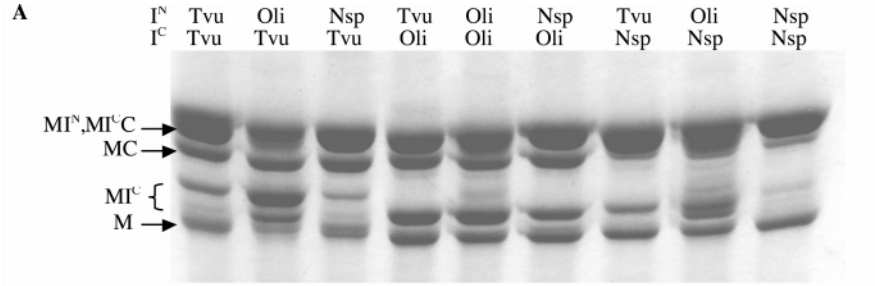
Another significant difference in splicing activity was observed between the endogenous combination of Nsp (I^N-Nsp/I^C-Nsp) and two of its exogenous variations (I^N-Tvu/I^C-Nsp and I^N-Oli/I^C-Nsp) in high urea concentrations (Figure 5B). The endogenous combination of Nsp underwent protein splicing in the presence of up to 6 M urea, but its exogenous combinations (which included the I^C-Nsp half) did so to a much lesser extent. For comparison, other exogenous combinations that included the I^N-Nsp half (e.g., I^N-Nsp/I^C-Oli and I^N-Nsp/I^C-Tvu) showed protein-splicing activity similar to their endogenous combinations under the same conditions (Figure 4B). Furthermore, the protein-splicing activity of the Nsp endogenous combination was slightly improved in the presence of urea.

Reversible Inhibition of Protein Splicing by Copper Ions. We found that the addition of divalent copper ions ($CuSO_4$ or $CuCl_2$) inhibited trans protein splicing and cleavage of all split-inteins combinations (Figure 6) to the same extent as zinc ion ($ZnSO_4$ or $ZnCl_2$) inhibition (not shown). While the effect of zinc ions is known to be reversed by chelating the ions with EDTA (*14*, *15*), copper inhibition was not reversible under the same conditions. Splicing activity was restored only upon the addition of a reducing agent, TCEP, together with EDTA. Use of DTT as a reducing agent resulted in the aggregation of the proteins, possibly because of the formation of metal−protein complexes.

DISCUSSION

We present an experimental system that compares the cross-activity of three naturally split inteins mixed in endogenous (wild-type) and exogenous combinations, examined under both moderate and extreme conditions. Protein splicing of all nine split-intein combinations was robust to their intrinsic sequence variations and to their diverse phylogenetic origin. The comparable splicing efficiencies of the endogenous and exogenous combinations under moderate conditions indicate that a proper and active intein fold was formed in all cases. The general approach in this study compares the activity of different split inteins when flanked with the same protein tags and uniformly expressed, thus examining only natural sequence variations of the intein domain itself, abolishing a possible bias of the heterologous protein expression. The fact that exogenous combinations had a splicing activity at all suggests that split-intein halves did not accumulate functional changes and were under a

A



B

Limiting factor	Condition	I^N: / I^C:	*Oli* / *Oli*	*Tvu* / *Oli*	*Nsp* / *Oli*	*Oli* / *Tvu*	*Tvu* / *Tvu*	*Nsp* / *Tvu*	*Oli* / *Nsp*	*Tvu* / *Nsp*	*Nsp* / *Nsp*
Time	0.5 or 16 hr		+++	+++	+++	+++	+++	+++	++	++	++
Temperature	4, 25 or 55 C; 24h		+++	+++	+++	+++	+++	+++	++	++	++
	55 C; 15'		++	++	+++	++	+++	+++	+	+	+
Ionic strength	NaCl 2M; 3h		+++	++	+++	++	++	++	++	++	+
	NaCl 2M; 30'		+++	+	+++	+	+	++	+	+	+
Viscosity	Glycerol 40%; 30'		+++	++	++	++	++	++	++	++	+
	1% Tween-20 30'		+++	+	+++	++	++	++	+	+	+
Denaturants	GuHCl 1M; 3hr		+++	+++	++	+++	+++	++	+++	+++	++
	6M Urea 30'		+++	++	++	+++	+++	+++	+	+	++
Reductant	0.2M DTT; 16hr		++	+++	+++	++	+++	+++	+	+++	++
Metal ions	10 mM $CuSO_4$		-	-	-	-	-	-	-	-	-
	10 mM $ZnSO_4$		-	-	-	-	-	-	-	-	-

