

Splicing of the Mycobacteriophage Bethlehem DnaB Intein

IDENTIFICATION OF A NEW MECHANISTIC CLASS OF INTEINS THAT CONTAIN AN OBLIGATE BLOCK F NUCLEOPHILE^{*[§]♦}

Received for publication, September 23, 2009, and in revised form, November 18, 2009. Published, JBC Papers in Press, November 22, 2009, DOI 10.1074/jbc.M109.069567

Kazuo Tori^{‡§}, Bareket Dassa[¶], Margaret A. Johnson^{||}, Maurice W. Southworth[‡], Lear E. Brace[‡], Yoshizumi Ishino[§], Shmuel Pietrokovski^{¶1}, and Francine B. Perler^{‡2}

From [‡]New England Biolabs, Ipswich, Massachusetts, 01938, [§]Kyushu University, Fukuoka 812-8581, Japan, the [¶]Department of Molecular Genetics, Weizmann Institute of Science, Rehovot 76100, Israel, and the ^{||}Department of Molecular Biology, The Scripps Research Institute, La Jolla, California 92037

Inteins are single turnover enzymes that splice out of protein precursors during maturation of the host protein (extein). The Cys or Ser at the N terminus of most inteins initiates a four-step protein splicing reaction by forming a (thio)ester bond at the N-terminal splice junction. Several recently identified inteins cannot perform this acyl rearrangement because they do not begin with Cys, Thr, or Ser. This study analyzes one of these, the mycobacteriophage Bethlehem DnaB intein, which we describe here as the prototype for a new class of inteins based on sequence comparisons, reactivity, and mechanism. These Class 3 inteins are characterized by a non-nucleophilic N-terminal residue that co-varies with a non-contiguous Trp, Cys, Thr triplet (WCT) and a Thr or Ser as the first C-extein residue. Several mechanistic differences were observed when compared with standard inteins or previously studied atypical K1bA Ala¹ inteins: (a) cleavage at the N-terminal splice junction in the absence of all standard N- and C-terminal splice junction nucleophiles, (b) activation of the N-terminal splice junction by a variant Block B motif that includes the WCT triplet Trp, (c) decay of the branched intermediate by thiols or Cys despite an ester linkage at the C-extein branch point, and (d) an absolute requirement for the WCT triplet Block F Cys. Based on biochemical data and confirmed by molecular modeling, we propose roles for these newly identified conserved residues, a novel protein splicing mechanism that includes a second branched intermediate, and an intein classification with three mechanistic categories.

Inteins are the protein equivalent of introns. These intervening sequences must be post-translationally removed from the surrounding host protein fragments (exteins) before the host protein becomes functional. The intein together with the first C-extein amino acid act as a single turnover enzyme that removes the intein from the precursor protein and seamlessly

joins the exteins. No external cofactors or energy sources are required for this self-catalytic protein splicing reaction. The standard intein-mediated protein splicing pathway consists of four coordinated nucleophilic displacements (Fig. 1) and has been extensively reviewed (1, 2). The first step is an acyl rearrangement in which the first residue of the intein (Cys¹ or Ser¹) reacts with the carbonyl carbon of the preceding extein residue, resulting in a (thio)ester bond at the N-terminal splice junction (II). The first C-extein residue (Cys⁺¹, Ser⁺¹, or Thr⁺¹) then attacks this (thio)ester bond, cleaving the N-terminal splice junction and transferring the N-extein to its side chain to form the Block G branched (thio)ester intermediate (III). The Block G branched intermediate (BI)³ is resolved by cyclization of the intein C-terminal Asn, which results in cleavage of the C-terminal splice junction and release of the intein (IV) from the (thio)ester linked exteins (V). A native peptide bond is formed between the exteins by a spontaneous acyl shift (VI). The succinimide at the end of the intein (IV) eventually hydrolyzes to form Asn (VII) or isoasparagine.

The catalytic pathway in all inteins is highly coordinated, but the precise mechanism that controls this series of reactions remains unclear. The mechanism may also vary among different inteins. For example, some inteins display coupling of N-terminal and C-terminal reactions (3), whereas others do not (4). Intein mutations and/or the presence of artificial, non-native exteins often disrupt the coordination of these steps, yielding off-pathway single or double splice junction cleavage products that are incapable of forming the mature host protein (Fig. 2). Off-pathway C-terminal splice junction cleavage is caused by Asn cyclization in precursors (I or II) prior to BI (III) formation, whereas off-pathway N-terminal splice junction cleavage is caused by cleavage of the (thio)ester bond in the linear (II) or branched (III) (thio)ester intermediates. Off-pathway splice junction cleavage can result from either an increase in the rate of cleavage at that splice junction

* This work was supported in part by Israel Science Foundation Grant 626/03 (to S. P.).

♦ This article was selected as a Paper of the Week.

‡ Author's Choice—Final version full access.

[§] The on-line version of this article (available at <http://www.jbc.org>) contains supplemental File 1, Tables 1 and 2, and Figs. 1–7.

¹ Holds a Hermann and Lilly Schilling Foundation chair.

² To whom correspondence should be addressed: New England Biolabs, 240 County Rd., Ipswich, MA 01938. Tel.: 978-380-7326; Fax: 978-921-1350; E-mail: perler@neb.com.

³ The abbreviations used are: BI, branched intermediate; HINT, Hedgehog/intein, a structural fold shared by intein splicing domains and Hedgehog autoprocessing domains; TDS, temperature-dependent splicing; WCT triplet, co-varying Trp, Cys and Thr residues; WDT triplet, a variation of the WCT triplet where Cys is replaced Asp; Mja, *M. jannaschii*; MP-Be, mycobacteriophage Bethlehem; Sce, *Saccharomyces cerevisiae*; I, intein; MBP, maltose-binding protein; MP, MBP-paramyosin; M, *E. coli* MBP; P, the Δ Sal fragment of *D. immitis* paramyosin; M + IP, N-terminal cleavage products; MI + P, C-terminal cleavage products; MIP, precursor; MIP*, branched intermediate; DTT, dithiothreitol.

