JACS
JOURNAL OF THE AMERICAN CHEMICAL SOCIETY

Communication

pubs.acs.org/JACS

Design of a Split Intein with Exceptional Protein Splicing Activity

Adam J. Stevens,[†] Zachary Z. Brown,[†] Neel H. Shah,[†] Giridhar Sekar,[‡] David Cowburn,[‡] and Tom W. Muir*[,†]

[†]Department of Chemistry, Princeton University, Frick Laboratory, Princeton, New Jersey 08544, United States

[‡]Department of Biochemistry, Albert Einstein College of Medicine, Bronx, New York 10461, United States

Supporting Information

ABSTRACT: Protein trans-splicing (PTS) by split inteins has found widespread use in chemical biology and biotechnology. Herein, we describe the use of a consensus design approach to engineer a split intein with enhanced stability and activity that make it more robust than any known PTS system. Using batch mutagenesis, we first conduct a detailed analysis of the difference in splicing rates between the Npu (fast) and Ssp (slow) split inteins of the DnaE family and find that most impactful residues lie on the second shell of the protein, directly adjacent to the active site. These residues are then used to generate an alignment of 73 naturally occurring DnaE inteins that are predicted to be fast. The consensus sequence from this alignment (Cfa) demonstrates both rapid protein splicing and unprecedented thermal and chaotropic stability. Moreover, when fused to various proteins including antibody heavy chains, the N-terminal fragment of Cfa exhibits increased expression levels relative to other N-intein fusions. The durability and efficiency of Cfa should improve current intein based technologies and may provide a platform for the development of new protein chemistry techniques.

Protein splicing is a post-translational autoprocessing event in which an intervening protein domain called an intein excises itself from a host protein in a traceless manner such that the flanking polypeptide sequences (exteins) are ligated together via a normal peptide bond (Figure S1A).[1] While protein splicing typically occurs spontaneously following translation of a contiguous polypeptide, some inteins exist naturally in a split form.[1] The two pieces of the split intein are separately expressed and remain inactive until encountering their complementary partner, upon which they cooperatively fold and undergo splicing *in trans* (Figure S1B). This activity has been harnessed in a host of protein engineering methods that provide control over the structure and activity of proteins both *in vitro* and *in vivo*.[1] The first two split inteins to be characterized, from the cyanobacteria *Synechocystis* species PCC6803 (Ssp) and *Nostoc punctiforme* PCC73102 (Npu), are orthologs naturally found inserted in the α subunit of DNA Polymerase III (DnaE).[2−4] Npu is especially notable due its remarkably fast rate of protein trans-splicing (PTS) ($t_{1/2}$ = 50 s at 30 °C).[5] This half-life is significantly shorter than that of Ssp ($t_{1/2}$ = 80 min at 30 °C),[5] an attribute that has expanded the range of applications open to PTS.[1]

Despite the ongoing discovery of new fast inteins,[6,7] little is known about what separates them from their slower homologues. Such an understanding should help identify new inteins that are likely to splice rapidly and potentially allow for the engineering of split inteins with superior PTS properties. We sought to investigate the basis of rapid protein splicing through a comparative study of Npu and Ssp. The substantial difference in splicing rate between these two proteins is especially puzzling given their highly similar sequences (63% identity) and near-superimposable active site structures (Figure S2). Previous mutagenesis studies on Npu and Ssp suggest that the difference in activity between the two is likely due to the combined effects of several residues, rather than a single site.[6,8] However, it remains unclear just how many residues are responsible for the fast versus slow reaction rates and, by extension, whether these "accelerator" residues contribute equally to the individual chemical steps in the overall protein splicing process. Consequently, we began our study by exploring these questions, in the hope that this would provide a starting point for developing an improved PTS system.

