Supporting Information

for

Extein Residues Play an Intimate Role in the Rate Limiting Step of Protein *Trans*-Splicing

Neel H. Shah,[†] Ertan Eryilmaz,[‡] David Cowburn,[‡] 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

Table of Contents: **Page #**

Materials and Methods

Figures and Tables

Materials and Methods

Materials

All buffering salts, isopropyl-β-D-thiogalactopyranoside (IPTG), and *N,N*-diisopropylethylamine (DIPEA) were purchased from Fisher Scientific (Pittsburgh, PA). Ampicillin (Amp), kanamycin sulfate (Kan), β-Mercaptoethanol (BME), DL-dithiothreitol (DTT), sodium 2-mercaptoethanesulfonate (MESNa), Coomassie brilliant blue, *N,N*-dimethylformamide (DMF), triisopropylsilane (TIS), L-cysteine, L-cysteine methyl ester hydrochloride, L-phenylalanine methyl ester hydrochloride, L-phenylalanine amide hydrochloride, methylamine, 1-ethyl-3-(3-dimethylaminopropyl)carbodiimide) (EDC), and diethyl ether were purchased from Sigma-Aldrich (St. Louis, MO). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Thermo Scientific (Rockford, IL). Bis-L-cysteine amide disulfide hydrochloride $(\text{H-Cys-NH}_2)_2\text{-HCl}$ was purchased from BACHEM (Bubendorf, Switzerland). Fmoc-Ala-OH, Fmoc-Phe-OH, Fmoc-Cys(Trt)-OH, Fmoc-Asn(Trt)-OH, Boc-Cys(Trt)-OH, and *O*-(Benzotriazol-1-yl)-*N,N,N',N'*-tetramethyluronium hexafluorophosphate (HBTU) were purchasd from Novabiochem (Läufelfingen, Switzerland). Piperidine was purchased from Alfa Aesar (Ward Hill, MA). Dichloromethane (DCM) and rink amide resin were purchased from EMD Chemicals (Billerica, MA). 1-Hydroxybenzotriazole hydrate (HOBt) was purchased from AnaSpec (Fremont, Ca). Trifluoroacetic acid (TFA) was purchased from Halocarbon (North Augusta, SC). Deuterated NMR solvents, $^{15}NH_4Cl$, and ^{13}C-glucose were purchased from Cambridge Isotope Laboratories (Andover, MA).

Complete protease inhibitor tablets were purchased from Roche Diagnostics (Mannheim, Germany). Nickel-nitrilotriacetic acid (Ni-NTA) resin was from Novagen (Gibbstown, NJ). The QuikChange XL II site directed mutagenesis kit was from Agilent (La Jolla, CA). DpnI and the Phusion High-Fidelity PCR kit were from New England Biolabs (Ipswich, MA). DNA purification kits (QIAprep spin minikit, QIAquick gel extraction kit, QIAquick PCR purification kit) were from Qiagen (Valencia, CA). Sub-cloning efficiency DH5α competent cells and One Shot BL21(DE3) chemically competent *E.*

coli were purchased from Invitrogen (Carlsbad, CA) and used to generate “in-house” high-competency cell lines. Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). All plasmids used in this study were sequenced by GENEWIZ (South Plainfield, NJ). Criterion XT Bis-Tris gels (12%) and Mini-PROTEAN TGX gels (12%) were purchased from Bio-Rad (Hercules, CA). 20x MES-SDS running buffer was purchased from Boston Bioproducts (Ashland, MA).

Equipment

Size-exclusion chromatography was carried out on an ÄKTA FPLC system from GE Healthcare. Preparative FPLC were carried out on a Superdex 75 16/60 column (column volume, CV, = 125 mL). For all runs, proteins were eluted over 1.35 CV of buffer (flow rate: 1 mL/min). Analytical RP-HPLC was performed on Hewlett-Packard 1100 and 1200 series instruments equipped with a C_{18} Vydac column (5 μm, 4.6 x 150 mm) at a flow rate of 1 mL/min. Preparative RP-HPLC was performed on a Waters prep LC system comprised of a Waters 2545 Binary Gradient Module and a Waters 2489 UV detector. Purifications were carried out on a C_{18} Vydac 218TP1022 column (10 μM; 22 x 250 mm) at a flow rate of 18 mL/min. All runs used 0.1 % TFA (trifluoroacetic acid) in water (solvent A) and 90 % acetonitrile in water with 0.1% TFA (solvent B). For all analytical runs, a two minute isocratic period in initial conditions was followed by a 30 minute linear gradient with increasing solvent B concentration. Electrospray ionization mass spectrometric analysis (ESI-MS) was performed on a Bruker Daltonics MicrOTOF-Q II mass spectrometer. Gels were quantified on a LI-COR Odyssey Infrared Imager. NMR experiments were carried out at either Varian Inova (600MHz) or Bruker (600, 800, or 900MHz) spectrometers equipped with cryogenic probes that can apply pulse field gradients along the z-axis.

Synthesis of C-extein molecules for expressed protein ligation

L-cysteine and L-cysteine methyl ester were used as purchased. Typically, these amino acids were dissolved in Milli-Q water to 1.2 M, then the solutions were neutralized to pH 7 with sodium hydroxide, and finally the volumes were raised to achieve a cysteine concentration of 1 M.

Preparation of L-Cys(NH_2)

L-Cysteine amide was purchased in a disulfide dimer hydrochloride salt form. Immediately prior to its use for thiolysis and ligation, a neutralized and reduced form was prepared as follows. First, (H-Cys-NH_2)$_2$-HCl was dissolved in a pH 8.0 phosphate buffered saline to a final concentration of 1 M and a final volume of 0.8 mL. Next, 0.8mL of a 1.1 M TCEP solution was prepared by dissolving TCEP-HCl in a 3:5 (v/v) mixture of 1 M Tris and 5 N NaOH to afford a solution with pH ~6.0. The two solutions were mixed and 300μL of 5N NaOH was added to raise the pH to 7. The mixture was incubated at room temperature for 10 minutes to reduce the cysteine amide. The resulting solution was directly used for thiolysis and ligation without any additional purification.

Synthesis of L-Cys($NHCH_3$)

Cysteine N-methyl amide was prepared in two steps from Boc-L-Cys(Trt)-OH. In the first step, 1 g of Boc-L-Cys(Trt)-OH (2.16 mmol), 364 mg of HOBt•H_2O (2.38 mmol), and 455mg of EDC•HCl (2.38 mmol) were dissolved in 35mL of DCM. The mixture was stirred under argon on ice until almost everything dissolved. 3.25 mL of 2 M methylamine in THF (6.48 mmol) was added, followed by 3 mL of DMF to help solubilize whatever was still left in solution. Finally, 413 μL of DIPEA were added to neutralize any excess HCl from the EDC. The reaction was warmed up to room temperature and stirred overnight. The next day, the solution was extracted with two 25 mL 5% citric acid washes, one water wash, then two 25 mL saturated bicarbonate washes. The organic phase was collected and dried by rotary evaporation. In the second step, 40 mL of deprotection cocktail (49:50:1 of DCM:TFA:TIS) was added to dissolve the residue, and the deprotection was carried out for 30 minutes at room temperature. The solvent was removed under vacuum, and the residue was resuspended in 20 mL of water then filtered. The filtrate was collected and analyzed by Ellman's test to determine free thiol concentration, indicating a 74% yield. The solution was frozen and lyophilized to give an amorphous solid, which was analyzed by ^{1}H-NMR to confirm the product (Figure S3A). The product was dissolved in water and stored at -80 °C.

Solution-phase synthesis of CF(OCH_3) and CF(NH_2)

Boc-L-Cys(Trt) (0.575 mmol, 267 mg), HOBT (0.575 mmol, 78 mg), and EDC•HCl (0.575 mmol, 110 mg) were dissolved in DCM (9 mL) in a 25-mL flask under argon atmosphere at 0 °C. DIPEA (0.575 mmol, 100 μL) was added to the mixture slowly and it was stirred for 15 min at 0 °C. Then phenylalanine methyl ester HCl (0.46 mmol, 100 mg) was dissolved in 1mL of DMF and added to the mixture, followed by more DIPEA (0.575 mmol, 100 μL). After 20 minutes, the reaction was allowed to proceed overnight under argon and warm slowly to room temperature as the ice melted. The next morning, 25mL of ether was added to the reaction solution, and the solution was transferred to a separatory funnel. The reaction solution was extracted twice with 25 mL of 5% citric acid, then twice with 25 mL of saturated sodium bicarbonate. The organic phase was removed leaving a damp precipitate which was redissolved in 20 mL of DCM. 20 mL of TFA and 500 μL of TIS were added and the deprotection was carried out for 30 minutes. The DCM/TFA was rotovapped off, then the precipitated material was resuspended in 25mL of aqueous 0.1% TFA and filtered. The dissolved crude dipeptide was purified by preparative RP-HPLC over a 10-25% B gradient in 60 minutes, preceded by a 5 minute isocratic phase in 10% B. The pure HPLC fractions were lyophilized then redissolved in water. The concentration was determined by Ellman's test and by phenylalanine absorbance at 257 nm (ε_{257} = 190 1/M*cm), then the peptide was diluted to 100 mM and stored at -80 °C. The CF(NH_2) peptide was prepared analogously to the CF(OCH_3) peptide, with isolated yields of 55% and 50%, respectively. The products were confirmed by ^{1}H-NMR (Figures S3B-C) and by ESI-MS (Table S1).