Class 3 Inteins

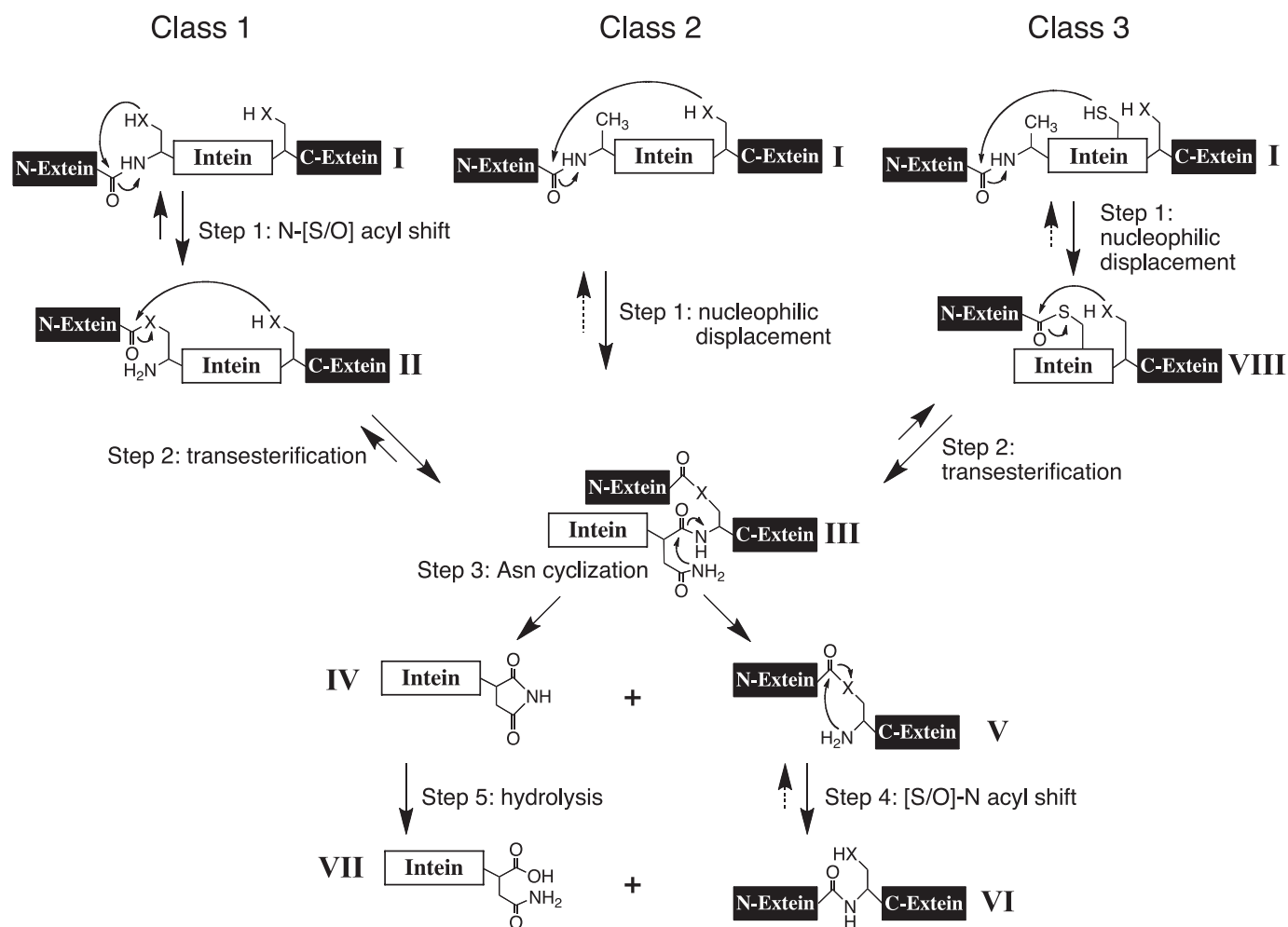


FIGURE 1. Three intein classes splice by different mechanisms. Most inteins follow the standard protein splicing pathway (*Class 1*). However, some inteins splice using modifications of this mechanism. The *Class 1* standard protein splicing pathway consists of four nucleophilic displacement reactions: *Step 1*, an *N*-(*S*/*O*) acyl shift at the *N*-terminal splice junction; *Step 2*, transesterification to form the Block G branched intermediate; *Step 3*, Asn cyclization to cleave the *C*-terminal splice junction; *Step 4*, a spontaneous (*S*/*O*)-*N* acyl shift to form a native peptide bond between the exteins; and *Step 5*, the slow hydrolysis of the succinimide ring. Although Thr could perform the initial acyl shift, it has not been observed in standard inteins (5, 9). *Class 2* and *Class 3* inteins lack an *N*-terminal Ser or Cys, so they cannot perform the acyl shift that initiates the splicing reaction in *Class 1* inteins. In *Class 2* inteins, Cys⁺¹ directly attacks an amide bond at the *N*-terminal splice junction to form the standard Block G branched intermediate (**III**). Known examples of *Class 2* inteins are the K1bA inteins. The *Class 3* protein splicing pathway also consists of four nucleophilic displacement reactions, including two branched intermediates. In *Step 1*, the Cys in Block F (Cys³²⁰ in the MP-Be DnaB intein) attacks the peptide bond at the *N*-terminal splice junction, forming the Block F branched intermediate (**VIII**), which is thiol labile. In *Step 2*, the *N*-extein is transferred from the side chain of the Block F Cys to the side chain of the +1 residue (Thr⁺¹ in the MP-Be DnaB intein) by a transesterification reaction resulting in the formation of the same Block G branched intermediate (**III**) as in *Class 1*; this is an ester linkage in the MP-Be DnaB intein and is proposed to be resistant to cleavage by thiols at room temperature. Once the Block G branched intermediate (**III**) is formed, the remainder of the splicing pathway is the same in all inteins. *Class 3* inteins contain the WCT triplet, whereas *Class 2* inteins do not. Residues within the intein assist these reactions. Tetrahedral intermediates are not shown. X represents the sulfur or oxygen atom in the side chain of Ser, Thr, or Cys.

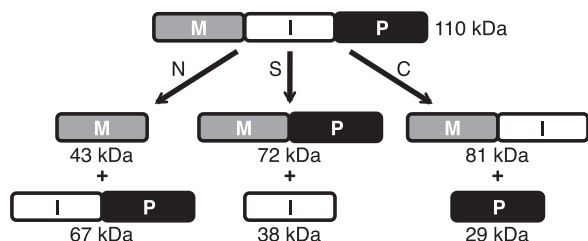


FIGURE 2. A schematic representation of splicing and off-pathway reactions. The MP-Be DnaB intein (I) was cloned into a model precursor consisting of MBP as the *N*-extein (M) and a fragment of paramyosin (P) as the *C*-extein, along with 5 native extein residues flanking each splice site. Protein splicing (S) yields MP and free intein, whereas off-pathway *N*-terminal splice junction cleavage (N) yields M + IP and off-pathway *C*-terminal splice junction cleavage (C) yields *C*-terminal cleavage products (MI + P). Cleavage of the (thio)ester linkage in branched intermediates also yields M + IP. Some mutations result in double cleavage at both splice junctions to yield M + I + P. Molecular weights of MP-Be DnaB intein MIP precursor and products are listed.

or a decrease in the reaction rate of another step. However, to our knowledge, neither splicing nor off-pathway cleavage of an amide bond at the *N*-terminal splice junction has been observed in standard inteins after non-conservative substitution of Ser¹ or Cys¹.

Over 450 inteins have been identified in archaea, bacteria, viruses, and unicellular eukaryotes (see InBase, the on-line intein data base) (5). Intein splicing domains contain four sequence motifs (see Fig. 3) that are not well conserved in their entirety but contain certain well conserved residues that form part of the protein splicing catalytic center and the hydrophobic core of the splicing domain. There are two concurrently used nomenclatures for these motifs: A or N1, B or N3, F or C2, and G or C1 (6–8). Block A begins at position 1 of the intein and contains the intein *N*-terminal nucleophile

(Ser¹ or Cys¹). Thr¹ has not been observed in inteins identified to date (5) but can perform the same acyl rearrangement as Ser or Cys and has been shown to allow splicing in the S1T mutant of the *Thermococcus litoralis* DNA pol-2 intein (9). Block B (N3) is usually 60–100 amino acids from the intein N terminus and contains a conserved Thr at position 7 and a His at position 10, both of which assist reactions at the N-terminal splice junction (1, 2, 6–8). Residues in Block F (C2), especially an Asp or a Trp in position 4, have been shown to be essential for N-terminal cleavage, leading many authors to conclude that Block F may be important in orienting the C-extein nucleophile for attack on the N-terminal scissile bond and/or may participate in catalysis at the N-terminal splice site (10–12). In the *Methanococcus jannaschii* (Mja) KlbA intein, Asp¹⁴⁷ likely assists catalysis by helping to activate the Cys⁺¹ nucleophile (12). Block G (C1) includes the last 7 amino acids of the intein and the first residue of the C-extein (termed the +1 amino acid). The intein penultimate His usually assists Asn cyclization, although some inteins use other residues for this purpose (1, 2, 12). Block F is generally <10 amino acids from Block G.

Inteins exhibit numerous polymorphisms in nucleophiles and assisting residues, with conserved positions populated by amino acids with similar functionalities (see Fig. 3) and positions of assisting residues varying among inteins (5). For example, the conserved C-terminal Asn is replaced by Asp or Gln in some inteins (2, 5, 13, 14). These residues are capable of similar cyclization reactions as Asn, although other mechanisms have been proposed (1, 2, 13, 14). The low sequence similarity between inteins makes the identification of assisting residues difficult and has limited generalization from one intein to another.

The first reported variation in the standard protein splicing pathway was necessitated by the absence of an N-terminal nucleophile in the Ala¹ inteins of archaeal KlbA proteins (15). These inteins have overcome the barrier to direct attack on an amide bond at the N-terminal splice junction that is present in standard inteins, likely due to a slight increase in the size of the splicing active site (12). In the Mja KlbA intein, the C-extein nucleophile (Cys⁺¹) directly attacks the amide bond at the N-terminal splice junction, yielding the standard Block G BI (III) (Fig. 1). The remainder of the splicing pathway is indistinguishable from standard Class 1 inteins (15).

This study examines the splicing of another intein that lacks an N-terminal nucleophile. The mycobacteriophage Bethlehem (MP-Be) DnaB intein begins with Pro¹ (16) and represents a new class of inteins that lack an N-terminal nucleophile, have a unique set of conserved residues, and can form a previously undetected additional BI at the Block F Cys.

EXPERIMENTAL PROCEDURES

Bioinformatic Sequence Analysis—The multiple sequence alignment of intein motifs is from the June 25, 2009 update of the InBase (5) splicing motifs page. The alignment included the 51 positions (columns) comprising the four conserved splicing motifs from 450 sequences in the data base after the removal of bacterial intein-like domains and joining of the N- and C-parts of each split intein. Forty-two inteins lacked an N-terminal Ser

or Cys (supplemental Table 1). The ShP-Sfv-2a-2457T-n_Primase and ShP-Sfv-2a-301-n_Primase inteins lack the standard N-terminal nucleophile but are partial inteins stopping before Block F and were not used in the co-variation analysis. This left 40 inteins lacking an N-terminal Ser or Cys, including the five KlbA intein orthologs. Co-variation analysis was performed on every pair of multiple sequence alignment columns using the Chi-square method. The co-variance is quantified by the standard Chi-square calculation method between the expected independent co-occurrence of the residues in two columns from a multiple sequence alignment relative to the observed occurrence of the residue pairs (17). To avoid possible bias by closely related sequences the alignment, sequences are weighted with the weights adjusted to sum to the number of the alignment sequences (18). For each pair of columns, only the sequences where the residues in both columns are defined (*i.e.* not gapped or unknown) are used in the calculation. Sequence logos were generated as described (6) with amino acids colored according to physico-chemical groups in Fig. 3: *red* for acidic, *blue* for basic, *black* for hydrophobic, *orange* for aromatic, *cyan* for Asn/Gln, and *green* for Thr/Cys/Ser.

Cloning and Mutagenesis—All clones were sequenced by the New England Biolabs Core facility, and all enzymes were obtained from New England Biolabs (Ipswich, MA) and used as described by the manufacturer. DNA encoding the MP-Be DnaB intein plus 5 native extein residues flanking each side was amplified by PCR using Phusion DNA polymerase (New England Biolabs) from mycobacteriophage Bethlehem genomic DNA (provided by Graham Hatfull, University of Pittsburgh) using the primers 5'-ATGTTTCTCGAGGTCTCCAGGAC-CAGCCT-3' and 5'-AATGAAACTAGTGAACGTGTTCT-TCGTGTT-3'. The MP-Be DnaB intein PCR product was digested with XhoI and SpeI, purified from an agarose gel with the Wizard SV gel and PCR clean-up system (Promega, Madison, WI), and ligated into pMP1, which was also digested with XhoI and SpeI. This resulted in pMIP, where the MP-Be DnaB intein (I) is flanked by the *Escherichia coli* maltose-binding protein (M) and the Δ Sal fragment of *Dirofilaria immitis* paramyosin (P) (15). pMP1 was derived from pMXP (19) by replacing the *Mycobacterium xenopi* GyrA intein with a cassette containing SpeI and SphI restriction sites inserted between XhoI and StuI.

Individual mutations were introduced into MIP using either the PhusionTM site-directed mutagenesis kit (New England Biolabs) or the QuikChange site-directed mutagenesis kit (Stratagene, La Jolla, CA) using the primers listed in supplemental Fig. 1. Multiple mutations were made sequentially by the same methods.

