

An artificially split class 3 intein

Tia M. Ariagno^{1*}, John S. Smetana^{1*}, Christopher W. Lennon^{1§}

¹Department of Biological Sciences, Murray State University, Murray, Kentucky, United States

§To whom correspondence should be addressed: clennon1@murraystate.edu

*These authors contributed equally.

Abstract

Inteins excise themselves from precursor polypeptides through protein splicing, joining N- and C-exteins with a peptide bond. Split inteins are expressed as separate polypeptides that undergo protein *trans* splicing (PTS). Here, we demonstrate PTS can be achieved using an artificially split class 3 intein. Because class 3 inteins use an internal initiating nucleophile near the C-extein junction, rather than the first residue of the intein, both catalytic nucleophiles are present on a single polypeptide. This results in a compact arrangement of catalytic nucleophiles for PTS compared to the standard arrangement for split class 1 inteins.

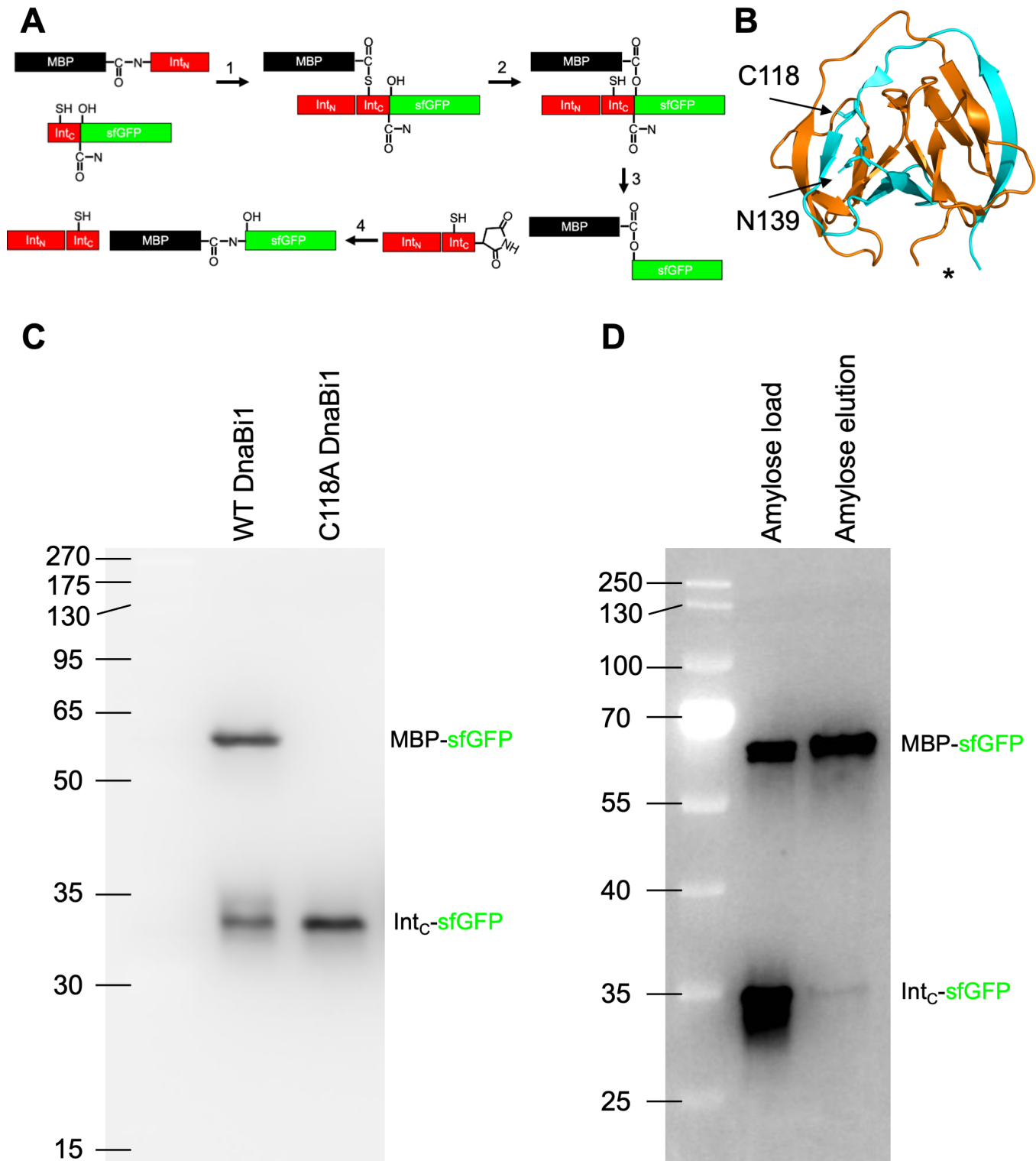


Figure 1. Protein trans splicing of split *M. smegmatis* DnaBi1:

A. Mechanisms of class 3 protein splicing with the Maltose Binding Protein (MBP) as the N-extein (black), superfolder GFP (sfGFP) as the C-extein (green), and the split intein (red). The peptide bonds between the N-extein and intein (prespliced) and N-extein and C-extein (spliced) are shown, as well as the side chains of the catalytic nucleophiles (Cys118 and Ser +1) and C-terminal intein asparagine. **B.** Structure of contiguous *Mycobacterium smegmatis* DnaBi1 generated using pymol from PDB 6BS8 (Kelley et al., 2018). Residues 1-88 are shown in orange and residues 101-139 are shown in cyan. Residues 89-100 are

unstructured and labeled with an asterisk (site of artificial split is between residues 94-95). Initiating nucleophile (C118) and terminal asparagine (N139) are labeled and shown as sticks. C. MBP-DnaBi1^N and DnaBi1^C-sfGFP or DnaBi1^C C118A-sfGFP were co-expressed at 21°C for 20 hours and in-gel fluorescence was used to measure GFP-containing products. D. MBP-sfGFP is both fluorescent and retained by amylose resin, demonstrating *trans* splicing between MBP-DnaBi1^N and DnaBi1^C-sfGFP. MBP-DnaBi1^N and DnaBi1^C-sfGFP were co-expressed at 16°C for 20 hours prior to amylose resin pull-down.