FIGURE 4: Trans protein splicing of split-intein halves in endogenous or exogenous combinations. (A) Protein precursors (MI^N and MI^CC) of nine split-intein combinations were incubated *in vitro* in equal molar ratios for 24 h at room temperature in moderate conditions. Proteins were separated on 10% SDS−PAGE and visualized with Coomassie Blue staining. (B) Relative splicing activity of all combinations (columns) under different conditions (rows). Splicing efficiencies were normalized for each condition relative to the most efficient combination in that condition and labeled as either relative high splicing (+++) for 70−100% efficiency rates, relative medium splicing (++) for 30−70% efficiency rates, relative low splicing (+) for 5−30% efficiency rates, or no splicing (−). All experiments were done at room temperature, unless otherwise indicated. Oli, Tvu, and Nsp cyanobacterial species. I^N is the N-terminal half, and I^C is the C-terminal half.

similar evolutionary selection pressure to maintain their splicing activity. Although exogenous combinations have no biological relevance, they may be of use as biotechnological tools, as was demonstrated in the recent finding that split inteins from *N. punctiforme* and Ssp6803 have an efficient cross-splicing activity when coexpressed in *E. coli* (*12*).

A strong interaction between the two split-intein halves is suggested from our demonstration that split inteins were able to protein-splice under a wide range of conditions, including 6 M urea or 2 M NaCl. This is consistent with previous structural and biochemical observations that split-intein halves form a tight fold (*10*) and stable complexes in SDS−PAGE (*7*). Other split inteins were shown to be active in extreme conditions, such as high temperatures, for example, the naturally split intein from *N. equitans*, which can splice at 80 °C (*5*), and an artificially split intein from *Pyrococcus furiosus*, which splices efficiently at 70 °C (*23*). Additionally, the analysis of the recent structure of artificially joined Ssp6803 split-intein halves shows a considerable interaction surface between the N- and C-terminal intein halves (*10*).

However, our observations show that the splicing activity of several split-intein combinations (endogenous and exogenous) was significantly better than others in high concentrations of salt or urea. It is possible that minor sequence changes of different split inteins affect other characteristics of trans protein splicing, such as the association and dissociation rates of the split-intein halves or its structure stability, which were not measured in the scope of this work. The C-terminal region of each I^N intein half is variable and is not conserved among the cyanobacterial split inteins (Figure 2A) (*3*). Previous mutational analysis of the assembly of the Ssp6803 split intein showed that this variable region is not necessary for the binding of I^N to I^C (*15*). It has yet to be examined whether this variable region determines the specificity of the split-intein interaction or contributes to the splicing activity under certain conditions. Additionally, it is proposed that sequence variations in the I^C half alone can contribute to the splicing activity based on our observations that all combinations containing the same I^C half shared distinct characteristics, such as a reduced amount of splicing products or appearance of additional cleavage products (Figure 4).

It was previously noted that the association of split inteins corresponds to the opposite charges of the intein halves, as derived from their calculated overall isoelectric points (*9*). However, in this work, all split-intein combinations showed splicing activity, whether the split-intein halves had opposite charges [i.e., splicing of I^N-Oli (−10.2) with I^C-Nsp (+2.7)] or when both halves were negative [i.e., I^N-Oli (−10.2) with I^C-Oli (−0.2)] (Table S1 in the Supporting Information). Additionally, our finding that the splicing activity of several split-intein combinations was significantly better than others in high concentrations of salt or urea cannot be attributed to the global charge differences of these combinations (Table S1 in the Supporting Information). We suggest that local (rather than global) charge distributions between split-intein halves are important for their association or activity. Our suggestion is based on the structural analysis of the Ssp6803 split intein [PDB codes 1ZDE and 1ZD7 (*10*)] and molecular modeling of the cyanobacterial split-intein combinations. Examining the intramolecular electrostatic interactions in the Ssp6803 split-intein structure, we observed a cluster of positively charged residues at the I^C half (Lys126, Arg130,