Expression, Purification, and Protein Characterization—All fusion proteins were expressed in *E. coli* NEB Turbo cells (New England Biolabs) by induction at A₆₀₀ 0.4–0.6 with 0.4 mM isopropyl- β -D-thiogalactoside for 2 h at 30 °C or overnight at 15 °C. Cells were sonicated in buffer A (20 mM NaPO₄ pH 8.0, 500 mM NaCl), and soluble lysates were either examined directly by gel electrophoresis or purified on amylose resin (New England Biolabs) in buffer A at pH 8.0 and eluted in buffer A containing 10 mM maltose. For analysis with Factor Xa protease, 60 μ g of purified MBP-paramyosin (MP) protein was digested with 2 μ g of Factor Xa protease at 37 °C as described by

Class 3 Inteins

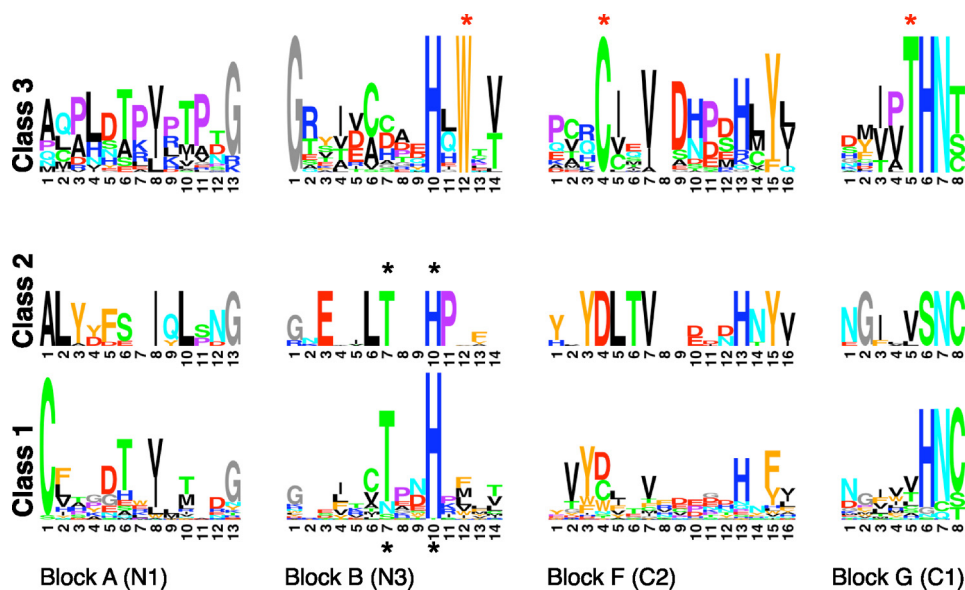


FIGURE 3. Comparison of conserved intein sequences. Intein splicing domains contain four conserved motifs consisting of 51 positions in this sequence logo representation (6) of the amino acids that populate these motifs in the three mechanistic classes of inteins. Class 3 inteins are described in this report as lacking the standard N-terminal Ser or Cys nucleophile and containing the WCT triplet. Class 2 inteins consist of K1bA-like inteins, and Class 1 inteins are the standard inteins, which comprise the largest class of inteins. The first 37 sequences in [supplemental Table 1](#) minus the Arsp-FB24 DnaB intein (WDT triplet) were used to generate the sequence logo for Class 3, the five K1bA inteins were used to generate the sequence logo for Class 2, and the remaining 408 inteins present in the June 25, 2009 version of InBase (5) were used in Class 1. Block F, position 8 in Class 3 and positions 8 and 9 in Class 2 represent gaps in this motif due to a variable sized loop. The absence of visible residues in other positions in the Class 2 sequence logo is due to a high degree of variability in these positions and the small class size (five inteins). Residues are colored according to physico-chemical groups: *red* for acidic, *blue* for basic, *black* for hydrophobic, *orange* for aromatic, *cyan* for Asn/Gln, and *green* for Thr/Cys/Ser. The *red* asterisks mark the positions of the WCT triplet, and the *black* asterisks mark the position of the conserved Thr and His in Block B. Block A is sometimes referred to as N1, B as N3, F as C2, and G as C1 (6–8).

the manufacturer (New England Biolabs). Samples were boiled for 5 min in sample buffer plus DTT (New England Biolabs), loaded onto 4–20% SDS-polyacrylamide gels (Invitrogen) and either stained with Coomassie Blue or transferred to nitrocellulose for Western blot analysis with an anti-MBP or anti-paramyosin antibody (15). Relative molecular masses were calculated in comparison with the New England Biolabs Broad Range Protein Marker (New England Biolabs). Coomassie Blue-stained gels were digitized with a Microtek Scanmaker III, and the signals were quantified with Quantity One (Bio-Rad). The values for at least two independent experiments were averaged. N-terminal protein sequencing was performed by the New England Biolabs core facility (15).

Analysis of the Denatured Branched Intermediate—The BI from the N341A mutant was purified on amylose resin. Samples were denatured by adding solid urea to a final concentration of 4.5 M, and then the buffer was exchanged to 6 M urea, 50 mM NaPO₄, and 0.5 M NaCl at either pH 5.0 or pH 9.0 using Microcon filtration devices (Millipore, Billerica, MA). The pH of each sample was checked prior to subsequent treatment. Samples were incubated at 50 °C for 0–6 h and electrophoresed as described above.

Comparative Modeling—A sequence alignment between the MP-Be DnaB HINT (Hedgehog/intein) domain and a mutated Sce VMA intein (20) was generated using the program FFAS (21). The Sce VMA intein included the following mutations in catalytically important residues, which were necessary to isolate the precursor: C1S (C284S in Ref. 20), the intein N termi-

nus; H79N (H362N in Ref. 20), Block B, position 10 His; N454S (N737S in Ref. 20), intein C terminus; and C+1S (C738S in Ref. 20). The alignment was submitted to the SWISS-MODEL web-based comparative modeling server (22, 23) in the alignment mode. Initial tests showed that the linker regions between the MP-Be endonuclease and HINT domains could not be modeled effectively by this method, and therefore we did not attempt to include the MP-Be endonuclease domain in the model. The model was evaluated using the programs ANOLEA (24), Verify3D (25), and GROMOS, as implemented in the SWISS-MODEL workspace, and in addition using the program WhatCheck. The protein geometry parameters calculated by WhatCheck were satisfactory, but the model exhibited a few regions with positive ANOLEA or GROMOS energies, indicating some defects in the local packing or electrostatic environment of some side chains, which are difficult to avoid when using a template with very

low sequence identity. The model exhibited a Verify3D cumulative score of 38.19 and average scores ranging from –0.02 to 0.57, which indicate a satisfactory overall structure (25).

RESULTS

The MP-Be DnaB Pro¹ Intein Splices *in Vivo*—The MP-Be DnaB helicase intein (341 amino acids) begins with Pro¹. It was cloned into a model MIP precursor with 5 native extein residues on each side (native N-extein: Val-Ser-Gln-Asp-Gln; native C-extein: Thr-Lys-Asn-Thr-Phe). The MIP precursor consists of the intein (I) inserted in-frame between the *E. coli* maltose-binding protein (MBP or M) and a fragment of *D. immitis* paramyosin (P) (15). Growth of *E. coli* transformants was very slow, suggesting toxicity potentially due to cleavage of the *E. coli* DnaB gene by the intein homing endonuclease domain. A healthy transformant was fortuitously obtained that had a spontaneous mutation in a motif known to be important for endonuclease activity (I224S). This endonuclease mutation was included in all subsequent constructs. Both the I224S (data not shown) and the wild type MP-Be DnaB intein MIP precursors spliced to completion *in vivo* after induction with isopropyl-β-D-thiogalactoside either at 30 °C for 2 h or overnight at 15 °C, as evidenced by the presence of spliced MP and excised MP-Be DnaB intein (I) and by the absence of MIP precursor or off-pathway cleavage products (Figs. 2 and 4, [Lane 3](#), [supplemental Fig. 2](#), and data not shown).

Soluble protein was affinity-purified on amylose resin yielding MP in the eluate and excised intein in the flow-through

(supplemental Fig. 2). N-terminal protein sequencing confirmed Pro¹ at the beginning of the intein. MP was characterized by N-terminal sequencing and reactivity with anti-MBP and anti-paramyosin sera (15) (data not shown). To study the splice junction after ligation, MP was subjected to Factor Xa protease treatment, which cleaves MP near the C terminus of the MBP subdomain to yield an N-terminal 43-kDa MBP fragment and a 29-kDa C-terminal fragment that begins 9 residues N-terminal to the splice junction. The experimentally determined sequence of the 29-kDa Factor Xa digestion product

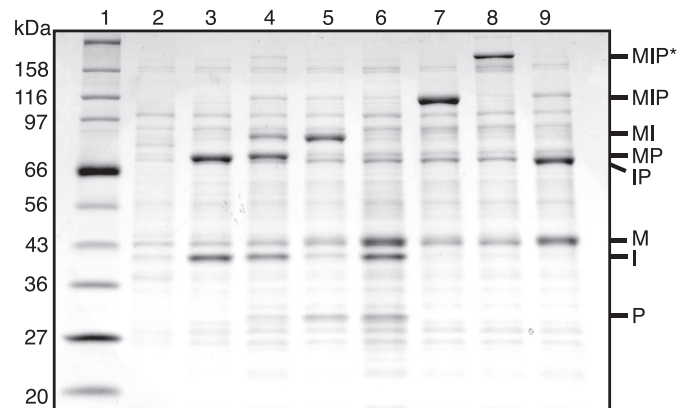


FIGURE 4. Splicing of wild type and mutant MP-Be DnaB inteins in a model precursor. MIP fusion proteins containing the indicated mutations were induced at 30 °C. Soluble protein was electrophoresed in SDS-PAGE followed by staining with Coomassie Blue. Lane 1, New England Biolabs Broad Range Protein Marker; Lane 2, uninduced cells; lane 3, wild type intein with I224S mutation in the endonuclease domain; lane 4, P1A mutant; lane 5, P1A plus T+1A double mutant; lane 6, T+1A mutant; lane 7, P1A plus N341A plus T+1A triple mutant; lane 8, N341A mutant; lane 9, N341A plus T+1A double mutant. Abbreviations are as in the legend for Fig. 2, with the addition of MIP*, which is the branched intermediate.

