Supporting Information

Lockless and Muir 10.1073/pnas.0902964106

SI Text

Construction of Plasmids. The KanR-Npu* (the "SPD" junction) selection plasmid was created from 4 PCR products encoding the KanR promoter and gene (amino acids 1–189) from pET28b, the N-terminus of Npu DnaE intein (Npu^N), followed by the KanR ribosomal binding site (RBS), the C-terminus of Ssp DnaE intein (Ssp^C), and the remaining portion of the KanR gene (amino acids 190–271). The resulting PCR fragment was digested with KpnI/SacI and cloned into pBluescript.

The plasmids encoding H_6-Ub-Int^N (Int^N = Ssp^N or Npu^N) were created by PCR using 2 overlapping PCR products encoding H_6-ubiquitin and Int^N. This fragment was cloned into a pMal vector using NsiI and HindIII.

The Crk-II N-terminal–intein construct was created by PCR using 3 overlapping PCR products encoding the murine Crk-II SH2 domain (amino acids 1–124), Npu^N, and ubiquitin. The final product was digested and cloned into pET28b using NcoI and SalI. The Crk-II C-terminal–intein construct was created by PCR using 3 overlapping PCR products encoding MBP, Ssp^C, and murine Crk-II SH3n-SH3c domains (amino acids 125–304). The final product was digested and cloned into pET29b using NdeI and KpnI. The Crk-II–intein coexpression plasmid was created from 2 overlapping PCR products of the split Crk-II genes described above with an intervening RBS. The final product was cloned into pET29b using NdeI and KpnI.

The eukaryotic expression plasmids were created as follows. N-terminal inteins with HA tags were cloned into pEB4 3′ of an in-frame MBP using EcoRI and XhoI. C-terminal inteins were PCR-amplified to add an N-terminal M2 tag, a C-terminal "CFN" motif, and HindIII and BamHI restriction enzyme sites for cloning into the GFP containing pMPG04 vector.

Peptide Synthesis. The Ssp^C-"SGV"-HA peptide was synthesized on a Rink-ChemMatrix resin (Matrix Innovation) on a 0.1-mmol scale using standard Fmoc synthesis protocols. All amino acid derivatives were purchased from Novabiochem. After chain assembly, the peptide was cleaved off the resin with 95:2.5:2.5 TFA:H2O:triisopropyl silane (v/v) and was purified by semi-preparative RP-HPLC (Vydac C18) at 50 °C using a linear gradient of 30–38% B (9:1 MeCN/water, 0.1% TFA) over 45 min at a flow rate of 4 mL/min. An RP-HPLC chromatograph and mass spectra of the purified peptide are shown in supporting information (SI) Fig. S10. The Ssp^C-"CFN"-HA peptide synthesis conditions and characterization can be found in the article by Vila-Perello et al. (1).

Isolation of Branched Intermediate. H_6-Ub-Npu^N was mixed with Ssp^C-"SGV"-HA peptide at 30 °C. Aliquots were removed at t = 0, 1, 2, and 3 h and mixed 1:1 with HPLC buffer A (0.1% TFA), and components were separated on a C18 analytical column over a 25–58% B gradient. The peak at 19.5 min was collected, dried, and resuspended in 50:49.9:0.1 MeCN:H2O:formic acid (v/v) for electrospray ionization (ESI) mass spectrometry.

1. Vila-Perello M, Hori Y, Ribo M, Muir TW (2008) Activation of protein splicing by protease- or light-triggered O to N acyl migration. *Angew Chem Int Ed* 47:7764–7767.

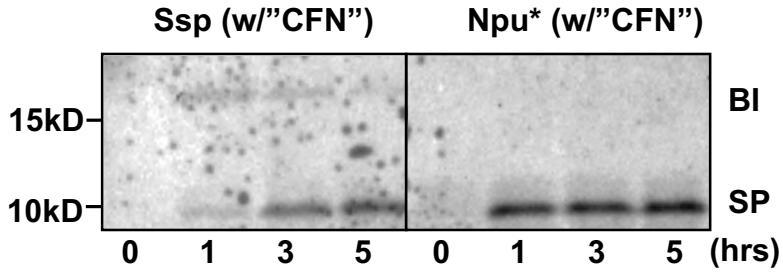


Fig. S1. In vitro protein *trans*-splicing of Ssp and Npu* DnaE inteins. Time course of in vitro splicing reactions employing either Ssp or Npu* split inteins using a "CFN" splice junction and the model extein indicated in Fig. 2*A*. Reaction mixtures were resolved by nonreducing SDS-PAGE and blotted using an anti-HA antibody. BI, branched intermediate (expected: 15.7 kDa); SP, spliced product (expected: 11.8 kDa).

