FEBS Letters 580 (2006) 1853–1858

Highly efficient protein *trans*-splicing by a naturally split DnaE intein from *Nostoc punctiforme*

Hideo Iwai[a,b,c,*], Sara Züger[a,1], Jennifer Jin[a], Pui-Hang Tam[a,2]

[a] *Department of Chemistry, 110 Science Place, Saskatoon, SK, Canada, S7N 5C9*
[b] *Department of Biochemistry, University of Saskatchewan, 110 Science Place, Saskatoon, SK, Canada, S7N 5C9*
[c] *Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki, P.O. Box 65, Helsinki, FIN-00014, Finland*

Received 20 January 2006; revised 7 February 2006; accepted 15 February 2006

Available online 24 February 2006

Edited by Christian Griesinger

Abstract **Protein *trans*-splicing by the naturally split intein of the gene *dnaE* from *Nostoc punctiforme* (*Npu* DnaE) was demonstrated here with non-native exteins in *Escherichia coli*. *Npu* DnaE possesses robust *trans*-splicing activity with an efficiency of >98%, which is superior to that of the DnaE intein from *Synechocystis* sp. strain PCC6803 (*Ssp* DnaE). Both the N- and C-terminal parts of the split *Npu* DnaE intein can be substituted with the corresponding fragment of *Ssp* DnaE without loss of *trans*-splicing activity. Protein splicing with the *Npu* DnaE$_N$ is also more tolerant of amino acid substitutions in the C-terminal extein sequence.**

Keywords: Protein splicing; Intein; Segmental isotopic labelling; NMR

1. Introduction

Protein splicing is an intriguing post-translational process [1,2]. During protein splicing, an intervening sequence (intein) auto-catalytically excises itself from the precursor protein, and concomitantly ligates the two flanking sequences (exteins) with a peptide bond [3]. This process does not require any external energy and cofactors. In nature, inteins have been found in two separate polypeptide chains. A naturally occurring split intein can catalyze protein ligation in *trans*, ligating the two exteins in the two polypeptide chains into one polypeptide chain [4]. Protein *trans*-splicing has become increasingly important in many biotechnological applications such as protein cyclization [5–7], segmental isotopic labelling [8–10], protein switches [11–13], in vivo protein engineering [14], and activation of genes in transgenic plants [15]. It has become essential to obtain the biochemical and biophysical properties of split inteins in detail for further applications of intein-based protein engineering [21]. Particularly, we are interested in identifying protein sequences that have robust properties such as efficient ligation activity. From the viewpoint of the costs of stable isotopes, an efficient protein splicing activity is highly desirable for the applications of segmental isotopic labelling of proteins, where stable isotopes are incorporated into only a part of the sequence of a protein. Several inteins have been artificially split into two fragments in order to obtain split inteins that can perform *trans*-splicing for such purposes [8,10,14]. However, *trans*-splicing using artificially split inteins requires denaturation/renaturation steps in vitro in order to reconstitute protein splicing activity [9,16,17]. Protein ligation using such artificially split inteins was limited largely due to the low yield of *trans*-splicing and its labour-intensive procedures in the optimization of their refolding and splicing conditions [9,10]. In contrast to artificially split inteins, the naturally occurring split intein *Ssp* DnaE has been used for spontaneous *trans*-splicing in vivo as well as in vitro without refolding procedures [3,6]. Recently, we have developed a convenient method to prepare segmentally isotope-labelled proteins using *Ssp* DnaE intein without any refolding process that could widen applications of segmental isotopic labelling for NMR analysis of large proteins [18]. This procedure could be easily extended to multiple protein ligation of three polypeptide fragments, thereby accessing even larger proteins, if one uses an additional split intein that is able to catalyze spontaneous *trans*-splicing efficiently and specifically. Indeed, two different artificially split inteins have been utilized simultaneously for a three-fragment protein ligation in which three extein sequences have been ligated into one polypeptide chain [10]. The practical applications of such multiple protein ligations using several split inteins can be limited by the efficiency of the *trans*-splicing. Hence, it is crucial to identify split inteins with more robust and specific *trans*-splicing activity.

In this report, we describe the cloning and characterization of the naturally occurring split intein DnaE from *Nostoc punctiforme* (*Npu* DnaE). Furthermore, we investigated the cross-reactivity between *Npu* DnaE and *Ssp* DnaE and the effects of the second residue of the C-extein on *trans*-splicing.

