FEBS Letters 585 (2011) 3471–3477

FEBS Letters

journal homepage: www.FEBSLetters.org

Evaluation and comparison of protein splicing by exogenous inteins with foreign exteins in *Escherichia coli*

Simo Ellilä, Jaana Maria Jurvansuu, Hideo Iwaï *

Research Program in Structural Biology and Biophysics, Institute of Biotechnology, University of Helsinki. P.O. Box 65, Helsinki, FIN-00014, Finland

ARTICLE INFO

Article history:
Received 3 August 2011
Revised 23 September 2011
Accepted 4 October 2011
Available online 10 October 2011

Edited by Gianni Cesareni

Keywords:
Protein splicing
Trans-splicing
Segmental labeling
Protein ligation
Intein

ABSTRACT

Protein splicing catalyzed by inteins has enabled various biotechnological applications such as protein ligation. Successful applications of inteins are often limited by splicing efficiency. Here, we report the comparison of protein splicing between 20 different inteins from various organisms in identical contexts to identify robust inteins with foreign exteins. We found that RadA intein from *Pyrococcus horikoshii* and an engineered DnaB intein from *Nostoc punctiforme* demonstrated an equally efficient splicing activity to the previously reported highly efficient DnaE intein from *Nostoc punctiforme*. The newly identified inteins with efficient *cis*-splicing activity can be good starting points for the further development of new protein engineering tools.

1. Introduction

Protein splicing reactions catalyzed by inteins have become important tools in biotechnological applications such as site-specific modification, segmental isotopic labeling, protein semi-synthesis, protein cyclization [1,2]. Protein splicing is a post-translational modification where an intervening sequence, termed intein, is self-catalytically removed from the host protein, accompanied by the ligation of the two flanking sequences, termed exteins (Fig. 1a) [3]. Inteins can also be split into two pieces and reassembly of the two fragments can catalyze protein splicing in *trans* (Fig. 1a) [3]. This protein *trans*-splicing (PTS) has great potential to manipulate primary structures of proteins *in vivo* as well as *in vitro* because protein splicing does not require any additional energy or co-factors [3]. However, low splicing efficiency and poor tolerance of sequence alternations near the splicing junctions, which are insufficiently characterized for most of inteins, hinder wider application of inteins [4,5]. High efficiency of protein splicing is desirable to maximize yields of protein ligation by PTS, particularly for multi-fragment ligation such as central fragment isotopic labeling [6,7]. Even though the previously reported DnaE intein is more tolerant of variations in the splicing junctions, it cannot splice with all possible sequences at the splicing junctions [4]. It is of practical importance to identify more inteins with robust splicing activity with various junction sequences for the development of biotechnological tools, which could alleviate the limitations imposed by the junction sequences [2,4]. Currently, more than 550 genes have been registered in the intein database (InBase) based on their characteristic sequence motif [8]. Only dozens of inteins have been experimentally tested for protein splicing. There has been no systematic comparison of protein splicing among different inteins in an identical extein context. In this report, we tested 20 naturally occurring and engineered inteins from various organisms connected with three identical foreign exteins to identify inteins with high splicing efficiency in *Escherichia coli*. The 20 inteins we have chosen are (1) KlbA intein and (2) Transcription factor IIB (TFIIB) intein from *Methanocaldococcus jannaschii* (*Mja*), (3) Cell division control protein 21 (CDC21-1) intein, (4) Second CDC21 (CDC21-2) intein, and (5) DNA polymerase II, DP2 subunit (Pol-II) intein from *Pyrococcus horikoshii* OT3 (*Pho*), (6) Cell division control protein 21 (CDC21-1) intein and (7) Pol-II intein from *Halobacterium* sp. NRC-1 (*Hsp*), (8) Vascular ATPase, subunit A (VMA) intein from *Thermoplasma volcanium* GSS1 (*Tvo*), (9) VMA intein from *Thermoplasma acidophilum* (*Tac*), (10) DnaE polymerase (DnaE) intein and (11) Minimized DnaB intein ($DnaB_{mini}$) from *Nostoc punctiforme* (*Npu*), (12) DNA gyrase subunit A (GyrA) from *Mycobacterium xenopi* (*Mxe*), (13) DnaE intein and (14) Minimized DnaB helicase ($DnaB_{mini}$) intein from *Synechocystis* sp. strain PCC6803, (15) Minimized Ribonucleoside-diphosphate reductase (RIR-1) intein from *Pyrococcus furiosus* (PI-*Pfu*I_{mini}), (16) RIR intein from *Methanothermobacter thermautotrophicus* (*Mth*),