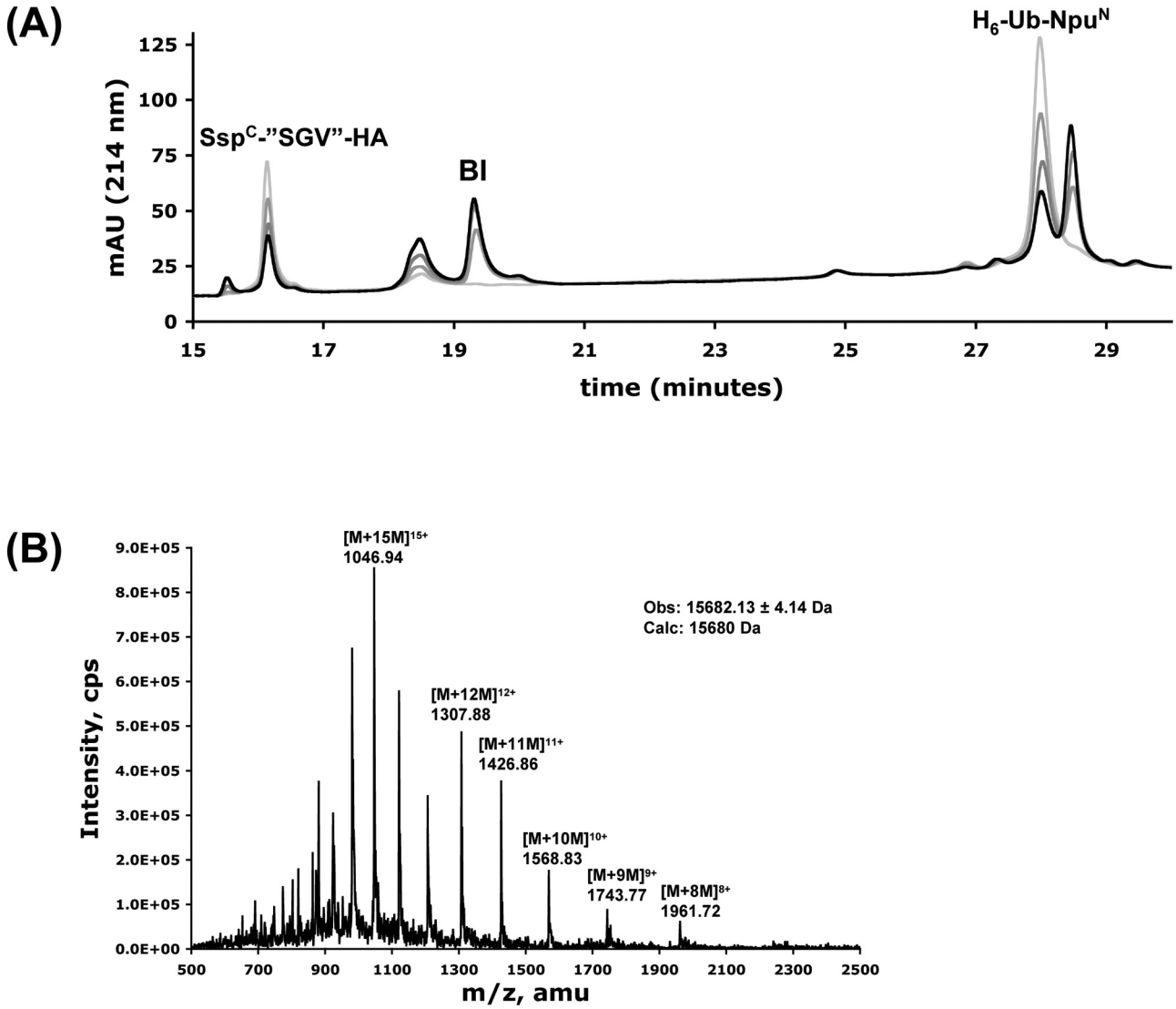
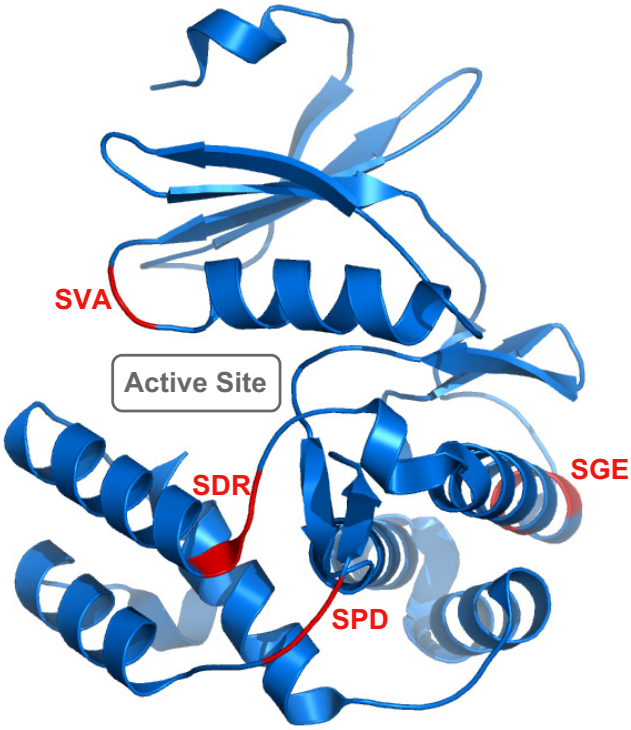