2. Materials and methods

2.1. Cloning of *Ssp DnaE$_N$* and *Ssp DnaE$_C$*

The N-terminal part of the *Ssp* DnaE intein (*Ssp* DnaE$_N$, residues 1–123) was amplified from pMSNd(16ex)23 [20] (a gift from Dr. Henry Paulus, Harvard Medical School, MA, USA) using the oligonucleotides

* Corresponding author. Fax: +358 9 191 59541.
E-mail address: hideo.iwai@helsinki.fi (H. Iwai).

[1] Present address: Biochemisches Institut der Universität Zürich, Winterthurerstrasse 190, Zürich CH-8057, Switzerland.
[2] Present address: Department of Chemistry, University of Alberta, Edmonton, Canada.

Abbreviations: IPTG, isopropyl β-D-thiogalactoside; SDS–PAGE, sodium dodecyl sulfate–polyacrylamide gel electrophoresis; MALDI-TOF, matrix-assisted laser desorption and ionization-time-of-flight; GB1, the B1 domain of IgG binding protein, protein G

0014-5793/$32.00
doi:10.1016/j.febslet.2006.02.045

#TT01: 5′-TCGGGATCCTGCCTCAGTTTTGGTAC and #TT02: 5′-TGCCAAGCTTTATTTATTTAATTG and cloned into pRSET$_A$ (Invitrogen). In order to synthesize the C-terminal part of *Ssp* DnaE$_C$ (*Ssp* DnaE$_C$, residues 1–36), the gene of *Ssp* DnaE$_N$ in pRSET$_A$ was extended using the following synthetic oligonucleotides in a stepwise approach: #TT01: 5′-TCGGGATCCTGCCTCAGTTTTGGTAC, #TT02: 5′-TGCCAAGCTTTATTTATTTAATTG, #TT05: 5′-TGGAAGCTTAACGACCGATAACTTTAACCATTTTAATTGTACC, #TT06: 5′-GCCAAGCTTAGCGTTGCACACCCAGGGATCGACGACCGATAAC, #TT07: 5′-TGGAAGCTTACGGCAGACCGATATCAAAGATGCGTTGCACACC, #TT08: 5′-GCCAAGCTTACAGCAGAAAATTATGGTCTTGCGGCAGCACGAT, #TT09: 5′-TGGAAGCTTAGGCGATAGCACCATTAGCCAGCAGAAAATT, and #TT10: 5′-TACAAGCTTATTTGTTAAAACAGTTGGCGGCGATAGCACC. The final plasmid contains the gene of the fused *Ssp* DnaE$_N$ and *Ssp* DnaE$_C$ together with the C-terminal four amino acids of the native extein sequence. The gene of *Ssp* DnaE$_C$ was amplified by PCR, fused with the the B1 domain of IgG binding protein, protein G (GB1) coding sequence into the pBAD vector, which resulted in pSZBAD1PG as previously described [18]. The native C-extein sequence CFNK in the pSZBAD1PG vector was shortened to CFN by using the two oligonucleotides #SK49: 5′-GGTACCCAGTACAAACTTATCCTGAACGG and #SK50: 5′-GTTTGTACTGGGTACCGTTAAAACAGTTGGCGGC, resulting in the vector, pSZBAD35.

2.2. Cloning of *Npu DnaE$_N$* and *Npu DnaE$_C$*

The N-terminal part of the *Npu* DnaE intein (*Npu* DnaE$_N$, residues 1–102) was amplified from the genomic DNA of *Nostoc punctiforme* (ATCC 29133) by using the two oligonucleotides #SK092: 5′-ACGGATCCTGTTTAAGCTATGAAACGGAAATATTG and #SK093: 5′-ATGAAGCTTAATTCGGCAAATTATCAACCCG. The PCR product obtained was cloned into the vector pJJDuet30 [18] using the *Bam*HI and *Hin*dIII restriction sites, resulting in the new expression vector pSKDuet01. This plasmid expresses the GB1/*Npu* DnaE$_N$ fusion protein proceeded by the N-terminal His-tag upon induction with isopropyl β-D-thiogalactoside (IPTG).

The C-terminal part of the split *Npu* DnaE intein (*Npu* DnaE$_C$, residues 1–36) was amplified from the aforementioned genomic DNA using the two oligonucleotides #SK094: 5′-TAACATATGATCAAAATAGCCACACG and #SK095: 5′-TAGGTACCATTGAAACAATTAGAAGCTATG. The PCR product was digested with *Nde*I and *Kpn*I, and cloned into the pSZBAD35 vector. The resulting expression vector pSKBAD02 codes for the gene of the *Npu* DnaE$_C$/GB1 fusion protein containing the first three amino acids of the native C-extein sequence, which can be expressed by addition of L-arabinose.