Solid-phase synthesis of CFN(NH_2), CFA(NH_2), and CAN(NH_2)

Fmoc-based solid phase peptide synthesis (SPPS) using standard Fmoc amino acids was used to produce the three C-extein tripeptides. The peptides were synthesized on Rink amide resin at a 0.25 mmol scale as follows. First, 20% piperidine in DMF was used for Fmoc deprotection of the resin using a one minute equilibration of the resin followed by a 20 minute incubation at room temperature with agitation by a stream of N_2 gas. After Fmoc deprotection and DMF washes, amino acids were coupled using HOBt/HBTU as activating agents. First, the amino acid (1.25 mmol) was dissolved to 0.5 M in a DMF solution containing 0.45 M HOBt and 0.45 M HBTU (2.5 mL). DIPEA (2.5 mmol) was added to the amino acid/activating agent solution, and the mixture was immediately added to the resin. The coupling reaction was carried out for 30 minutes at room temperature with agitation by a stream of N_2 gas. After substantial washing with DMF, subsequent deprotections and couplings were carried out identically as the initial steps.

The peptides were cleaved off the resin using 95% TFA, 2.5% TIS and 2.5% H_2O (4 mL) for two hours at room temperature. After cleavage, the crude peptides were precipitated with 45 mL of cold diethyl ether. The precipitates were redissolved in water with 0.1 % TFA and purified by preparative RP-HPLC on C_{18} prep column over a 0-15% B gradient in 60 minutes. The purified peptides were analyzed by analytical RP-HPLC, ^{1}H-NMR (Figures S3D-F) and ESI-MS (Table S1) to confirm their identity and purity. Peptide concentrations were determined by the Ellman's test for free thiols and phenylalanine absorbance when applicable.

Cloning of N- and C-intein plasmids

The N-inteins were cloned into a pET vector with a His_6-SUMO tag at the 5' end of the multiple cloning site by overlap-extension PCR.[1] First, the NpuN and SspN genes were amplified from our previously used expression vectors[2] with an N-terminal AEY N-extein sequence and 5' and 3' overhangs complementary to appropriate regions of the pET-SUMO vector. Next, the resulting amplicons were inserted downstream of the SUMO gene by PCR using the Phusion polymerase. The new plasmids, pET-SUMO-AEY-NpuN and pET-SUMO-AEY-SspN, encoded for the following protein sequences (the product of SUMO removal is shown in bold):

H$_6$-SUMO-AEY-NpuN
MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKI
KKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG**A**
EYCLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEY
CLEDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN

H$_6$-SUMO-AEY-SspN
MGSSHHHHHHGSGLVPRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKI
KKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGG**A**
EYCLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGEQEVLEY
ELEDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFPLLD
AGTIK

The C1A mutation (red and underlined) was introduced into this plasmid using the QuikChange Site-Directed Mutagenesis kit with the manufacturer's recommended protocol.

The C-inteins were also cloned by overlap-extension PCR[1] into a modified pTXB1 intein vector containing the GyrA intein mutated for EPL followed by a C-terminal His$_6$-tag. First, the NpuC and SspC genes were amplified from previously used expression vectors[2] with the appropriate overhangs. Then the resulting amplicons were inserted upstream of GyrA in the pTXB1 vector to yield the plasmids pTXB1-NpuC-GyrA-H$_6$ and pTXB1-SspC-GyrA-H$_6$, which encode for the following protein sequences:

NpuC-GyrA-H$_6$
MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASNCITGDALVALPEGESVRIADIVPG
ARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHPLLCLVDVA
GVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPGLVRFLEAHH
RDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHAHHHHHH

SspC-GyrA-H$_6$*
***M*VKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN**CITGDALVALPEGESVRIADIVPG
ARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYTVRTVEGLRVTGTANHPLLCLVDVA
GVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGKPEFAPTTYTVGVPGLVRFLEAHH
RDPDAQAIADELTDGRFYYAKVASVTDAGVQPVYSLRVDTADHAFITNGFVSHAHHHHHH

The D124Y, H125N, and N137A mutations (red and underlined) were introduced into this plasmid using the QuikChange Site-Directed Mutagenesis kit with the manufacturer's recommended protocol.

* Note that for the SspC fusion, the N-terminal methionine (italics and underlined) was removed >50% by endogenous methionine amino-peptidases when over-expressed in *E. coli* BL21(DE3) cells, whereas for NpuC it is only cleaved roughly 10-20%.

Semisynthesis and purification of C-intein constructs

E. coli BL21(DE3) cells transformed with a C-intein plasmid were grown in 2 L of LB medium containing ampicillin (100 µg/mL) at 37 °C until OD_{600} = 0.6. Then, expression was induced by addition of 0.5 mM IPTG for 3 hours at 37 °C. After harvesting the cells by centrifugation (10,500 rcf, 30 min), the cell pellets were transferred to 50 mL conical tubes with 5 mL of lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, no thiols, pH 8.0) and stored at -80°C. The cell pellets were resuspended by adding an additional 35 mL of lysis buffer supplemented with Complete protease inhibitor cocktail. Cells were lysed by sonication (35% amplitude, 10x 20 second pulses separated by 30 seconds on ice). The soluble fraction was recovered by centrifugation (35,000 rcf, 30 min). The soluble fraction was mixed with 4 mL of Ni-NTA resin and incubated at 4 °C for 30 minutes. After incubation, the slurry was loaded onto a fritted column. After discarding the flow-through, the column was washed with 5 column volumes (CV) of lysis buffer, 5 CV of wash buffer 1 (lysis buffer with 20 mM imidazole), and 3 CV of wash buffer 2 (lysis buffer with 50 mM imidazole). The protein was eluted with elution buffer (lysis buffer with 250 mM imidazole) in four 1.5 CV elution fractions. The wash and elution fractions were analyzed by SDS-PAGE on Mini-PROTEN TGX gels run in Tris/glycine/SDS running buffer (Figure S4).

The cleanest fractions containing the desired fusion protein were treated in one of two ways depending on the C-extein. For single amino acid C-exteins, the eluted protein from the Ni column (typically 20-30 mL at 50-100 µM) was treated with 10 mM TCEP and the Roche Complete protease inhibitor cocktail. Then the cysteine derivative (free carboxylate, methyl ester, carboxamide, or N-methyl amide) was added to the solution to a final concentration of 100 mM. The direct thiolysis/ligation reaction solution was kept at room temperature for 15-20 hours. For the di- and tri-peptide C-exteins, the eluted protein was first dialyzed at 4 °C against lysis buffer to remove excess imidazole. The dialyzed solution was then treated with 10 mM TCEP and the Roche Complete protease inhibitor cocktail. Next, 100 mM MESNa was added, followed immediately by 1-5 mM di- or tri-peptide. These one-pot thiolysis/ligation reaction solutions were kept at room temperature for 15-20 hours. For both reaction types, if the pH dropped after addition of all reagents, it was raised to pH 7.5 by addition of sodium hydroxide. Note that if the MESNa thiolysis and ligation reaction were not done in one pot, the C-terminal asparagine thioester would cyclize, then hydrolyze, resulting in a dead-end side product carboxylic acid.

After analysis by analytical RP-HPLC to confirm reaction completion, HPLC solvent B was added to achieve a final concentration of 10% B. Then, neat TFA was added to 0.5% to acidify the solution. Upon acidification, the cleaved GyrA-H_6 protein typically precipitated, but no C-intein adduct was lost. This solution was centrifuged then filtered for purification by preparative RP-HPLC. All C-intein constructs (typically 20-30 mL) were loaded on a C_{18} preparative column over a 10 minute isocratic phase in 10% B. Then the system was raised to the gradient starting conditions (20% or 25% B) over a 5 minute phase. Finally, the proteins were purified over a 20-35% B or 25-40% B gradient in 60 minutes. Fractions were analyzed by analytical RP-HPLC, and the purest fractions were lyophilized, then redissolved in water and pooled. Concentrations (50-400 μM) were determined by A_{280} using a calculated extinction coefficient (ProtParam tool on the Expasy server).[3] Purity of the constructs was assessed by analytical RP-HPLC (Figure S7), and their identities were confirmed by ESI-MS (Table S2).

Purification of N-intein constructs

E. coli BL21(DE3) cells transformed with each N-intein plasmid were grown in 2 L of LB containing 50 μg/mL of kanamycin at 37°C until OD_{600} = 0.6. The cells were then cooled down to 18 °C, and expression was induced by addition of 0.5 mM IPTG for 16 hours at 18 °C. After harvesting the cells by centrifugation (10,500 rcf, 30 min), the cell pellets were transferred to 50 mL conical tubes with 5 mL of lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, 2 mM BME, pH 8.0) and stored at -80°C. The cell pellets were resuspended by adding an additional 35 mL of lysis buffer supplemented with Complete protease inhibitor cocktail. Cells were lysed by sonication (35%

amplitude, 10x 20 second pulses separated by 30 seconds on ice). The soluble fraction was recovered by centrifugation (35,000 rcf, 30 min). The soluble fraction was mixed with 4 mL of Ni-NTA resin and incubated at 4°C for 30 minutes. After incubation, the slurry was loaded onto a fritted column. After discarding the flow-through, the column was washed with 5 column volumes (CV) of lysis buffer, 5 CV of wash buffer 1 (lysis buffer with 20 mM imidazole), and 3 CV of wash buffer 2 (lysis buffer with 50 mM imidazole). The protein was eluted with elution buffer (lysis buffer with 250 mM imidazole) in four 1.5 CV elution fractions. The wash and elution fractions were analyzed by SDS-PAGE on Mini-PROTEN TGX gels run in Tris/glycine/SDS running buffer (Figure S5A). The purest fractions of eluted protein were dialyzed at 4 °C against lysis buffer to remove excess imidazole.