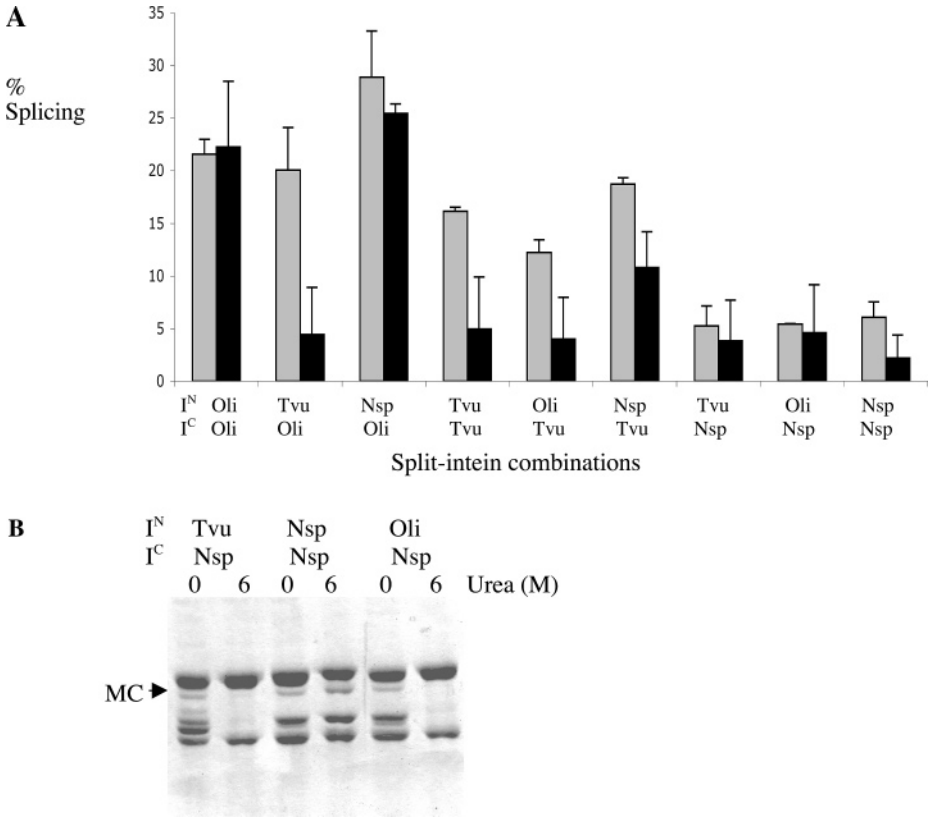


FIGURE 5: Trans protein splicing of different split-intein combinations in extreme conditions. (A) Splicing activity of all split-intein combinations incubated *in vitro* in equal molar ratios with (black) or without (gray) 2 M NaCl, for 30 min at room temperature. Proteins were separated on 10% SDS−PAGE and quantified by densitometry (NIH image). Splicing activity was calculated ($n \geq 3$, with error bars) as the fraction of the splicing product MC divided by half of the precursors and partial cleavage products. (B) Splicing activity of three split-intein combinations with or without 6 M Urea. Protein precursors were mixed *in vitro* in equal molar ratios at room temperature overnight, separated on 10% SDS−PAGE, and visualized by Coomassie Blue staining.

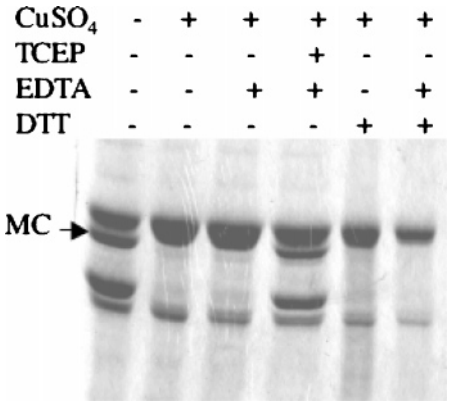
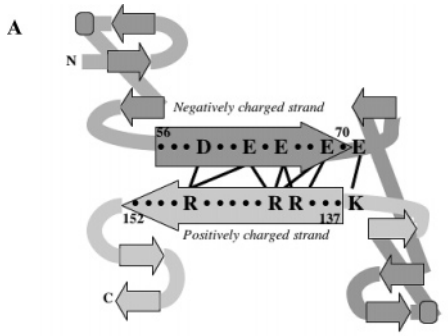


FIGURE 6: Reversible inhibition of trans protein splicing by copper ions. Inhibition is demonstrated for Tvu protein precursors mixed *in vitro* in equal molar ratios at room temperature overnight, in the presence of $CuSO_4$. Reversion of inhibition was observed upon additional incubation of the reaction with TCEP and EDTA at room temperature overnight. Substituting DTT for TCEP did not reverse the inhibition. Proteins were separated on 10% SDS−PAGE and visualized by Coomassie Blue staining.