The high level of conservation within the active sites of Npu and Ssp suggests that differences in distal amino acids account for the disparity in splicing rate between the two. Thus, we focused our attention on "second shell" residues, those directly adjacent to the active site. To simplify this analysis, we employed a batch mutagenesis strategy in conjunction with a previously reported *in vitro* PTS assay.[5] This assay uses split intein constructs with short native extein sequences and allows the rates of branched intermediate formation (k_1,k_2) and its resolution to final splice products (k_3) to be determined using a three-state kinetic model (Figure S3).

The known cross-reactivity of Npu and Ssp intein fragments served as a convenient platform to assess which half of the split intein contributes most significantly to the difference in activity.[3] Both the Ssp^N-Npu^C (chimera 1) and Npu^N-Ssp^C (chimera 2) chimeras show a decrease in the rates of branch formation and resolution compared to that of native Npu (Figure S4C, S4D). This indicates that residues on both the N- and C-intein fragments of Npu and Ssp contribute to the difference in their splicing rate. Next, four groups of second shell positions on each of these chimeras were chosen based on their proximity to key catalytic residues, and the corresponding Ssp residues were mutated to those in Npu (Figure S4A, S4B). From the chimera 1 mutants, Batch 2 (L56F, S70K, A83P,

Received: December 28, 2015
Published: February 8, 2016

DOI: 10.1021/jacs.5b13528
J. Am. Chem. Soc. 2016, 138, 2162−2165

E85D) completely restored branch formation activity to that of native Npu (Figure S4C), while Batch 1 (R73K, L75M, Y79G, L81M) restored the majority of branch resolution activity (Figure S4D). The effects of mutations on the chimera 2 background were more prosaic, with no single batch able to restore splicing activity to that of native Npu (Figures S4C, S4D). Lastly, the A136S mutation on Ssp^C has previously been shown to accelerate protein splicing and was examined separately.[8] This A136S mutation increases the rate of branch resolution 2-fold, but has no impact on branch formation (Figures S4C, S4D).

Next we sought to better understand the individual contributions of residues within batch mutants 1 and 2, since these had the most profound effect on splicing activity. For Batch 2, further mutagenesis shows that the interaction between F56, K70, and D85 is likely responsible for the increased rate of branch formation in Npu^N (Figure S5A). Structural evidence supports these data, as K70 is a part of the highly conserved TXXH block B loop in Npu^N (residues 69−72) that catalyzes the initial N-to-S acyl shift in protein splicing.[9] Thus, the position and dynamics of K70 (packed against F56 and D85) should directly impact the catalytic residues T69 and H72 (Figure S5B).[10−12] From Batch 1, K73, M75, and M81 are responsible for the faster rate of branch resolution in Npu^N (Figure S6A). These residues pack against the terminal asparagine of the C-intein, which must undergo succinimide formation in the final step of protein splicing (Figure S6B). Taken together, the mutagenesis data point to the key role that second shell "accelerator" residues play in tuning the activity of split inteins.

The "accelerator" residues found to affect the splicing rate allow for an activity-guided approach to engineer a consensus DnaE intein. Consensus protein engineering is a tool applied to a homologous set of proteins in order to create a thermostable variant derived from the parent family.[13,14] A multiple sequence alignment (MSA) is first generated from homologues of a particular protein, from which the most frequent residue at each position is chosen as the representative in the consensus sequence. For the DnaE inteins, 105 sequences were identified through a BLAST[15] search of the JGI[16] and NCBI[17] databases (Figure S7A). Next, the alignment was filtered to only contain sequences bearing the second shell indicators of fast splicing: K70, M75, M81, and S136. The 73 theoretically fast inteins left in the MSA (Figure S7B) were then used to generate a consensus fast DnaE intein sequence (Cfa) (Figure 1). The Cfa intein has high sequence similarity to Npu (82%), and the nonidentical residues are spread throughout the 3D structure of the protein.