As an alternative, cleaner and higher-yielding enrichment method, the N-intein could be extracted from the insoluble *E. coli* inclusion bodies as follows. First, the lysate pellet was resuspended in 40 mL of Triton wash buffer (lysis buffer with 0.1% Triton X-100) and incubated at room temperature for 30 minutes. The Triton wash was centrifuged at 35,000 rcf for 30 minutes, and the supernatant was discarded. Next, the pellet was resuspended in 40 mL of lysis buffer containing 6 M urea, and the mixture was incubated overnight at 4°C. The mixture was centrifuged at 35,000 rcf for 30 minutes, and then the supernatant was mixed with 4 mL of NiNTA resin. The Ni column was run identically as for the native purifications described above, except that every buffer had a background of 6 M urea (Figure S5B). Following enrichment over a Ni-NTA column, the protein was refolded into lysis buffer (without urea) by step-wise dialysis removal of the urea and excess imidazole at 4°C. It is noteworthy that the flow-through from this denaturing Ni column could be reapplied to Ni resin to purify even more of the desired protein.

After enrichment over the Ni-NTA columns, the native or refolded proteins were treated with 10 mM TCEP and Ulp1 (SUMO-specific protease) overnight at room temperature. The cleavage was confirmed by RP-HPLC/MS, after which the reaction solution was incubated with Ni-NTA resin at room temperature for 30 min. The flow-through and two 1.5 CV washes with wash buffer 1 were collected and pooled (Figure S6). The protein was then concentrated to 10 mL, treated with 10 mM TCEP, injected onto the S75 16/60 gel filtration column, and eluted over 1.35 CV in freshly prepared, degassed splicing assay buffer (100 mM sodium phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.2) supplemented with 1mM DTT. FPLC fractions were analyzed by SDS-PAGE, and the purest fractions were pooled and analyzed by analytical RP-HPLC (Figure S7) and ESI-MS (Table S2). The concentration of pure protein was determined by UV A_{280nm} using the calculated extinction coefficient.[3] The N-inteins could be stored at -80 °C after adding buffered glycerol to a final concentration of 20% and flash-freezing in liquid N_2.

RP-HPLC and ESI-MS analysis of protein *trans*-splicing

Prior to any splicing assay, the N-intein solutions were dialyzed against splicing assay buffer (100 mM sodium phosphates, 150 mM NaCl, 1 mM EDTA, pH 7.2) containing no thiols overnight at 4°C. If thiols were present, substantial N-extein cleavage could be observed for reactions with a slow k_3, however 1 mM DTT had no impact on reactions with $k_3 > 1.5 \times 10^{-3}$ s^{-1}. Note that for reaction 15, even in the absence of thiols, some N-extein cleavage was observed due to the extremely slow k_3 (see Figure S12. N-extein cleavage is characterized by a re-emergence of the starting material, species **1**). After removing thiols, N-inteins were diluted to 15 μM in assay buffer, C-inteins were diluted to 10 μM, and both solutions were treated with 2 mM TCEP. The solutions were incubated at 30 °C for 10 minutes, then reactions were initiated by mixing equal volumes of N- and C-inteins and continuing to incubate at 30 °C. During the reaction, aliquots of the solution were removed and mixed 3:1 (v/v) with quenching solution (8 M guanidine hydrochloride and 4% trifluoroacetic acid).

For RP-HPLC analysis, reactions were typically carried out on a 1.3-1.4 mL scale, where 100 μL were removed at each time point and mixed with 33 μL of quenching solution. 100 μL of the quenched solutions were separated 0-73% B gradient in 30 minutes on a C_{18} analytical column, recording absorbance at 214 nm (Figure S9). The major peaks were collected and identified by ESI-MS (Table

S3). For direct ESI-MS analyses, reactions were typically carried out on a 600 μL scale, where 30 μL were removed at each time point and mixed with 10 μL of quenching solution. 20 μL of the quenched solutions were desalted using Millipore C_{18} Zip-Tips as follows. First, the Zip-Tip was wetted by aspirating and ejecting 10 μL of dilution buffer (50% acetonitrile, 49.9% H_2O, 0.1% formic acid) 5 times. Next, the tip was equilibrated by aspirating and ejecting 10 μL of wash buffer (99.9% H_2O, 0.1% formic acid) 5 times. Then, the quenched mixture was loaded onto the Zip-Tip by pipetting 10 μL up and down 10 times in a 20 μL aliquot of the time point solution. The salts were washed away by aspirating and ejecting 10 μL of wash buffer 7 times. Finally, the proteins were eluted by pipetting up and down 10 times in 10 μL of elution buffer (70% acetonitrile, 29.9% H_2O, 0.1% formic acid). The eluates were diluted 20-fold in dilution buffer and loaded on the mass spectrometer by direct infusion. The complex mixture of multiply-charged states of each species were deconvoluted using the Maximum Entropy algorithm (Spectrum Square Associates, Ithaca, NY) into spectra depicting a well-defined mixture of singly-charged species (Figure S10 and Table S3).

Kinetic analysis of *trans*-splicing reactions

The RP-HPLC and ESI-MS data were quantified to yield reaction progress curves. First, using the manufacturer's analytical software, the peak areas for all relevant chromatographic or mass spectrometric peaks were calculated. For RP-HPLC, the relevant peaks came from species **1**-**5**, and for ESI-MS, the relevant peaks came from species **1**-**4** (Figure S8). Then, for each time point in a reaction, the area of an individual peak was expressed as a fraction of the total peak area. Reaction progress curves for each species were generated by expressing this normalized intensity as a function of time (Figures S11 and S12). Note that species **3**-**5** are all products that irreversibly form after branched intermediate resolution (Figures 1 and S1). Since **3** and **5** are direct products of *trans*-splicing, and **3** converts into **4** over time:

$$\frac{d[3+4+5]}{dt}=\frac{d[3+4]}{dt}=\frac{d[5]}{dt}$$

To account for changes in absorbance, in the RP-HPLC reactions, the peak area of the product is considered to be the total peak areas **3** + **4** + **5**. On the other hand for ESI-MS, based on size and composition, we assume that species **1**-**4** all have similar ionization efficiencies by ESI-MS, whereas **5** is a small molecule with dramatically different ionization properties. Thus the product from ESI-MS reactions is simply **3** + **4**. Note that N-intein species (AEY-IntN and IntN) do not separate under any RP-HPLC conditions tested (not shown), thus they were not used for any quantitative analysis. It is noteworthy that the two assay formats gave extremely similar quantitative results (Figure S13).

Our data were fit to the simplified three-state kinetic model shown in Figure 3C and below:

$$\underset{\mathbf{1}}{\text{Starting Mat. / Linear Int.}} \quad \underset{k_2}{\overset{k_1}{\rightleftharpoons}} \quad \underset{\mathbf{2}}{\text{Branched Int.}} \quad \xrightarrow{k_3} \quad \underset{\mathbf{3}+\mathbf{4}\ (+\ \mathbf{5})}{\text{Product}}$$

For each reaction, the normalized reaction progress curves for **1**, **2**, and **3** + **4** (+ **5**) were collectively fit to a system of equations that are the analytical solution to the coupled differential rate equations for those species:

$$\begin{aligned}
p &= k_1 + k_2 + k_3 \\
q &= \sqrt{p^2 - 4(k_1 k_3)} \\
a &= \tfrac{1}{2}(p+q) \\
b &= \tfrac{1}{2}(p-q) \\
[S/L](t) &= [S/L]_0\left[\left(\frac{k_1(a-k_3)}{a(a-b)}\right)e^{-at} + \left(\frac{k_1(k_3-b)}{b(a-b)}\right)e^{-bt}\right] \\
[B](t) &= [S/L]_0\left[\left(\frac{-k_1 a}{a(a-b)}\right)e^{-at} + \left(\frac{k_1 b}{b(a-b)}\right)e^{-bt}\right] \\
[P](t) &= [S/L]_0\left[\left(\frac{k_1 k_3}{ab}\right) + \left(\frac{k_1 k_3}{a(a-b)}\right)e^{-at} - \left(\frac{k_1 k_3}{b(a-b)}\right)e^{-bt}\right]
\end{aligned} \qquad (1)$$

In these equations, *p*, *q*, *a*, and *b* are algebraic combinations of rate constants k_1, k_2, and k_3 that simplify expression of the rate equations. [*S*/*L*] is the normalized intensity of the **S**tarting Material / **L**inear intermediate (**1**), [B] is the normalized intensity of the **B**ranched Intermediate (**2**), and [P] is the normalized intensity of the **P**roducts (**3** + **4** + **5** for RP-HPLC analysis and **3** + **4** for ESI-MS analysis). The variable t is the reaction time in seconds, and $[S/L]_0$ can be considered a normalization factor that represents the extent of the reaction (the fraction of starting material that gets converted to product).