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

0014-5793/$36.00
doi:10.1016/j.febslet.2011.10.005

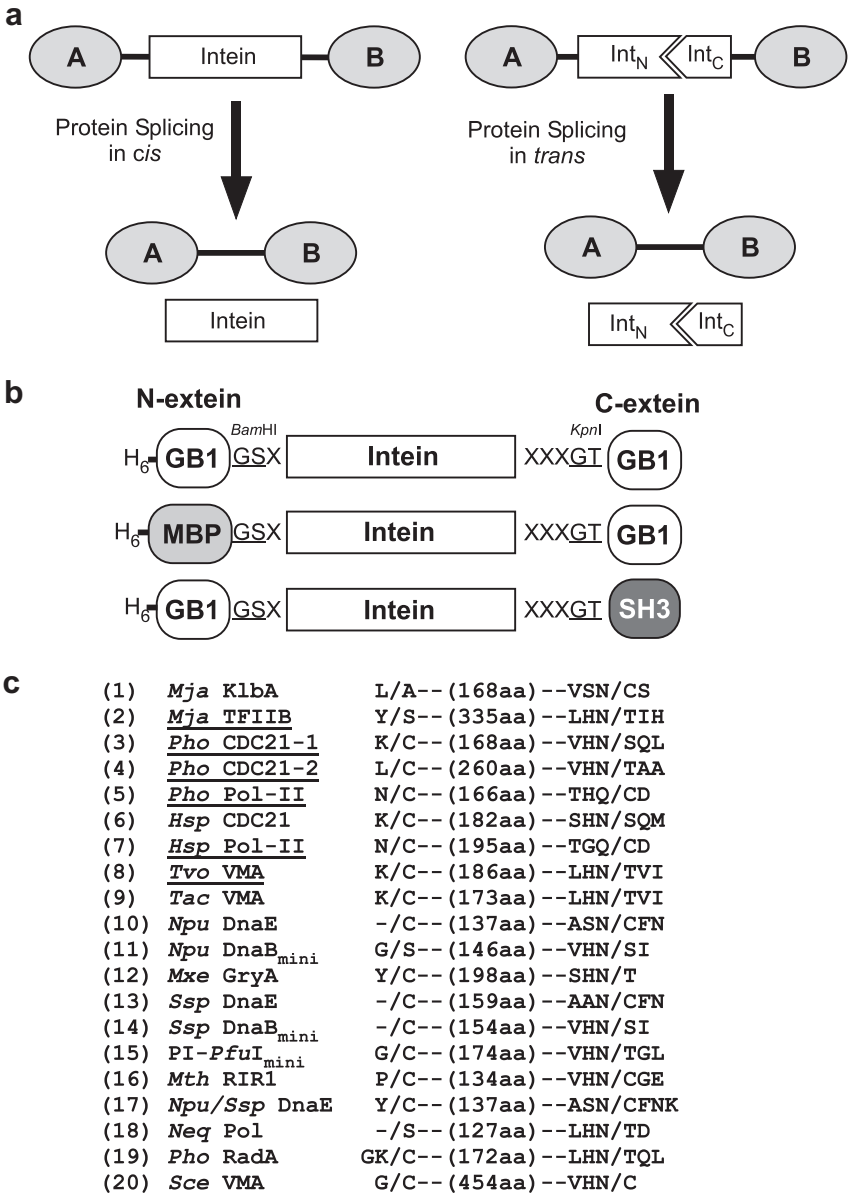

Fig. 1. (a) Protein splicing in *cis* and in *trans*. (b) Schematic drawings of the three precursors used in this study. The corresponding amino-acid sequence of the cloning sites of *Bam*HI and *Kpn*I are underlined. Xs stand for amino acid residues representing the included native junction sequences. (c) A list of the 20 inteins tested in this work. N- and C-terminal junction sequences are shown ("/" indicates the cleavage sites). The native extein sequences are also included, which are present after *Bam*HI and before *Kpn*I sites. The lengths of the inteins are indicated in brackets. Inteins that have not been experimentally tested in the literature are marked with underlines.

(17) *Npu*/*Ssp* chimeric DnaE intein, (18) DNA polymerase (Pol) intein from *Nanoarchaeum equitans* Kin4-M (*Neq*), (19) RadA DNA repair protein intein (RadA) from *P. horikoshii* OT3 (*Pho*), and (20) VMA intein from *Saccharomyces cerevisiae* (Sce). These inteins were primarily chosen because they are so-called mini-inteins without endonuclease domains or previously tested inteins. As functional split inteins cannot be derived from poor *cis*-splicing inteins, we compared their *cis*-splicing activities under identical conditions with three different foreign exteins to find potentially robust inteins for advancing the intein-based applications.

2. Materials and methods

2.1. Construction of plasmids for GB1-intein-GB1 precursors

The genes of the 20 inteins were amplified by polymerase chain reaction (PCR) from various sources (Table 1). *Cis*-splicing vectors were constructed by cloning them between *Bam*HI and *Kpn*I sites of pJJDuet63 (Fig. 1) [9], which was derived from pJJDuet30 by removing an *Nde*I site at the front of B1 domain of IgG binding protein G (GB1) using two primers SK096 and SK097 and by cloning an C-terminal GB1 domain between *Nde*I and *Xho*I sites with two primers SZ001 and SK099. Table 1 summarizes the genes, plasmids, and the oligonucleotides used for cloning into the vectors. Other primer sequences are listed in Supporting Information (Table S1). The 13 genes for (1)–(9), (12), (14), (16), and (20) in Table 1 were obtained by PCR from genomic DNAs (ATCC) or commercial plasmids (New England Biolabs) using the listed oligonucleotides. The first residue of (1) *Mja* KlbA intein in pSKDuet19, which is mutated to serine, was changed back to alanine by QuikChange® protocol (Stratagene) using the two oligonucleotides HK401 and HK402, resulting in pSEDuet33. (10) *Npu* DnaE was amplified from the previously created gene, resulting in pSKDuet16 [10]. The gene for (11) the minimized *Npu* DnaB intein was constructed by ampli-

Table 1
Summary of the 20 intein constructs.