We generated the Cfa intein fragments fused to model exteins and measured their PTS activity using the aforementioned in vitro assay (Figure 2). This revealed that the Cfa intein splices 2.5-fold faster at 30 °C than Npu ($t_{1/2}$ 20 s vs 50 s), a notable enhancement in activity since the latter is the fastest characterized DnaE split intein (Figure 2A). Cfa demonstrates faster equilibrium rates of branch formation (3-fold) and rate branch resolution (2-fold) (Figure S8). We note that, in line with parent DnaE inteins, Cfa retains the preference for a bulky hydrophobic residue at the +2 position of the C-extein (Figure S9). Strikingly, Cfa shows an increased splicing rate as a function of temperature and is consistently faster than Npu (Figure 2A). The Cfa intein even maintains activity at 80 °C, albeit with reduced yield of splice products, while Npu is inactive at this temperature. These results

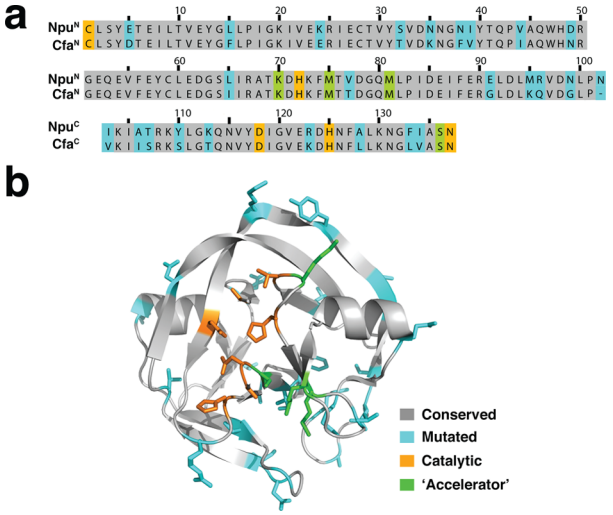

Figure 1. Design of the Cfa split intein. (a) Sequence alignment of Npu DnaE and Cfa DnaE. The sequences share 82% identity with the differences (cyan) evenly distributed through the primary sequence. Catalytic residues and second shell "accelerator" residues are shown in orange and green, respectively. (b) The same residues highlighted in panel a mapped on to the Npu structure (PDB ID 4Kl5).

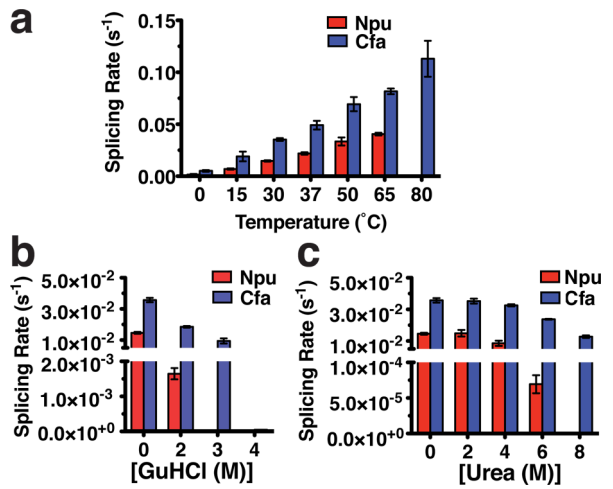

Figure 2. Characterization of the Cfa intein. (a) Splicing rates for Cfa and Npu as a function of temperature. Npu is inactive at 80 °C. Error = SD (n = 3). (b, c) Splicing rates for Cfa and Npu as a function of added chaotrope. Npu is inactive in 3 M GuHCl or 8 M Urea. Note: Cfa has residual activity in 4 M GuHCl ($k = 7 \times 10^{-5}$). Error = SD (n = 3).

demonstrate that consensus engineering is effective in producing an intein that is highly active across a broad range of temperatures.

