

# Sequential formation of two branched intermediates during protein splicing of class three inteins

Kazuo Tori<sup>1,2</sup> · Francine Perler<sup>1,3,4</sup>

Received: 15 July 2016 / Accepted: 24 September 2016 / Published online: 4 October 2016  
© The Author(s) 2016. This article is published with open access at Springerlink.com

**Abstract** Inteins are the protein equivalent of introns. They are seamlessly removed during post-translational maturation of their host protein (extein). Inteins from extremophiles played a key role in understanding intein-mediated protein splicing. There are currently three classes of inteins defined by catalytic mechanism and sequence signatures. This study demonstrates splicing of three class 3 mini-inteins: *Burkholderia vietnamiensis* G4 Bvi IcmO intein, *Mycobacterium smegmatis* MC2 155 Msm DnaB-1 intein and *Mycobacterium leprae* strain TN Mle DnaB intein. *B. vietnamiensis* has a broad ecological range and remediates trichloroethene. *M. smegmatis* is a biofilm forming soil bacteria. Although other intein classes have only a single branched intermediate at the C-terminal splice junction, the class 3 intein reaction pathway includes two branched intermediates. The class 3 specific branched intermediate is formed by an internal cysteine, while the C-terminal branch

intermediate is at a serine or threonine in all class 3 inteins except the Bvi IcmO intein, where it is a cysteine. This latter cysteine was unable to compensate for mutation of the class 3-specific internal catalytic cysteine despite the Bvi IcmO intein having an N-terminal splice junction naturally tuned for a cysteine nucleophile, demonstrating the mandatory order of branch intermediates in class 3 inteins.

**Keywords** Intein · Protein splicing · Branched intermediate · Enzyme mechanism · Cysteine

## Abbreviations

aa	Amino acid
BI	Branched intermediate
BI <sup>F</sup>	Branched intermediate with Cys <sup>F:4</sup> as the branch point
BI <sup>G</sup>	Branched intermediate with the +1 (G8) amino acid as the branch point
M or MBP	Maltose binding protein
P	Paramyosin $\Delta$ Sal fragment
I	Intein
MIP, MSP, MLP and MVP	Model precursors with MBP and P flanking either any intein, the Msm DnaB-1 intein, the Mle DnaB intein or the Bvi IcmO intein, respectively

Communicated by H. Atomi.

This article is part of a special feature based on the 11th International Congress on Extremophiles held in Kyoto, Japan, September 12–16, 2016.

✉ Francine Perler  
PerlsOfWisdom@icloud.com

Kazuo Tori  
hiris5211@yahoo.co.jp

- <sup>1</sup> New England Biolabs, Inc., Ipswich, MA 01938, USA
- <sup>2</sup> Present Address: Takara Bio USA, Inc., 1290 Terra Bella Ave., Mountain View, CA 94043, USA
- <sup>3</sup> Present Address: Perls of Wisdom Biotech Consulting, Brookline, MA 02446, USA
- <sup>4</sup> Department of Molecular Biophysics and Biochemistry, Yale University, New Haven, CT 06520, USA

## Introduction

Inteins are protein splicing elements that are removed from precursor proteins by a self-catalytic mechanism. Inteins from extremophiles were key to defining all three currently

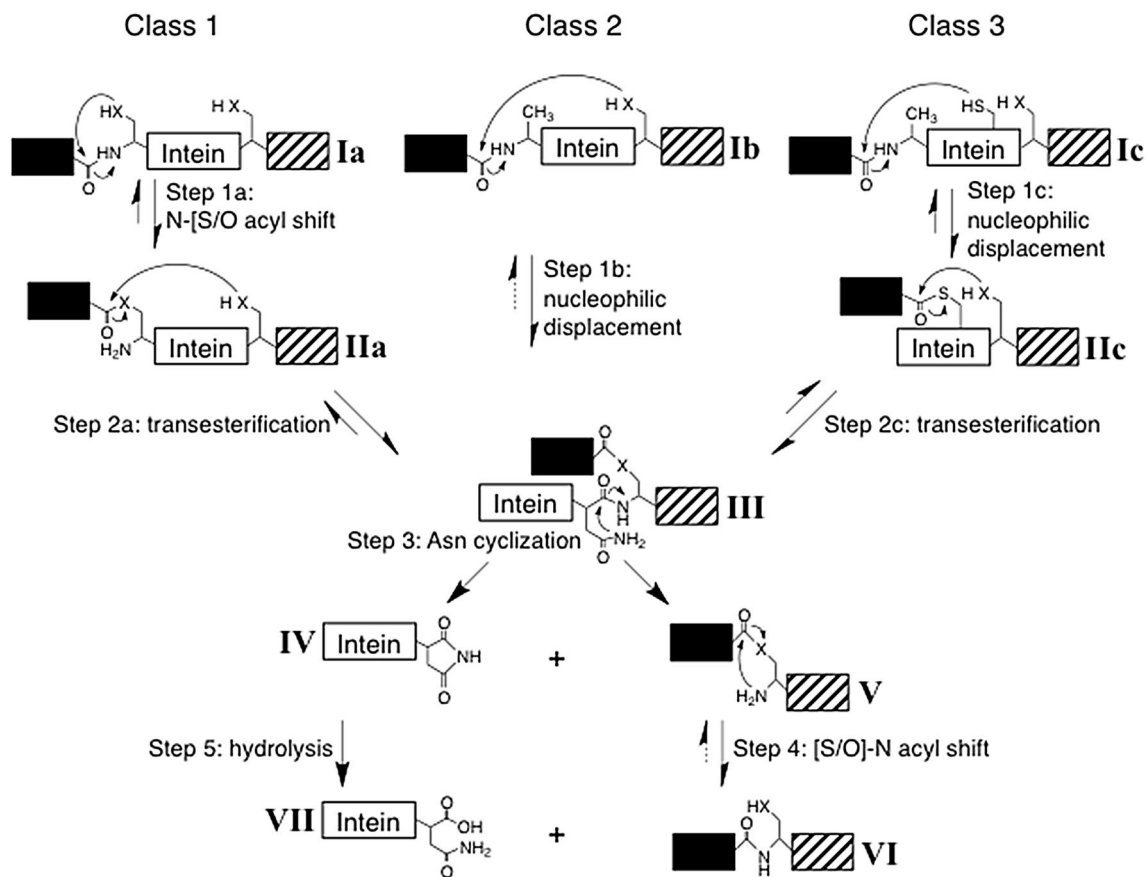
known mechanisms of intein-mediated protein splicing (Brace et al. 2010; Eryilmaz et al. 2014; Mills et al. 2014; Southworth et al. 2000; Tori et al. 2010; Volkmann and Mootz 2013; Xu et al. 1994; Xu and Perler 1996). Over 500 intein genes have been identified in numerous archaea, eubacteria, single cell eukaryotes and viruses (Perler 2002), and the numbers are growing rapidly with the explosion of archaeal and bacterial genome sequences. Inteins disrupt the function of their host protein (termed an extein) and thus protein splicing is required for survival if the intein is present in an essential protein (Dalgaard et al. 1997; Novikova et al. 2014; Perler 2002). The intein and extein are translated as a single, fused precursor protein. During post-translational maturation, the intein removes itself from this precursor while joining the flanking extein fragments with a native peptide bond. No external enzyme or cofactor is required. The majority of inteins are bifunctional enzymes that have a homing endonuclease domain as well as a protein splicing domain. The homing endonuclease is responsible for lateral transmission of intein genes, making them parasitic mobile genetic elements (Barzel et al. 2011; Novikova et al. 2014). Mini-inteins do not have an endonuclease domain, but retain the core protein splicing domain. Modern day mini-inteins are thought to be the descendants of inteins that lost their homing endonuclease domain (Barzel et al. 2011; Novikova et al. 2014).