The data were fit to equations (1) using the multiple fit function in the Pro Fit data analysis software (QuantumSoft, Switzerland). In these fits, k_1, k_2, k_3, and $[S/L]_0$ were allowed to vary, and best-fit values were identified using the Levenberg-Marquardt algorithm. For all reactions except reaction 15 (which had significant N-extein cleavage), $[S/L]_0$ was between 0.81 and 0.95. This fitting process was independently carried out for individual replicates, and the best-fit values from three or more replicates were averaged. These averages and the standard deviation between these replicates are given in Table 1 of the main text. Finally, the overall rate of splicing (k_{splice}) was determined by treating product formation, **3** + **4** (+ **5**), as a single-step first order reaction and fitting those reaction progress curves to the following equation:

$$[P](t) = [P]_{\max} \cdot \left(1 - e^{-k_{splice} \cdot t}\right) \qquad (2)$$

where $[P]_{\max}$ is analogous to $[S/L]_0$ from the previous equations. The average and standard deviation for k_{splice} from three independently fit reactions can be found in Table 1 of the main text.

There were five exceptions to our general curve-fitting protocol. For reactions 4 and 5, branched intermediate resolution was not observed (Figures S11 and S12). Thus, the reactions were treated as an equilibrium between two states and the curves for **1** and **2** were simultaneously fit to these equations:

$$\begin{aligned} [S/L](t) &= \frac{1}{(k_1 + k_2)} \cdot \left(k_2 + k_1 e^{-(k_1 + k_2)t}\right) \\ [B](t) &= \frac{k_1}{(k_1 + k_2)} \cdot \left(1 - e^{-(k_1 + k_2)t}\right) \end{aligned} \qquad (3)$$

For reaction 15, N-extein cleavage caused a re-emergence of starting material **1** (Figure S12), thus the equations for the three-state model could not be readily applied for global fitting. Since BI resolution was extraordinarily slow for this reaction, the reaction was comprised of an initial pre-equilibrium step and a product formation step that were de-coupled. Thus, the data for **1** and **2** from the first 10 minutes of the reaction curve (Figure S11) were fit to the two-state model described by equations (3) to extract k_1 and k_2. The data for product accumulation (**3** + **4**) over the course of the entire reaction (Figure S12), were fit to equation (2) to extract k_{splice}, which we assumed to be equal to k_3. The last two exceptions were reactions 3 and 13, reactions with the C1A mutation. For these reaction, only C-extein cleavage was observed, thus they were treated as a first-order reactions. For reaction 3 the curve for **3** + **4** + CFN(NH_2) was fit to equation (2). For reaction 13, the curve for **3** + **4** was fit to equation (2). Since reaction 13 was extremely slow and did not plateau after a few days, $[P]_{\max}$ was constrained to 0.95, which was the plateau value for the analogous reaction 3.

Double-mutant cycle analysis of the His$_{125}$ and Phe$_{+2}$ interaction

Reactions 1, 12, 14, and 15 collectively make up a double-mutant cycle that probes the importance of Phe$_{+2}$ in assisting His$_{125}$ in intein chemistry. Both His$_{125}$ and Phe$_{+2}$ were shown to be important for catalysis. In reaction 12, the F+2A mutation demonstrates that the steric bulk of Phe$_{+2}$ is necessary. In reaction 14, the H125N mutation demonstrates that His$_{125}$ is a proton donor or acceptor involved in catalysis. Thus, if the effect of the double mutant in reaction 15 is energetically additive of the two individual mutations, the interaction of these two amino acids is not important for protein splicing. This can be expressed mathematically as follows. The change in activation energy upon mutating a system from state *a* to state *b* can be calculated by equation (4):

$$\Delta\Delta G^{\neq}_{a\rightarrow b} = -RT\ln\frac{k_b}{k_a} \qquad (4)$$

As defined in Figure S14, state *a* contains the native His$_{125}$ and Phe$_{+2}$ (reaction 1), state *b* contains His$_{125}$ but Ala$_{+2}$ (reaction 12), state *c* contains Asn$_{125}$ and native Phe$_{+2}$ (reaction 14), and state *d* is the double-mutant with Asn$_{125}$ and Ala$_{+2}$ (reaction 15). Then, the coupling or interaction energy between His$_{125}$ and

Phe_{+2} lost upon double mutation that pertains to BI resolution kinetics can be calculated by applying equation (4) with k_3 for each reaction and equation (5) given below:

$$\Delta G_{coup} = \Delta\Delta G^{\neq}_{a\to b} - \Delta\Delta G^{\neq}_{c\to d} = \Delta\Delta G^{\neq}_{a\to c} - \Delta\Delta G^{\neq}_{b\to d} \qquad (5)$$

This value is +1.07 kcal/mol for the double-mutant cycle presented herein (Figure S14).

Preparation of complexes for NMR spectroscopy

^{13}C,^{15}N-labeled $\mathrm{NpuC_{N137A}}$ was expressed analogously to its unlabeled counterpart, except that the expression cultures were grown in minimal M9 medium supplemented with ^{13}C-glucose and $^{15}NH_4Cl$ as the only carbon and nitrogen sources. After expression, the fusion protein was enriched over a Ni column, and the thiolysis/ligation to the unlabeled CFN(NH_2) and CAN(NH_2) peptides was carried out as written above. The C-intein constructs were purified and their identities were confirmed by RP-HPLC and ESI-MS (Figure S7 and Table S2).

To prepare the complexes, 1.2 eq of C-intein were added to 10 mL of 35 μM N-intein. The complex was treated with 5 mM TCEP for 10 minute at room temperature then injected on a Superdex 75 16/60 size exclusion column. The complex was eluted over 1.35 CV in NMR buffer (25 mM sodium phosphates, 100 mM NaCl, 1 mM DTT, pH 6.5) and separated from excess C-intein (Figure S15). Fractions from the size exclusion purification were analyzed by SDS-PAGE on a 12% Bis-Tris gel to determine what region of the major chromatographic peak had a constant N-intein to C-intein ration, and thus 1:1 stoichiometry (Figure S16). These fractions were pooled and concentrated for NMR analysis.

NMR spectroscopy of split intein complexes

Typically 250μM samples were used for all NMR experiments in NMR Buffer. NMR experiments were carried out at either Varian Inova (600MHz) or Bruker (600, 800 or 900MHz) spectrometers equipped with cryogenic probes that can apply pulse field gradients along the z-axis. The data were collected at 25 °C and processed by NMRPipe,[4] and NMRViewJ was used for the analysis.[5] In all constructs the N-intein and N-extein were unlabeled. Only the C-intein was uniformly ^{15}N, ^{13}C labeled; the residues comprising the C-terminal extein, CFN(NH_2) or CAN(NH_2), were chemically synthesized using unlabeled amino-acids and ligated to the C-intein using expressed protein ligation.

Resonance Assignments

The backbone assignments of NpuC$_{N137A}$ with a CFN(NH_2) C-extein and in complex with AEY-NpuN$_{C1A}$ (Figure S17) were done with ^{15}N, ^{13}C labeled fully protonated samples using standard triple resonance experiment pairs: HNCO/HN(CA)CO, HNCA/HN(CO)CA, HNCACB/CBCA(CO)NH.[6] The HNCO/HN(CA)CO pair employed 1024, 42 and 60 complex points recorded with 12.6 ppm, 34 ppm and 20 ppm sweep widths in ^{1}H, ^{15}N and ^{13}C dimensions respectively; the HNCA/HN(CO)CA pair used 1024, 42 and 60 complex points recorded with 12.6 ppm, 34 ppm and 32 ppm sweep widths in ^{1}H, ^{15}N and ^{13}C dimensions respectively, and; the HNCACB/CBCA(CO)NH pair employed 1024, 42 and 60 complex points recorded with 12.6 ppm, 34 ppm and 58 ppm sweep widths in ^{1}H, ^{15}N and ^{13}C dimensions respectively.

Both CFN and CAN constructs of NpuC possess only a single histidine (His$_{125}$). The imidazole ring carbon resonances, Cδ_2 and Cε_1, were resolved by aromatic ^{13}C,^{1}H HSQC experiments[7-9] which were run under the same buffer conditions at 25 ^{o}C using a 600MHz Bruker spectrometer; the experiment employed 1024 and 64 complex points recorded with 12.6 ppm and 30 ppm sweep widths in ^{1}H and ^{13}C dimensions (Figure 5D in the main text).