Arg131, and Arg137) directly interacting with a cluster of negatively charged residues at the I^N half (Asp49, Glu52, Glu57, Glu59, and Glu61) (Figure 7A). These electrostatic interactions are distributed along the two longest β strands (and their immediate +1 position) of the structure. The atomic contact area between these two strands is significant, comprising half of the contact area between the I^N and I^C intein parts (918.1/1865.9 Å^2; Table S2 in the Supporting Information). While some residues in these salt bridges are naturally substituted in cyanobacterial split inteins (Figure 2), their long strands always remain oppositely charged. For example, the long β strand of the Tvu, Oli, and Nsp I^C half is positively charged (+3), and the long β strand of their I^N half is negatively charged (−1, −3, or −2, respectively). For comparison, the artificially split Sce VMA intein halves have little or no affinity for each other (*9*) and their I^N long strand is neutral, while the I^C long strand is negatively charged. We therefore believe that local charge distributions within each pair of split inteins are involved in their specific interactions and may contribute to their activity under different conditions.

Molecular modeling of all nine cyanobacterial split-intein combinations support our experimental finding that all combinations can protein-splice. The surface contact area between the I^N half and the I^C half was similar in all combinations, as well as their binding energy and number of salt bridges (Figure 7B). No significant differences in the main parameters of the energy function were observed between endogenous and exogenous combinations. Finding that all different combinations can be accommodated without major clashes indicates a large degree of promiscuity for this system: all N-terminal halves can bind to all C-terminal halves. This promiscuity probably stems from the common origin of all cyanobacterial split inteins from an intact progenitor (*3*) and the necessity to keep the two intein parts compatible within each species. Thus, all cyanobacterial split inteins are closely inter-related, relative to other inteins, and



B

Combination	Number of Salt Bridges	Buried Surface area (Å^2)	Binding Energy
I^N-Oli + I^C-Oli	18	1507	-73
I^N-Tvu + I^C-Oli	13	1515	-76
I^N-Nsp + I^C-Oli	16	1479	-74
I^N-Oli + I^C-Tvu	15	1437	-72
I^N-Tvu + I^C-Tvu	13	1389	-72
I^N-Nsp + I^C-Tvu	16	1403	-72
I^N-Oli + I^C-Nsp	19	1512	-76
I^N-Tvu + I^C-Nsp	13	1537	-77
I^N-Nsp + I^C-Nsp	19	1508	-76

FIGURE 7: Structural analysis of electrostatic interactions in the Ssp6803 split-intein crystal structure (PDB code 1ZDE). (A) Schematic two-dimensional representation of the Ssp6803 split-intein structure, showing the secondary structure elements and the interacting polar residues in the two long antiparallel β strands of the molecule. The N-terminal split-intein half is shown in dark gray, and the C-terminal split-intein half is shown in light gray. β strands are represented as arrows. Interacting residues were identified by the CSU program (*34*). (B) Calculated parameters of split-intein combinations modeled structures. The number of electrostatic interactions and binding energy are calculated between each I^N and I^C half. The surface area is the atomic contact area between the I^N and I^C intein halves. The models were created on the basis of the structure of the Ssp6803 split intein (PDB code 1ZDE) with *Rosetta* (see the Experimental Procedures) The models span the aligned split-intein region, but no attempt was made to model the C-terminal end of variable length in the I^N intein sequences (Figure 2).

the residues necessary for the successful interaction between the intein parts are likely to be conserved because of (i) the requirement for compensatory mutations in the two independent loci coding for the intein parts and (ii) the absolute necessity for the replicative DNA polymerase splicing product of the split inteins at every cell generation.