Fig. S2. Isolation and identification of the branched intermediate. (*A*) RP-HPLC chromatograph showing components of splicing mixture as a function of time. The black trace is t = 3 h, and the light gray trace is t = 0, whereas intermediate time points are intermediate shades of gray. The branched intermediate (BI) and precursor fragments (H_6-Ub-Npu^N and Ssp^C-"SGV"-HA) are labeled. (*B*) ESI mass spectrum of the BI peak (expected *m/z* 15,680 Da).



MSHIQRETSCSRPRLNSNMDADLYGYKWARDNVGQ
SGATIYRLYGKPDAPELFLKHGKGSVANDVTDEMV
RLNWLTEFMPLPTIKHFIRTPDDAWLLTTAIPGKT
AFQVLEEYPDSGENIVDALAVFLRRLHSIPVCNCP
FNSDRVFRLAQAQSRMNNGLVDASDFDDERNGWPV
EQVWKEMHKLLPFSPDSVVTHGDFSLDNLIFDEGK
LIGCIDVGRVGIADRYQDLAILWNCLGEFSPSLQK
RLFQKYGIDNPDMNKLQFHLMLDEFF

Fig. S3. Intein splice site junctions in KanR. The 4 test sites (in red) within the KanR gene are mapped onto the crystal structure (pdb1ND4) and protein sequence (gi no. 4378798). SGE was used in previous studies, although the remaining 3 begin with a serine, are found in loops, and are not conserved across many KanR proteins. The "SPD" site was used in this study.