Applications of PTS typically require fission of a target protein and fusion of the resulting fragments to the appropriate split intein segments.[1] As a consequence, the solubility of these fusion proteins can sometimes be poor. Because protein denaturants such as guanidine hydrochloride (GuHCl) and urea are frequently used to keep these less soluble fragments in solution, we tested the ability of Cfa to splice in the presence of these chaotropic agents. We found that the Cfa intein can splice in the presence of up to 4 M GuHCl (with little decrease in activity seen up to 3 M), while no activity was observed for Npu in ≥3 M GuHCl (Figure 2B). Remarkably, the splicing of Cfa is

DOI: 10.1021/jacs.5b13528
J. Am. Chem. Soc. 2016, 138, 2162−2165

largely unaffected up to 8 M urea, while splicing of Npu falls off dramatically above 4 M urea (Figure 2C).

The unprecedented tolerance of Cfa to high concentrations of GuHCl and urea suggests the intein might retain activity directly following chaotropic extraction of insoluble proteins from bacterial inclusion bodies, thereby expediting PTS-based studies. Accordingly, we overexpressed the model fusion protein, His_6-Sumo-Cfa^N, in *E. coli* cells and extracted the protein from inclusion bodies with 6 M urea. The protein was purified from this extract by nickel affinity chromatography and then directly, and efficiently, modified by PTS under denaturing conditions, i.e. without the need for any intervening refolding steps (Figure S10). In general, we expect that the robust activity of Cfa in the presence of chaotropic agents will prove useful when working with protein fragments that demonstrate poor solubility under native conditions.

Fusing a protein of interest to a split intein can result in a marked reduction in cellular expression levels compared to the protein alone.[6] This situation is more frequently encountered for fusions to N-inteins than to C-inteins, which is likely due to the larger size of the former and their partially folded state.[18] We therefore wondered whether the improved thermal and chaotropic stability of Cfa would translate to increased expression levels of Cfa^N fusions. Indeed, model studies in *E. coli* revealed a significant (30-fold) increase in soluble protein expression for a Cfa^N fusion compared to the corresponding Npu^N fusion (Figure S11). Given this result, we wondered whether Cfa^N fusions would also exhibit increased protein expression levels in mammalian cells. In particular, intein fusions to the heavy chain (HC) of monoclonal antibodies (mAbs) have emerged as a powerful tool for site-specific conjugation of synthetic cargoes.[19,20] We compared the expression levels in HEK293 cells of an mAb (αDec205) as a function of the N-intein fused to its HC. Gratifyingly, and consistent with the bacterial expression results, production of the HC-Cfa^N fusion was significantly higher than that for the other inteins examined; for example, the secreted levels of the mAb-Cfa construct were ~10-fold higher than those for the corresponding Npu fusion (Figure 3A,B). Importantly, mAb-Cfa retained PTS activity and could be site-specifically modified with a synthetic peptide by splicing directly in the growth medium following the four-day expression at 37 °C (Figure S12).

Finally, to further explore the utility of the Cfa intein in the context of antibody conjugation, we asked whether the PTS system could be used to attach multiple copies of a synthetic cargo to the heavy chain of the mAb. Accordingly, we used semisynthesis to prepare a construct in which the C-terminal half of Cfa (Cfa^C) was fused to a C-extein containing a dendrimeric scaffold allowing multimeric attachment of cargo, in this case fluorescein (Figure 3C). This dendritic cargo was successfully linked to the αDec205 antibody via Cfa-mediated PTS, again performed directly in situ within the cellular growth medium (Figure 3D, E). To our knowledge, this represents the first time that PTS has been used to attach a branched extein construct to a target protein, highlighting the potential of the system for manipulating the payload quantity of antibody drug conjugates.[21]

The discovery of fast split inteins has revolutionized the applications of protein trans splicing. The remarkable robustness of the Cfa intein described in this study should extend the utility of many of these technologies by allowing PTS to be performed in a broader range of reaction conditions. Moreover,