TABLE 1

The effects of mutating splice junction proximal residues

Mutations were made at the indicated positions in Blocks A and G. The products observed in samples expressed at either 30 °C or 15 °C are indicated. The single-letter amino acid code was used. N-cleavage and C-cleavage refer to off-pathway single splice junction cleavage reactions. Some mutations led to double cleavage (DC) at both splice junctions to yield M+I+P in the presence or absence of additional single splice junction cleavage. Abbreviations are as in the legend for Fig. 2. Key: 15/30, no difference observed at the two temperatures; blank, no detectable MIP substrate or product; +, <10% of total MIP substrate and products; ++, 10–50% of total MIP substrate and products; +++, 51–90% of total MIP substrate and products; + + + +, >90% of total MIP substrate and products.

| Position/substitution | Temperature | Splice (MP+I) | Precursor (MIP) | N-cleavage (M+IP) | C-cleavage (MI+P) | DC M+I+P | Branch (MIP*) |
|---------------------------------|-------------|---------------|-----------------|-------------------|-------------------|----------|---------------|
| °C | | | | | | | |
| Block A | | | | | | | |
| P1A | 15/30 | +++ | + | | ++ | | + |
| P1S | 30 | ++ | ++ | | + | | |
| P1S | 15 | +++ | + | | ++ | | |
| P1C | 30 | ++ | | | + | +++ | + |
| P1C | 15 | ++ | | | ++ | +++ | |
| Block G | | | | | | | |
| T339A | 30 | +++ | ++ | | | | |
| T339A,T339S,T339V | 15 | ++++ | | | | | |
| T339S,T339V | 30 | ++++ | + | | | | |
| H340A | 15/30 | ++++ | + | | | | |
| N341A | 15/30 | | | ++ | | | +++ |
| T+1A | 30 | | | ++ | | +++ | |
| T+1A | 15 | | | | | ++++ | |
| T+1C,T+1S | 15/30 | ++++ | | | | | |
| Double or triple mutants | | | | | | | |
| P1A,N341A | 15/30 | | ++ | ++ | | | |
| P1A,T+1A | 30 | | + | + | | | |
| P1A,T+1A | 15 | | | | +++ | ++ | |
| T+1A,N341A | 15/30 | | + | ++++ | | | |
| P1A,N341A,T+1A | 30 | | ++++ | + | | | |
| P1A,N341A,T+1A | 15 | | +++ | ++ | | | |
| S62A,S64A | 30 | ++++ | + | | | | |

began at the Factor Xa cleavage site in MBP and spanned the predicted splice junction (GTLEVSQDQ/TKNTFSLFPI). These results indicate that the products produced by this Pro¹ intein were the same as those produced by standard inteins.

Activity after Mutation of Splice Junction Residues—Gln⁻¹ and Pro¹ form the MP-Be DnaB intein N-terminal splice junction; Asn³⁴¹ and Thr⁺¹ form the C-terminal splice junction. The effects of mutating Pro¹, Asn³⁴¹, and Thr⁺¹ on splicing *in vivo* after expression at 30 or 15 °C were determined by analyzing soluble protein in SDS-PAGE stained with Coomassie Blue (Fig. 4 and Table 1). No notable differences were observed with mutants expressed at the two temperatures except for small differences in some Pro¹ substitutions.

The Thr⁺¹ C-extein nucleophile is required for Block G BI (III) formation (Fig. 1). Splicing was observed after substitution of Thr⁺¹ with functionally similar residues (Ser or Cys), but non-conservative substitution with Ala completely blocked splicing and yielded almost complete cleavage at both splice junctions (double cleavage yielding M, I, and P). Double cleavage was unexpected because no nucleophile known to attack the N-terminal scissile bond was still present in the T+1A precursor, and the absence of Ser¹/Thr¹/Cys¹ precluded the formation of a labile (thio)ester bond at this position. To our knowledge, hydrolysis of an amide bond at an N-terminal splice junction has not been observed in standard or Klba inteins. However, Pro¹ might impart or require increased lability of the N-terminal splice junction, so the effects of mutating Pro¹ were explored. Splicing was reduced and C-terminal cleavage was observed in Ser, Cys, and Ala substitutions of Pro¹, which is similar to what occurred with Klba Ala¹ inteins (15). Double cleavage predominated in the P1C mutant. Combining T+1A with P1A still resulted in cleavage at both splice junctions, prov-

Class 3 Inteins

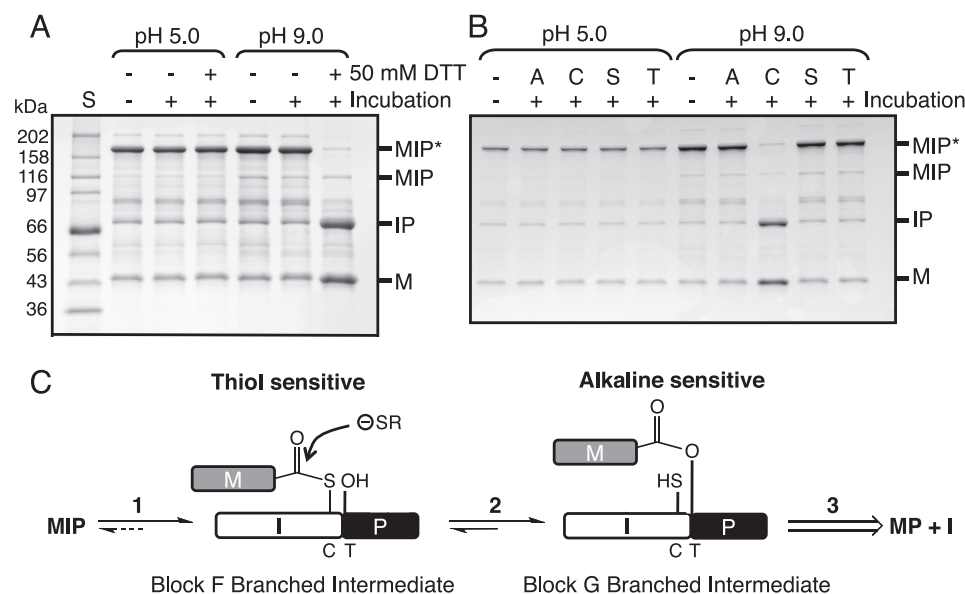


FIGURE 5. The MIP* branched intermediate decays when treated with thiols. The N341A mutant yielded large amounts of MIP*, which was purified by affinity chromatography on amylose resin and adjusted to either pH 5 or pH 9. MIP* was then incubated at room temperature for 14 h in the presence (+) or absence (-) of 50 mM DTT (A) or in the presence (+) or absence (-) of 50 mM Ala, Cys, Ser, or Thr (B). A Coomassie Blue-stained SDS-PAGE is shown. Abbreviations are as in the legend for Fig. 2, with the addition of the use of the single letter amino acid codes and S for New England Biolabs Broad Range Protein Marker. C, proposed pathway for thiol-induced cleavage of the Block G BI. Steps 1 and 2 are reversible, with the equilibrium in step 2 favoring the Block G BI. Step 3, which includes Asn cyclization and the O-N acyl shift, is irreversible. In the N341A mutant, Block G BI can be isolated and unfolded with urea, making Step 2 irreversible. The observed decay of the Block G BI in response to thiols is due to elimination of the Block F BI by thiols and the equilibrium between the 2 BI. The Cys branch point in Block F (Cys³²⁰ in the MP-Be DnaB intein) and the Thr⁺¹ branch point in Block G are indicated below the rectangles as C and T, respectively.

ing that Pro¹ is not required for the off-pathway N-terminal cleavage observed in T+1A mutants. Combining T+1A with N341A yielded >90% N-terminal cleavage, indicating that the unknown mechanism of N-terminal cleavage in the T+1A mutant does not require Asn³⁴¹ or C-terminal cleavage. Remarkably, the P1A,N341A,T+1A triple mutant was still capable of generating a small amount of N-terminal cleavage products. These results demonstrate that either the MP-Be DnaB intein N-terminal splice junction is highly activated and thus easily hydrolyzes or a previously unrecognized nucleophile is capable of attacking it.

The intein C-terminal Asn is required for resolution of the Block G BI (III) by cleaving the C-terminal splice junction during succinimide ring formation (Fig. 1). Mutation of Asn³⁴¹ to Ala blocked C-terminal cleavage, resulting in accumulation of the MIP* Block G BI (III) and its breakdown N-terminal cleavage products. Neither Thr⁺¹ nor Pro¹ is required for Asn cyclization and cleavage at the C-terminal splice junction as C-terminal cleavage was observed in the P1A plus T+1A double mutant.

Analysis of the Branched Intermediate—Several lines of evidence indicated that MIP* was the BI. (a) Mobility in SDS-PAGE was slower than MIP, (b) it reacted with anti-MBP and anti-paramyosin sera (data not shown), (c) the expected pair of residues from the two predicted N termini were present in each cycle of Edman degradation, and (d) it had an alkaline-labile linkage connecting M and IP, which is indicative of an ester bond. Alkaline lability of the bond between M and IP was evident by the first time point (1 h) at pH 9.0 after purified MIP* was denatured to prevent reversion to the linear MIP precursor

and then incubated at 50 °C for 0–6 h (supplemental Fig. 3). MIP* was stable at 50 °C and pH 5.

Under milder conditions (room temperature for 14 h), MIP* was stable at pH 9 (Fig. 5A) and exhibited a half-life for decay of 47 h (data not shown). However, the addition of 50 mM DTT at pH 9 resulted in the rapid decay of MIP* to M + IP with only 59% MIP* remaining after 1 h at room temperature. After 14 h at room temperature and pH 9, MIP* completely decayed to M + IP when treated with 50 mM DTT or 50 mM Cys but not with 50 mM Ala, Ser, or Thr (Fig. 5). When MIP* was denatured at room temperature by 6 M urea and incubated at pH 9, there was <5% cleavage in the presence of 50 mM DTT and no detectable cleavage in the presence of Cys, Ala, Ser, or Thr (data not shown). Decay of MIP* was slower at pH 8. A 48-h time course at room temperature and pH 8 demonstrated that MIP* decayed with a half-life of 7.7 h in the presence of 50 mM DTT and 330 h in the absence of DTT, with

rates of $2.5 \times 10^{-5} \text{ s}^{-1}$ and $5.8 \times 10^{-7} \text{ s}^{-1}$, respectively.