We show that copper ions inhibit protein splicing. However, unlike previously discovered metal-ion inhibitors, reversion of copper inhibition is obtained in a two-step procedure. Divalent metal ions, such as zinc and, to less extent, cadmium ions (*5*, *14*, *15*), block the splicing reaction at low micromolar concentrations, while other previously tested divalent ions, such as cobalt, nickel, and magnesium, have a much weaker inhibitory effect. Since the initial discovery of zinc in the crystal structure of the *Saccharomyces cerevisiae* VMA intein (*35*), additional structural studies (*10*) and mutation analyses (*6*, *7*) provide evidence for the mechanism for metal-ion inhibition of protein splicing. However, different zinc-coordinating residues are proposed for different types of inteins [Cys455, Glu80, and His453 for PI-SceI intein (*35*) and His48, Asp140, His110, and Cys160 for the Ssp DnaE intein (*10*)]. All models share the highly conserved C-terminal cysteine (Cys +1) as a zinc-chelating residue. It is suggested that zinc interacts with the catalytic Cys +1 residue and thereby affects the N−S acyl rearrangement and transesterification reactions of protein splicing. However, zinc-mediated inhibition is reported even when substituting the Cys +1 to Ser (*6*, *7*) and in the split intein from *N. equitans* (*5*), which naturally contains a Thr +1 residue, suggesting that the mechanism of metal-ion inhibition is not yet fully understood and that additional residues may play an important role.

We propose that copper ions interact with either the catalytic N-terminal and/or the catalytic C-terminal cysteine residues of the split intein, forming a protein−metal complex that structurally modifies the active site and blocking the splicing and cleavage reactions. The Cu^{+2} ion is a stronger oxidizing agent than Zn^{+2} and thus may not only be chelated by the cysteine sulfhydryl but also oxidize it. Indeed, the recruitment of splicing activity was not regained by just chelating the copper with EDTA but only upon additional reduction of the protein. It is therefore likely that cysteines are oxidized by copper, which inactivated the sulfhydryls necessary for the catalytic nucleophilic attacks. This double-lock mechanism of copper-mediated inhibition would be biotechnologically advantageous for controlling protein-ligation reactions in two steps (chelation and reduction).

The relatively moderate level of splicing activity of the split inteins in this work can be attributed to the deviations from the native split-intein expression and reaction conditions and to the replacement of the natural DnaE protein flanks with other domains. For example, overexpression and purification of the split intein may affect its folding and *in vitro* activity. Previous studies of natural and artificial split inteins reported similar or higher maximal splicing efficiencies. This could be the result of *in vivo* coexpression and nonuniform definitions of splicing efficiency in different studies. The splicing extent of the naturally split Ssp6803 intein was reported as 48−54% for *in vitro* (*6*, *14*) and as 80% for *in vivo* coexpression (*2*). Artificial split inteins, however, reconstitute their splicing activity only upon prior unfolding, heating, or addition of fused affinity domains, reaching levels of 74% for the split *Psp* Pol-1 (*17*) and 64% for the split *Mtu RecA* (*14*).

The experimental system reported in this work shows great potential toward understanding the sequence−function relation of split inteins. Knowing the conditions in which split inteins are active or inhibited allows for better control of the splicing reaction *in vitro* and perhaps also *in vivo*. The modularity of split inteins is useful when designing expression systems based on split inteins. The wide range of conditions that allow protein splicing may be an advantage under extreme conditions (of heat or denaturants) over other systems of assembled protein fragments.

ACKNOWLEDGMENT

We thank Prof. Israel Pecht, Prof. Ed Bayer, and Prof. Gideon Schreiber from the Weizmann Institute of Science for their discussions and supportive suggestions.

SUPPORTING INFORMATION AVAILABLE

Calculated charge differences between split-intein halves at pH 7.5 (Table S1); migration of the protein-splicing product of Tvu split inteins, compared with a control fusion construct MBP−CBD (Figure S2); and solvent accessible surfaces of areas calculated for the listed structure regions with the NACCESS program (Table S2). This material is available free of charge via the Internet at http://pubs.acs.org.