Gene	Plasmid[a]	Template	Oligos[b]
(1) *Mja* KlbA	pSEDuet33 (pSKDuet19) pSEDuet60 pSEDuet86	*Methanocaldococcus jannaschii* genomic DNA (ATCC 43067)	SK167: 5′-AA GGA TCC TTA GCT TAT GAT GAA CCT ATT TAT TTA TC SK168: 5′-GAT GGT ACC AGA ACA GTT TGA GAC AGC AAA AC
(2) *Mja* TFIIB	pSKDuet20 pSEDuet34 pSEDuet67	*Methanocaldococcus jannaschii* genomic DNA (ATCC 43067)	SK169: 5′-TT GGA TCC TAT AGT GTT GAT TAC AACG AAC C SK170: 5′-CT GGT ACC GTG GAT GGT GTT G TGT AAT ACAAATC
(3) *Pho* CDC21-1	pSKDuet21 pSEDuet35 pSEDuet68	*P. horikoshii* genomic DNA (ATCC 700860)	SK171: 5′-AA GGA TCC AAG TGC GTT GAC TAC GAT ACA G SK172: 5′-AT GGT ACC GAG TTG GCT ATT GTG CAC G
(4) *Pho* CDC21-2	pSKDuet22 pSEDuet36 pSEDuet69	*P. horikoshii* genomic DNA (ATCC 700860)	SK173: 5′-AA GGA TCC CTC TGC GTG GCC CCG GAT AC SK174: 5′-AA GGT ACC AGC AGC TGT GTT GTG GAC TAT G
(5) *Pho* Pol-II	pSKDuet23 pSEDuet37 pSEDuet70	*P. horikoshii* genomic DNA (ATCC 700860)	SK175: 5′-CA GGA TCC AAT TGC TTC CCG GGA GAT ACA AG SK176: 5′-AA GGT ACC ATC GCA CTG ATG CGT CAC
(6) *Hsp* CDC21	pSKDuet24 pSEDuet38 pSEDuet71	*Halobacterium* salinarum NRC-1 genomic DNA (ATCC 700922)	SK177: 5′-AA GGA TCC AAG TGC GTG CGG GGC GAC ACC SK178: 5′-TA GGT ACC CAT CTG GGA GTT GTG CGA GAC C
(7) *Hsp* Pol-II	pSKDuet25 pSEDuet39 pSEDuet72	*Halobacterium* salinarum NRC-1 genomic DNA (ATCC 700922)	SK179: 5′-AC GGA TCC AAC TGC TTC CAC CCG GAG ACG SK180: 5′-AT GGT ACC GTC GCA CTG CCC GGT GAA CAG
(8) *Tvo* VMA	pSKDuet26 pSEDuet40 pSEDuet73	*Thermoplasma volcanium* genomic DNA (ATCC 51530)	SK181: 5′-AA GGA TCC AAG TGC GTA TCA GGT GAA ACA C SK182: 5′AT GGT ACC TAT AAC CGT GTT GTG AAG TAC GAG
(9) *Tac* VMA	pSKDuet27 pSEDuet41 pSEDuet74	*Thermoplasma acidophilum* genomic DNA (ATCC 25905)	SK183: 5′-AA GGA TCC AAA TGT GTG TCT GGC GAT ACA C SK184: 5′-AT GGT ACC TAT CAC GGT GTT GTG AAG GAC TAT G
(10) *Npu* DnaE	pSKDuet16 pSEDuet42 pSEDuet75	*Nostoc* sp. genomic DNA from strain PCC 73102 (ATCC 29133)	SK092: 5′-AC GGA TCC TGT TTA AGC TAT GAA ACG GAA ATA TTG SK095: 5′-TAG GTA CCA TTG AAA CAA TTA GAA GCT ATG
(11) *Npu* $DnaB_{mini}$	pMMDuet19 pSEDuet43 pSEDuet76	*Nostoc* sp. genomic DNA from strain PCC 73102 (ATCC 29133)	HK151: 5′-TA GGA TCC GGT TGT TTA GCA GGC GAT AGT C HK212: 5′-GA GGT ACC AAT GGA ATT GTG AAC AAT
(12) *Mxe* GyrA	pAHDuet12 pSEDuet44 pSEDuet77	pTXB1 (New England BioLabs)	HK295: 5′-CGG GAT CCT ATT GCA TCA CGG GAG ATG C SZ021: 5′-GC GGT ACC TGT GTT GTG GCT GAC GAA CCC
(13) *Ssp* DnaE	pNPDuet16 pSEDuet45 pSEDuet78	pTTDuet20 [5]	HK294: 5′-AT GGA TCC TGC CTC AGT TTT GGA ACC GAA ATT SZ002: 5′-TTG GGT ACC TTT GTT AAA ACA GTT GGC
(14) *Ssp* $DnaB_{mini}$	pSEDuet5 pSEDuet46 pSEDuet79	pTWIN2 (New England Biolabs)	HK39: 5′-CT GGA TCC TGT ATC TCT GGC GAT AGT C HK42: 5′-TC GGT ACC AAT AGA GTT GTG TAC AAT GAT GTC
(15) PI-*Pfu*I_{mini}	pAHDuet6+7 pSEDuet47 pSEDuet80	pIWT5563his [11]	HK298: 5′-AA GGA TCC GGG TGC ATA GAC GGA AAG GCC HK299: 5′-AC GGT ACC AAG TCC G GTG TTG TGG ACG AAA ATC
(16) *Mth* RIR1	pHYDuet130 pSEDuet48 pSEDuet81	pTWIN2 (New England Biolabs)	#106: 5′-CT GGA TCC CCA TGC GTA TCC GGT GAC HK285: 5′-AT GGT ACC CTC GCC GCA ATT GTG TAC AAT G
(17) *Npu*/*Ssp* DnaE	pSEDuet20 pSEDuet49 pSEDuet82	pSKDuet1 [9] pSZBAD1PG [5]	SK092: 5′-AC GGA TCC TGT TTA AGC TAT GAA ACG GAA ATA TTG SZ002: 5′-TTG GGT ACC TTT GTT AAA ACA GTT GGC
(18) *Neq* Pol	pSEDuet16 pSEDuet50 pSEDuet83	Synthetic	SZ058: 5′-CA GGT ACC TAA TGA ATC GGT ATT GTG TAA AAC SK111: 5′-TT GGA TCC TCT ATA ATG GAT ACT G
(19) *Pho* RadA	pHYDuet183 pSEDuet51 pSEDuet84	*P. horikoshii* genomic DNA (ATCC 700860 D)	HK375: 5′-AAG GAT CCG GGA AGT GCT TTG CTA GGG ATA CCG AA HK376: 5′-TTG GTA CCT AGC TGA GTA TTA TGG AGA ACA AGT
(20) *Sce* VMA	pSEDuet30 pSEDuet52 pSEDuet85	*S. cerevisae* genomic DNA (ATCC 9763 D)	#10: 5′-GGG GGA TCC GGG TGC TTT GCC AAG GGC ACC AAT GTT TTA ATG GCG G HK397: 5′-TT GGT ACC GCA ATT ATG GAC GAC AAC