Decay of the BI by 50 mM DTT or free Cys would be more likely to occur if a thioester intermediate were present. To our knowledge, cleavage of the ester bond between M and the side chain oxygen of Thr⁺¹ (III) would be unprecedented under such mild conditions. The requirement of a native structure for decay of MIP* by thiols suggests the possibility of a thioester-containing intermediate in equilibrium with the standard Block G BI.

Identification of Unique Sequence Features—Standard inteins are unable to splice or undergo N-terminal cleavage when non-conservative substitutions of Ser¹ or Cys¹ are made. Inteins naturally lacking this N-terminal nucleophile have overcome the barrier to direct attack on an N-terminal splice junction amide bond. In an attempt to identify possible residues specific to these atypical inteins that might be involved in catalysis, we analyzed a multiple sequence alignment of the intein protein splicing motifs to identify co-varying positions, *i.e.* residues that experience simultaneous substitutions in different intein sequences (Fig. 3). The coupling of amino acid sequence changes may indicate linked functions for these residues or structural interactions. Position 12 in Block B (N3:12), position 4 in Block F (C2:4), and position 5 in Block G (C1:5) showed strong co-variance (Chi-square analysis upper probabilities for chance occurrence of 2×10^{-31} to 5×10^{-46}). These extremely low values are a consequence of the almost invariance of a specific triplet at these positions in a subset of sequences relative to the absence of conservation in these positions in the majority of the other sequences analyzed. Examining the source of this co-

TABLE 2

The effects of mutating residues in Blocks B and F

Mutations were made at the indicated positions. The products observed in samples expressed at either 30 °C or 15 °C are indicated. The single-letter amino acid code was used. N-cleavage and C-cleavage refer to off-pathway single splice junction cleavage reactions. Abbreviations as in the legend for Fig. 2. Key: 15/30, no difference observed at the two temperatures; blank, no detectable MIP substrate or product; +, <10% of total MIP substrate and products; ++, 10–50% of total MIP substrate and products; +++, 51–90% of total MIP substrate and products; +++++, >90% of total MIP substrate and products.

| Position/substitution | Temperature | Splice (MP+I) | Precursor (MIP) | N-cleavage (M+IP) | C-cleavage (MI+P) | Branch (MIP*) |
|-------------------------------|-------------|---------------|-----------------|-------------------|-------------------|---------------|
| | °C | | | | | |
| Block B | | | | | | |
| T60A | 15/30 | ++++ | | | | |
| S62A | 15/30 | ++++ | | | | |
| S64A | 15/30 | ++++ | + | | | |
| H65A,H65K | 15/30 | | ++++ | | + | |
| W67A,W67L,W67H,W67F,W67Y,W67Q | 30 | + | ++++ | | | |
| W67A,W67L,W67N | 15 | +++ | ++ | | | |
| W67H,W67F,W67Y,W67Q | 15 | ++++ | + | | | |
| W67K | 15/30 | + | ++++ | | | |
| T68A | 30 | ++++ | + | | | |
| T68V | 30 | +++ | ++ | | | |
| T68V | 15 | ++++ | | | | |
| T69A | 30 | +++ | ++ | | | |
| T69A | 15 | ++++ | + | | | |
| T69V | 15/30 | ++++ | + | | | |
| H75A | 15 | ++++ | | | | |
| Block F | | | | | | |
| K319A | 30 | ++ | +++ | | | |
| K319A | 15 | ++++ | | | | |
| C320A,C320S | 15/30 | | ++++ | | + | |
| C320D | 30 | | ++++ | | | |
| C320D | 15 | | ++ | | ++ | |
| I321A | 30 | + | ++++ | | | |
| I321A | 15 | ++++ | | | | |
| D324A | 30 | +++ | ++ | | | |
| D324A | 15 | ++++ | | | | |
| H328A | 15/30 | ++++ | + | | | |
| F330A | 30 | | ++++ | | | + |
| F330A | 15 | ++ | + | ++ | | + |
| F330L | 30 | + | +++ | + | | + |
| F330L | 15 | ++ | | ++ | | + |
| F330Y | 15/30 | ++++ | + | | | |

variation identified the sequence Trp, Cys, and Thr in these positions in 37 inteins, which we refer to as the WCT triplet. This triplet appeared much more often than would be expected by the individual occurrences of each residue in these positions. Of the 37 inteins containing the WCT triplet, 34 lack an N-terminal nucleophile (supplemental Table 1) and three have an N-terminal Ser or Cys (supplemental Table 2). The only inteins lacking an N-terminal nucleophile that are not in the WCT triplet group are the five known archaeal Klba intein orthologs that begin with Ala¹ (15) and the DNA helicase intein from *Arthrobacter* species FB24 (Arsp-FB24) that begins with Gly¹ and has a WDT triplet. The *Burkholderia vietnamiensis* G4 (Bvi) IcmO intein is the only intein possessing the WCT triplet that lacks the N-terminal nucleophile and has Cys⁺¹ instead of Thr⁺¹ or Ser⁺¹. These findings specifically link non-Klba inteins lacking N-terminal nucleophiles with a previously unrecognized pattern of intein sequence motifs (Fig. 3). It is interesting to note that all inteins lacking an N-terminal nucleophile except the Klba inteins are present in proteins with putative helicase activity (e.g. DnaB, RecG, SNF2, and terminase large subunits of bacteriophage), and most are inserted in an ATPase P-loop motif (5).

Mutations in the WCT Triplet and Surrounding Block B, F, and G Residues—Ala scanning was performed on the WCT triplet and proximal residues to examine their role in the splicing pathway (Table 2 and supplemental Figs. 4–6). Residues identified as being important for splicing were examined fur-

ther (Table 2 and supplemental Fig. 6). Splicing *in vivo* at 30 or 15 °C was analyzed by quantifying MIP-related proteins in soluble cell lysates after SDS-PAGE. Some mutations yielded a similar product profile when expressed at both temperatures, whereas others did not. In the latter case, splicing or cleavage was reduced at 30 °C when compared with 15 °C. Such temperature-dependent enzyme activity generally reflects conformational instability or effects on protein folding caused by the mutations.

The intein Block B typically contains a Thr at position 7 and His at position 10, with position 7 being less conserved and sometimes substituted with other residues capable of hydrogen-bonding interactions (1, 2, 5, 6, 8). Previous studies indicated that this His is essential for splicing and N-terminal cleavage, whereas mutation of this Thr has a more limited effect on these reactions (1, 2, 15). Several Block B residues were examined in the MP-Be DnaB intein (Table 2 and supplemental Figs. 4 and 6). His⁶⁵ at position 10 is essential for splicing of the MP-Be DnaB intein, whereas mutation of Ser⁶² at position 7 does not appreciably inhibit splicing. Ala substitution of Thr⁶⁰, Ser⁶⁴, Thr⁶⁸, and His⁷⁵ had a minimal effect on splicing. T69A (position 14) displayed temperature-dependent splicing (TDS), with complete splicing at 15 °C but only 50% spliced product at 30 °C. The Trp in the WCT triplet (Trp⁶⁷ in the MP-Be DnaB intein) at position 12 in Block B is 2 residues after the conserved His⁶⁵. Mutation of Trp⁶⁷ to Ala, Leu, His, Tyr, Gln, or Lys yielded >90% inactive MIP precursor at 30 °C, whereas Phe

Class 3 Inteins

yielded equal amounts of precursor and spliced products. Splicing was rescued at 15 °C in the Trp⁶⁷ to Ala, Leu, His, Tyr, Phe, Asn, or Gln mutations, but not in the Lys mutant. These results indicate that of all the Block B residues examined, only His⁶⁵ and Trp⁶⁷ are essential at 30 °C. The inactivation pattern of the Trp⁶⁷ mutants is consistent with a structural role for this residue.

Block F contains Cys³²⁰ at position 4, which is part of the conserved WCT triplet. Ala scanning was performed on several Block F residues: Lys³¹⁹, Cys³²⁰, Ile³²¹, Asp³²⁴, His³²⁸, and Phe³³⁰ (Table 2 and supplemental Figs. 5 and 6). At 30 °C, Ala substitution of Cys³²⁰, Ile³²¹, and Phe³³⁰ abrogated splicing, whereas Ala substitution of Lys³¹⁹ and Asp³²⁴ significantly reduced splicing, and Ala substitution of His³²⁸ had no effect. Splicing could be rescued by expression at 15 °C for Ala substitutions of Lys³¹⁹, Ile³²¹, and Asp³²⁴, whereas N-terminal cleavage was the predominant reaction for F330A. Substitution of Phe³³⁰ with Leu yielded similar results to the Ala substitution, but substitution by Tyr yielded >95% spliced product at both temperatures, suggesting that an aromatic residue is important in this position, as opposed to any hydrophobic residue. A similar conclusion was reached for the Mja KlbA intein (12). Moreover, Phe and Tyr frequently occur at this position in all inteins (Fig. 3), suggesting that the requirement for an aromatic residue at position 15 of Block F is probably general to all intein classes. The most critical residue in Block F was Cys³²⁰ because neither splicing nor cleavage could be rescued at 15 °C. Similar results were obtained with the C320S mutant. However, substitution by Asp, which is a very common residue found at this position in standard inteins, resulted in nearly equal amounts of precursor and C-terminal cleavage products at 15 °C.