Very few inteins have been characterized biochemically or even shown to be functional (Perler 2002). In this study we characterized the activity of three mini-inteins: the *Burkholderia vietnamiensis* G4 Bvi IcmO intein (Nordberg et al. 2014; Perler 2002), the *Mycobacterium smegmatis* MC2 155 Msm DnaB-1 intein (Mohan et al. 2015; Perler 2002), and the *Mycobacterium leprae* TN Mle DnaB intein (Eiglmeier et al. 1993; Perler 2002). DnaB is a replicative helicase in bacteria. The function of IcmO has yet to be established. The Bvi IcmO gene is present on a *B. vietnamiensis* plasmid, not the main chromosome (Nordberg et al. 2014). *B. vietnamiensis* G4 has a broad ecological range, fixes nitrogen, remediates trichloroethene and is found in the lungs of Cystic Fibrosis patients (Nordberg et al. 2014). *M. smegmatis* is a soil bacteria that forms biofilms and *M. leprae* is a slow growing human pathogen.

Inteins are currently divided into three classes based on differences in their protein splicing mechanism (Fig. 1) and conserved signature sequences (Tori et al. 2010). All inteins have at least four conserved motifs (Blocks A, B, F and G) in the splicing domain (Perler 2002; Perler et al. 1997; Pietrokovski 1994, 1998). Amino acids (aa) within these conserved motifs are numbered using the block designation and the position within the block, separated by a colon (Perler 2002; Tori et al. 2010). For example, the fourth amino acid in Block F is referred to as F:4. Several amino acids in these conserved motifs are present at or near

the intein active site where they directly act as nucleophiles and facilitating residues for catalysis or they assist in proper packing of the intein active site to align catalytic residues (Brace et al. 2010; Eryilmaz et al. 2014; Kawasaki et al. 1997; Mills et al. 2014; Romanelli et al. 2004; Southworth et al. 2000; Tori et al. 2010; Volkmann and Mootz 2013; Xu et al. 1994; Xu and Perler 1996). Catalytically important residues common to all classes of inteins include His<sup>B:10</sup> (essential for N-terminal splice junction reactions), the intein penultimate His<sup>G:6</sup> (assists in C-terminal splice junction reactions), the intein C-terminal Asn<sup>G:7</sup> (responsible for C-terminal splice junction cleavage during resolution of the Block G branched intermediate, BI<sup>G</sup>) and Ser<sup>+1</sup>, Cys<sup>+1</sup> or Thr<sup>+1</sup> [the G:8 residue, responsible for BI<sup>G</sup> formation and conversion of a (thio)ester bond to an amide peptide bond between the ligated exteins]. Each intein has a larger set of facilitating residues that are tuned to its specific set of nucleophiles (Eryilmaz et al. 2014; Kawasaki et al. 1997; Mills et al. 2014; Romanelli et al. 2004; Tori et al. 2010; Volkmann and Mootz 2013; Wu et al. 2014; Xu and Perler 1996). Moreover, amino acids in the extein, especially proximal ones, influence protein splicing by directly participating in catalysis or by affecting the architecture of the intein active site (Amitai et al. 2009; Cheriyan et al. 2013, 2014; Chong et al. 1997; Eryilmaz et al. 2014; Iwai et al. 2006; Liu et al. 2014; Muona et al. 2010; Perler et al. 1994, 1997; Southworth et al. 1999; Volkmann and Mootz 2013; Xu et al. 1994; Xu and Perler 1996). Class 3 inteins have an additional signature sequence (Table 1) consisting of a dispersed triplet: Trp<sup>B:12</sup>, Cys<sup>F:4</sup>, and Thr<sup>G:5</sup> (Tori et al. 2010). Previous studies indicated that mutation of Thr<sup>G:5</sup> can have a minimal effect, while mutation of Trp<sup>B:12</sup> more significantly impairs splicing, and mutation of Cys<sup>F:4</sup> totally blocks splicing (Brace et al. 2010; Tori et al. 2010; Tori and Perler 2011). Another feature of class 3 inteins is Ser<sup>+1</sup> or Thr<sup>+1</sup> at the G:8 position instead of Cys<sup>+1</sup>, which dominates this position in other intein classes (Brace et al. 2010; Perler 2002; Southworth et al. 2000; Tori et al. 2010; Tori and Perler 2011). The Bvi IcmO intein is the only currently known class 3 intein with Cys<sup>+1</sup>.