Chemical Shift Perturbations: CFN vs. CAN

^{15}N, ^{1}H HSQC spectra of AEY-NpuN$_{C1A}$: NpuC$_{N137A}$($^{13}C,^{15}N$)-CFN(NH_2) and AEY-NpuN$_{C1A}$: NpuC$_{N137A}$($^{13}C,^{15}N$)-CAN(NH_2) complexes were recorded at 25 ^{o}C using a 900 MHz Bruker spectrometer. The chemical shift perturbations ($\Delta\delta_i$) of each assigned residue were calculated using equation (6) and can be found in Figure 5C in the main text:

$$\Delta\delta_i = \sqrt{\left(\Delta\delta_{CFN,H} - \Delta\delta_{CAN,H}\right)_i^2 + 0.11\left(\Delta\delta_{CFN,N} - \Delta\delta_{CAN,N}\right)_i^2} \qquad (6)$$

Spin Relaxation Measurements

Relaxation experiments were run at 900 MHz at 25 °C and the analyses were performed using NMRViewJ. Spin-lattice, spin-spin relaxation rates (R_1, R_2) and ^{15}N-^{1}H *NOE* measurements of both complexes were performed by using standard pulse sequences (Figure S18). R_1 and R_2 rates were measured using recycle delays of 1.5 s and the following relaxation delays were used:

R_1: 10.0, 160, 310, 610 (x2), 760, 910, and 1210 ms
R_2: 0, 16.3, 32.6, 48.9 65.2 (×2), 81.5, and 130.4 ms

All-atom molecular dynamics simulations

The simulations of DnaE intein from *Nostoc punctiforme* (NpuDnaE) were carried out using the molecular dynamics (MD) software package AMBER11.[10,11] The first alternative NMR solution structure within PDB 2KEQ was selected.[12] The selected fused structure was cut between Asn102 and Ile103 to mimic the split form of the intein. The sequence was modified *in silico* using Chimera[13] to generate constructs of interest: native Npu with the canonical N-extein ($A_{-3}E_{-2}Y_{-1}$) and C-extein ($C_{+1}F_{+2}N_{+3}$) sequences, a C-extein variant ($C_{+1}A_{+2}N_{+3}$), or a D124Y mutation combined with the same C-extein mutation. The C-terminal ends of the C-exteins were charge capped using a special C-terminal residue, NHE, defined in the AMBER ff99sb force field.[14]

For each construct, identical all-atom molecular dynamics (MD) simulations were run. Each system was solvated in a box with sides measuring 74 Å, 61|62 Å and 71 Å, charge neutralized with sodium ions, and modeled using the AMBER ff99SB force field. The TIP3P force field was used to model water.[15] We used the SHAKE algorithm to constrain all bonds involving hydrogen atoms.[16] An 8 Å cutoff radius was used for range-limited interactions, utilizing Particle Mesh Ewald electrostatics for long-range interactions.[17] The initial structures were minimized to clear any steric clashes. Periodic boundary with constant volume was used. Harmonic positional restraints of strength 10 kcal/mol/Å^2

were applied to backbone atoms. After the initial minimization step, each system was heated from 0 K to 300 K as a function of time over the course of 50 ps using the Langevin thermostat[18] with a collision frequency of 2 ps while the volume was kept constant. Harmonic positional restraints of strength 2 kcal/mol/Å^2 were used for backbone atoms. Following the heating step, a density equilibration step of 50ps was performed for each system with constant isotropic pressure of 1 atm. Temperature was maintained at 300 K using the Langevin thermostat with 2 ps collision frequency. Harmonic positional restraints of strength 2 kcal/mol/Å^2 were used for backbone atoms. Prior to the production run, a conventional equilibration step of 100 ps was run at 1 atm constant isotropic pressure and 300 K using Langevin thermostat with 2 ps collision frequency with relaxation time of 2 ps without any positional restraints. The conventional production run for each system was carried out at constant isotropic pressure of 1 atm at 300 K maintaining the temperature using a Langevin thermostat with 5 ps collision frequency. The relaxation times for barostat and thermostat were set to 2 ps. The simulation time step was 1 fs and snap shots were saved at every 5 ps. The distances and dihedral angles were extracted from these snap shots over the MD trajectories (Figures S19-S22 and Figure 6 and 7 in the main text).

References

(1) Bryksin, A. V.; Matsumura, I. *BioTechniques* **2010**, *48*, 463-5.

(2) Shah, N. H.; Dann, G. P.; Vila-Perelló, M.; Liu, Z.; Muir, T. W. *J. Am. Chem. Soc.* **2012**, *134*, 11338-41.

(3) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. *The proteomics protocols handbook* **2005**, 571-607.

(4) Delaglio, F.; Grzesiek, S.; Vuister, G. W.; Zhu, G.; Pfeifer, J.; Bax, A. *J. Biomol. NMR* **1995**, *6*, 277-93.

(5) Johnson, B. A. *Methods Mol. Biol.* **2004**, *278*, 313-52.

(6) Sattler, M.; Schleucher, J.; Griesinger, C. *Prog. Nucl. Mag. Res. Sp.* **1999**, *34*, 93-158.

(7) Palmer, A. G.; Cavanagh, J.; Wright, P. E.; Rance, M. *J. Mag. Reson.* **1991**, *93*, 151-170.

(8) Kay, L. E.; Keifer, P.; Saarinen, T. *J. Am. Chem. Soc.* **1992**, *114*, 10663-10665.

(9) Schleucher, J.; Schwendinger, M.; Sattler, M.; Schmidt, P.; Schedletzky, O.; Glaser, S. J.; Sørensen, O. W.; Griesinger, C. *J. Biomol. NMR* **1994**, *4*, 301-6.

(10) Case, D. A.; Darden, T. A.; Cheatham, T. E.; Simmerling, C. L.; Wang, J.; Duke, R. E.; Luo, R.; Walker, R. C.; Zhang, W.; Merz, K. M.; Roberts, B.; Hayik, S.; Roitberg, A.; Seabra, G.; Swails, J.; Goetz, A. W.; Kolossváry, I.; Wong, K. F.; Paesani, F.; Vanicek, J.; Wolf, R. M.; Liu, J.; Wu, X.; Brozell, S. R.; Steinbrecher, T.; Gohlke, H.; Cai, Q.; Ye, X.; Wang, J.; Hsieh, M.-J.; Cui, G.; Roe, D. R.; Mathews, D. H.; Seetin, M. G.; Salomon-Ferrer, R.; Sagui, C.; Babin, V.; Luchko, T.; Gusarov, S.; Kovalenko, A.; Kollman, P. A. *AMBER 12.* **2012**, *University of California, San Francisco.*

(11) Götz, A. W.; Williamson, M. J.; Xu, D.; Poole, D.; Le Grand, S.; Walker, R. C. *J. Chem. Theory. Comput.* **2012**, *8*, 1542-1555.

(12) Oeemig, J. S.; Aranko, A. S.; Djupsjöbacka, J.; Heinämäki, K.; Iwaï, H. *FEBS Lett.* **2009**, *583*, 1451-1456.

(13) Pettersen, E. F.; Goddard, T. D.; Huang, C. C.; Couch, G. S.; Greenblatt, D. M.; Meng, E. C.; Ferrin, T. E. *J. Comput. Chem.* **2004**, *25*, 1605-12.

(14) Lindorff-Larsen, K.; Piana, S.; Palmo, K.; Maragakis, P.; Klepeis, J. L.; Dror, R. O.; Shaw, D. E. *Proteins* **2010**, *78*, 1950-8.

(15) Jorgensen, W. L.; Madura, J. D. *J. Am. Chem. Soc.* **1983**, *105*, 1407-1413.

(16) Lippert, R. A.; Bowers, K. J.; Dror, R. O.; Eastwood, M. P.; Gregersen, B. A.; Klepeis, J. L.; Kolossvary, I.; Shaw, D. E. *J. Chem. Phys.* **2007**, *126*, 046101.

(17) Darden, T.; York, D.; Pedersen, L. *J. Chem. Phys.* **1993**, *98*, 10089-10092.

(18) Grest, G.; Kremer, K. *Phys. Rev. A* **1986**, *33*, 3628-3631.