[a] The upper name is for GB1-GB1 construct. The middle is for MBP-GB1 construct. The lower is for GB1-SH3 construct.
[b] The cloning sites of *Bam*HI or *Kpn*I are indicated in underlines.

fying the N- and C-terminal fragments from the genomic DNA with the oligonucleotides HK151, HK213, HK211, and HK212, and assembled for the cloning. The gene of *cis*-splicing (13) *Ssp* DnaE intein was amplified from pTTDuet20 [5]. The gene for (15) PI-*Pfu*I_{mini}, which is a minimized variant of PI-*Pfu*I, was constructed by removing the endonuclease domain. PI-*Pfu*I intein was

amplified from plasmid pIWT5563his [11] with oligonucleotides, HK298, HK296, HK300 and HK299, and cloned after the assembly. The gene for (17) *Npu*/*Ssp* chimeric DnaE intein was constructed by assembling the N-intein from *Npu* DnaE and the C-intein of *Ssp* DnaE by PCR, using pairs of the primers SK092 and HK372, or HK371 and SZ002, respectively. The gene of (18) *Neq* Pol intein was synthesized by PCR. The N-terminal part of *Neq* Pol intein was synthesized from the following oligonucleotides, SK111, SK112, SK113, SK114, SK115, SK116, SK117, SK118, SK119, SK120, SK121, and SK122. The PCR product was cloned into vector pJJDuet30, resulting in pJJDuet58. The C-intein was also synthesized by PCR with the following oligonucleotides, SZ055, SZ056, SZ057, and SZ058 and cloned into pSFBAD09 [9], resulting in pJJ-BAD69. The two split fragments of *Neq* Pol inteins were then assembled by PCR with the following primers, SZ058, HK286, SK111, and SK120. The gene of (19) *Pho* RadA intein was amplified from the genomic DNA with two oligonucleotides, HK376 and HK377, which was further extended to remove *Kpn*I site within the gene with two additional oligonucleotides HK379 and HK375. All the plasmids described in this article will be available from addgene inc. together with their DNA sequences (Cambridge, USA, www.addgene.org/Hideo_Iwai). The primary structures of the 20 inteins are also listed in Supporting Information (S2).

2.2. Construction for MBP-intein-GB1 precursors

The complete set of pDuet plasmids harbouring the 20 inteins with two GB1s as exteins was used to construct the plasmids with maltose binding protein (MBP) replacing the N-terminal GB1. A set of plasmids for expressing H_6-MBP-Intein-GB1 precursors was constructed by digesting the original plasmids harbouring the precursor of H_6-GB1-Intein-GB1 with *Bam*HI and *Hind*III when this was possible. The excised genes containing the intein genes with attached C-terminal GB1 were then cloned into pSARSF40-13 just after the MBP gene. pSARSF40-13 is a derivative of pHYRSF-1 [12], in which *Spe*I site was introduced after H_6-tag using QuikChange® protocol with the two primers HK123 and HK124, resulting in pHYRSF34. The gene of GB1 between *Spe*I and *Bam*HI sites in pHYRSF34 was further replaced by the gene of MBP obtained from pMYB5 (New England Biolabs) using the two primers, HK222 and HK220. The excised gene fragments corresponding to the twenty intein-GB1 fragments were ligated between *Bam*HI and *Hind*III in pSARSF40-13, resulting plasmids pSERSF34-52 and pSERSF60 (Table 1).

2.3. Construction of plasmids for GB1-intein-SH3 precursors

A third set of plasmids with Src homology 3 (SH3) domain replacing the C-terminal GB1 extein was constructed. The plasmids for expressing H_6-GB1-Intein-GB1 precursors were digested with the restriction enzymes *Nco*I and *Kpn*I to excise the genes of H_6-GB1-Intein. These fragments were then ligated into the vector pMMRSF1-16 [13], harbouring the C-terminal SH3 domain resulting in the constructs pSERSF67-86 (Fig. 2b, Table 1).

2.4. Evaluation cis-splicing activity of the inteins

The constructed plasmids were transformed into *E. coli* ER2566 (New England Biolabs) for testing *in vivo* protein splicing. The *E. coli* cells harbouring each plasmid were grown in 5 ml of Luria Broth media supplemented with 25 μg/ml kanamycin at either 20, 30, or 37 °C to a density of OD_{600} = 0.5–0.6 and induced with isopropyl-P-D-thiogalactoside (IPTG) to a final concentration of 1 mM. Protein expression was then allowed to proceed for 5 h. The cells were then harvested by spinning down at 4500×*g* for 10 min and stored at −20 °C for further purification. The harvested cells were lysed with 150 μl of CelLytic™B Cell Lysis Reagent (Sigma–Aldrich). The cell debris was removed from the protein solution by centrifugation at 18 000×*g* for 15 min. The entire amount of the resulting supernatant was loaded on a Ni-NTA spin column and proteins were isolated according to the manufacturer protocol (Qiagen). The concentrations of the eluted proteins were adjusted to an apparent equal amount based on the absorbance values of A280. The splicing was evaluated from the intensities of the bands in the SDS-gels stained with PhastGel Blue R (GE Healthcare) by quantifying the scanned gels with ImageJ (NIH). The amounts of proteins were calculated with the assumption that the staining dye binds to the proteins equally.