Protein splicing requires two or three catalyzed nucleophilic displacement reactions, depending on the intein class, followed by a spontaneous acyl rearrangement that results in a peptide bond between the ligated exteins (Fig. 1). Inteins perform these reactions by functioning as single turnover enzymes. Single turnover enzymes use the same methods as traditional enzymes to achieve catalysis, but do not act on multiple substrates. Although an intein was originally defined as the sequence that is removed from a precursor protein (Perler et al. 1994), when discussing the enzymatic properties of an intein we include the +1 aa (G:8) because it actively participates in splicing. The majority of inteins belong to class 1, which splices in four



**Fig. 1** Inteин-mediated protein splicing mechanisms. The majority of inteins follow the class 1 inteин-mediated protein splicing mechanism, which consists of four coordinated nucleophilic displacements and requires Ser<sup>1</sup>, Thr<sup>1</sup> or Cys<sup>1</sup> as the inteин N-terminal residue. Step 1a results in a linear (thio)ester intermediate and step 2a results in BI<sup>G</sup> with Cys<sup>+1</sup>, Ser<sup>+1</sup> or Thr<sup>+1</sup> as the branch point. Class 2 and 3 inteins do not require an inteин N-terminal nucleophile. Class 2 inteins directly form BI<sup>G</sup> when the +1 residue attacks the N-terminal splice junction peptide bond. Class 3 inteins use a conserved Cys at Block F position 4 (Cys<sup>F:4</sup>) to initiate protein splicing resulting in formation of the class-specific BI<sup>F</sup>. Once BI<sup>G</sup> is formed, the remaining

reactions are the same for all inteins. The acyl shift in step 4 is rapid and spontaneous. Step 5 is also spontaneous, but is often slow. *Solid arrows* represent steps that have been experimentally verified while *dashed arrows* represent theoretical steps. Note that steps 1 and 2 are reversible; the forward reactions are driven by kinetic rates, equilibrium positions toward the forward reaction, and substrate/intermediate elimination as the protein moves toward the final products, among other factors. Inteин residues and flanking exteин residues that assist these reactions are not shown, nor are tetrahedral intermediates. ‘X’ represents the sulfur or oxygen atom in the side chain of Cys, Ser or Thr

**Table 1** Inteин conserved motifs

Inteин	Block A	Block B	Block F	Block G
Bvi IcmO	1	10	4	67/8
Mle DnaB	/PQPLHSLVRMADG	GRSVEAARVHHWPV	PARCLVVADEMHCYI	HDIVTHN/ <u>C</u>
Msm DnaB-1	/ALALDTPLPTPTG	GTIVIVADAQHQPWT	PVRCVEVDNAAHLYL	GMVPTHN/ <u>S</u>
	/ALALDTPLPTPSG	GTAIVADAQHQPWT	PVRCVEVDNPEHLYL	GMVPTHN/ <u>S</u>
	*	*	*	*

The sequence of each motif is listed with the inteин class 3 signature positions marked by an asterisk. The slash denotes each splice junction. The position number within each block of selected catalytically important amino acids is listed. The class 3 specific BI<sup>F</sup> branch point (F:4) is shaded gray and the BI<sup>G</sup> branch point common to all inteins (G:8 or +1) is underlined

well-known steps (Eryilmaz et al. 2014; Mills et al. 2014; Perler 2002; Volkmann and Mootz 2013). Class 1 is defined by an intein N-terminal Ser<sup>1</sup>, Thr<sup>1</sup> or Cys<sup>1</sup> that forms a linear (thio)ester intermediate (IIa, Fig. 1) prior to forming BI<sup>G</sup> (III). Class 2 inteins do not have a Cys<sup>1</sup>, Thr<sup>1</sup> or Ser<sup>1</sup> (Perler 2002; Southworth et al. 2000). Instead, the +1 aa directly attacks the peptide bond at the N-terminal splice junction (step 1b) to yield a standard BI<sup>G</sup> (Southworth et al. 2000). Only class 2 inteins can perform step 1b. Class 3 inteins are similar to class 2 inteins since an intein N-terminal Cys<sup>1</sup>, Thr<sup>1</sup> or Ser<sup>1</sup> is not required for splicing (Brace et al. 2010; Perler 2002; Tori et al. 2010; Tori and Perler 2011). However, class 3 inteins initiate splicing when the class-specific conserved Cys<sup>F:4</sup> attacks the peptide bond at the N-terminal splice junction (step 1c) resulting in the formation of the class 3-specific Block F BI (BI<sup>F</sup>, IIc) (Brace et al. 2010; Tori et al. 2010; Tori and Perler 2011). BI<sup>G</sup> (III) is then formed by a transesterification reaction (step 2c). All inteins follow the same pathway for BI<sup>G</sup> resolution and formation of the peptide bond between the exteins.

This study examined whether three mini-inteins lacking an N-terminal Ser<sup>1</sup>, Thr<sup>1</sup> or Cys<sup>1</sup> are functional and the mechanism by which they splice. For the first time we were able to test whether there is a mandatory order of BI formation in a class 3 intein that naturally has Cys<sup>+1</sup> and an N-terminal splice junction already tuned for attack by a Cys nucleophile.

## Materials and methods

### Cloning, mutagenesis, and protein expression

All clones were sequenced by the New England Biolabs core facility. The genes for the Msm DnaB-1 and Mle DnaB inteins with flanking DnaB extein sequences and appropriate restriction enzyme sites were synthesized by Invitrogen (Carlsbad, CA, USA). The gene encoding the Bvi IcmO intein with flanking IcmO extein residues was amplified by PCR from *B. vietnamiensis* G4 genomic DNA using Phusion DNA polymerase and primers containing *Xho*I and *Spe*I restriction enzyme sites. The Msm DnaB-1 and Mle DnaB intein precursors included native DnaB flanking residues Phe-Gly-Val-Gly-Lys (N-extein) and Ser-Thr-Leu-Gly-Leu (C-extein), while the Bvi IcmO intein precursor included native flanking IcmO residues Ala-Arg-Ser-Leu-Gly-Phe (N-extein) and Cys-Ile-Thr-Phe-Ala (C-extein). The DNAs were digested by *Xho*I and *Spe*I, agarose gel purified and ligated into pMP1 (Southworth et al. 2000; Tori et al. 2010) previously digested with the same enzymes. This resulted in pM<sub>SP</sub> with the M<sub>sm</sub> DnaB-1 intein, pM<sub>LP</sub> with the M<sub>le</sub> DnaB intein and pM<sub>VP</sub> with the B<sub>vi</sub> IcmO intein, where the intein was flanked with the *Escherichia*

*coli* maltose-binding protein (M or MBP) and the  $\Delta$ *Sal* fragment of *Diroflaria immitis* paramyosin (P).