2.3. Mutations of the second residue (Phe+2) in the C-terminal extein sequence

The second residue (Phe+2) of the native C-terminal extein sequence was mutated according to the protocol of QuikChange® Site-Directed Mutagenesis (STRATAGENE) using the corresponding synthetic oligonucleotides encoding 19 amino acid mutations (see supplemental materials). All the DNA sequences of the mutations were confirmed by DNA sequencing carried out at the NRC Plant Biotechnology Institute (Saskatoon, Canada).

2.4. In vivo trans-splicing and purification of the ligated product

25 ml of Luria broth supplemented with 100 μg/ml ampicillin and 50 μg/ml kanamycin was inoculated with *E. coli* ER2566 harboring the plasmids carrying the genes of the two precursor proteins containing DnaE$_N$ and DnaE$_C$. The culture was grown at 37 °C to a density of OD_{600} = 0.4–0.6 and induced with 0.04% L-arabinose for the expression of the C-terminal precursor protein bearing DnaE$_C$. After 35 min, the second expression of the N-terminal precursor containing DnaE$_N$ was induced with a final concentration of 0.5 mM IPTG. Expression was carried out for 6 h starting from the first induction of L-arabinose and the cells were then harvested by centrifugation at 3700 × *g* for 10 min. The cells were stored at −80 °C.

The harvested cells from the 25 ml cultures were thawed and re-suspended in 600 μl of buffer A (50 mM sodium phosphate, 300 mM NaCl, 10 mM imidazole, pH 8.0) and lysed by sonication. The cell debris was removed from the protein solution by centrifugation for 15 min at 18 000 × *g*. The entire amount of the supernatant was loaded on a Ni-NTA spin column (Qiagen) equilibrated with buffer A and centrifuged for 2 min at 700 × *g*. The column was washed twice with 600 μl buffer B (50 mM sodium phosphate, 300 mM NaCl, 20 mM imidazole, pH 8.0). The bound protein was eluted from the spin column by washing twice with 200 μl buffer C (50 mM sodium phosphate, 300 mM NaCl, 250 mM imidazole, pH 8.0).

The molecular mass of the purified ligated product was determined by matrix-assisted laser desorption and ionization-time-of-flight (MALDI-TOF) mass-spectrometry using the Applied Biosystem QSTAR® XL System at the Saskatchewan Structural Science, University of Saskatchewan.

2.5. Quantification of the efficiency of protein ligation

The efficiency of the ligation by *trans*-splicing with DnaE$_C$-GB1 variants were estimated from the intensities of the bands from the sodium dodecyl sulfate–polyacrylamide gel electrophoresis (SDS–PAGE) colored with Coomassie brilliant blue. The scanned image was analyzed by NIH image software (NIH) for quantification. The amounts of proteins were calculated using the intensities of the protein molecular weight marker (MBI Fermentas) as a reference, assuming that the staining dye binds to the proteins equally. The efficiencies of the ligation was derived by comparing the molar amount of the ligated product and the smaller molar amount of either the residual N-terminal precursor fragment or the residual C-terminal precursor fragment.