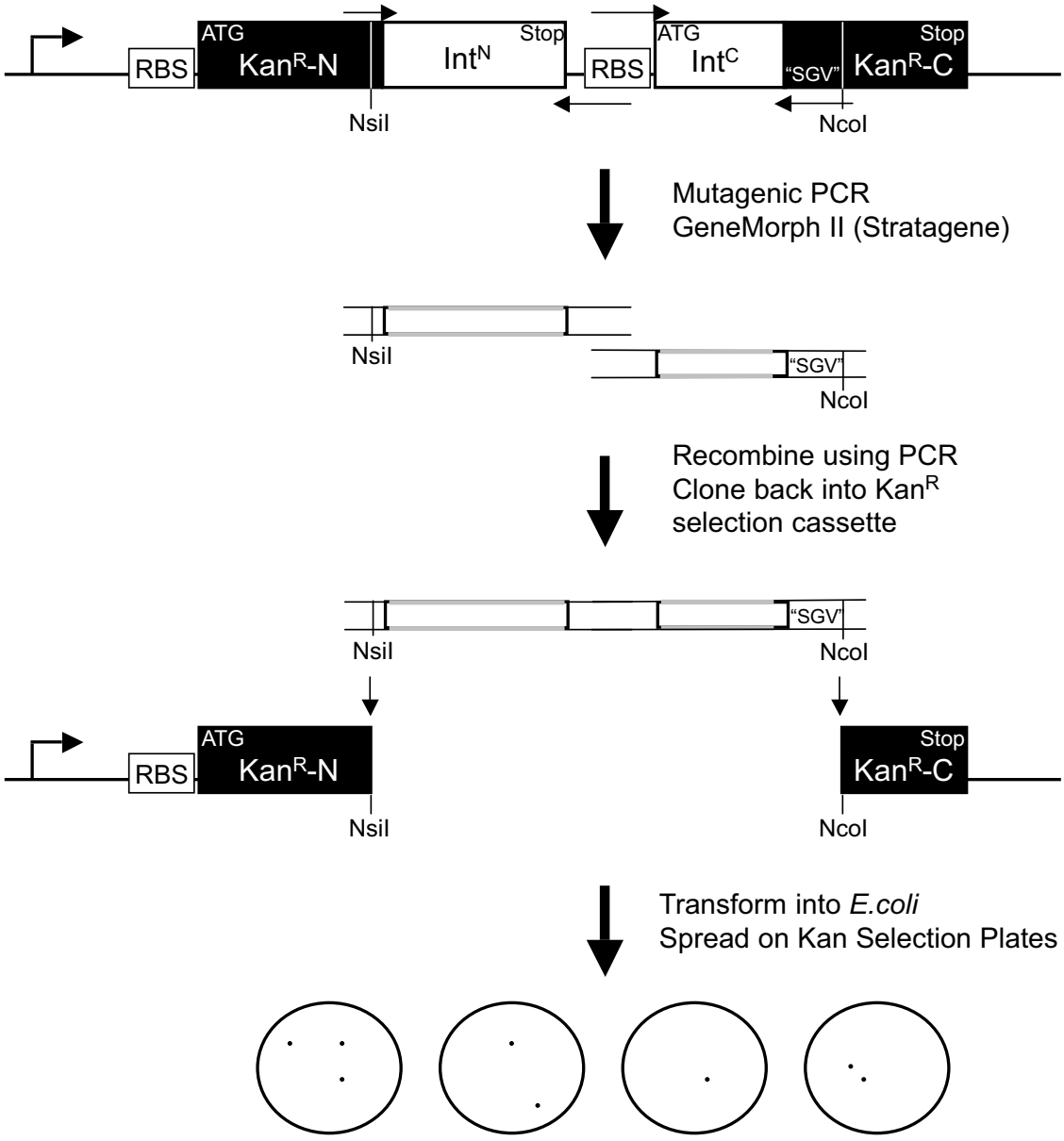


Fig. S4. Random mutagenesis strategy for KanR-Npu*. GeneMorph II (Stratagene) was used to mutagenize the N- and C-terminal fragments of the split intein randomly. Oligonucleotides flanking each intein fragment are shown as thin arrows and are designed to preserve the cloning sites, start and stop codons, RBS, and extein junction sequences. PCR products from the N- and C-terminal parts of the intein are recombined in a subsequent nonmutagenic PCR to give a full-length product. This product is digested with NsiI and NcoI, ligated into the Kan^R selection vector, transformed into DH5α *E. coli*, and selected on LB plates containing kanamycin.

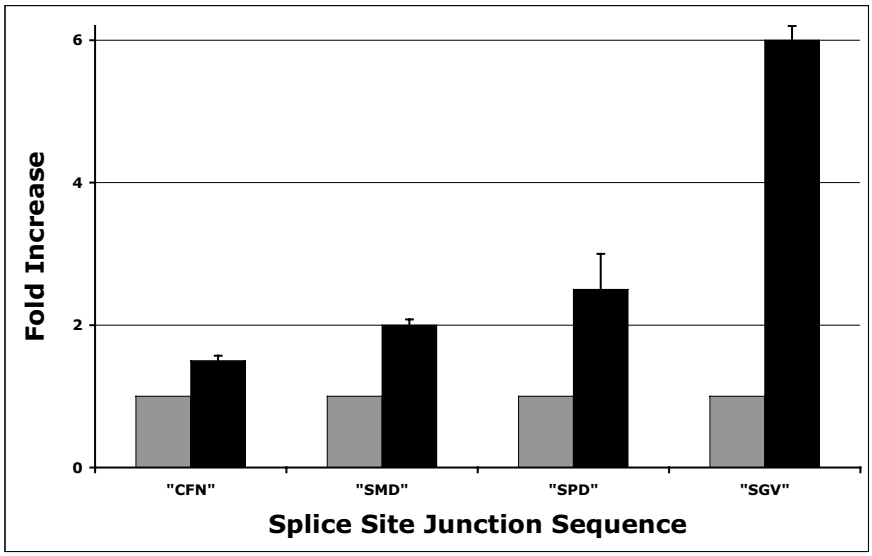


Fig. S5. mNpu* activity with different splice site junctions. Graph showing the fold change in activity of Npu* vs. mNpu* inteins on 4 splice junction sequences. The data are normalized to the mean activity of Npu* with the respective sequence. The standard deviation after error propagation is shown.