By convention, amino acids in the intein are numbered beginning with the intein N-terminus and residues in the C-extein are numbered similarly, but contain a plus sign to denote the C-extein. His<sup>B:10</sup>, Trp<sup>B:12</sup>, Cys<sup>F:4</sup> and Asn<sup>G:7</sup> are, respectively, His<sup>65</sup>, Trp<sup>67</sup>, Cys<sup>118</sup> and Asn<sup>139</sup> in the Msm DnaB-1 intein, His<sup>65</sup>, Trp<sup>67</sup>, Cys<sup>124</sup> and Asn<sup>145</sup> in the Mle DnaB intein and His<sup>65</sup>, Trp<sup>67</sup>, Cys<sup>122</sup> and Asn<sup>142</sup> in the Bvi IcmO intein (Perler 2002).

All mutations were made using the Phusion site-directed mutagenesis kit (New England Biolabs) with primers that introduced the desired mutation. For protein expression, freshly transformed *E. coli* NEB Turbo cells were grown in LB media containing 100  $\mu$ g/ml ampicillin at 37 °C until reaching an OD<sub>600</sub> of ~0.5 and then induced with 0.4 mM IPTG at room temperature, 30 °C and 37 °C for 2 h or at 15 °C overnight. Protein splicing was assessed using soluble lysates after electrophoresis in 4–20 % SDS-PAGE (Invitrogen, Carlsbad, CA, USA). Proteins were detected by either Simply Blue Safe Stain (Invitrogen) or fluorescent Western Blot as described previously (Cheriyian et al. 2013; Southworth et al. 2000; Tori and Perler 2011). Briefly, nitrocellulose filters were concurrently probed with mouse anti-MBP sera and rabbit anti-paramyosin sera, and then developed concurrently with IR-Dye 680 anti-mouse secondary antibody (green) or IR-Dye 800 anti-rabbit secondary antibody (red) (LI-COR, Lincoln, NE, USA).

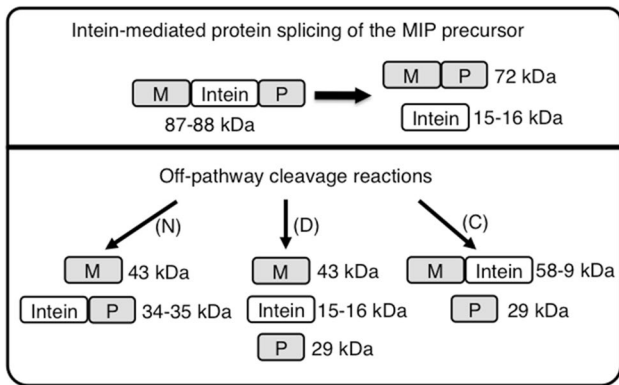
### Purification and characterization of branched intermediates

BIs of Msm DnaB-1 and Mle DnaB inteins with Asn<sup>G:7</sup> mutated to alanine in MVP and MLP were purified by affinity chromatography over amylose resin. The pH of an aliquot of each purified BI was changed by addition of sodium phosphate buffer. Samples were incubated overnight at room temperature in the absence or presence of 50 mM DTT. Time zero ( $T_0$ ) samples were heated at 100 °C for 5 min in SDS-PAGE sample buffer (New England Biolabs) without incubation. An aliquot of purified MLP BI was denatured by adding solid urea to a final concentration of 8 M. The pH of the denatured sample was checked prior to subsequent treatment as above for native samples.

## Results

### Splicing of three mini-inteins in vivo

DNAs encoding the Msm DnaB-1 (139 aa) and Mle DnaB (145 aa) mini-inteins along with 5 DnaB extein residues flanking the intein on each side (Perler 2002) were



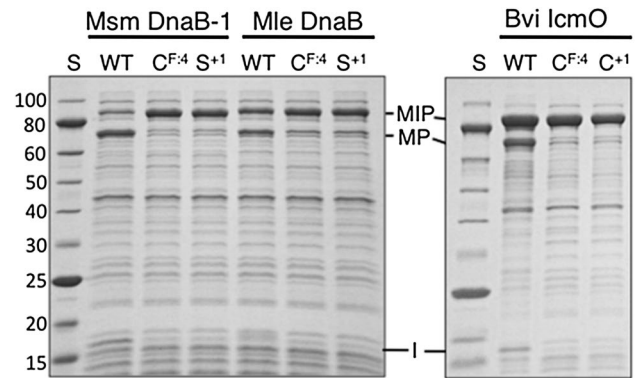
**Fig. 2** Splicing and cleavage schemes for the MIP precursor with either the Bvi IcmO, Mle DnaB or Msm DnaB-1 inteins. Off-pathway cleavage reactions can occur at the N-terminal splice junction (N), the C-terminal splice junction (C) or both splice junctions (D). Branched intermediate decay yields the same products as direct N-terminal splice junction cleavage (N)

synthesized and cloned in the MIP model precursor system (Xu et al. 1994) between the *E. coli* maltose binding protein (MBP or M) and the *D. immitis* paramyosin  $\Delta$  Sal fragment (P) generating precursors MSP and MLP, respectively. The coding sequence for the Bvi IcmO intein (142 aa) was amplified by PCR with flanking IcmO extein residues and likewise cloned into MIP generating MVP. Splicing of these model precursors (Fig. 2) results in production of MP (72 kDa) plus free intein (I, 14.7–16.0 kDa). Off-pathway cleavage reactions would result in production of M (43 kDa) and IP (34–35 kDa) after N-terminal cleavage or BI decay, and MI (58–59 kDa) and P (29 kDa) after C-terminal cleavage.