3. Results

3.1. In vivo ligation by Npu DnaE

In vivo protein ligation by the naturally split *Npu* DnaE intein with non-native exteins in *Escherichia coli* was tested using our previously described model system of the dual expression system [18]. *Npu* DnaE is homologous to *Ssp* DnaE with a protein sequence identity of 67% (68/102) for the N-terminal intein (DnaE$_N$) and 53% (19/36) for the C-terminal part (DnaE$_C$), respectively (Fig. 1). The N-terminal precursor protein contains an N-terminal hexahistidine tag (his-tag), the B1 domain of protein G (GB1), and *Npu* DnaE$_N$ (Fig. 2A). The N-terminal precursor protein can be overexpressed under the control of the *T*7 promoter upon induction with IPTG (Fig. 2A). The C-terminal precursor protein contains *Npu* DnaE$_C$, a linker containing 3 amino acids (CFN) from the native C-terminal extein sequence of DnaE intein, and GB1. The expression of the C-terminal precursor protein can be induced upon addition of L-arabinose using the *araBAD* promoter (Fig. 2A). The simultaneous expression of these two precursor proteins in *E. coli* cells is expected to produce the N-terminally his-tagged GB1-GB1 dimer (his-GB1-GB1) upon protein splicing triggered by the spontaneous association of the DnaE$_N$ and DnaE$_C$ (Fig. 2B). In Fig. 3, the time course of the expression of the two split *Npu* DnaE fragments has been analyzed by SDS–PAGE. After the induction of the C-terminal precursor protein *Npu* DnaE$_C$-GB1 by L-arabinose, a ca. 15 kDa protein was expressed (Fig. 3, lane 2). 0.5 h after the induction with L-arabinose only, the N-terminal precursor protein his-GB1-*Npu* DnaE$_N$ was subsequently induced by the addition of IPTG. Upon expression of the N-terminal precursor protein, protein *trans*-splicing was induced, resulting in the rapid disappearance of the DnaE$_C$-GB1 fragment by the excision of DnaE$_N$ and DnaE$_C$ and the ligation of the two exteins (Fig. 3, lane 3). The ligated product of his-GB1–GB1 at the position of around 18 kDa was detected 2 h after the induction by IPTG. After additional 2 h of expression, large amounts of the N-terminal precursor proteins in addition to the ligated product

```
A
Ssp DnaE_N    AEY/CLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGEQEVLEYEL 60
Npu DnaE_N    AEY/CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL 60
              *** ***  ********* *******   * * *****  *  ***  ********** ** *

Ssp DnaE_N    EDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFPLLDAGTIK 123
Npu DnaE_N    EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN---------------------- 102
              **** **** ** * * * * * * *** * ***    *

B
Ssp DnaE_C  1 MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN/CFN
Npu DnaE_C  1 MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN/CFN
              * *   *  ** *   ***   **** * ** ** * ***
```

Fig. 1. Sequence comparison between *Npu* DnaE and *Ssp* DnaE. (A) Sequence alignment of *Npu* $DnaE_N$ and *Ssp* $DnaE_N$ including the three amino acid residues of the N-terminal extein sequences. (B) Sequence alignment of *Npu* $DnaE_C$ and *Ssp* $DnaE_C$ including the three amino acids of the C-terminal extein sequences. The identical amino acids are marked with asterisks.

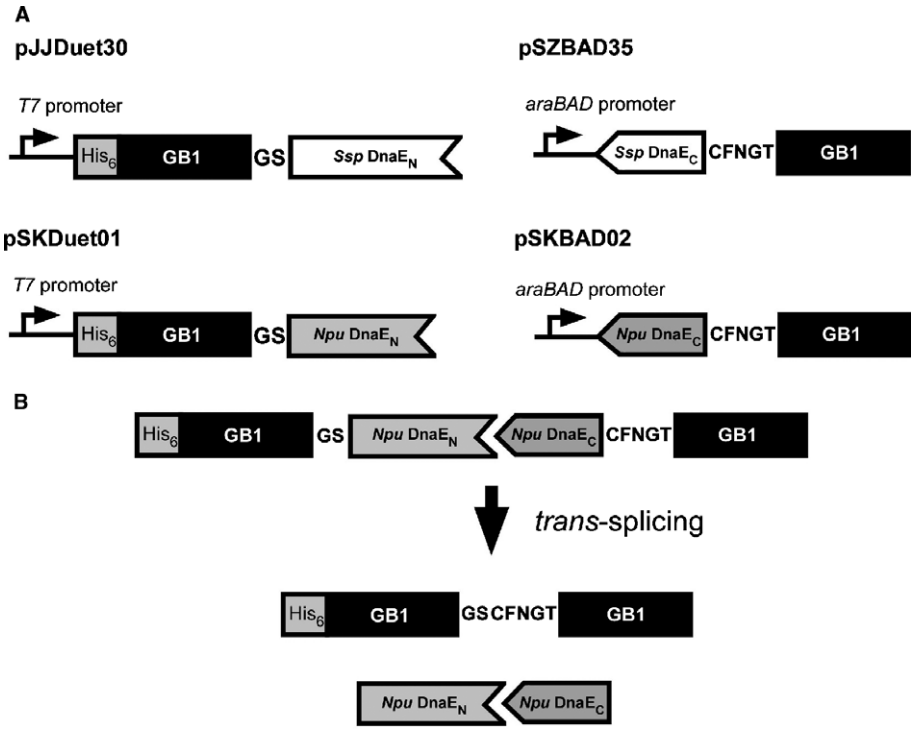

Fig. 2. Schematic drawing of the expression vectors and protein constructs. (A) Schematic drawing of the vectors used in this work depicting the promoters and the proteins that are expressed under those promoters. (B) Schematic drawing of the *trans*-splicing process. The linker sequence is shown in one code letters for amino acids.