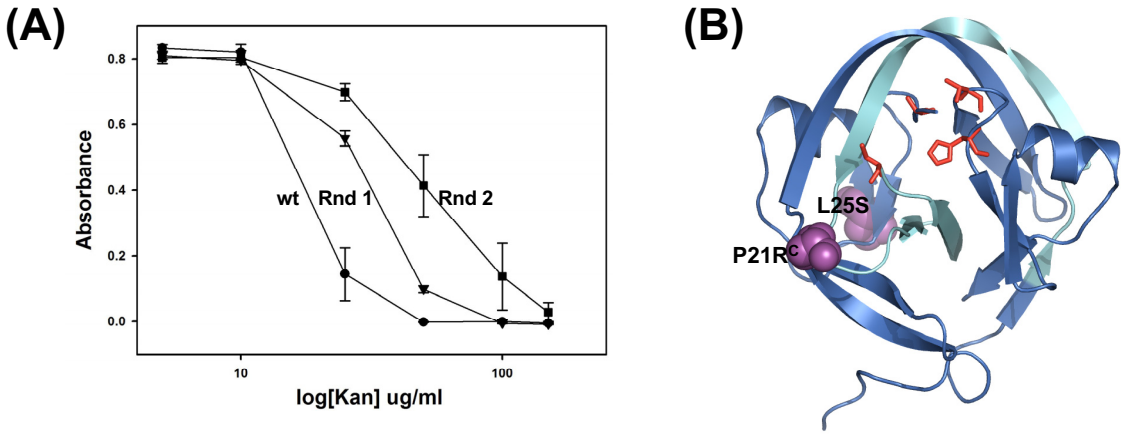


Fig. S6. Evolution of split inteins at 37 °C. (*A*) Growth of transformed *E. coli* cells after 16 h at 37 °C in LB containing different concentrations of kanamycin. Each curve is the growth profile for cells expressing the Npu*-KanR mutant with the highest activity (as determined by this assay) from each round of evolution. The WT Npu* starting point is shown for comparison. Errors = SEM ($n = 3$). (*B*) Mutations in the final selected Npu* intein (mNpu37*) are mapped onto the Ssp intein structure (pdb1DE3). The blue- and light blue-colored ribbons indicate the Int^N and Int^C portions of the intein, respectively. Essential catalytic residues are rendered as red sticks, and mutated amino acids are shown in purple space-filling representation.

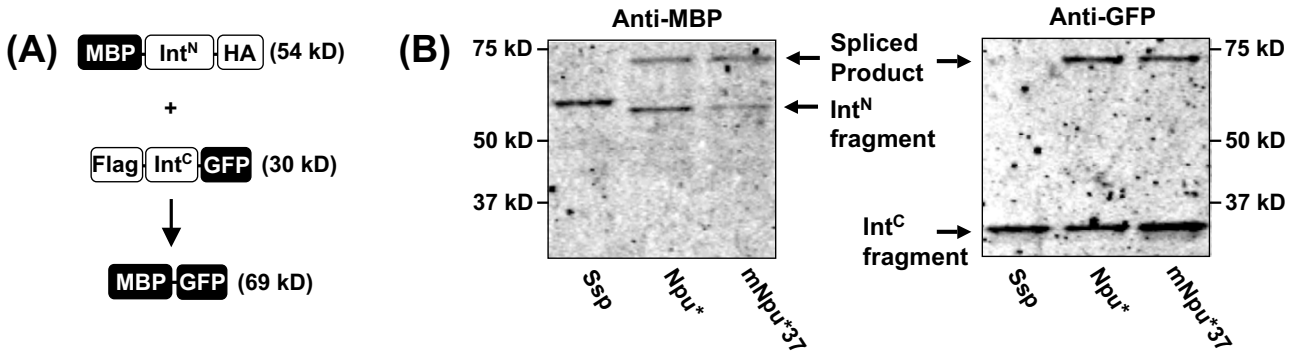


Fig. S7. Splicing of evolved intein in mammalian cells. (*A*) Schematic showing the starting materials and expected MBP-GFP product (with an intervening "CFN" splice junction) of the splicing reaction in mammalian cells. (*B*) Anti-MBP or anti-GFP Western blots of HeLa cell extracts after 8 h of expression of indicated split intein pairs. The Ssp^N intein is slightly larger than the Npu^N intein, which explains the shift in the molecular weight of the precursor band. The spliced product of all 3 reactions is the same size.