3. Results

3.1. Cis-splicing with two GB1s as exteins

Intein with robust splicing activity is a key to develop novel split inteins [5,17]. As a first step to develop new intein-based protein engineering tools, we compared *cis*-splicing of the selected 20 inteins with identical foreign extein sequences, i.e., two B1 domains of IgG binding protein G (GB1) as exteins (Fig. 1b) [4]. The comparison of *cis*-splicing allows us to evaluate splicing efficiencies directly as compared with *trans*-splicing in which an additional association step is involved [14–16]. The *cis*-splicing precursors with 20 different inteins are flanked by two GB1s as N- and C-exteins, together with 0-2 residues of native N-extein and 1–4 residues of native C-exteins at the splicing junctions (Fig. 1b and c). Since extein sequences around the junctions affect the splicing activity significantly, we tried to keep a few residues of the natural extein sequences, based on the assumption that the native junction sequences are optimal for protein splicing. The two GB1 domains and the intein sequences were joined via the amino-acid sequences of GS and GT, translated from the two restrictions sites of *Bam*HI and *Kpn*I respectively. A hexahistidine tag (H_6) was introduced at the N-terminus of the precursor proteins so that immobilized metal ion affinity chromatography (IMAC) purification using Ni-NTA spin column will pull down all the H_6-tagged proteins and their associated proteins. If there were no protein splicing activity at all, only the full-length precursor should be eluted. In the case of 100% splicing efficiency, only a single band of the *cis*-spliced product is expected to be in the elution fractions [4]. Since the precise estimation of the splicing kinetics *in vivo* is not feasible, we simply quantified the intensity of the spliced product band to the rest of other protein bands in SDS–PAGE for the comparison. The elution fractions from the Ni-NTA columns could contain unspliced precursors, intermediates with H_6-tag, cleaved products with H_6-tag, and inteins excised during the SDS–PAGE sample preparation. The SDS–PAGE gels of the elution fractions from the 20 precursors with the 20 different inteins are shown in Fig. 2. The spliced products are indicated inside a dotted rectangle. The intensities of these bands were compared with the remaining protein bands to derive *cis*-splicing efficiencies. In Fig. 3, the *cis*-splicing activities of the 20 inteins are plotted according to their estimated splicing efficiencies. The 20 inteins can be classified into three groups, (1) inteins with high splicing efficiency (>85%), (2) inteins with no or very little splicing (<5%), and (3) the intermediate group between (1) and (2). The high efficiency group includes *Pho* RadA, *Npu* DnaE, *Npu*/*Ssp* DnaE, and *Npu* $DnaB_{mini}$ inteins. *Npu* DnaE and *Npu*/*Ssp* DnaE inteins are previously reported to be extremely efficient split inteins, which have been successfully used for protein ligation [4,18]. Not surprisingly, these inteins also turned out to be very efficient in *cis*-splicing. In addition, this study revealed that *Pho* RadA and *Npu* $DnaB_{mini}$ inteins have similar *cis*-splicing efficiency to *Npu* DnaE intein. No or very little (<5%)

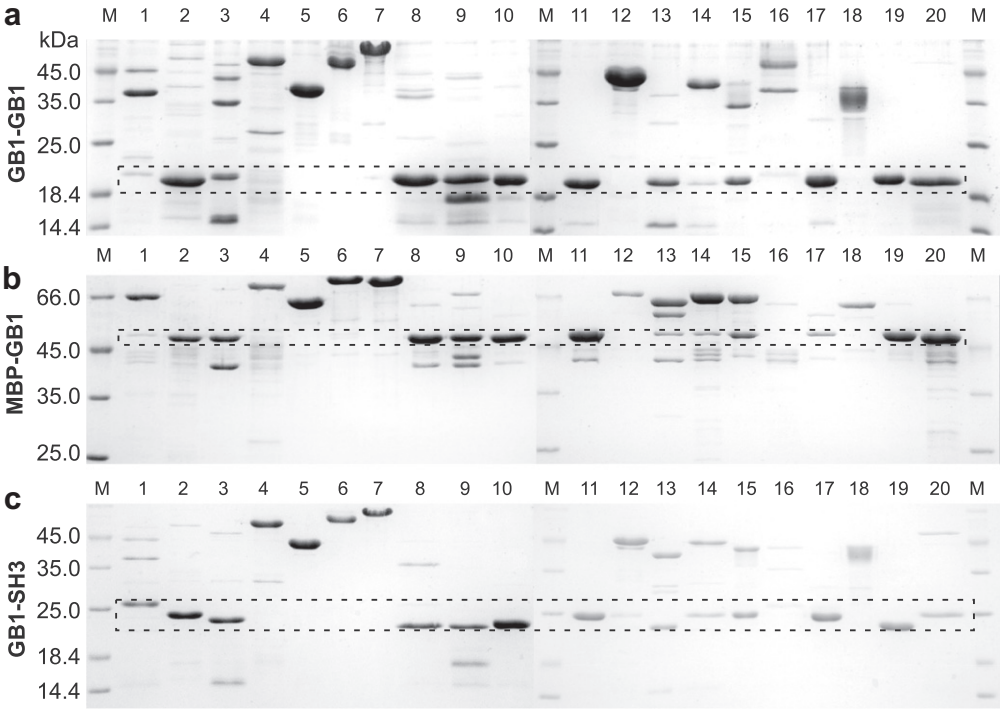

Fig. 2. Analysis of *cis*-splicing by IMAC. Typical SDS–PAGES analysis of the elution fractions from the expression of three precursors at 37 °C with (a) two GB1s, (b) MBP and GB1, and (c) GB1 and SH3, as exteins. M stands for Molecular Weight Marker. Lanes 1–20 correspond to (1) *Mja* KlbA, (2) *Mja* TFIIB, (3) *Pho* CDC21-1, (4) *Pho* CDC21-2, (5) *Pho* Pol-II, (6) *Hsp* CDC21, (7) *Hsp* Pol-II, (8) *Tvo* VMA, (9) *Tac* VMA, (10) *Npu* DnaE, (11) *Npu* $DnaB_{mini}$, (12) *Mxe* GyrA, (13) *Ssp* DnaE, (14) *Ssp* $DnaB_{mini}$, (15) PI-*Pfu*I_{mini}, (16) *Mth* RIR1, (17) *Npu*/*Ssp* DnaE, (18) *Neq* Pol, (19) *Pho* RadA, and (20) *Sce* VMA inteins. The dotted rectangles indicate the spliced products.