were accumulated. The ligated product and the residual N-terminal precursor protein could be easily purified using immobilized metal affinity chromatography with the incorporated N-terminal his-tag. The elution from the Ni-chelating column contained a major product of ca. 18 kDa and a minor product of the unprocessed precursor proteins (Fig. 4, lane 3). The molecular mass of the purified major product was determined to be 15027.6 Da by MALDI-TOF mass-spectrometry (data not shown). The molecular mass is in good agreement with the theoretical molecular weight of the his-GB1-GB1 (15027.2 Da). This confirms that the naturally split *Npu* DnaE can indeed perform spontaneous *trans*-splicing in vivo in a foreign context. Compared with the naturally split *Ssp* DnaE intein (Fig. 4, lane 1), the amount of the unprocessed precursor proteins with the *Npu* DnaE intein is approximately 20-fold less in the same extein format (Fig. 5). This result demonstrates that *Npu* DnaE shows superior *trans*-splicing activity to *Ssp* DnaE. We could typically purify ca. 0.4 mg of the ligated GB1-GB1 from 25 ml culture with a purity of >95%, which corresponds to ca. 15 mg/L. This yield is more than 10-fold higher compared to the previously reported yield using *Ssp* DnaE intein with a similar system [18].

3.2. Cross-splicing between Npu DnaE and Ssp DnaE

The specific interaction between $DnaE_N$ and $DnaE_C$ is crucial for three-fragment ligation using a tandem *trans*-splicing,

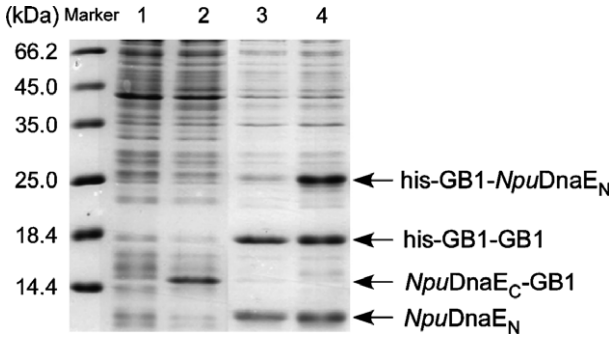

Fig. 3. SDS–PAGE analysis of the expression of the precursor proteins containing split *Npu* DnaE. Lane 1: Before induction. Lane 2: 0.5 h after the induction with L-arabinose. Lane 3: 2 h after the induction with IPTG (2.5 h after the induction with L-arabinose). Lane 4: 4 h after induction with IPTG (4.5 h after the induction with L-arabinose). All the samples were prepared by boiling the cells in SDS sample buffer.

in which two different inteins are used for ligation of three fragments [10,19]. Non-specific interactions between the homologues and their *trans*-splicing could result in undesired ligation of the two fragments. Therefore, it is necessary to characterize the specificity of the interaction between DnaE$_N$ and DnaE$_C$. *Npu* DnaE$_N$ is 21 residues shorter than *Ssp* DnaE$_N$ (Fig. 1A). If this C-terminal region of *Npu* DnaE$_N$ is involved in the interation between DnaE$_N$ and DnaE$_C$, *Npu* DnaE$_N$ might interact specifically with *Npu* DnaE$_C$, but not with *Ssp* DnaE$_C$. In order to investigate possible cross-splicing activity between *Npu* DnaE and *Ssp* DnaE, we have replaced the *Npu* DnaE$_C$ part with *Ssp* DnaE$_C$ simply by substituting the plasmid containing *Npu* DnaE$_C$ with the plasmid containing *Ssp* DnaE$_C$ in the same context. All of the four possible combinations of the different N- and C-terminal precursor proteins containing *Npu* DnaE and *Ssp* DnaE were tested in the dual expression system with the same extein sequence [18]. As seen in Fig. 4, *Ssp* DnaE$_C$ can replace *Npu* DnaE$_C$ without