Description

Intervening proteins (inteins) are translated within host proteins and excised via protein splicing (Paulus, 2000). In this reaction, the intein rearranges the two flanking peptide bonds into one, in the process ligating the adjacent N- and C-exteins and escaping the host protein (ligated exteins) without a trace. The ability of inteins to perform this chemistry has been exploited for a variety of protein engineering applications.

The class 1 mechanism of protein splicing occurs in four steps (Mills et al., 2014). First, an N-terminal intein residue (Cysteine or Serine) performs a nucleophilic attack on the peptide bond between the N-extein and intein, resulting in a (thio)ester linkage. Second, the first residue of the C-extein, a Cysteine, Serine, or Threonine, performs a second nucleophilic attack on the (thio)ester resulting from step 1, forming a branched intermediate. Third, the last residue of the intein, often a conserved asparagine, cyclizes to release the intein. Lastly, a peptide bond is formed between the N- and C-exteins through (thio)ester rearrangement. Class 2 and 3 splicing mechanisms have also been described. Class 3 inteins utilize a very similar strategy as class 1 inteins, except for the position of the initiating nucleophile that begins the splicing reaction - the initiating nucleophilic attack is carried out by an internal residue near the C-terminal intein-extein junction.

Inteins are particularly abundant in archaea and bacteria, but are also present in phages, viruses, and unicellular eukaryotes (Novikova et al., 2016). Inteins can be full-length, meaning the intein has an internal homing endonuclease domain (HEN), which assists in the invasion of intein-free alleles and makes some inteins mobile genetic elements. Mini-inteins lack the HEN domain, but retain conserved splicing blocks necessary for splicing. Interestingly, some inteins have been split in nature and are synthesized as separate polypeptides that undergo protein *trans* splicing (PTS).