Block G contains Thr³³⁹ (part of the WCT triplet) in position 5, the intein penultimate His³⁴⁰ in position 6, the intein C-terminal Asn³⁴¹ in position 7, and the C-extein Thr⁺¹ in position 8. The results of mutations of the latter two residues were presented above (Fig. 4). Substitution of the WCT triplet Thr³³⁹ caused TDS, with Ala having the greatest effect at 30 °C (Table 2 and supplemental Fig. 6C). The minimal effect of mutating the penultimate His was unexpected because this His has been shown to assist C-terminal cleavage in other inteins (1, 2).

Examination of Temperature-dependent Splicing Mutants—Mutations that yield temperature-dependent enzyme activity normally affect folding, which in turn affects catalysis. Many of the Trp⁶⁷ and Block F substitutions yielded precursors that failed to splice at 30 °C but spliced at 15 °C. A temperature shift assay was performed to determine the degree of misfolding or aggregation when precursors were expressed at the non-permissive temperature. Cultures were induced normally at 30 °C for 2 h, and then an aliquot was shifted to 15 °C overnight for *in vivo* splicing analysis, whereas a second aliquot was harvested, and splicing *in vitro* at 15 °C overnight was examined using amylose-purified protein. Precursors with only minor or local structural defects should be active *in vitro* when shifted to the permissive temperature, whereas aggregated precursors or more severely misfolded proteins should remain inactive. However, the refolding machinery of the cell may permit the more severely impaired proteins to be activated during the temperature shift *in vivo*.

All of the following mutants were able to recover activity in the *in vivo* temperature shift assay, yielding similar products as precursors initially synthesized at 15 °C: P1C; His, Leu, and Tyr substitutions of Trp⁶⁷; K319A and I321A, which flank the critical Cys³²⁰; C320D; Block F mutants D324A, F330A, and F330L; and Block G mutant T339A (data not shown). No rescue of splicing activity was observed in the *in vitro* temperature shift assay with these same mutants (data not shown).

Comparative Modeling of the MP-Be DnaB Intein—Experimental data suggested novel reactivities in the MP-Be DnaB intein, which led to the new protein splicing mechanism proposed in under “Discussion.” A structural model of this intein was generated to determine whether the structure would support these conclusions. All intein splicing domains have a common fold termed the HINT (Hedgehog/intein) domain, which they share with Hedgehog autoprocessing domains (26). The HINT domain excludes the endonuclease domain and other domains inserted within the intein splicing domain. We performed comparative modeling of the MP-Be DnaB intein HINT domain based on the structure of a precursor of the Sce VMA intein ((20) Protein Data Bank (PDB) code 1JVA), which was hypothesized to most closely resemble an intein reactive conformation (12) despite the presence of mutations in 4 active site residues (20). Sequence identity between the MP-Be DnaB HINT domain and the HINT domains of other structurally characterized inteins ranges from 10 to 17%, which presents a significant challenge for homology modeling. A sequence alignment between the HINT domains of the MP-Be DnaB and Sce VMA intein was constructed using the program FFAS (Fold & Function Assignment System), which is a sequence profile alignment method that takes into account weak sequence similarity over the entire protein that is difficult to detect by other methods (21). The alignment contains 2 N-extein residues and 4 C-extein residues (Fig. 6). It has a sequence identity of 17% and is reasonable based on the conserved positions of the catalytic residues. The resulting model has a root mean square deviation with respect to the template Sce VMA intein of 1.7 Å for the backbone N, C^α, and C' atoms of the 117 aligned residues. A PDB file of the modeled structure is included as supplemental File 1 along with Ramachandran data (supplemental Fig. 7A). A superposition of the model and template is shown in supplemental Fig. 7B, and a view of the MP-Be intein active site is shown in Fig. 7. The model shows that Ser⁶² (Block B, position 7) and His⁶⁵ (Block B, position 10) are located near the N-terminal scissile bond between Gln⁻¹ and Pro¹. No unusual activating interactions of the N-terminal scissile bond are observed that could explain the lability of the N-terminal splice junction in the absence of Thr⁺¹. The model reveals that the WCT triplet Cys³²⁰ residue (Block F, position 4) is located close to both the N-terminal and C-terminal splice junctions. The Cys³²⁰ side chain sulfur atom is located 4.9 Å from the Gln⁻¹ carbonyl carbon and 2.6 Å from the Thr⁺¹ side chain oxygen. This is slightly closer to the N-terminal scissile bond than the Thr⁺¹ side chain oxygen atom, which is 5.2 Å from the Gln⁻¹ carbonyl carbon. Thus, the model indicates that Cys³²⁰ is in position to (a) cleave the N-terminal splice junction equally well as Thr⁺¹, (b) cleave the N-terminal splice junction in the absence of Thr⁺¹, (c) transfer the N-extein from the side chain of Cys³²⁰ to

| | Block A | Block B |
|------------|--|---|
| MP-Be DnaB | dqPLALNTEVPT-PSGWTTVGDLVSGDYVLGSDGQPHRVQRET FPVLEGLATYVVRF -----DD--GTEITASASHGWT TQRL ----- | |
| Scce VMA | vgSFAKGTNVLMADGSIECIENIEVGNKVMGKDRPREVIKLRPGRETMYSVVQKSQHRAHKSDSSREVPE LLKFTCNATNELVVVRT PRSVRRL | * |
| MP-Be DnaB | ----- | ----- |
| Scce VMA | SRTIKGVEYFEVITFEMGQKAPDGRIVELVKEVSKSYPISEGPERANELVESYRKASNKAYFEWTEARDLSLLGSHVRKATYQTYAPILYEN | |
| MP-Be DnaB | ----- | ----- |
| Scce VMA | DHFFDYMQKSKFHLTIEGPKVLAYLLGLWIGDGLSDRATFVSDSRDTSLMERVTEYAEKLNLCAEYKDRKEPQVAKTVNLYSKVVRGNGIRNNL | |
| MP-Be DnaB | ----- | ----- |
| Scce VMA | NTENPLWDAIVGLGFLKDGKVNIPSLFDNIGTRETFLAGLIDSDGYVTDEHGKATIKTIHTSVRDGLVSLARSLGLVSVNAEPAKVDMMG | |
| | | Block F Block G |
| MP-Be DnaB | ----- | TQHRWVESVTPVESVPVKCIGIDTED-HLFQVSRSRILTHNtknt |
| Scce VMA | TKHKISYAIYMSGGDVLLNVLSKACGSKKFRPAPAAFARECRGFYFELQELKEDDYGGITLSDDSDHQFLLANQ-VVHVSsger | ** |

FIGURE 6. **Sequence alignment of the MP-Be DnaB intein and the Scce VMA intein.** The endonuclease domain of the MP-Be DnaB intein was deleted between Thr⁷³ and Gln³⁰³, and the remaining splicing domain was aligned with a mutated version of the Scce VMA intein (20) using the program FFAS. The mutated Scce VMA intein included the following changes in catalytically important residues, which were necessary to isolate the precursor: C15, the intein N terminus; H79N, Block B, position 10 His; N454S, intein C terminus and C+15. The Scce VMA mutations are marked by *asterisks* below the Scce VMA intein sequence. Extein residues are in *lowercase*. The position of intein Blocks A, B, F, and G are indicated by the *line* above the MP-Be DnaB intein sequence and *below* the Scce VMA intein sequence.

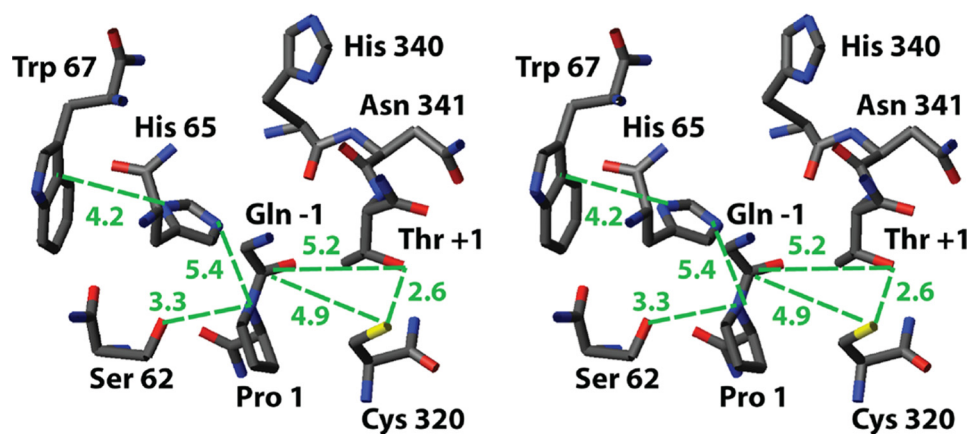


FIGURE 7. **Stereo view of the arrangement of residues near the splice junctions in the theoretical model of the MP-Be DnaB intein.** The heavy atoms of the backbone and side chains of the residues are shown in *stick representations*, with carbon atoms colored *gray*, nitrogen atoms in *blue*, and oxygen atoms in *red*. Distances between atoms of neighboring side chains are indicated by *green dashed lines* and labeled with the distance in Å. Cys³²⁰ and Trp⁶⁷ are WCT triplet residues.

the side chain of Thr⁺¹, and (*d*) activate Thr⁺¹ for attack on the N-terminal scissile bond. Any of these potential roles would explain the essential nature of Cys³²⁰.

The Trp⁶⁷ side chain is found near the side chain of His⁶⁵, with a distance of 4.2 Å between His⁶⁵ ND1 and Trp⁶⁷ CD2. Trp⁶⁷ is involved in hydrophobic packing interactions with Tyr⁴⁹, Ile⁷⁹, and Ala⁴⁷. Despite the limitations of this model, the positioning of Trp⁶⁷ near His⁶⁵ is likely to be correct as several other intein structures contain a hydrophobic residue (Val, Leu, or Phe) at this position, which is located on the same side of the β -strand as the Block B His and packs against hydrophobic residues from neighboring strands. This packing suggests a possible structural role for the Trp in maintaining the proper arrangement of the intein fold for reactivity. Trp⁶⁷ may play an additional catalytic role. The positioning of Trp⁶⁷ and His⁶⁵ suggests a potential cation- π interaction to stabilize a positive charge on His⁶⁵ that might be produced during splicing.