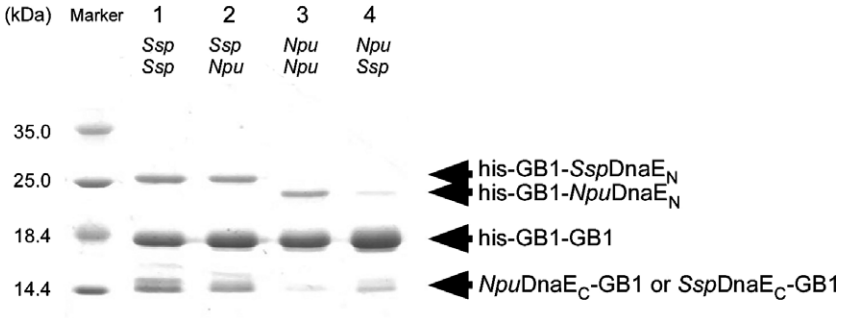

Fig. 4. SDS–PAGE analysis of *trans*- and scross-splicing. Lane 1: Combination of *Ssp* DnaE$_N$ and *Ssp* DnaE$_C$. Lane 2: combination of *Ssp* DnaE$_N$ and *Npu* DnaE$_C$. Lane 3: combination of *Npu* DnaE$_N$ and*Npu* DnaE$_C$. Lane 4: combination of *Npu* DnaE$_N$ and*Ssp* DnaE$_C$. All the samples were prepared by boiling the elution fractions from the IMAC in SDS sample buffer.

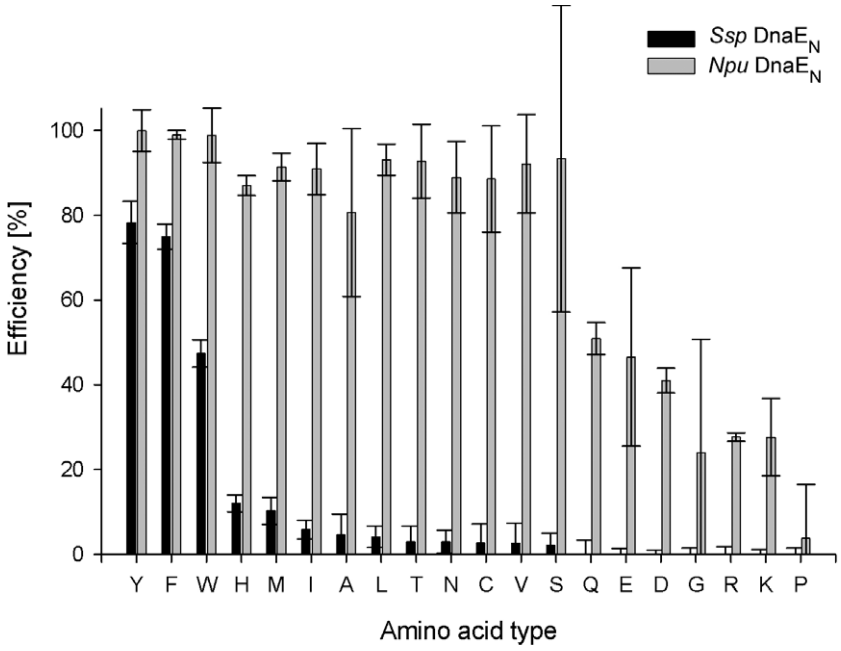

Fig. 5. Graphical representation of the ligation efficiencies of the 20 variants of *Ssp* DnaE$_C$-GB1. Filled and grey bars indicate the efficiency of *trans*-splicing for the variants of *Ssp* DnaE$_C$ and *Ssp* DnaE$_N$, and the variants of *Ssp* DnaE$_C$ and *Npu* DnaE$_N$, respectively. The amino acid types in the linker CXNGT connecting *Ssp* DnaE$_C$ and GB1 are shown at the bottom, where X stands for the substituted amino acid type. The error bars indicate the error in the estimation of the protein amounts by image analysis, except for Phe (F) in which the error was estimated from the three independent experiments.

loss of protein splicing activity. Similarly, *Npu* $DnaE_N$ can substitute *Ssp* $DnaE_N$. Thus, the interaction between the N- and C-terminal portions of the split intein *Npu* DnaE is not highly specific in regards to one another and can be substituted by one of the counterparts of *Ssp* DnaE, although the affinity of *Ssp* $DnaE_N$ towards *Ssp* $DnaE_C$ has been reported to be around a nanomolar range [19]. Interestingly, the efficiency of *trans*-splicing with the combination of *Npu* $DnaE_N$ and *Ssp* $DnaE_C$ (Fig. 4, lane 4) is similar to that of *Npu* $DnaE_N$ and *Npu* $DnaE_C$ (Fig. 4, lane 3). On the other hand, the combination with *Ssp* $DnaE_N$ and *Npu* $DnaE_C$ (Fig. 4, lane 2) gives a similar result as the combination with *Ssp* $DnaE_N$ and *Ssp* $DnaE_C$ (Fig. 4, lane 1), suggesting that the *trans*-splicing activity is dominantly influenced by the N-terminal fragment of the split intein rather than the C-terminal fragment.