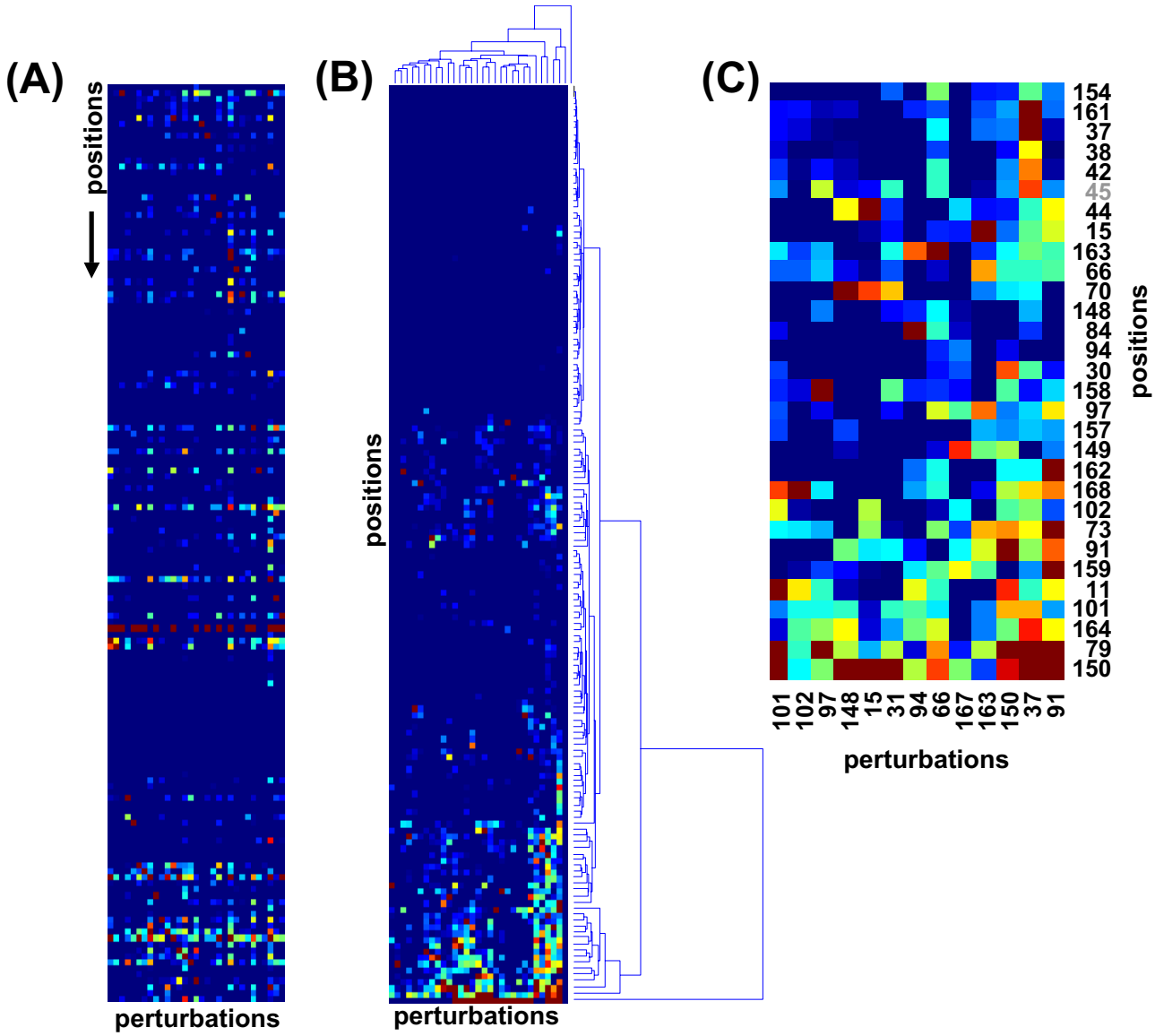


Fig. S8. SCA of the intein protein family. (*A*) A matrix representation of the statistical coupling between perturbations and positions in the alignment that correspond to the Ssp DnaE intein. The scale is from low coupling values in blue to high coupling values in red. (*B*) A 2-D clustering of the matrix in *A* is used to group positions (*y* axis) and perturbations (*x* axis) with similar patterns of coupling values together. (*C*) The final group of residues shown in Fig. 6*B* is a group of self-consistent perturbations and positions. Position 45 is shown in gray because too few sequence data exist to evaluate the significance of its inclusion in the group.