Mutations of Phe³³⁰ and the WCT triplet residue, Thr³³⁹, both display TDS. In this model, their main chain atoms hydro-

gen bond as part of a β -hairpin that contains the +1 nucleophile (Thr⁺¹), the terminal Asn (Asn³⁴¹), and the C-terminal scissile bond. Their side chains are within 5 Å of each other and are directed toward the preceding β -strand that includes Cys³²⁰, suggesting a structural role. We therefore conclude that they are important to the packing and positioning of the C-terminal β -hairpin and thus help to align the +1 nucleophile for attack on the N-terminal splice junction.

DISCUSSION

We initially wondered whether Pro¹ inteins would splice at all because Pro¹ imposes chemical and structural limitations on the N-terminal splice junction. However, some previous studies of standard inteins showed a strained, twisted, or *cis* conformation of the N-terminal splice junction scissile bond that was hypothesized to provide an energetic driving force for the splicing reaction (1, 2, 27). The additional steric repulsion in the *trans* isomer of X-Pro peptide bonds leads to a reduced energy barrier for *cis-trans* isomerization, which could facilitate splicing if ground state destabilization of the scissile bond is a major contributor to the energetics of splicing. Additionally, the pyrrolidine side chain of Pro renders the backbone nitrogen atom a tertiary amide, which will affect hydrogen bonding and leaving group properties during the splicing reactions. Because some of these factors are expected to impede splicing, whereas others would assist splicing, it is difficult to predict whether Pro¹ inteins would be more or less efficient than other inteins. Here we report experimental data showing that the MP-Be DnaB intein splices very efficiently.

Class 3 Inteins

Absence of an Intein N-terminal Cys¹ or Ser¹—Standard inteins cannot attack an amide bond at the N-terminal splice junction when Ser¹ or Cys¹ are mutated, suggesting that inteins naturally lacking these N-terminal nucleophiles have unique strategies for initiating splicing. Johnson *et al.* (12) proposed a steric explanation for the ability of KlbA inteins (Ala¹) to splice by a three-step mechanism that is initiated by a direct attack on the N-terminal scissile bond by the +1 nucleophile (Fig. 1, Class 2), with no obvious differences in chemical activating interactions at the scissile bond when compared with standard inteins. In this study, we examined 40 full-length inteins (supplemental Table 1) present in InBase as of June 25, 2009 (5) that lacked an N-terminal Ser¹, Thr¹, or Cys¹ to identify residues that might also contribute alternative strategies for initiating splicing. We found that 34 of them include 3 invariant interspersed residues (the WCT triplet, Fig. 3), 1 had a WDT triplet, and 5 KlbA orthologs lacked the WCT triplet. The WCT triplet co-varied with the absence of the standard N-terminal nucleophile with very high statistical significance. All of these inteins except the *B. vietnamiensis* G4 IcmO intein also have Ser or Thr at the +1 position (the first amino acid in C-extein), instead of the more common Cys⁺¹. Nevertheless, this association is less statistically relevant because Ser⁺¹ and Thr⁺¹ are found in many standard inteins.

Temperature-dependent Splicing after Mutation—Mutation of the MP-Be DnaB intein WCT triplet residues Trp⁶⁷ and Cys³²⁰ drastically reduced splicing at 30 °C, whereas mutation of Thr³³⁹ had a lesser effect. Only Cys³²⁰ was also essential for splicing at 15 °C. Similar TDS activity (inhibited at 30 °C, active at 15 °C) was seen with other mutations. TDS is not likely due to the fact that mycobacteriophage Bethlehem grows in soil at temperatures below 30 °C because it is normally propagated at 37 °C (16) and the wild type intein is active at 30 °C. Temperature-dependent enzyme activity generally reflects structural issues that can range from minor or local folding perturbations to severely misfolded or aggregated proteins. Activity comparable with that observed when initially synthesized at 15 °C was rescued during *in vivo* temperature shifts but not *in vitro* with Trp⁶⁷ and Thr³³⁹ mutants or the seven other positions tested. This indicates that these mutations significantly altered the protein conformation such that the refolding machinery of the cell was necessary to recover function at 15 °C. In support of this conclusion, the modeled structure indicates that Trp⁶⁷ participates in hydrophobic packing with surrounding residues. The increased structural rigidity at lower temperatures could possibly overcome the effect of the loss of Trp⁶⁷ on hydrophobic packing. Trp⁶⁷ substitutions that were more hydrophobic had the least negative impact. Mutation of residues known to be catalytically important in other inteins (Block B His, C-terminal Asn, and Thr⁺¹) did not show a TDS phenotype in the MP-Be DnaB intein, suggesting a catalytic role for Cys³²⁰.

Activation of the N-terminal Splice Junction—The MP-Be DnaB intein utilized different residues to assist N-terminal cleavage when compared with standard or KlbA inteins. Standard and KlbA inteins use the Block B His in position 10 and the Block B, position 7 residue (most often Thr, Ser, or Asn) to activate the N-terminal splice junction, although the residue at position 7 is not essential (1, 2, 10, 15). Ala scanning of 8 Block

B residues demonstrated that only His⁶⁵ (position 10) and Trp⁶⁷ (position 12) were important for splicing in the MP-Be DnaB intein. Although amino acid substitutions, TDS, and the modeled structure suggest that Trp⁶⁷ has a structural role, this Block B Trp may have an additional catalytic role in protein splicing. The modeled structure suggests that Trp⁶⁷ can participate in cation- π interactions with the Block B His⁶⁵, which could stabilize the positive charge generated when His⁶⁵ is protonated during the splicing reaction. Thus, the WCT triplet residue Trp⁶⁷ seems to have both structural and catalytic roles.

The Importance of Block F Residues in the Splicing Reaction—This study supports recent reports positing the importance of Block F while also showing that some previous findings appear to be intein-specific (10–12, 28). Block F encodes parts of two β -strands separated by a loop of variable length (11, 12, 20, 28). The Block F N-terminal β -strand is at the C terminus of one of the two large β -strands in the HINT domain that form its main horseshoe structure. The Block F C-terminal β -strand forms a twisted β -hairpin with the last intein β -strand that contains the C-terminal nucleophiles and C-terminal splice junction. The Block F loop could act as a hinge for aligning C-terminal residues in the active site, which may be why mutations in this motif often caused TDS in the MP-Be DnaB intein. A previous study (28) proposed that the Block F His at position 13 played a role in activating the penultimate His in the *Synechocystis* DnaB intein, but our study and a previous study of the Mja KlbA intein (12) show no major role for this residue or for Asp³²⁴ (position 9). However, the aromatic residue near the end of Block F (position 15) is important for splicing of the MP-Be DnaB intein (Phe³³⁰), the *Synechocystis* DnaB intein (Phe¹⁴⁵) (28), and the Mja KlbA intein (Tyr¹⁵⁶) (12). Structural modeling shows Phe³³⁰ in the intein C-terminal β -hairpin with its side chain directed toward the preceding β -strand containing Cys³²⁰, suggesting that Phe³³⁰ is important for the packing and positioning of the C-terminal β -hairpin. A similar conclusion was also reached for Tyr¹⁵⁶ in the Mja KlbA intein (12). The frequent occurrence of Phe and Tyr at this position supports an important structural role for Block F, position 15 in all classes of inteins. The positioning of the intein C-terminal β -hairpin is essential for the alignment of the C-extein nucleophile (Thr⁺¹) during attack on the N-terminal splice junction and for the alignment of assisting groups. It is interesting to note that the backbone atoms of Phe³³⁰ and Thr³³⁹ (part of the WCT triplet) hydrogen-bond to each other in the β -hairpin.

In standard inteins, position 4 of Block F is usually Asp but is Trp in the PRP8 inteins (5, 10) and Cys in WCT triplet inteins. This position is important for reactions at both the N-terminal and the C-terminal splice junctions in the atypical Mja KlbA intein (12) and in standard inteins, such as the *Cryptococcus neoformans* PRP8 intein (10) and the *Mycobacterium tuberculosis* (Mtu) RecA intein (11). In the Mja KlbA intein, Asp¹⁴⁷ (Block F, position 4) is in position to interact with both splice junctions and likely assists catalysis by helping to activate the Cys⁺¹ nucleophile (12). Similar results were seen in the Mtu RecA intein structure, where this Asp was shown to be required for both N-terminal and C-terminal cleavage, leading the authors to hypothesize that it is both catalytically important and involved in coordinating the steps in the splicing pathway

(11). In the MP-Be DnaB intein, Cys³²⁰ is also essential for splicing and cleavage at both splice junctions. In the *C. neoformans* PRP8 intein, the authors proposed that Block F and especially the position 4 Trp form a switch that coordinates the steps in the protein splicing pathway (10).

Unique Reactivities of the MP-Be DnaB Intein—A unique feature of the MP-Be DnaB intein is the lability of the N-terminal splice junction in T+1A mutants. The analogous mutation in the Ala¹ Mja KlbA intein (C+1A) yielded predominantly precursor *in vivo* and C-terminal cleavage after overnight incubation *in vitro* but no N-terminal cleavage (15). In standard inteins, off-pathway N-terminal cleavage occurs by hydrolysis or thiolysis of the (thio)ester bond in the linear (II) or branched (III) (thio)ester intermediates. The linear (thio)ester intermediate (II) cannot form in Pro¹ inteins, and the Block G BI (III) cannot form in the T+1A mutant. To our knowledge, there are no examples of off-pathway cleavage of an amide bond at an intein N-terminal splice junction. Mutational analysis indicates that neither Pro¹, Asn³⁴¹, Thr⁺¹, the Block G BI (III), nor C-terminal cleavage are required for off-pathway N-terminal cleavage in the MP-Be DnaB intein. At 15 °C, the only mutations that blocked splicing (on-pathway N-terminal cleavage) and/or off-pathway N-terminal cleavage were mutations of Cys³²⁰, His⁶⁵, and Trp⁶⁷. Could Cys³²⁰ be responsible for the unusual lability of the N-terminal splice junction?