3.3. Influence of the amino acid sequence near the splicing junction

The importance of the C-terminal two amino acids of the native N-extein and the N-terminal three amino acids of the native C-extein sequences in *cis*-splicing of *Ssp* DnaE have been speculated [6]. The five native N-extein and the three native C-extein sequences have often been utilized previously for the applications of *Ssp* DnaE intein [13,15]. However, we have discovered that the replacement of the native N-extein sequence Glu-Tyr (EY) to Gly-Ser (GS) prior to the ligation junction has little effect on *trans*-splicing of *Ssp* DnaE in the presence of the native C-extein sequence of CFN [18]. Currently, there is no other information about the required sequence of the N-extein for protein splicing. The sequence of GS encoded by the *Bam*HI restriction site prior to the $DnaE_N$ was used in this study thoroughly. The importance of the first C-extein residue (Cys+1) has been well-studied as it is known that its side-chain thiol group serves as a nucleophile during protein splicing [21]. However, the second residue of the C-extein (Phe+2) of the DnaE intein has not been characterized. In order to obtain molecular basis of the effects from the second residue of the C-extein, we constructed the 20 plasmids bearing all of the 20 amino acid types at the position of Phe+2 fused to *Ssp* $DnaE_C$. Most of the amino acid substitutions at the Phe+2 position abolished *trans*-splicing completely in the case of *Ssp* DnaE (Fig. 5). It remains unclear why Phe+2 influences the splicing activity of DnaE intein. We also tested the effect of these substitutions on cross-splicing activity between *Npu* $DnaE_N$ and *Ssp* $DnaE_C$. Interestingly, even if there was no splicing activity observed with the combination of *Ssp* $DnaE_N$ and *Ssp* $DnaE_C$ due to the mutation of Phe+2, cross-splicing could be observed in some cases when *Ssp* $DnaE_N$ was replaced with *Npu* $DnaE_N$ (Fig. 5). This demonstrates that *Npu* $DnaE_N$ can accommodate amino acid variations at the position of Phe+2 more readily than *Ssp* $DnaE_N$.

4. Discussion

In this article, we demonstrated that the naturally split DnaE intein from *Nostoc punctiforme*, indeed performs *trans*-splicing in a foreign context in *E. coli*. 20-fold less residual precursor proteins containing *Npu* DnaE was detected after the purification, compared with *Ssp* DnaE intein in the same context. This ligation efficiency with *Npu* DnaE was estimated to be >98% for the model system, removing almost all the precursor proteins. This demonstrates the robustness of the *trans*-splicing by *Npu* DnaE intein. The N-terminal part, i.e., $DnaE_N$, seems to determine the *trans*-splicing efficiency of the split DnaE inteins dominantly, since the replacement of $DnaE_N$ but not $DnaE_C$ significantly influenced the yield of our model system. We also found that *Npu* DnaE and *Ssp* DnaE have cross-splicing activity, meaning that one of the N- or C-terminal fragments of both *Npu* DnaE and *Ssp* DnaE can be substituted with each other without loss of their protein splicing activity. This indicates that the interaction between $DnaE_N$ and $DnaE_C$ is not highly specific although *Npu* $DnaE_N$ lacks 21 amino acids at the C-terminus compared with *Ssp* $DnaE_N$ (Fig. 1A). In contrast to *Ssp* $DnaE_N$, the *trans*-splicing with *Npu* $DnaE_N$ can accommodate many amino acid types at the second residue of the C-terminal extein (Phe+2) with no or modest reduction of *trans*-splicing activity. This tolerance of *Npu* $DnaE_N$ for the substitutions of Phe+2 can widen the application of *trans*-splicing, because it has often been limited by the inevitable insertion of the native extein sequence at the ligation junction. It is yet unknown why the extein sequences near the splicing junctions that are not directly involved in the catalytic reaction affect the splicing activity significantly. The effect of the amino acid substitutions at the Phe+2 position on splicing is not obvious. However, the mutations of the residue Phe+2 in the C-terminal extein have revealed that aromatic residues and hydrophobic residues seems to be relatively more preferred for protein splicing activity and that hydrophilic side-chains tend to lower the splicing activity significantly. The flexibility of the ligation junction has been suggested to be an important factor for protein splicing of the PI-*Pfu*I intein from *Pyrococcus furiosus* [10]. In the case of DnaE intein, the replacement of Phe+2 to Gly reduced the *trans*-splicing dramatically, although a glycine linker was successfully used with PI-*Pfu*I to facilitate its *trans*-splicing activity. We conclude that the flexibility at the junction has little importance for the splicing activity of the split DnaE inteins. Thus, the requirements for protein splicing seem to be specific to individual inteins. For the DnaE intein, there could be two possible factors originating from the extein sequence near the splicing junction that has influence over the protein splicing activity. One could be due to the structural requirements for protein splicing, because the interactions between N- and C-extein sequences near the splicing junction is likely to determine the active conformation of the intein. The other might be due to the inductive effects from the second residue of the C-extein that might affect the nucleophilicity of the active site. The information of the acceptable amino acid types in the extein sequence as we described in this article will be very useful for designing experiments with *trans*-splicing. However, the rationale of the influences from the extein sequence on protein splicing activity of different inteins remains elusive. Understanding how the exteins influence the protein splicing activity and the structural basis of the difference in protein splicing activities between *Npu* DnaE and *Ssp* DnaE will be helpful for broadening applications of *trans*-splicing. The structural comparison of *Npu* DnaE and *Ssp* DnaE at atomic resolution and the dynamics of their protein structures could provide a structural basis for protein splicing regulation and remains to be elucidated.