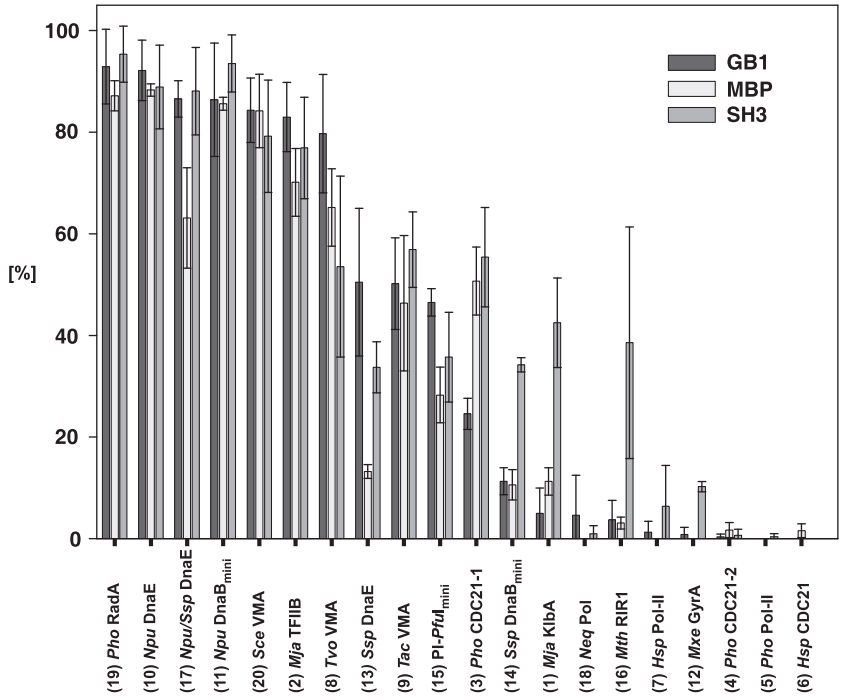

Fig. 3. Comparison of *cis*-splicing for the 20 inteins with two GB1 as exteins (filled bars) and with two different extein sequences, SH3 (light grey bars) and MBP (dark grey bars) at 37 °C. The error bars were estimated from at least three independent experiments. The inteins are sorted after the average splicing efficiencies in percentage. The names of the 20 inteins are listed at the bottom of the graph.

splicing was detected for *Pho* Pol-II, *Pho* CDC21-2, *Mxe* GyrA, *Hsp* Pol-II, *Hsp* CDC21, and *Neq* Pol inteins with the model system of two GB1s as exteins, even though their splicing were either previously reported or they are registered in InBase based on their characteristic sequence motif [19–21]. Our results indicate that these inteins cannot tolerate changes in the splicing junction regions and exteins, suggesting that they are unlikely to be useful for developing general protein ligation tools.