Splicing in model systems can vary at different temperatures due to differences in expression rates, precursor folding and aggregation. Therefore, in vivo splicing activity of all three inteins was examined after each precursor was expressed in *E. coli* at 37 °C, 30 °C and room temperature for 2 h or at 15 °C overnight. Msm DnaB-1 and Mle DnaB inteins yielded spliced products only at 15 °C (Fig. 3 and data not shown). MVP spliced poorly with less than half of the MVP precursor converted to spliced product at all temperatures tested (Fig. 3 and data not shown). These results demonstrate that all three mini-inteins are active, although the degree of splicing in these model precursors varied with the intein. Low levels of spliced product generally reflect misfolding of model precursors, especially when the precursor accumulates as a nonreactive component.

#### Dissecting the splicing pathway by mutation of conserved amino acids

Although all three inteins have the class 3 signature motif (Table 1), they can potentially splice by either the class 2

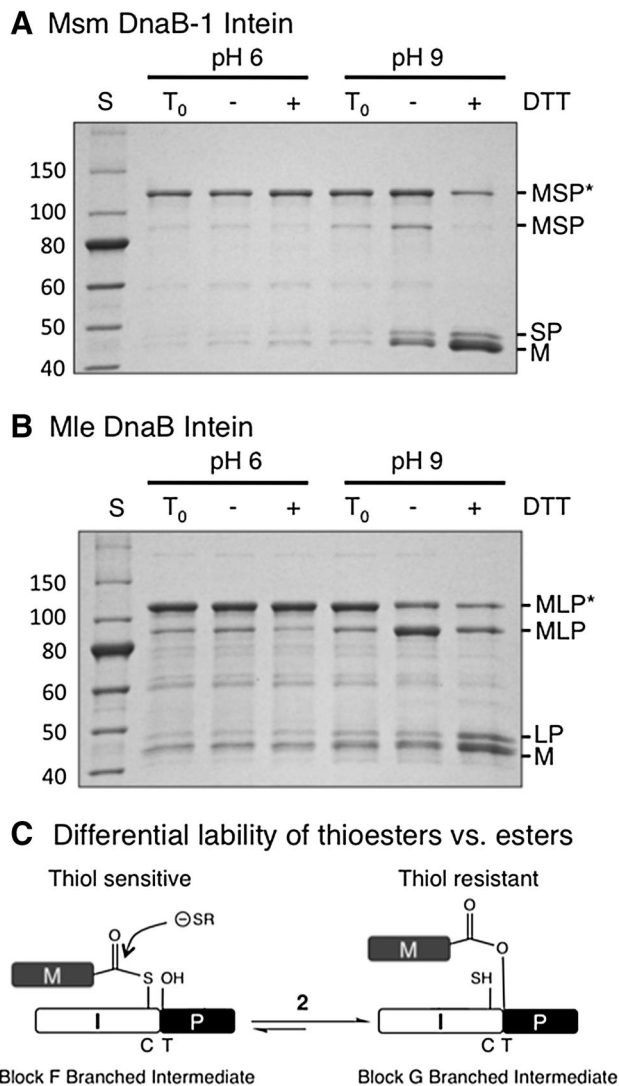


**Fig. 3** Splicing of wild type and mutant inteins in model precursors. MSP (Mle DnaB intein) and MLP (Msm DnaB-1 intein) were expressed at 15 °C and MVP (Bvi IcmO intein) was expressed at 30 °C. In vivo splicing of soluble proteins was examined after electrophoresis in SDS-PAGE followed by staining with Simply Blue Safe Stain. Residues mutated to alanine are listed above each lane. Lane S contains the NEB 10–250 kDa Protein Ladder with molecular masses (kDa) listed on the left. MIP precursor with the indicated inteins, MP spliced product, I excised intein, WT wild type intein

or class 3 mechanism. Mutation of Cys<sup>F:4</sup> can distinguish between these splicing mechanisms because it blocks the first step in the class 3 splicing pathway, but not in the class 2 pathway (Fig. 1). No splicing or cleavage was observed in MLP, MSP or MVP after substitution of Cys<sup>F:4</sup> with alanine when analyzed by Simply Blue Safe Stained SDS-PAGE or fluorescent Western Blot (Fig. 3 and data not shown). Only precursor was observed when His<sup>B:10</sup> or Trp<sup>B:12</sup> was mutated to alanine in all three inteins, which is consistent with their importance for splicing of class 3 inteins. Nonconservative substitution of Ser<sup>+1</sup> or Cys<sup>+1</sup> to alanine yielded unreacted precursor in Simply Blue Safe Stained SDS-PAGE and by fluorescent Western Blot analysis (Fig. 3 and data not shown). However, splicing was observed after the conservative substitution of Ser<sup>+1</sup> with Cys in MSP and MLP, while substitution of Cys<sup>+1</sup> with Ser blocked splicing and yielded only unreacted MVP precursor as analyzed in fluorescent Western Blots and stained SDS-PAGE (data not shown). These results are consistent with data from many other inteins where cysteine could substitute for a catalytic serine or threonine, but serine could not substitute for a catalytic cysteine; these effects are attributed to differences in pKa and the greater need to activate serine or threonine side chain hydroxyls to increase their nucleophilicity compared to cysteine side chain thiols (Eryilmaz et al. 2014; Mills et al. 2014; Volkmann and Mootz 2013). Such differences in reactivities and overall amino acid size are factors contributing to the tuning of an enzyme to its specific catalytic residues. Taken together, the mutation data indicate that all three mini-inteins follow the class 3 splicing pathway (Fig. 1).