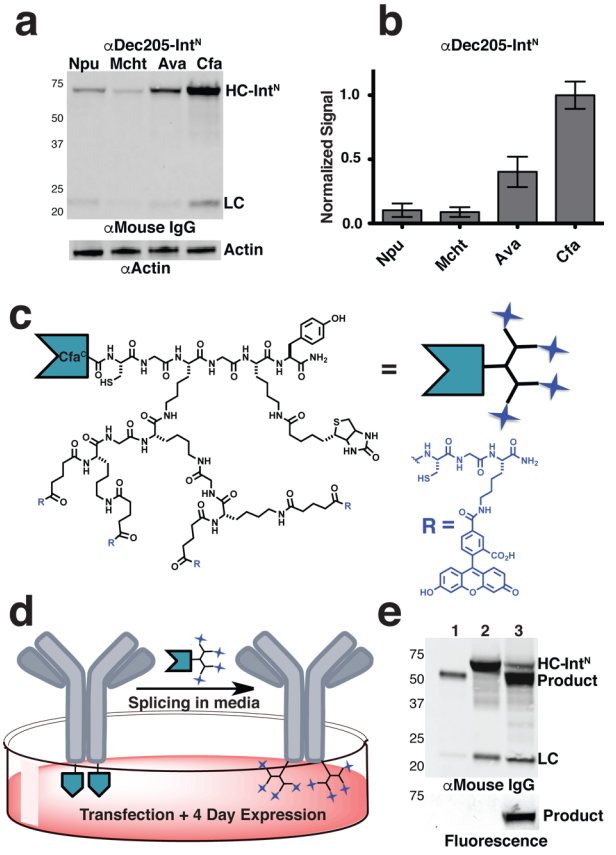

Figure 3. Expression and modification of a mouse monoclonal antibody using the Cfa intein. (a) Test expression in HEK293T cells of various Int^N homologues (Npu, Mcht, Ava, and Cfa) fused to the C-terminus of the heavy chain of a mouse αDec205 monoclonal antibody. Top: Western blot analysis (αMouse IgG) of antibody levels present in the medium following the 96 h expression. Bottom: α-actin Western blot of cell lysate as a loading control. (b) Quantification of normalized expression yield by densitometry of αDEC205 HC-Int^N signal in panel a. Error = SD (n = 4). (c) Structure of the Cfa^C-dendrimer construct used in PTS reactions with the αDEC205 HC-Int^N fusion. For simplicity the Cfa^C peptide sequence and a short peptide linker are depicted symbolically in teal. (d) Schematic of the *in situ* PTS approach used to modify the HC of an mAb with a multivalent cargo. (e) SDS-PAGE analysis of PTS reaction. Lane 1: Wild type mouse αDEC205 mAB. Lane 2: Mouse αDEC205-Cfa^N mAB fusion. Lane 3: addition of the Cfa^C-dendrimer to the media containing the αDEC205-Cfa^N mAB. The splicing reaction was analyzed after two hours by fluorescence (bottom) and Western blot (top, αMouse IgG).

the ability of Cfa to increase the expression yields of N-intein fusions should encourage further use of split inteins for protein semisynthesis. The activity-guided approach we use to engineer this intein may be applied to other intein families or act as a general strategy for the refinement of multiple sequence alignments used for consensus engineering.

■ ASSOCIATED CONTENT

Supporting Information

The Supporting Information is available free of charge on the ACS Publications website at DOI: 10.1021/jacs.5b13528.

Full methods and experimental data (PDF)

DOI: 10.1021/jacs.5b13528
J. Am. Chem. Soc. 2016, 138, 2162−2165

AUTHOR INFORMATION

Corresponding Author

*muir@princeton.edu

Notes

The authors declare no competing financial interest.

ACKNOWLEDGMENTS

The authors thank the members of the Muir laboratory for valuable discussions. This work was supported by the U.S. National Institutes of Health (R37-GM086868) and National Science Foundation Graduate Research Fellowship under Grant No. (DGE-1148900).