A second unique property of the MP-Be DnaB intein is the apparent thiol sensitivity of the Thr⁺¹ MIP* Block G BI (III), which accumulates in the N341A mutant. Both 50 mM DTT and 50 mM Cys cleaved the BI to yield M + IP under mild conditions (room temperature, pH 9). The thiolate group of molecular cysteine, rather than the amino group, was most likely the reactive agent because other amino acids had no effect. Thiol-induced MIP* cleavage required a natively folded precursor, with only a few percent of denatured MIP* being cleaved by 50 mM DTT. A native structure should not be necessary if the thiols were merely attacking the ester linkage in the Block G BI. However, cleavage by free Cys would be more plausible and a native structure would be required if the Thr⁺¹ Block G BI was in equilibrium with a thioester containing intermediate. There are only 2 cysteines in the MP-Be DnaB intein that could form this hypothetical thioester intermediate. One Cys is in the endonuclease domain, which is distant from the splicing active site. The other is Cys³²⁰, which is clearly in or near the active site.

A New Variation in the Protein Splicing Mechanism—Based on these experimental results, we propose the following new variation of the protein splicing mechanism that accounts for all of our unusual results (Fig. 1, Class 3). In step 1, Cys³²⁰ attacks the N-terminal splice junction to form a Block F BI (VIII) with a thioester linkage between M and IP at the Cys³²⁰ branch point. Transesterification rapidly transfers the N-extein (M) to Thr⁺¹, which results in formation of the standard Block G BI (III). The remainder of the pathway is the same as in other inteins. Transesterification is energetically favored from Cys³²⁰ to Thr⁺¹ because sulfur is a better leaving group than oxygen, which may account for the prevalence of Ser or Thr at the +1 position in WCT-containing inteins. This mechanism is sup-

ported by the MP-Be DnaB intein modeled structure. The sulfur atom of Cys³²⁰ is 4.9 Å from the carbonyl carbon of Gln⁻¹ at the N-terminal splice junction. Given the nature of the model and possible conformational dynamics in the precursor, this distance is reasonable. Cys³²⁰ is slightly closer to the Gln⁻¹ carbonyl carbon in our model than is Thr⁺¹ at 5.2 Å. The Cys³²⁰ side chain sulfur atom is 2.6 Å from the Thr⁺¹ side chain oxygen, which would allow it to rapidly transfer the N-extein to Thr⁺¹ in step 2. This model also explains the thiol lability of the Thr⁺¹ BI if the transesterification step is reversible, as is the conversion of linear and branched (thio)ester intermediates in standard inteins. Even if <5% of the BI in N341A mutants is the thiol-labile Block F BI (as suggested by the thiol cleavage experiments with denatured BI), product consumption would drive the decay of the Thr⁺¹ BI (III) in response to thiol reagents. Maintaining the equilibrium between the two branches drives more protein toward the thiol-sensitive Cys³²⁰ form until all of the BI is converted to M + IP (Fig. 5C). This proposed splicing mechanism explains (a) the essential nature of Cys³²⁰ for both N-terminal cleavage and splicing, (b) N-terminal cleavage in the absence of the Thr⁺¹ nucleophile, and (c) cleavage of only properly folded Block G BI by thiols. Unfortunately, the small percentage of thiol-labile Block F BI in N341A mutants, as indicated in lability studies with denatured BI, makes it difficult to isolate and characterize, especially because both BIs have the same mass. However, we will seek out this new branched intermediate in future studies.

A New Classification of Inteins—This study describes a new class of inteins that can be identified based on the absence of an N-terminal Ser, Thr, or Cys and the presence of the WCT triplet. We propose calling them Class 3 inteins, with standard inteins being Class 1. KlbA-like inteins represent Class 2, which lack both the standard N-terminal nucleophile and the WCT triplet. Although its splicing mechanism has not been experimentally determined, the DNA helicase intein from *Arthrobacter* species FB24 is also a putative member of Class 2 based on sequence signatures. The three classes represent different splicing mechanisms (Fig. 1). Although Class 1 and Class 2 inteins use the typical Thr (position 7) and His (position 10) in Block B to activate the intein N-terminal splice junction, Class 3 inteins require the Block B Trp (position 12) to activate N-terminal cleavage. The MP-Be DnaB intein does not require the Block B, position 7 residue (Ser⁶²) for N-terminal cleavage. Except for one Class 3 intein, all Class 3 inteins that we identified have Thr or Ser at the +1 position, which energetically favors the transesterification step to generate the Block G BI. We are currently examining splicing of the three known WCT inteins that begin with Ser¹ or Cys¹ (supplemental Table 2) and the one known WCT intein that has Cys⁺¹ (supplemental Table 1) to determine whether they fall into Class 3 based on differences in reactivity when compared with standard inteins, such as cleavage of the N-terminal splice junction in +1 mutants and the absolute requirement of the Block F Cys for splicing. These future studies will allow us to further refine the requirements for inclusion into Class 3. As more inteins are identified, the range of splicing mechanisms continues to

Class 3 Inteins

expand, reflecting the robust nature of intein-mediated protein splicing.

Acknowledgments—We thank Dr. Graham Hatfull (University of Pittsburgh) for kindly providing us with mycobacteriophage Bethlehem genomic DNA. We thank Rowena Matthews (University of Michigan) and Manoj Cheriyan (New England Biolabs) for helpful discussions, Manoj Cheriyan, Chris Noren, and Bill Jack (New England Biolabs) for helpful comments and review of this article, Jack Benner and Shelley Cushing (New England Biolabs) for protein sequencing, and Don Comb (New England Biolabs) and Kurt Wüthrich (The Scripps Research Institute) for support and encouragement.

REFERENCES

1. Perler, F. B. (2005) *IUBMB Life* **57**, 469–476
2. Mills, K. V., and Perler, F. B. (2005) *Protein Pept. Lett.* **12**, 751–755
3. Martin, D. D., Xu, M. Q., and Evans, T. C., Jr. (2001) *Biochemistry* **40**, 1393–1402
4. Xu, M. Q., and Perler, F. B. (1996) *EMBO J.* **15**, 5146–5153
5. Perler, F. B. (2002) *Nucleic Acids Res.* **30**, 383–384
6. Pietrokovski, S. (1998) *Protein Sci.* **7**, 64–71
7. Perler, F. B., Olsen, G. J., and Adam, E. (1997) *Nucleic Acids Res.* **25**, 1087–1093
8. Pietrokovski, S. (1994) *Protein Sci.* **3**, 2340–2350
9. Hodges, R. A., Perler, F. B., Noren, C. J., and Jack, W. E. (1992) *Nucleic Acids Res.* **20**, 6153–6157
10. Pearl, E. J., Tyndall, J. D., Poulter, R. T., and Wilbanks, S. M. (2007) *FEBS Lett.* **581**, 3000–3004
11. Van Roey, P., Pereira, B., Li, Z., Hiraga, K., Belfort, M., and Derbyshire, V. (2007) *J. Mol. Biol.* **367**, 162–173
12. Johnson, M. A., Southworth, M. W., Herrmann, T., Brace, L., Perler, F. B., and Wüthrich, K. (2007) *Protein Sci.* **16**, 1316–1328
13. Mills, K. V., Manning, J. S., Garcia, A. M., and Wuerdeman, L. A. (2004) *J. Biol. Chem.* **279**, 20685–20691
14. Amitai, G., Dassa, B., and Pietrokovski, S. (2004) *J. Biol. Chem.* **279**, 3121–3131
15. Southworth, M. W., Benner, J., and Perler, F. B. (2000) *EMBO J.* **19**, 5019–5026
16. Hatfull, G. F., Pedulla, M. L., Jacobs-Sera, D., Cichon, P. M., Foley, A., Ford, M. E., Gonda, R. M., Houtz, J. M., Hryckowian, A. J., Kelchner, V. A., Namburi, S., Pajcini, K. V., Popovich, M. G., Schleicher, D. T., Simanek, B. Z., Smith, A. L., Zdanowicz, G. M., Kumar, V., Peebles, C. L., Jacobs, W. R., Jr., Lawrence, J. G., and Hendrix, R. W. (2006) *PLoS Genet* **2**, e92
17. Kass, I., and Horovitz, A. (2002) *Proteins* **48**, 611–617
18. Henikoff, S., and Henikoff, J. G. (1994) *J. Mol. Biol.* **243**, 574–578
19. Southworth, M. W., Amaya, K., Evans, T. C., Xu, M. Q., and Perler, F. B. (1999) *BioTechniques* **27**, 110–114, 116, 118–120
20. Mizutani, R., Nogami, S., Kawasaki, M., Ohya, Y., Anraku, Y., and Satow, Y. (2002) *J. Mol. Biol.* **316**, 919–929
21. Rychlewski, L., Jaroszewski, L., Li, W., and Godzik, A. (2000) *Protein Sci.* **9**, 232–241
22. Guex, N., and Peitsch, M. C. (1997) *Electrophoresis* **18**, 2714–2723
23. Schwede, T., Kopp, J., Guex, N., and Peitsch, M. C. (2003) *Nucleic Acids Res.* **31**, 3381–3385
24. Melo, F., and Feytmans, E. (1998) *J. Mol. Biol.* **277**, 1141–1152
25. Eisenberg, D., Lüthy, R., and Bowie, J. U. (1997) *Methods Enzymol.* **277**, 396–404
26. Hall, T. M., Porter, J. A., Young, K. E., Koonin, E. V., Beachy, P. A., and Leahy, D. J. (1997) *Cell* **91**, 85–97
27. Romanelli, A., Shekhtman, A., Cowburn, D., and Muir, T. W. (2004) *Proc. Natl. Acad. Sci. U.S.A.* **101**, 6397–6402
28. Ding, Y., Xu, M. Q., Ghosh, I., Chen, X., Ferrandon, S., Lesage, G., and Rao, Z. (2003) *J. Biol. Chem.* **278**, 39133–39142