## In vitro analysis of the Msm DnaB-1 and Mle DnaB intein branched intermediates

Asn<sup>G:7</sup> to alanine mutations in MSP and MLP resulted in BI accumulation (Fig. 4), which is normally seen as a slowly migrating band in SDS-PAGE compared to the initial precursor. These results are consistent with the role of Asn<sup>G:7</sup> in branch resolution (step 3, Fig. 1). The nature of the BIs that accumulated in vivo was then examined



**Fig. 4** Investigation of MSP and MLP branched intermediates in vitro. MSP (a) and MLP (b) BIs accumulated in vivo when the intein C-terminal Asn<sup>G:7</sup> was mutated to alanine. Purified BIs (MSP\* or MLP\*) were incubated in vitro at pH 6 or 9 in the absence (–) or presence (+) of 50 mM DTT at room temperature. Time zero samples ( $T_0$ ) were not incubated in vitro. The SDS-PAGE gel was stained with Simply Blue Safe Stain. Lane S NEB 10–250 kDa Protein Ladder with molecular masses (kDa) listed on the left of the gels. c Step 2 in the splicing pathway of class 3 inteins reversibly converts BI<sup>F</sup> to BI<sup>G</sup>. Under the conditions of these experiments, thiol reagents can attack a thioester bond, but not an ester bond

in vitro. Purified BI samples from MSP and MLP after Asn<sup>G:7</sup> to alanine substitution were incubated overnight at room temperature at either pH 6 or pH 9 in the presence or absence of 50 mM DTT (Fig. 4). The BIs were stable at pH 6 under all conditions tested. At pH 9, in the presence of DTT the BIs decayed to form M + SP or LP, while in the absence of DTT the BIs partially reverted back to MSP or MLP precursors. Stability was also tested with denatured MLP BI. No change was observed upon incubation of the denatured MLP BI sample overnight at room temperature with 50 mM DTT at either pH (data not shown), indicating that thiol induced BI decay requires a properly folded protein and that the steady state BI molecules do not have detectable amounts of a thiol sensitive bond.

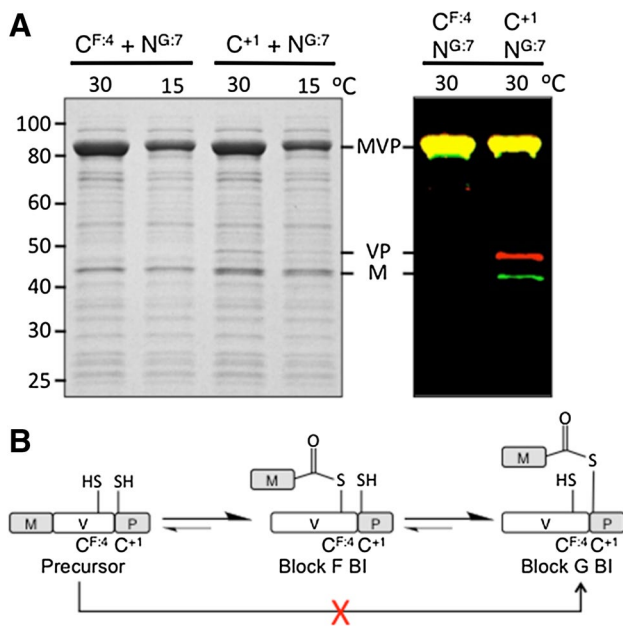
## Characterization of catalytically important cysteines in the Bvi IcmO intein

The Bvi IcmO intein is the only currently identified class 3 intein with Cys<sup>+1</sup> and thus both BI<sup>F</sup> and BI<sup>G</sup> have a thioester linkage. Mutation of the Bvi IcmO intein Asn<sup>G:7</sup> did not result in BI accumulation in vivo and instead yielded N-terminal splice junction cleavage products (M + VP). This is consistent with previous studies that demonstrated in vivo lability of BI thioester linkages (Brace et al. 2010; Mills et al. 2014; Southworth et al. 2000; Tori et al. 2010; Tori and Perler 2011; Volkmann and Mootz 2013).

The Bvi IcmO intein provides a unique opportunity to examine any potential competition between the catalytic cysteines at F:4 and G:8 (+1), and to determine if Cys<sup>+1</sup> can directly attack an N-terminal splice junction that is already tuned for cleavage by a cysteine. Single alanine substitutions of Cys<sup>F:4</sup> or Cys<sup>+1</sup> yielded only unreacted precursor in both stained gels and fluorescent Western Blots (Fig. 3 and data not shown). However, when alanine substitutions of each cysteine were combined with alanine substitutions of Asn<sup>G:7</sup>, the Cys<sup>+1</sup> double mutant produced N-terminal cleavage products while the Cys<sup>F:4</sup> double mutant did not, as assayed in stained SDS-PAGE and confirmed in the more sensitive fluorescent Western Blot (Fig. 5). Further experimentation is necessary to determine why cleavage products were not observed with the single MVP Cys<sup>+1</sup> mutant, especially structural studies of wild type and mutant inteins in MVP. It is possible that the Asn<sup>G:7</sup> mutation opens up the active site for in vivo cleavage of the thioester linkage in BI<sup>F</sup>.

## Discussion

The Msm DnaB-1, Mle DnaB and Bvi IcmO inteins should be class 3 inteins based on their sequence signatures (Tori et al. 2010) and phylogenetic analysis (Tori and Perler



**Fig. 5** Analysis of alanine substitutions at conserved cysteines in the Bvi IcmO intein. **a** Double mutations were made in the MVP precursor substituting alanine for Asn<sup>G:7</sup> and either Cys<sup>F:4</sup> or Cys<sup>+1</sup>. The Asn<sup>G:7</sup> substitution prevents on-pathway BI<sup>G</sup> resolution. MVP precursors with the indicated residues mutated to alanine were expressed at 30 °C for 2 h or 15 °C overnight. *Left panel* SDS-PAGE stained with Simply Blue Safe Stain. *Right panel* Western Blots were probed using IR-dye tagged secondary antibodies with M shown in green and P shown in red. Yellow bands indicate an equal reaction with both secondary antibodies. **b** The forward and reverse reaction scheme for MVP precursor to BI<sup>G</sup>. The red 'X' indicates that this reaction was not detected

2011). This study proved that all three inteins splice using the class 3 mechanism as demonstrated by (1) the absolute requirement of Cys<sup>F:4</sup> for splicing and N-terminal splice junction cleavage, (2) mutation of the class 3 specific signature residue Trp<sup>B:12</sup> blocked splicing, and (3) the inferred presence of BI<sup>F</sup> by thiol induced decay of purified BIs from MSP and MLP. It is interesting to note that C-terminal cleavage by Asn<sup>G:7</sup> in all three inteins appears to be strongly coupled to earlier steps in the splicing reaction as observed previously in some other inteins (Martin et al. 2001; Mills et al. 2014; Volkmann and Mootz 2013), since His<sup>B:10</sup>, Trp<sup>B:12</sup>, Cys<sup>F:4</sup>, Ser<sup>+1</sup> and Cys<sup>+1</sup> mutations did not yield C-terminal splice junction cleavage products.