REFERENCES

1. Paulus, H. (2000) Protein splicing and related forms of protein autoprocessing, *Annu. Rev. Biochem. 69*, 447−496.
2. Wu, H., Hu, Z., and Liu, X. Q. (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.
3. Caspi, J., Amitai, G., Belenkiy, O., and Pietrokovski, S. (2003) Distribution of split DnaE inteins in cyanobacteria, *Mol. Microbiol. 50*, 1569−1577.
4. Waters, E., Hohn, M. J., Ahel, I., Graham, D. E., Adams, M. D., Barnstead, M., Beeson, K. Y., Bibbs, L., Bolanos, R., Keller, M., Kretz, K., Lin, X., Mathur, E., Ni, J., Podar, M., Richardson, T., Sutton, G. G., Simon, M., Soll, D., Stetter, K. O., Short, J. M., and Noordewier, M. (2003) The genome of *Nanoarchaeum equitans*: Insights into early archaeal evolution and derived parasitism, *Proc. Natl. Acad. Sci. U.S.A. 100*, 12984−12988.
5. Choi, J. J., Nam, K. H., Min, B., Kim, S. J., Soll, D., and Kwon, S. T. (2005) Protein trans-splicing and characterization of a split family B-type DNA polymerase from the hyperthermophilic archaeal parasite *Nanoarchaeum equitans*, *J. Mol. Biol. 356*, 1093−1106.
6. Nichols, N. M., and Evans, T. C., Jr. (2004) Mutational analysis of protein splicing, cleavage, and self-association reactions mediated by the naturally split Ssp DnaE intein, *Biochemistry 43*, 10265−10276.
7. Nichols, N. M., Benner, J. S., Martin, D. D., and Evans, T. C., Jr. (2003) Zinc ion effects on individual Ssp DnaE intein splicing steps: Regulating pathway progression, *Biochemistry 42*, 5301−5311.
8. Martin, D. D., Xu, M. Q., and Evans, T. C., Jr. (2001) Characterization of a naturally occurring trans-splicing intein from *Synechocystis* sp. PCC6803, *Biochemistry 40*, 1393−1402.
9. Shi, J., and Muir, T. W. (2005) Development of a tandem protein trans-splicing system based on native and engineered split inteins, *J. Am. Chem. Soc. 127*, 6198−6206.
10. 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.
11. Wei, X. Y., Sakr, S., Li, J. H., Wang, L., Chen, W. L., and Zhang, C. C. (2006) Expression of split DnaE genes and trans-splicing of DnaE intein in the developmental cyanobacterium *Anabaena* sp. PCC7120, *Res. Microbiol. 157*, 227−234.
12. Iwai, H., Zuger, S., Jin, J., and Tam, P. H. (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*, *FEBS Lett. 280*, 1853−1858.
13. Xu, M. Q., and Evans, T. C., Jr. (2005) Recent advances in protein splicing: Manipulating proteins in vitro and in vivo, *Curr. Opin. Biotechnol. 16*, 440−446.
14. Mills, K. V., and Paulus, H. (2001) Reversible inhibition of protein splicing by zinc ion, *J. Biol. Chem. 276*, 10832−10838.
15. Ghosh, I., Sun, L., and Xu, M. Q. (2001) Zinc inhibition of protein trans-splicing and identification of regions essential for splicing and association of a split intein, *J. Biol. Chem. 276*, 24051−24058.
16. Evans, T. C., Jr., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L., Benner, J., Liu, X. Q., and Xu, M. Q. (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.
17. Southworth, M. W., Adam, E., Panne, D., Byer, R., Kautz, R., and Perler, F. B. (1998) Control of protein splicing by intein fragment reassembly, *EMBO J. 17*, 918−926.
18. Mills, K. V., Lew, B. M., Jiang, S., and Paulus, H. (1998) Protein splicing in trans by purified N- and C-terminal fragments of the *Mycobacterium tuberculosis* RecA intein, *Proc. Natl. Acad. Sci. U.S.A. 95*, 3543−3548.
19. Sun, W., Yang, J., and Liu, X. Q. (2004) Synthetic two-piece and three-piece split inteins for protein trans-splicing, *J. Biol. Chem. 279*, 35281−35286.
20. Iwai, H., Lingel, A., and Pluecktthun, A. (2001) Cyclic green fluorescent protein produced in vivo using artificially split PI-PfuI intein from *Pyrococcus furiosus*, *J. Biol. Chem. 276*, 16548−16554.