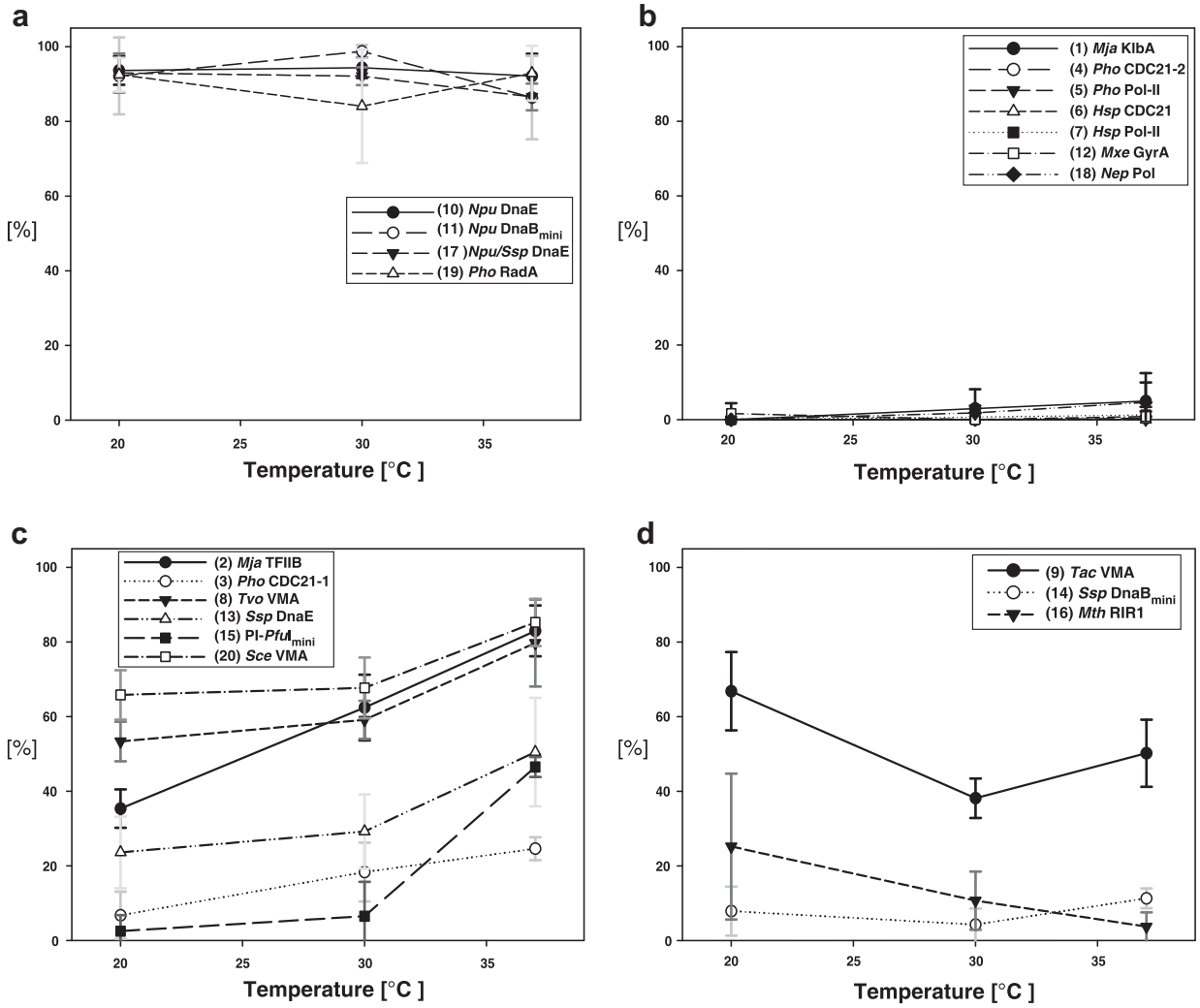

Fig. 4. The effect of the protein expression temperature on *cis*-splicing efficiency. Plots show splicing efficiency versus expression temperature (a) for the highly efficient inteins, (b) for the 'splicing-incompetent' group, (c, d) for the intermediate group. The errors were obtained from three independent experiments.

3.2. Effects from extein sequences

Not only splicing junction sequences but also extein sequences can be detrimental for protein *trans*-splicing [5]. To examine how extein sequences would affect *cis*-splicing with identical splicing junction sequences, we tested two other proteins as exteins, i.e., SH3 and MBP, as illustrated in Fig. 1b. One of the exteins in the vectors was replaced by utilizing the restriction sites of *Bam*HI and *Kpn*I, thereby retaining the same splicing junction sequences. The results are illustrated in the graph (Fig. 3). In general the replacement of the extein did not have notable influences on the splicing activity of highly efficient inteins except for *Npu*/*Ssp* DnaE intein with MBP, which showed about 20% decrease. Replacing the exteins did not revert the functionality of the 'splicing-incompetent' inteins such as *Hsp* CDC21, *Hsp* Pol-II, *Pho* Pol-II, and *Pho* CDC21-2. This is very different from the effect of exteins on PTS, where the exteins could inhibit *trans*-splicing activity of highly efficient split *Npu* DnaE intein [5]. *Hsp* Pol-II and *Pho* Pol-II have the atypical C-terminal glutamine instead of the highly conserved asparagine, which cleaves off C-extein by side-chain cyclization as the last step of protein splicing process. Intein with the C-terminal glutamine have been previously demonstrated to be splicing-competent [22,23]. Our experiments with *Hsp* Pol-II and *Pho* Pol-II indicated that these inteins showed little or very poor splicing activity, suggesting that not all the inteins with C-terminal glutamine are splicing-capable in *E. coli* with foreign exteins at 20–37 °C. This could be because both *Hsp* Pol-II and *Pho* Pol-II are from extremophilic organisms and likely to require high salt concentration or elevated temperature for splicing as observed for *Pab* Pol-II intein [23]. The replacement of exteins affected the inteins in the intermediate group more notably, but did not improve them into the high efficiency group (>85%).

3.3. Temperature effect

Next, we examined how the temperature during the expression in *E. coli* could influence the protein splicing efficiency. We observed a similar trend as observed for the extein replacements, that is, the temperature does not visibly affect the splicing of highly efficient inteins. The reaction rates for this class of inteins are probably too fast to observe any differences in their efficiencies with the procedure we used. On the other hand, inteins in the intermediate group were visibly affected with two trends. One group had reduced yields of splicing at lower temperature, for example, *Mja* TFIIB and *Pho* CDC21-1 inteins (Fig. 4c). The lowering temperature had positive effects on yields of the other group as observed for *Tac* VMA and *Mth* RIR1 inteins (Fig. 4d). This result points out that inteins in the intermediate efficiency group require more optimizations of the splicing conditions such as temperature for maximizing the yield of spliced product.

4. Discussion

In this study, we found that *Pho* RadA and a minimized *Npu* DnaB inteins, termed *Npu* $DnaB_{mini}$ intein, have very efficient *cis*-splicing activities (>85% efficiency) in all the three foreign contexts, in addition to the highly efficient split *Npu* DnaE and *Npu*/*Ssp* DnaE inteins [4,18]. The effects of extein sequences and temperature on splicing efficiency of various inteins emphasizes the importance of having highly efficient inteins with fast splicing kinetics for protein engineering because they are less affected by extein sequences as well as by the expression temperature. Exteins seem to disturb or improve proper folding of inteins in the intermediate group. The effects of exteins on *cis*-splicing seem to be moderate in contrast to *trans*-splicing where an extein could even inhibit the splicing (Fig. 3). The different effect from the attached exteins between *cis*- and *trans*-splicing suggests that extein sequences mainly modulate the association step of the two precursors in *trans*-splicing before the protein splicing reaction. This is possibly because attached exteins could change solubility or net charges of individual precursors, which could significantly influence the association step. Whereas lowering temperature could presumably increase correctly folded structures for some inteins, elevated temperature can promote the splicing reaction for some inteins such as inteins from thermophilic organisms. These temperature effects are likely to be further complicated by the attached exteins. Thus, it is necessary to assess the splicing conditions particularly for the inteins from the intermediate group.

It is noteworthy that some inteins registered in InBase as protein splicing units based on their characteristic sequence motif were unable to splice in foreign contexts. Further characterization of these inteins needs to be performed to answer whether this is due to the foreign exteins or subtle changes near the splicing junctions. *Hsp* CDC21 intein was reported to be splicing-incompetent in *E. coli* even in the native context, in agreement with our result [24]. The same protein was able to splice to a mature protein in the host organism, *Halobacterium* sp. NRC-1 implying a few possible scenarios [24]. For example, exogenous inteins might not be able to fold into a functional structure in *E. coli*, thereby resulting in a 'splicing-incompetent' form. It could also be due to the fact that some inteins might require special co-factor such as specific chaperons or ligands to induce protein splicing, even though it has been considered that protein splicing is self-catalytic not requiring any additional co-factors or metabolic energy [3]. Splicing of inteins has often been tested in *E. coli* system because of the convenient recombinant expression systems available. However, a few inteins in this study spliced very poorly or not at all in *E. coli*. This observation obligates further assessments of the splicing in the native context and environment. The characterization of 'splicing-incompetent' inteins in the host organisms might shed light on new aspects of protein splicing.