Figures and Tables

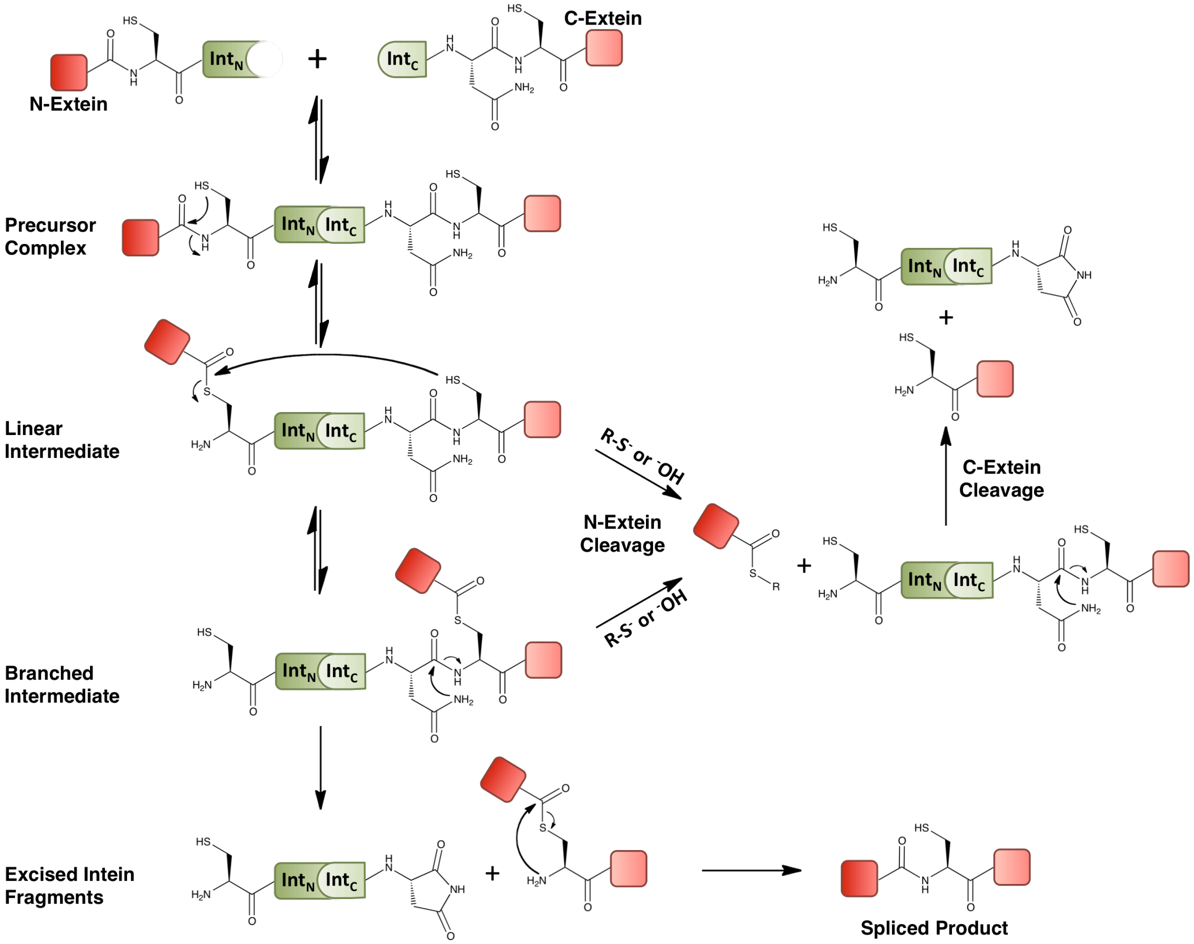

Figure S1. Mechanism of protein *trans*-splicing and associated side reactions.

```
           -3   1       10        20        30        40        50        60
NpuN        AEY-CLSYETEILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCL
            AEY-CLS+ TEILTVEYG LPIGKIV + I C+VYSVD  G +YTQ +AQWHDRGEQEV EY L
SspN        AEY-CLSFGTEILTVEYGPLPIGKIVSEEINCSVYSVDPEGRVYTQAIAQWHDRGEQEVLEYEL

           61       70        80        90       100
NpuN        EDGSLIRATKDHKFMTVDGQMLPIDEIFERELDLMRVDNLPN
            EDGS+IRAT DH+F+T D Q+L I+EIF R+LDL+ ++N+
SspN        EDGSVIRATSDHRFLTTDYQLLAIEEIFARQLDLLTLENIKQTEEALDNHRLPFPLLDAGTIK

          103    110       120       130        +1
NpuC        MIKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN-CFN
            M+K+  R+ LG Q ++DIG+ +DHNF L NG IA+N-CFN
SspC        MVKVIGRRSLGVQRIFDIGLPQDHNFLLANGAIAAN-CFN
```

Figure S2. Sequence alignment of Npu and Ssp. Conserved residues are indicated by their one letter code in the middle row, and similar residues are indicated by a +. The key catalytic residues mutated in this study are shown in orange, the critical +2 residue is red, and other significant non-catalytic residues highlighted in this study are shown in green. For consistency with the NMR structure of Npu (PDB 2KEQ), the N-terminal methionine of Npu$_C$ (shown in italics) is omitted from the numbering scheme. This methionine is absent from the Ssp$_C$-CFN(NH$_2$) construct used in this study (reaction 2).

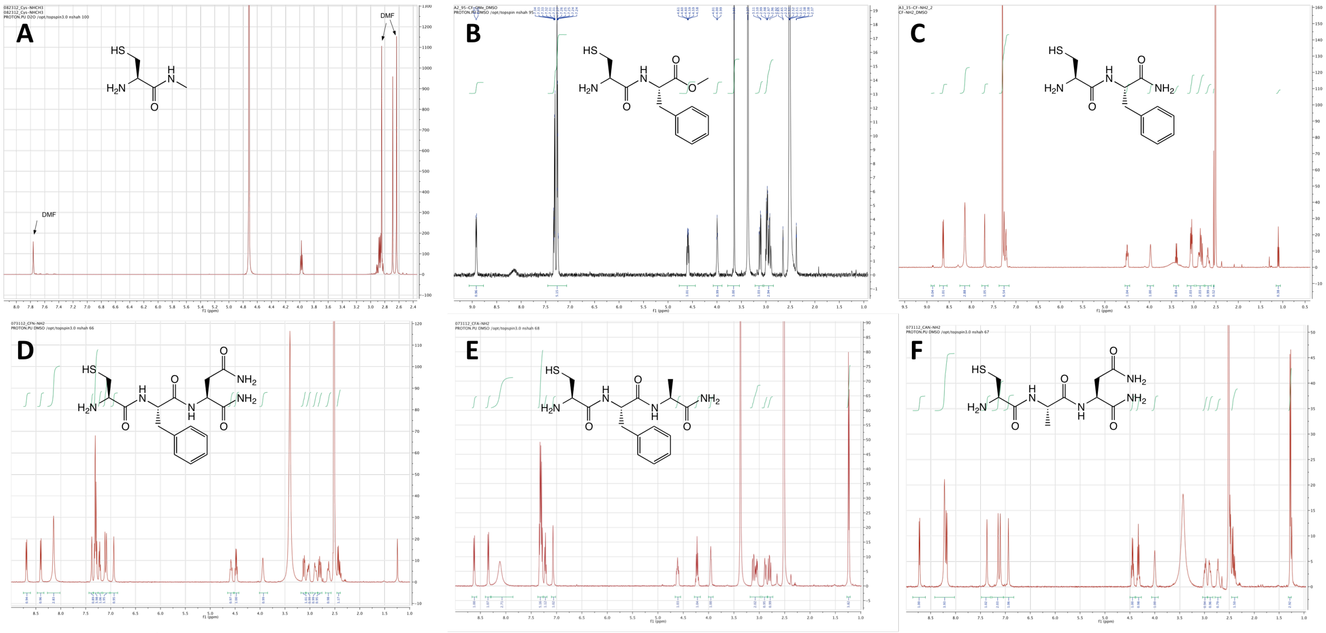

Figure S3. ^{1}H-NMR of synthetic C-extein molecules. A. C($NHCH_3$) in D_2O, **B.** CF(OCH_3) in DMSO-*d_6*, **C.** CF(NH_2) in DMSO-*d_6*, **D.** CFN(NH_2) in DMSO-*d_6*, **E.** CFA(NH_2) in DMSO-*d_6*, and **F.** CAN(NH_2) in DMSO-*d_6*. Note that resonances for residual DMF can be seen in panel A.

Table S1. Expected and observed monoisotopic masses of C-extein peptides.

Peptide	**Expected $[M+H]^+$ (Da)**	**Observed $[M+H]^+$ (Da)**
CF(OCH_3)	283.11	283.1
CF(NH_2)	268.11	268.1
CFN(NH_2)	382.15	382.2
CFA(NH_2)	339.14	339.1
CAN(NH_2)	306.12	306.1

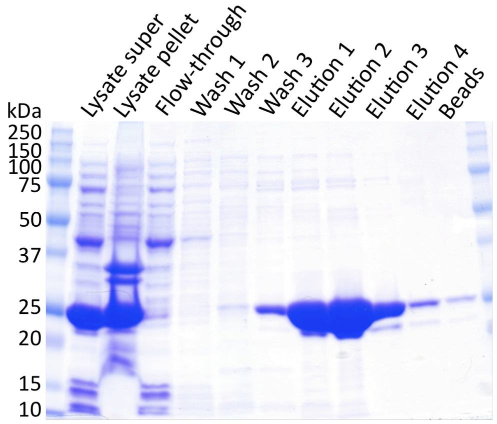

Figure S4. Representative SDS-PAGE of Int$_C$-GyrA-H$_6$ enrichment from cell lysates. The Npu$_C$-GyrA-H$_6$ protein was expressed in *E. coli*, and cells were lysed by sonication. The lysates were clarified by centrifugation then the soluble fraction ("lysate super") was loaded on a Ni-NTA column. The flow through was collected, then the column was washed with 5 CV of 5 mM imidazole buffer (Wash 1), 5 CV of 20 mM imidazole buffer (Wash 2), and 3 CV of 50 mM imidazole buffer (Wash 3). The protein was eluted with four 1.5 CV washes with 250 mM imidazole buffer (Elutions 1-4), and the Ni resin was boiled in gel loading dye after the elution to visualize any residual protein (Beads).