Inteins can be thought of as self-contained, single-turnover enzymes with a unique ability to break and remake peptide bonds at designated positions. As such, this ability has been exploited in numerous protein engineering applications including bioseparations, bioconjugation, biosensing, and protein cyclization (Wood and Camarero, 2014; Sarmiento and Camarero, 2019). The impact of intein-based technologies on protein engineering has been, and continues to be, profound. Split inteins provide a unique avenue to control the protein splicing reaction by physical separation of intein fragments. Split inteins typically divide into larger N-intein (~100 residues) and smaller C-intein (~40 residues) domains (Eryilmaz et al., 2014; Aranko et al., 2014). Upon reassociation, split inteins adopt a highly similar fold to *cis* splicing inteins.

Numerous split inteins using a class 1 splicing mechanism have been extensively engineered to improve rate, accuracy, and controllability of PTS (Eryilmaz et al., 2014; Aranko et al., 2014), and only recently has there been the demonstration of *trans* splicing using alternative chemistries (Pinto et al., 2020; Hoffmann et al., 2021). In this work, we artificially split the class 3 DnaB1 intein from *Mycobacterium smegmatis* (DnaBi1) and observe efficient splicing in *trans*. The major difference between this system and traditional split intein systems is that both the initiating and +1 nucleophiles, as well as conserved terminal asparagine of the intein, are on the same Int^C-C-extein fragment. The expected mechanism of class 3 *trans* splicing is shown (Figure 1A).

The replicative helicase DnaB from *M. smegmatis* houses two inteins (DnaBi1 and DnaBi2). DnaBi1 is a mini-intein (139-residues) that utilizes a class 3 splicing mechanism, while DnaBi2 is a full-length intein that employs class 1 chemistry. The crystal structure of DnaBi1 shows a disordered loop between conserved splicing blocks, consistent with the expected Hedgehog/Intein (Hint) domain fold (Kelley et al., 2018). This disordered loop, located ~40 residues from the intein-C-extein junction, is where most naturally *trans* splicing inteins are split. Several class 1 *cis* splicing inteins have been divided at this loop, resulting in engineered *trans* splicing (Aranko et al., 2014). We sought to determine whether the class 3 *M. smegmatis* DnaBi1 could splice in *trans*. As the initiating nucleophile (C118) for *M. smegmatis* DnaBi1 is located after the disordered HEN loop, both catalytic nucleophiles would be housed on the C-intein-extein fragment. This design results in protein *trans* splicing with a compact arrangement of catalytic nucleophiles.

We introduced a split in DnaBi1 within the disordered HEN loop between residues 94 and 95 (Figure 1B) and measured *trans* splicing using a modified version of an established splicing reporter, MIG (Kelley et al., 2018). In the MIG reporter (**MBP-Intein-GFP**), the intein (surrounded by 10 native extein residues) is flanked by the non-native exteins MBP (Maltose Binding Protein; N-extein) and superfolder GFP (sfGFP; C-extein). sfGFP-containing products can be detected using in-gel

fluorescence following semi-native PAGE (Weinberger II and Lennon, 2021). MBP-DnaBi1^N and DnaBi1^C-sfGFP were co-expressed from different promoters on the same plasmid in *E. coli* and sfGFP-containing products were measured. Two products were observed, the smaller of which is consistent in size with DnaBi1^C-sfGFP (32.8 kDa) and the larger with the expected product from *trans* splicing between MBP-DnaBi1^N and DnaBi1^C-sfGFP, MBP-sfGFP (69.7 kDa) (Figure 1C).

Previously, it was demonstrated that the initiating nucleophile C118 of DnaBi1 was required for splicing in *cis* (Kelley et al., 2018; Woods et al., 2020; Lennon et al., 2021). We next tested whether this residue is also required for splicing in *trans*. As expected, and consistent with splicing between MBP-DnaBi1^N and DnaBi1^C-sfGFP, the larger product is completely absent when the C118A substitution is made on DnaBi1^C-sfGFP (Figure 1C). The requirement of both of C118 for splicing in *cis* and loss of the band consistent in size with MBP-sfGFP only when C118 is mutated strongly suggests protein *trans* splicing between MBP-DnaBi1^N and DnaBi1^C-sfGFP. It may be possible for the larger sfGFP-containing product to be a dimer of Int^C-sfGFP (65.6 kDa), rather than MBP-sfGFP, as the initiating nucleophile C118 is free to form a disulfide. Additionally, while samples were not boiled to maintain sfGFP structure prior to semi-native SDS-PAGE, β-mercaptoethanol was added to reduce any potential disulfide bonds and eliminate this possibility.