It is difficult to experimentally distinguish between BI<sup>G</sup> (III, Fig. 1) and BI<sup>F</sup> (IIc) because they have the same mobility in SDS-PAGE, the same mass and have not been detected by Mass Spectrometry. Instead, indirect approaches have been employed to identify BIs, especially the difference in lability of BI thioester vs. ester linkages in vivo and in vitro. Thioester-linked BIs rarely accumulate in *E. coli*, while ester-linked BIs are often detected (Brace et al. 2010; Mills et al. 2014; Southworth et al. 2000; Tori

et al. 2010; Tori and Perler 2011; Volkmann and Mootz 2013; Xu et al. 1994; Xu and Perler 1996). The in vivo instability of thioester linked BIs was observed in this study since only products of BI thiolysis accumulated in MVP (both BIs have a thioester linkage) after BI<sup>G</sup> resolution was prevented by mutation of Asn<sup>G:7</sup>. Based on (thio) ester stability, the MSP and MLP BIs that accumulated in vivo after alanine substitution of Asn<sup>G:7</sup> are most likely the ester-linked BI<sup>G</sup> rather than the labile thioester linked BI<sup>F</sup> (Fig. 4c). These ester-linked BIs should be stable under the mild conditions used for in vitro thiolysis experiments. However, at pH 9 both the MSP and MLP BIs decayed during DTT treatment. At first glance, these results seem inconsistent because the in vivo data suggest accumulation of the ester linked BI<sup>G</sup>, while the in vitro data suggest the presence of the thioester containing BI<sup>F</sup>. The reversibility of step 2c (Figs. 1, 5b), explains this conundrum (Brace et al. 2010; Tori et al. 2010; Tori and Perler 2011). The reverse reaction was directly observed when MSP and MLP BIs converted back to linear precursors (Ic) at pH 9 in the absence of DTT (Fig. 4). If DTT is present during the reverse reaction, BI<sup>F</sup> is eliminated by DTT and purified BI<sup>G</sup> continually converts to BI<sup>F</sup> to maintain the equilibrium between the two BIs, eventually resulting in the complete disappearance of both BIs due to thiolysis of BI<sup>F</sup>. The observation that DTT was unable to cleave denatured MLP BI confirms the hypothesis that BI<sup>G</sup> and not BI<sup>F</sup> accumulates in vivo because direct cleavage of BI<sup>G</sup> by thiols would not occur under the experimental conditions (Fig. 4c) and denaturation only prevents forward or reverse reactions while leaving any (thio)ester bond intact.

The Bvi IcmO intein provided a sensitive means of analyzing the mandatory order of BI formation in class 3 inteins. It is possible that Cys<sup>+1</sup> can compete with Cys<sup>F:4</sup> for attack on the peptide bond at the N-terminal splice junction because this residue normally attacks the N-terminal splice junction in other intein classes during steps 2a or 1b (Eryilmaz et al. 2014; Mills et al. 2014; Southworth et al. 2000; Volkmann and Mootz 2013). Although previous studies with other class 3 inteins demonstrated that Ser<sup>+1</sup> cannot initiate the splicing reaction, they all required Ser<sup>+1</sup> to attack an N-terminal splice junction that was primed for attack by a cysteine nucleophile and thus may not work because of mechanistic differences caused by the different nucleophiles (Brace et al. 2010; Tori et al. 2010; Tori and Perler 2011). The Bvi IcmO intein allows testing of a naturally occurring Cys<sup>+1</sup> in a class 3 intein without the problems associated with changing the nucleophile. Cys<sup>+1</sup> did not generate any BI or N-terminal splice junction cleavage products in the sensitive fluorescent Western Blot assay of the MVP Cys<sup>F:4</sup> plus Asn<sup>G:7</sup> double mutant under conditions where cleavage products were produced by Cys<sup>F:4</sup> in the MVP Cys<sup>+1</sup> plus Asn<sup>G:7</sup> double mutant. It is unlikely

that undetected BI<sup>G</sup> formed or decayed in this experiment because the IR dyes used can detect picogram amounts of protein and microgram amounts of MVP precursor were queried. These results demonstrate that (1) the presence of a natural Cys<sup>+1</sup> cannot substitute for the loss of Cys<sup>F:4</sup> by mutation and (2) the order of BI formation in class 3 inteins must be BI<sup>F</sup> followed by BI<sup>G</sup> (Fig. 5b). This reaction order is likely maintained by local changes at the intein active site after formation of BI<sup>F</sup> that are required to properly align or activate Cys<sup>+1</sup> for the next catalytic step.

In summary, this study provides experimental evidence that all three class 3 mini-inteins are active. We conclude that there is no competition between the two catalytic cysteines for attack on the N-terminal splice junction and that the Bvi IcmO intein is unable to splice by the class 2 mechanism. The Bvi IcmO intein expands the repertoire of potential insertion sites for class 3 inteins in target proteins to include Cys for the numerous in vivo and in vitro applications based on intein technology (Aranko et al. 2014; Topilina and Mills 2014; Wood and Camarero 2014). Inteins continue to prove to be intriguing and robust single turnover enzymes.

**Acknowledgments** This work was funded by New England Biolabs, Inc. We thank Dr. Don Comb (Founder of NEB) for support and encouragement. We thank Dr. James Tiedje (Michigan State University) for kindly providing *Burkholderia vietnamiensis* G4 genomic DNA.

**Open Access** This article is distributed under the terms of the Creative Commons Attribution 4.0 International License (<http://creativecommons.org/licenses/by/4.0/>), which permits unrestricted use, distribution, and reproduction in any medium, provided you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made.