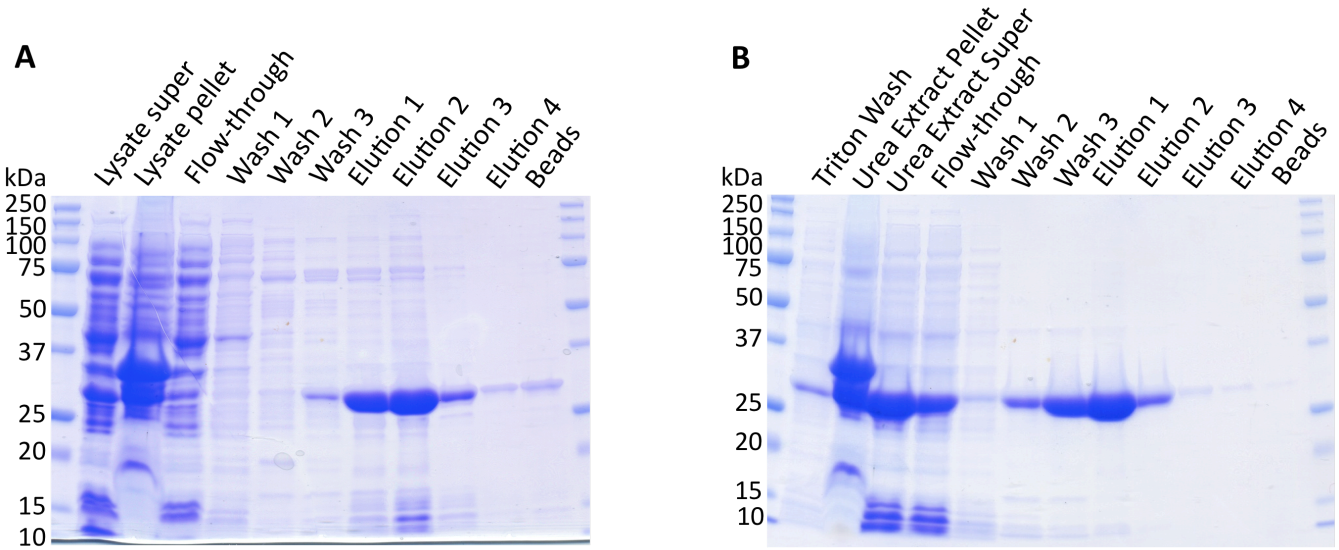

Figure S5. Representative SDS-PAGE of H_6-SUMO-AEY-Int_N enrichment from cell lysates. The H_6-SUMO-AEY-Npu_N fusion was extracted from **A.** the soluble fraction or **B.** the insoluble fraction of the cell lysate. The indicated fractions were obtained identically as for the Npu_C-GyrA-H_6 fusion in Figure S4. Note that for the urea extraction, additional protein could be recovered by re-passing the flow-through over the Ni resin, washing, and eluting as before.

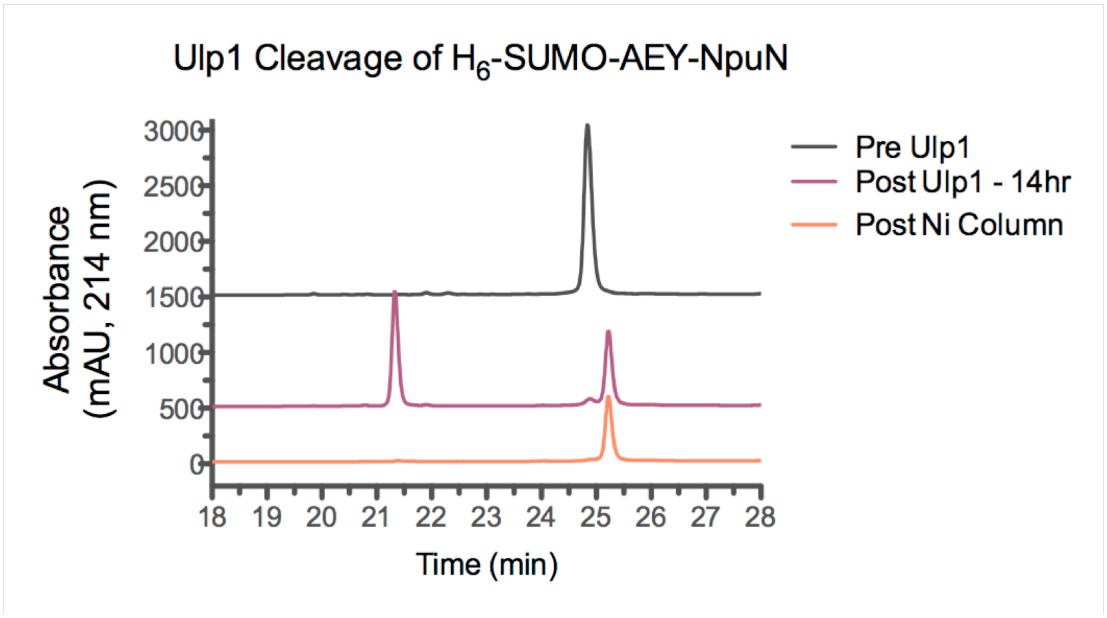

Figure S6. Representative analytical RP-HPLC analysis of Ulp1 cleavage to yield AEY-Int_N. The H_6-SUMO-AEY-NpuN protein, extracted from inclusion bodies, was enriched over Ni-NTA resin and refolded to yield the "Pre-Ulp1" sample. This was treated with Ulp1 to yield the "Post-Ulp1 - 14hr" sample. Finally, this sample was passed over a Ni column, and the flow-through and washes with <20 mM imidazole were collected to yield the "Post Ni Column" sample. These samples were analyzed on a C_{18} analytical RP-HPLC column at 1 mL/min over a 0-73% B linear gradient in 30 minutes, preceded by a 2 minute isocratic phase in 0% B (Solvent A: 99.9% H_2O, 0.1% TFA & Solvent B: 90% acetonitrile, 9.9% H_2O, 0.1% TFA).

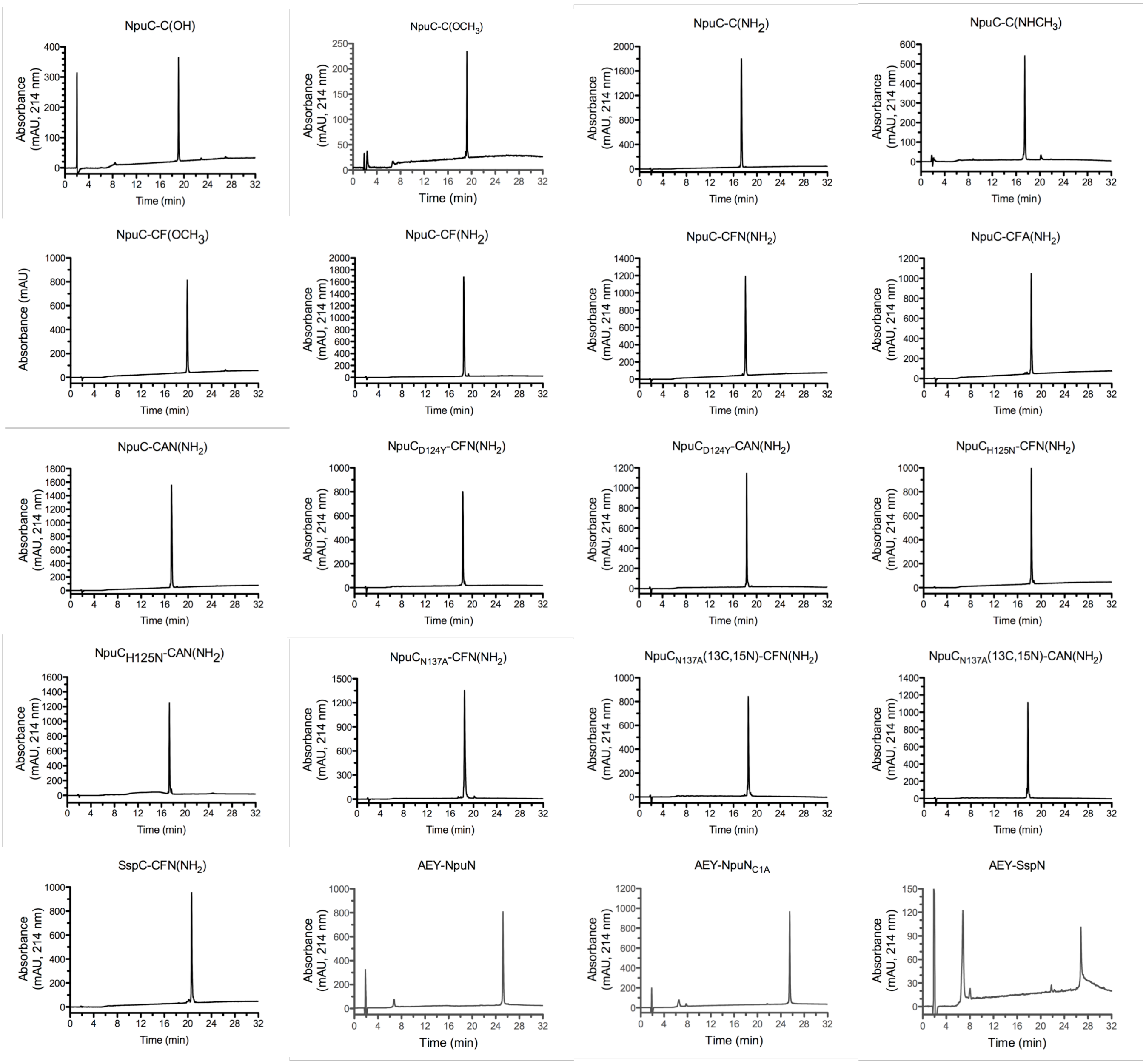

Figure S7. Analytical RP-HPLC analysis of purified proteins. Each purified protein was analyzed on a C_{18} analytical RP-HPLC column at 1 mL/min over a 0-73% B linear gradient in 30 minutes, preceded by a 2 minute isocratic phase in 0% B (Solvent A: 99.9% H_2O, 0.1% TFA & Solvent B: 90% acetonitrile, 9.9% H_2O, 0.1% TFA). Note that for the last three chromatographs, corresponding to AEY-Int$_N$ constructs, peaks around 6.5 and 8 minutes correspond to a DTT monomer and disulfide dimer, respectively. Additionally, the last chromatograph, for AEY-Ssp$_N$, shows that this protein smears on the C_{18} column. This was observed irrespective of gradient or sample loading conditions. The sharp peak and the trailing region were found to have the same mass by ESI-MS.