In summary, this study identified two new inteins, *Pho* RadA and *Npu* $DnaB_{mini}$ as inteins with efficient *cis*-splicing activity, which could be potentially good candidates for creating split inteins for the further development of protein ligation tools. In particular, they have serine or threonine at the +1 residue site, alleviating the limitation of splicing junction sequences. Protein splicing of exogenous inteins in *E. coli* suggests that further characterization of protein splicing in the host organisms may be required to understand their biological functions, which have not yet been clearly identified.

Acknowledgments

The authors thank J. Jin, S. Züger, P.H. Tam, A.L. Bucher, A.S. Aranko, and M. Muona for their contributions at the early stage of this project. This work is supported by the Academy of Finland (131413, 137995, 1128207) and Biocenter Finland.

Appendix A. Supplementary data

Supplementary data associated with this article can be found, in the online version, at doi:10.1016/j.febslet.2011.10.005.

References

[1] Saleh, L. and Perler, F.B. (2006) Protein splicing in *cis* and in *trans*. Chem. React. 6, 183–193.
[2] Volkmann, G. and Iwaï, H. (2010) Protein trans-splicing and its use in structural biology: opportunities and limitations. Mol. Biosyst. 6, 2110–2121.
[3] Paulus, H. (2000) Protein splicing and related forms of protein autoprocessing. Annu. Rev. Biochem. 69, 447–496.
[4] Iwai, H., Züger, S., Jin, J. and Tam, P.H. (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. FEBS Lett. 580, 1853–1858.
[5] Aranko, AS., Züger, S., Buchinger, E. and Iwaï, H. (2009) In vivo and in vitro protein ligation by naturally occurring and engineered split DnaE inteins. PLoS ONE 4, e5185.
[6] 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-Us 38, 16040–16044.
[7] Busche, A.E.L., Aranko, A.S., Talebzadeh-Farooji, M., Bernhard, F., Dötsch, V. and Iwaï, H. (2009) Segmental isotopic labeling of a central domain in a multidomain protein by protein trans-splicing using only one robust DnaE intein. Angew. Chem. Int. Ed. Engl. 48, 6128–6131.
[8] Perler, F.B. (2002) InBase: the Intein Database. Nucleic Acids Res. 30, 383–384.
[9] Züger, S. and Iwaï, H. (2005) Intein-based biosynthetic incorporation of unlabeled protein tags into isotopically labeled proteins for NMR studies. Nat. Biotechnol. 23, 736–740.
[10] Heinämäki, K., Oeemig, J.S., Pääkkonen, K., Djupsjöbacka, J. and Iwaï, H. (2009) NMR resonance assignment of DnaE intein from *Nostoc punctiforme*. Biomol NMR Assign 3, 41–43.
[11] 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.
[12] Muona, M., Aranko, A.S., Raulinaitis, V. and Iwaï, H. (2010) Segmental isotopic labeling of multi-domain and fusion proteins by protein *trans*-splicing in vivo and in vitro. Nat. Protoc. 5, 574–587.
[13] Muona, M., Aranko, A.S. and Iwaï, H. (2008) Segmental isotopic labelling of a multidomain protein by protein ligation by protein *trans*-splicing. Chembiochem 9, 2958–2961.
[14] Dassa, B., Amitai, G., Caspi, J., Schueler-Furman, O. and Pietrokovski, S. (2007) *Trans* protein splicing of cyanobacterial split inteins in endogenous and exogenous combinations. Biochemistry-Us 46, 322–330.
[15] 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.
[16] 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.
[17] Oeemig, J.S., Aranko, A.S., Djupsjöbacka, J., Heinämäki, K. and Iwaï, H. (2009) Solution structure of DnaE intein from *Nostoc punctiforme*: structural basis for the design of a new split intein suitable for site-specific chemical modification. FEBS Lett. 583, 1451–1456.
[18] Zettler, J., Schutz, V. and Mootz, H.D. (2009) The naturally split *Npu* DnaE intein exhibits an extraordinarily high rate in the protein *trans*-splicing reaction. FEBS Lett. 583, 909–914.
[19] Southworth, M.W., Benner, J. and Perler, F.B. (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. EMBO J. 19, 5019–5026.
[20] Choi, J.J., Nam, K.H., Min, B., Kim, S., Soil, D. and Kwon, S. (2006) 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.
[21] Telenti, A., Southworth, M., Alcaide, F., Daugelat, S., Jacobs, W.R. and Perler, F.B. (1997) The Mycobacterium xenopi GyrA protein splicing element: characterization of a minimal intein. J. Bacteriol. 179, 6378–6382.
[22] Amitai, G., Dassa, B. and Pietrokovski, S. (2004) Protein splicing of inteins with atypical glutamine and aspartate C-terminal residues. J. Biol. Chem. 279, 3121–3131.
[23] Mills, K.V., Manning, J.S., Garcia, A.M. and Wuerdeman, L.A. (2004) Protein splicing of a Pyrococcus abyssi intein with a C-terminal glutamine. J. Biol. Chem. 279, 20685–20691.
[24] Sakakibara, N., Han, M., Rollor, C.R., Gilson, R.C., Busch, C., Heo, G. and Kelman, Z. (2008) Cloning, Purification, and Partial Characterization of the Halobacterium sp. NRC-1 Minichromosome Maintenance (MCM) Helicase. Open Microbiol. J. 2, 13–17.