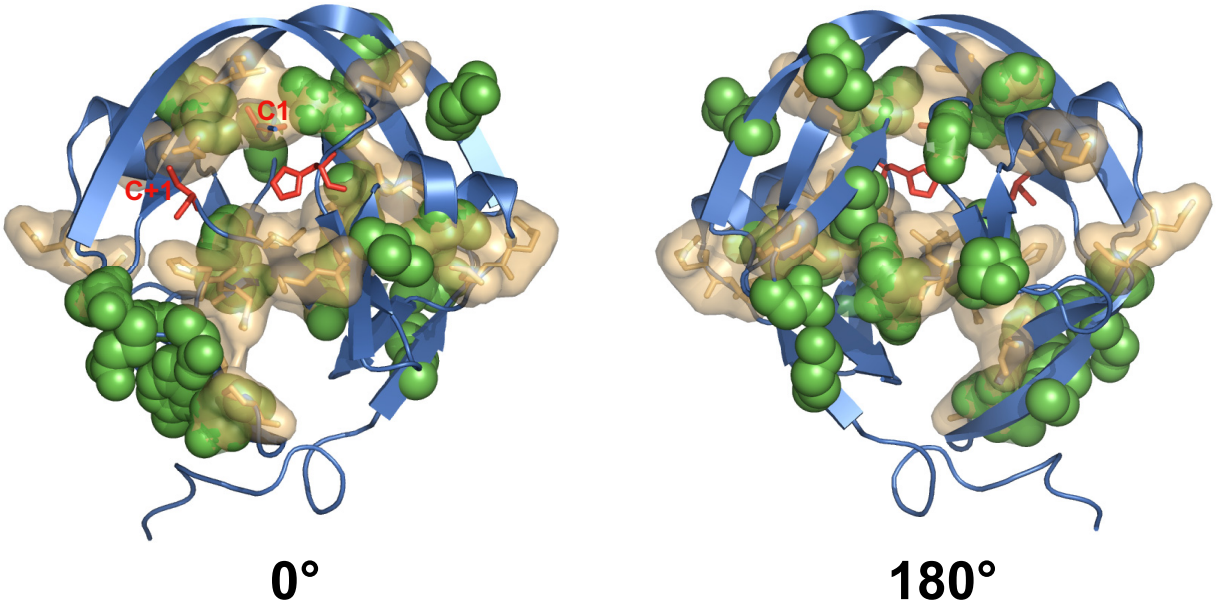


Fig. S9. Mutations from intein evolution experiments. The SCA network of coevolving residues (tan surface) is mapped onto the Ssp DnaE intein structure. The essential catalytic intein residues (C1, H82, and C+1) are in red, whereas sites of mutation from intein evolution experiments are shown in green (refs. 1–3 and this work).

1. Iwai H, Zuger S, Jin J, Tam PH (2006) Highly efficient protein trans-splicing by a naturally split DnaE intein from Nostoc punctiforme. *FEBS Lett* 580:1853–1858.
2. Wood DW, Wu W, Belfort G, Derbyshire V, Belfort M (1999) A genetic system yields self-cleaving inteins for bioseparations. *Nat Biotechnol* 17:889–892.
3. Adam E, Perler FB (2002) Development of a positive genetic selection system for inhibition of protein splicing using mycobacterial inteins in Escherichia coli DNA gyrase subunit A. *J Mol Microbiol Biotechnol* 4:479–487.