21. Williams, N. K., Prosselkov, P., Liepinsh, E., Line, I., Sharipo, A., Littler, D. R., Curmi, P. M., Otting, G., and Dixon, N. E. (2002) In vivo protein cyclization promoted by a circularly permuted *Synechocystis* sp. PCC6803 DnaB mini-intein, *J. Biol. Chem. 277*, 7790−7798.
22. Mootz, H. D., Blum, E. S., Tyszkiewicz, A. B., and Muir, T. W. (2003) Conditional protein splicing: A new tool to control protein structure and function in vitro and in vivo, *J. Am. Chem. Soc. 125*, 10561−10569.
23. Otomo, T., Ito, N., Kyogoku, Y., and Yamazaki, T. (1999) NMR observation of selected segments in a larger protein: Central-segment isotope labeling through intein-mediated ligation, *Biochemistry 38*, 16040−16044.
24. Shingledecker, K., Jiang, S.-q., and Paulus, H. (1998) Molecular dissection of the *Mycobacterium tuberculosis* RecA intein: Design of a minimal intein and of a trans-splicing system involving two intein fragments, *Gene 207*, 187−195.
25. Dassa, B., Haviv, H., Amitai, G., and Pietrokovski, S. (2004) Protein splicing and auto-cleavage of bacterial intein-like domains lacking a C′-flanking nucleophilic residue, *J. Biol. Chem. 279*, 32001−32007.
26. Kaneko, T., Nakamura, Y., Wolk, C. P., Kuritz, T., Sasamoto, S., Watanabe, A., Iriguchi, M., Ishikawa, A., Kawashima, K., Kimura, T., Kishida, Y., Kohara, M., Matsumoto, M., Matsuno, A., Muraki, A., Nakazaki, N., Shimpo, S., Sugimoto, M., Takazawa, M., Yamada, M., Yasuda, M., and Tabata, S. (2001) Complete genomic sequence of the filamentous nitrogen-fixing cyanobacterium *Anabaena* sp. strain PCC7120, *DNA Res. 8*, 205−213.
27. Honda, D., Yokota, A., and Sugiyama, J. (1999) Detection of seven major evolutionary lineages in cyanobacteria based on the 16S rRNA gene sequence analysis with new sequences of five marine *Synechococcus* strains, *J. Mol. Evol. 48*, 723−739.
28. Wei, X. Y., Sakr, S., Li, J. H., Wang, L., Chen, W. L., and Zhang, C. C. Expression of split DnaE genes and trans-splicing of DnaE intein in the developmental cyanobacterium *Anabaena* sp. PCC7120, *Res. Microbiol. 157*, 227−234.
29. Duan, X., Gimble, F. S., and Quiocho, F. A. (1997) Crystal structure of PI-SceI, a homing endonuclease with protein splicing activity, *Cell 89*, 555−564.
30. Pietrokovski, S. (1998) Modular organization of inteins and C-terminal autocatalytic domains, *Protein Sci. 7*, 64−71.
31. Pietrokovski, S. (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins, *Protein Sci. 3*, 2340−2350.
32. Kapust, R. B., and Waugh, D. S. (1999) *Escherichia coli* maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused, *Protein Sci. 8*, 1668−1674.
33. Amitai, G., Belenkiy, O., Dassa, B., Shainskaya, A., and Pietrokovski, S. (2003) Distribution and function of new bacterial intein-like protein domains, *Mol. Microbiol. 47*, 61−73.
34. Sobolev, V., Sorokine, A., Prilusky, J., Abola, E. E., and Edelman, M. (1999) Automated analysis of interatomic contacts in proteins, *Bioinformatics 15*, 327−332.
35. Poland, B. W., Xu, M. Q., and Quiocho, F. A. (2000) Structural insights into the protein splicing mechanism of PI-SceI, *J. Biol. Chem. 275*, 16408−16413.
36. Rohl, C. A., Strauss, C. E., Misura, K. M., and Baker, D. (2004) Protein structure prediction using Rosetta, *Methods Enzymol. 383*, 66−93.
37. Kuhlman, B., and Baker, D. (2000) Native protein sequences are close to optimal for their structures, *Proc. Natl. Acad. Sci. U.S.A. 97*, 10383−10388.
38. Hubbard, S. J., and Thornton, J. M. (1993) NACCESS Computer Program London, Department of Biochemistry and Molecular Biology, University College, London, U.K.

BI0611762