Acknowledgements: We thank Dr. Henry Paulus for providing the plasmid containing *Ssp* DnaE, Ken Thoms for the measurement of mass-spectrometry at the Saskatchewan Structural Science Centre, and Justin Tse for his technical assistance. This work is supported by NSERC (298493-05) and the Sigrid Jusélius Foundation.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2006.02.045.

References

[1] Hirata, R., Ohsumk, Y., Nakano, A., Kawasaki, H., Suzuki, K. and Anraku, Y. (1990) Molecular structure of a gene, *VMA1*, encoding the catalytic subunit of H^+-translocating adenosine triphosphatase from vacuolar membranes of *Saccharomyces cerevisiae*. J. Biol. Chem. 265, 6726–6733.

[2] Kane, P.M., Yamashiro, C.T., Wolczyk, D.F., Neff, N., Goebl, M. and Stevens, T.H. (1990) Protein splicing converts the yeast *TFP1* gene product to the 69-kD subunit of the vacuolar H^+-adenosine triphosphatase. Science 250, 651–657.

[3] Perler, F.B. (2000) InBase, the Intein Database. Nucleic Acids Res. 28, 344–345.

[4] 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. USA 95, 9226–9231.

[5] Scott, C.P., Abel-Santos, E., Wall, M., Wahnon, D.C. and Benkovic, S.J. (1999) Production of cyclic peptides and proteins in vivo. Proc. Natl. Acad. Sci. USA 96, 13638–13643.

[6] Evans, T.C., Martin, D., Kolly, R., Panne, D., Sun, L., Ghosh, I., Chen, L.X., 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.

[7] Iwai, H., Lingel, A. and Plückthun, A. (2001) Cyclic green fluorescent protein produced in vivo using an artificially split PI-*Pfu*I intein from *Pyrococcus furiosus*. J. Biol. Chem. 276, 16548–16554.

[8] Yamazaki, T., Otomo, T., Oda, N., Kyogoku, Y., Uegaki, K., Ito, N., Ishino, Y. and Nakamura, H. (1998) Segmental isotope labeling for protein NMR using peptide splicing. J. Am. Chem. Soc. 120, 5591–5592.

[9] Otomo, T., Teruya, K., Uegaki, K., Yamazaki, T. and Kyogoku, Y. (1999) Improved segmental isotope labeling of proteins and application to a larger protein. J. Biomol. NMR 14, 105–114.

[10] 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.

[11] 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.

[12] Buskirk, A.R., Ong, Y.C., Gartner, Z.J. and Liu, D.R. (2004) Directed evolution of ligand dependence: small-molecule-activated protein splicing. Proc. Natl. Acad. Sci. USA 101, 10505–10510.

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

[14] Giriat, I. and Muir, T.W. (2003) Protein semi-synthesis in living cells. J. Am. Chem. Soc. 125, 7180–7181.

[15] Chin, H.G., Kim, G.D., Marin, I., Mersha, F., Evans Jr., T.C., Chen, L., Xu, M.Q. and Pradhan, S. (2003) Protein *trans*-splicing in transgenic plant chloroplast: reconstruction of herbicide resistance from split genes. Proc. Natl. Acad. Sci. USA 100, 4510–4515.

[16] 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. USA 95, 3543–3548.

[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] Züger, S. and Iwai, H. (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat. Biotechnol. 23, 736–740.

[19] 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.

[20] Mills, K.V. and Paulus, H. (2001) Reversible inhibition of protein splicing by zinc ion. J. Biol. Chem. 276, 10832–10838.

[21] Nichols, N.M. and Evans Jr., T.C. (2004) Mutational analysis of protein splicing, cleavage, and self-association reactions mediated