Finally, to further confirm that *trans* splicing had occurred, we reasoned that the larger sfGFP-containing band must also contain MBP. To determine whether the larger sfGFP-fluorescent product also contained MBP, we tested capture by amylose resin, which will pull-down any MBP-tagged products. Indeed, we found that the larger sfGFP-containing band was retained by the resin (Figure 1D). Taken together, these results clearly demonstrate the formation of MBP-sfGFP from coordinated MBP-DnaBi1^N and DnaBi1^C-sfGFP expression, a product that can only result from class 3 *trans* splicing.

Here, we show that a well-studied, small, conditionally-regulated (Kelley et al., 2018; Woods et al., 2020; Lennon et al., 2021) class 3 intein DnaBi1 from *M. smegmatis* can be artificially split for PTS applications. Due to the catalytic nucleophile arrangement of a class 3 intein, this design results in a compact arrangement of nucleophiles for *trans* splicing. As split inteins technologies based on class 1 inteins have been exceptionally useful in protein engineering, alternative intein chemistries may prove useful for certain applications.

Methods

Plasmids and strains

Plasmids used in this study (pET-Duet1 MBP-DnaB1 Int_N / DnaB1Int_C-sfGFP; pET-Duet1 MBP-DnaB1 Int_N / DnaB1Int_C-sfGFP C118A) were synthesized (Genscript) and were transformed into *Escherichia coli* BL21(DE3) for protein expression.

Protein expression

Cells were grown in LB with 100 μg/mL ampicillin at 37°C to mid-log (OD₆₀₀ ~0.5), and protein expression was induced by addition 1 mM Isopropyl β- d-1-thiogalactopyranoside (Gold Bio). Protein expression proceeded for either 20 hours at either 16°C or 21°C after which cells were pelleted by centrifugation.

Amylose resin purification

Following protein expression (MBP-DnaBi1^N or MBP-DnaBi1_{STOP}), cells were resuspended in 20 mM Tris (pH 7.5), 200 mM NaCl (Amylose Column Buffer) with 1X HALT Protease Inhibitor Cocktail (Thermo Scientific), and lysed by sonication. Crude lysate was centrifuged at 20,000 x g, 4°C and supernatant was applied to Amylose Resin (New England Biolabs) equilibrated in Amylose Column Buffer, washed with ~20 column volumes of Amylose Column Buffer, and eluted with Amylose Column Buffer plus 10 mM maltose.

SDS-PAGE

For in-gel fluorescence, cells were resuspended in 50 mM Tris-HCl (pH 8.0), 20 mM NaCl, 10% glycerol and lysed by sonication. Insoluble material was removed by centrifugation and soluble lysate was mixed with 4X Laemmli Sample Buffer with 4% β-mercaptoethanol (Bio-Rad). Samples were separated using 8-16% Tris-Glycine TGX gels (Bio-Rad). Samples were not heated to maintain sfGFP structure and fluorescent products were detected with an Amersham Imager 680 (GE Healthcare).

Reagents

--

Strain	Genotype	Source
<i>Escherichia coli</i> BL21(DE3)	<i>fhuA2 [lon] ompT gal (λ DE3) [dcm] ΔhsdS</i> λ DE3 = λ <i>sBamHI</i> Δ <i>EcoRI-B int::(lacI::PlacUV5::T7 gene1) i21 Δnin5</i>	NEB
Plasmid	Description	Source
pET-Duet1 MBP-DnaB1 Int _N /DnaB1Int _C -GFP	Plasmid encoding for DnaB1 intein with MBP N-extein and GFP C-extein.	GenScript
pET-Duet1 MBP-DnaB1 Int _N /DnaB1 Int _C -GFP C118A	Plasmid encoding for DnaB1 intein with MBP N-extein and GFP C-extein. Has a substitution on C118A.	GenScript