REFERENCES

(1) Shah, N. H.; Muir, T. W. *Chem. Sci.* **2014**, *5*, 446.
(2) Wu, H.; Hu, Z.; Liu, X. Q. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95*, 9226.
(3) Iwai, H.; Zuger, S.; Jin, J.; Tam, P. H. *FEBS Lett.* **2006**, *580*, 1853.
(4) Zettler, J.; Schutz, V.; Mootz, H. D. *FEBS Lett.* **2009**, *583*, 909.
(5) Shah, N. H.; Eryilmaz, E.; Cowburn, D.; Muir, T. W. *J. Am. Chem. Soc.* **2013**, *135*, 5839.
(6) Shah, N. H.; Dann, G. P.; Vila-Perello, M.; Liu, Z.; Muir, T. W. *J. Am. Chem. Soc.* **2012**, *134*, 11338.
(7) Carvajal-Vallejos, P.; Pallisse, R.; Mootz, H. D.; Schmidt, S. R. *J. Biol. Chem.* **2012**, *287*, 28686.
(8) Wu, Q.; Gao, Z.; Wei, Y.; Ma, G.; Zheng, Y.; Dong, Y.; Liu, Y. *Biochem. J.* **2014**, *461*, 247.
(9) Aranko, A. S.; Oeemig, J. S.; Kajander, T.; Iwai, H. *Nat. Chem. Biol.* **2013**, *9*, 616.
(10) Pietrokovski, S. *Protein Sci.* **1994**, *3*, 2340.
(11) Dearden, A. K.; Callahan, B.; Roey, P. V.; Li, Z.; Kumar, U.; Belfort, M.; Nayak, S. K. *Protein Sci.* **2013**, *22*, 557.
(12) Du, Z.; Shemella, P. T.; Liu, Y.; McCallum, S. A.; Pereira, B.; Nayak, S. K.; Belfort, G.; Belfort, M.; Wang, C. *J. Am. Chem. Soc.* **2009**, *131*, 11581.
(13) Lehmann, M.; Kostrewa, D.; Wyss, M.; Brugger, R.; D'Arcy, A.; Pasamontes, L.; van Loon, A. P. *Protein Eng., Des. Sel.* **2000**, *13*, 49.
(14) Steipe, B. *Methods Enzymol.* **2004**, *388*, 176.
(15) Altschul, S. F.; Gish, W.; Miller, W.; Myers, E. W.; Lipman, D. J. *J. Mol. Biol.* **1990**, *215*, 403.
(16) Grigoriev, I. V.; Nordberg, H.; Shabalov, I.; Aerts, A.; Cantor, M.; Goodstein, D.; Kuo, A.; Minovitsky, S.; Nikitin, R.; Ohm, R. A.; Otillar, R.; Poliakov, A.; Ratnere, I.; Riley, R.; Smirnova, T.; Rokhsar, D.; Dubchak, I. *Nucleic Acids Res.* **2012**, *40*, D26.
(17) Tatusova, T.; Ciufo, S.; Fedorov, B.; O'Neill, K.; Tolstoy, I. *Nucleic Acids Res.* **2014**, *42*, D553.
(18) Shah, N. H.; Eryilmaz, E.; Cowburn, D.; Muir, T. W. *J. Am. Chem. Soc.* **2013**, *135*, 18673.
(19) Mohlmann, S.; Bringmann, P.; Greven, S.; Harrenga, A. *BMC Biotechnol.* **2011**, *11*, 76.
(20) Barbuto, S.; Idoyaga, J.; Vila-Perello, M.; Longhi, M. P.; Breton, G.; Steinman, R. M.; Muir, T. W. *Nat. Chem. Biol.* **2013**, *9*, 250.
(21) Shah, N. D.; Parekh, H. S.; Steptoe, R. J. *Pharm. Res.* **2014**, *31*, 3150.

DOI: 10.1021/jacs.5b13528
J. Am. Chem. Soc. 2016, 138, 2162−2165