## References

- Amitai G, Callahan BP, Stanger MJ, Belfort G, Belfort M (2009) Modulation of intein activity by its neighboring extein substrates. *Proc Natl Acad Sci USA* 106:11005–11010
- Aranko AS, Wlodawer A, Iwai H (2014) Nature's recipe for splitting inteins. *Protein Eng Des Sel* 27:263–271
- Barzel A, Naor A, Privman E, Kupiec M, Gophna U (2011) Homing endonucleases residing within inteins: evolutionary puzzles awaiting genetic solutions. *Biochem Soc Trans* 39:169–173
- Brace LE, Southworth MW, Tori K, Cushing ML, Perler F (2010) The *Deinococcus radiodurans* Snf2 intein caught in the act: detection of the Class 3 intein signature Block F branched intermediate. *Protein Sci* 19:1525–1533
- Cheriyian M, Pedomallu CS, Tori K, Perler F (2013) Faster protein splicing with the *Nostoc punctiforme* DnaE intein using non-native extein residues. *J Biol Chem* 288:6202–6211
- Cheriyian M, Chan SH, Perler F (2014) Traceless splicing enabled by substrate-induced activation of the *Nostoc punctiforme* Npu DnaE intein after mutation of a catalytic cysteine to serine. *J Mol Biol* 426:4018–4029
- Chong S et al (1997) Single-column purification of free recombinant proteins using a self-cleavable affinity tag derived from a protein splicing element. *Gene* 192:271–281
- Dalgaard JZ, Moser MJ, Hughey R, Mian IS (1997) Statistical modeling, phylogenetic analysis and structure prediction of a protein splicing domain common to inteins and hedgehog proteins. *J Comput Biol* 4:193–214
- Eiglmeier K, Honore N, Woods SA, Caudron B, Cole ST (1993) Use of an ordered cosmid library to deduce the genomic organization of *Mycobacterium leprae*. *Mol Microbiol* 7:197–206
- Eryilmaz E, Shah NH, Muir TW, Cowburn D (2014) Structural and dynamical features of inteins and implications on protein splicing. *J Biol Chem* 289:14506–14511
- Iwai H, Zuger S, Jin J, Tam PH (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from *Nostoc punctiforme*. *FEBS Lett* 580:1853–1858
- Kawasaki M, Nogami S, Satow Y, Ohya Y, Anraku Y (1997) Identification of three core regions essential for protein splicing of the yeast Vma1 protozyme. A random mutagenesis study of the entire Vma1-derived endonuclease sequence. *J Biol Chem* 272:15668–15674
- Liu Z, Frutos S, Bick MJ, Vila-Perello M, Debelouchina GT, Darst SA, Muir TW (2014) Structure of the branched intermediate in protein splicing. *Proc Natl Acad Sci USA* 111:8422–8427
- Martin DD, Xu MQ, Evans TC Jr (2001) Characterization of a naturally occurring trans-splicing intein from *Synechocystis* sp. PCC6803. *Biochemistry* 40:1393–1402
- Mills KV, Johnson MA, Perler FB (2014) Protein splicing: how inteins escape from precursor proteins. *J Biol Chem* 289:14498–14505
- Mohan A, Padiadpu J, Baloni P, Chandra N (2015) Complete Genome Sequences of a *Mycobacterium smegmatis* Laboratory Strain (MC2 155) and Isoniazid-Resistant (4XR1/R2) Mutant Strains. *Genome Announc* 3(1):e01520-14. doi:10.1128/genomeA.01520-14
- Muona M, Aranko AS, Raulinaitis V, Iwai 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
- Nordberg H et al (2014) The genome portal of the Department of Energy Joint Genome Institute: 2014 updates. *Nucleic Acids Res* 42:D26–D31
- Novikova O, Topilina N, Belfort M (2014) Enigmatic distribution, evolution, and function of inteins. *J Biol Chem* 289:14490–14497
- Perler FB (2002) InBase: the intein database. *Nucleic Acids Res* 30:383–384
- Perler FB et al (1994) Protein splicing elements: inteins and exteins—a definition of terms and recommended nomenclature. *Nucleic Acids Res* 22:1125–1127
- Perler FB, Olsen GJ, Adam E (1997) Compilation and analysis of intein sequences. *Nucleic Acids Res* 25:1087–1093
- Petrokovski S (1994) Conserved sequence features of inteins (protein introns) and their use in identifying new inteins and related proteins. *Protein Sci* 3:2340–2350
- Petrokovski S (1998) Modular organization of inteins and C-terminal autocatalytic domains. *Protein Sci* 7:64–71
- Romanelli A, Shekhtman A, Cowburn D, Muir TW (2004) Semisynthesis of a segmental isotopically labeled protein splicing precursor: NMR evidence for an unusual peptide bond at the N-extein-intein junction. *Proc Natl Acad Sci USA* 101:6397–6402
- Southworth MW, Amaya K, Evans TC, Xu MQ, Perler FB (1999) Purification of proteins fused to either the amino or carboxy terminus of the *Mycobacterium xenopi* gyrase A intein. *Biotechniques* 27:110–114 (116, 118–120)
- Southworth MW, Benner J, Perler FB (2000) An alternative protein splicing mechanism for inteins lacking an N-terminal nucleophile. *EMBO J* 19:5019–5026



- Topilina NI, Mills KV (2014) Recent advances in in vivo applications of intein-mediated protein splicing. *Mob DNA* 5:5
- Tori K, Perler FB (2011) Expanding the definition of class 3 inteins and their proposed phage origin. *J Bacteriol* 193:2035–2041
- Tori K et al (2010) Splicing of the Mycobacteriophage Bethlehem DnaB intein: identification of a new mechanistic class of inteins that contain an obligate block F nucleophile. *J Biol Chem* 285:2515–2526
- Volkman G, Mootz HD (2013) Recent progress in intein research: from mechanism to directed evolution and applications. *Cell Mol Life Sci* 70:1185–1206
- Wood DW, Camarero JA (2014) Intein applications: from protein purification and labeling to metabolic control methods. *J Biol Chem* 289:14512–14519
- Wu Q, Gao Z, Wei Y, Ma G, Zheng Y, Dong Y, Liu Y (2014) Conserved residues that modulate protein trans-splicing of Npu DnaE split intein. *Biochem J* 461:247–255
- Xu MQ, Perler FB (1996) The mechanism of protein splicing and its modulation by mutation. *EMBO J* 15:5146–5153
- Xu MQ, Comb DG, Paulus H, Noren CJ, Shao Y, Perler FB (1994) Protein splicing: an analysis of the branched intermediate and its resolution by succinimide formation. *EMBO J* 13:5517–5522