References

- Aranko AS, Wlodawer A, Iwai H. 2014. Nature's recipe for splitting inteins. *Protein Eng Des Sel* 27(8): 263-71. PubMed ID: [25096198](#)
- Eryilmaz E, Shah NH, Muir TW, Cowburn D. 2014. Structural and dynamical features of inteins and implications on protein splicing. *J Biol Chem* 289(21): 14506-11. PubMed ID: [24695731](#)
- Hoffmann S, Terhorst TME, Singh RK, Kümmel D, Petrokovski S, Mootz HD. 2021. Biochemical and Structural Characterization of an Unusual and Naturally Split Class 3 Intein. *Chembiochem* 22(2): 364-373. PubMed ID: [32813312](#)
- Kelley DS, Lennon CW, Li Z, Miller MR, Banavali NK, Li H, Belfort M. 2018. Mycobacterial DnaB helicase intein as oxidative stress sensor. *Nat Commun* 9(1): 4363. PubMed ID: [30341292](#)
- Lennon CW, Wahl D, Goetz JR, Weinberger J 2nd. 2021. Reactive Chlorine Species Reversibly Inhibit DnaB Protein Splicing in Mycobacteria. *Microbiol Spectr* 9(2): e0030121. PubMed ID: [34549994](#)
- Mills KV, Johnson MA, Perler FB. 2014. Protein splicing: how inteins escape from precursor proteins. *J Biol Chem* 289(21): 14498-505. PubMed ID: [24695729](#)
- Novikova O, Jayachandran P, Kelley DS, Morton Z, Merwin S, Topilina NI, Belfort M. 2016. Intein Clustering Suggests Functional Importance in Different Domains of Life. *Mol Biol Evol* 33(3): 783-99. PubMed ID: [26609079](#)
- Paulus H. 2000. Protein splicing and related forms of protein autoprocessing. *Annu Rev Biochem* 69: 447-96. PubMed ID: [10966466](#)
- Pinto F, Thornton EL, Wang B. 2020. An expanded library of orthogonal split inteins enables modular multi-peptide assemblies. *Nat Commun* 11(1): 1529. PubMed ID: [32251274](#)
- Sarmiento C, Camarero JA. 2019. Biotechnological Applications of Protein Splicing. *Curr Protein Pept Sci* 20(5): 408-424. PubMed ID: [30734675](#)
- Weinberger Ii J, Lennon CW. 2021. Monitoring Protein Splicing Using In-gel Fluorescence Immediately Following SDS-PAGE. *Bio Protoc* 11(16): e4121. PubMed ID: [34541040](#)
- Wood DW, Camarero JA. 2014. Intein applications: from protein purification and labeling to metabolic control methods. *J Biol Chem* 289(21): 14512-9. PubMed ID: [24700459](#)
- Woods D, Vangaveti S, Egbunum I, Sweeney AM, Li Z, Bacot-Davis V, et al., Lennon CW. 2020. Conditional DnaB Protein Splicing Is Reversibly Inhibited by Zinc in Mycobacteria. *mBio* 11(4). PubMed ID: [32665276](#)
- Funding:** This work was supported by National Institutes of Health grants P20GM103436 to (through KY INBRE) and 1R15GM143662 to CWL

Supported by National Institute of General Medical Sciences (United States) 1R15GM143662 to Christopher W. Lennon.

Supported by National Institute of General Medical Sciences (United States) P20GM103436 to KY INBRE.

Author Contributions: Tia M. Ariagno: validation, writing - review editing, investigation. John S. Smetana: investigation, validation, writing - review editing. Christopher W. Lennon: conceptualization, funding acquisition, investigation, project

9/22/2023 - Open Access

administration, validation, supervision.

Reviewed By: Anonymous

History: Received September 5, 2023 **Revision Received** September 20, 2023 **Accepted** September 20, 2023 **Published Online** September 22, 2023 **Indexed** October 6, 2023

Copyright: © 2023 by the authors. This is an open-access article distributed under the terms of the Creative Commons Attribution 4.0 International (CC BY 4.0) License, which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Citation: Ariagno, TM; Smetana, JS; Lennon, CW (2023). An artificially split class 3 intein. microPublication Biology. [10.17912/micropub.biology.000977](https://doi.org/10.17912/micropub.biology.000977)