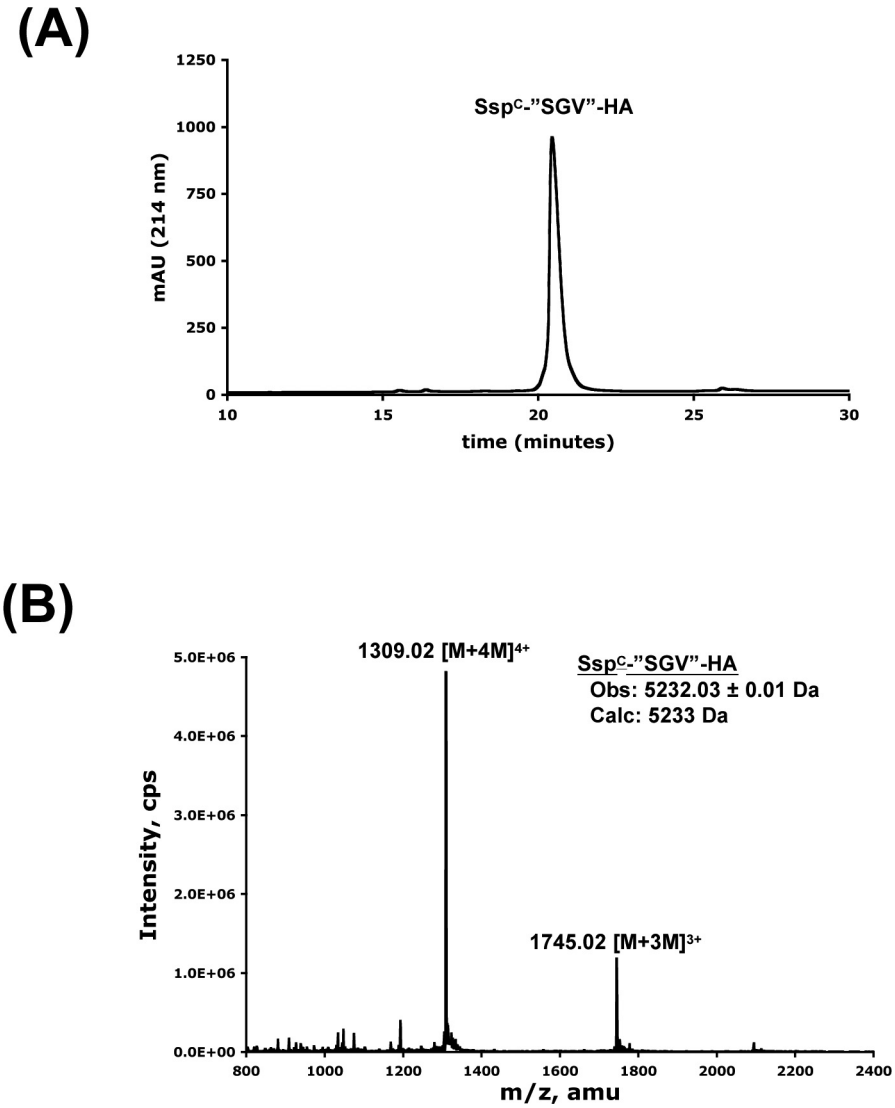
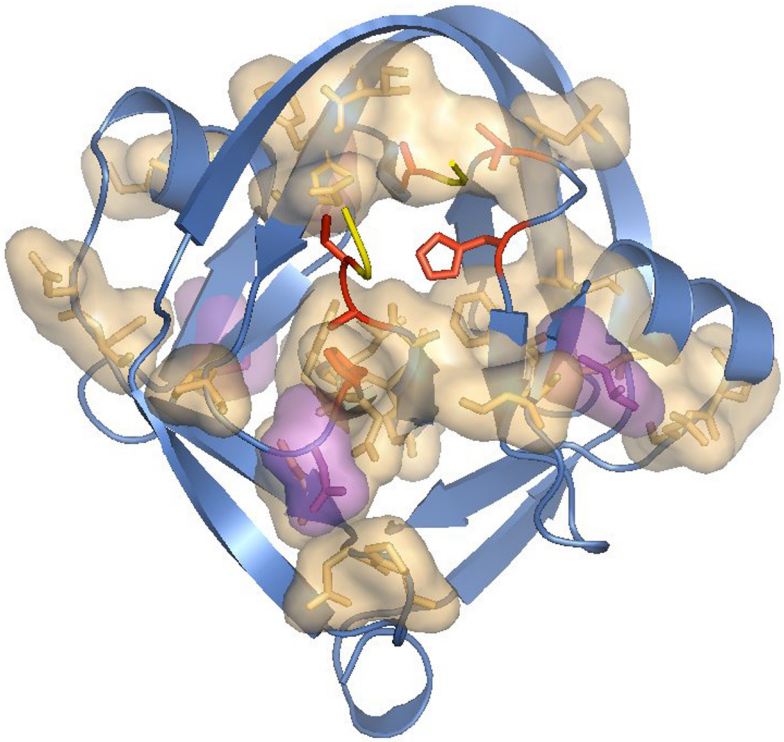


Fig. S10. Characterization of Ssp^{C}-"SGV"-HA peptides. (*A*) RP-HPLC chromatograph of purified Ssp^{C}-"SGV"-HA peptide on a gradient of 25–45% B. (*B*) ESI mass spectra of the purified Ssp^{C}-"SGV"-HA peptide.



Movie S1. The SCA network of coevolving residues (tan surface) is mapped onto a blue ribbon representation of the Ssp DnaE intein structure (pdb1DE3). The essential catalytic intein residues (C1, T79, H82, H24C, N36C, and C+1) are in red, whereas sites of mutation from the "SGV" splice junction and 37°C evolution experiments are shown in purple. The surface of the intein appears after one revolution of the structure to illustrate the sparseness of the SCA network.

Movie S1 (GIF)