Table S2. Expected and observed masses of purified proteins.

Peptide	Expected Mass (Da)	Observed Mass (Da)
NpuC$_{WT}$-C(OH)	4226.19	4226.2
NpuC$_{WT}$-C(OCH$_3$)	4240.21	4240.2
NpuC$_{WT}$-C(NH$_2$)	4225.21	4225.2
NpuC$_{WT}$-C(NHCH$_3$)	4239.23	4239.2
NpuC$_{WT}$-CF(OCH$_3$)	4387.28	4387.3
NpuC$_{WT}$-CF(NH$_2$)	4372.28	4372.3
NpuC$_{WT}$-CFN(NH$_2$)	4486.32	4486.4
NpuC$_{WT}$-CFA(NH$_2$)	4443.32	4443.4
NpuC$_{WT}$-CAN(NH2)	4410.29	4410.4
NpuC$_{D124Y}$-CFN(NH$_2$)	4534.36	4534.4
NpuC$_{D124Y}$-CAN(NH$_2$)	4458.33	4458.3
NpuC$_{H125N}$-CFN(NH$_2$)	4463.31	4463.3
NpuC$_{H125N}$-CAN(NH$_2$)	4387.28	4387.3
NpuC$_{N137A}$-CFN(NH$_2$)	4443.32	4443.4
NpuC$_{N137A}$(^{13}C,^{15}N)-CFN(NH$_2$) [a]	4679.76	4675.8
NpuC$_{N137A}$(^{13}C,^{15}N)-CAN(NH$_2$) [a]	4603.73	4599.7
SspC$_{WT}$-CFN(NH2) [b]	4135.25	4135.3
AEY-NpuN [c]	12219.79	12219.5
AEY-NpuN$_{C1A}$ [c]	12187.73	12187.4
AEY-SspN [c]	14331.06	14330.6

[a] Based on the difference in expected and observed masses, the ^{13}C and ^{15}N isotopic enrichment for these samples is >98%.

[b] SspC$_{WT}$ was isolated without its N-terminal methionine due to significant *in vivo* processing during expression in *E. coli*.

[c] The expected and observed masses for the N-intein constructs are the average mass. The expected and observed masses of C-intein constructs are the monoisotopic masses.

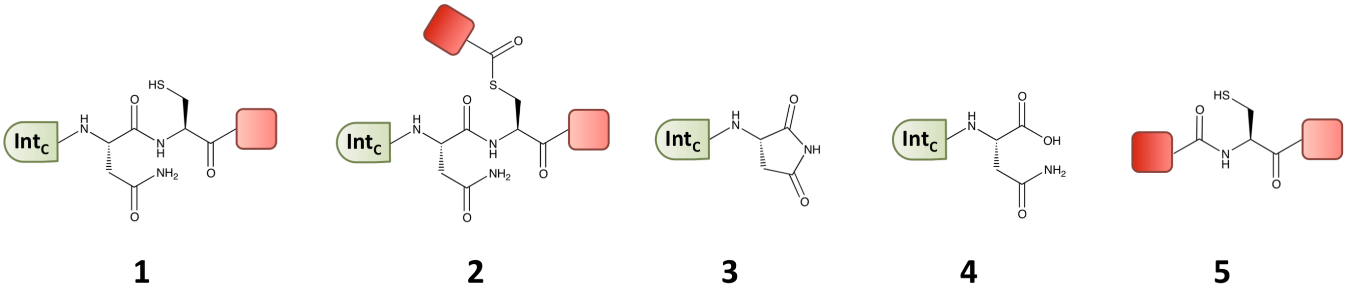

Figure S8. Scheme of observable species (1-5) in kinetic assays. Molecule **1** is found before association of the split intein fragments, in the precursor complex, and in the linear thioester intermediate. Molecule **2** is unique to the branched intermediate state of the *trans*-splicing reaction. Molecules **3-5** are products of *trans*-splicing. Specifically, molecule **3** comes from the excised intein complex, however it slowly hydrolyzes to **4**. Spliced product **5** initially exists in a thioester form immediately upon intein excision, however this is not detectable, as it rapidly isomerizes to the amide form sown in the scheme above.

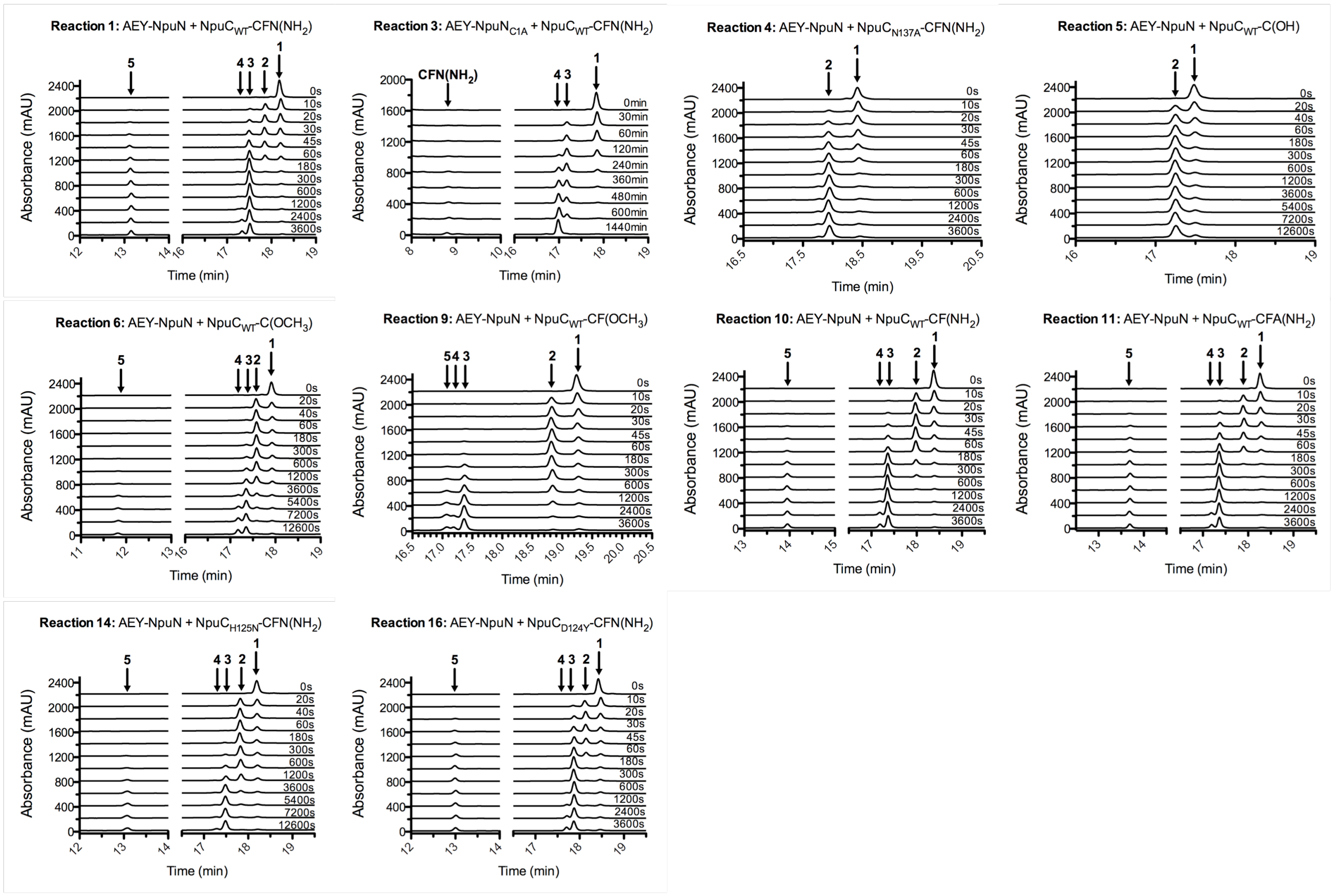

Figure S9. Representative RP-HPLC chromatographs from splicing assays. Reactions 1, 3, 4, 5, 6, 9, 10, 11, 14, and 16 were analyzed by RP-HPLC. Each time point was analyzed on a C_{18} analytical RP-HPLC column at 1 mL/min over a 0-73% B linear gradient in 30 minutes, preceded by a 2 minute isocratic phase in 0% B (Solvent A: 99.9% H_2O, 0.1% TFA & Solvent B: 90% acetonitrile, 9.9% H_2O, 0.1% TFA). Absorbance was measured at 214 nm.

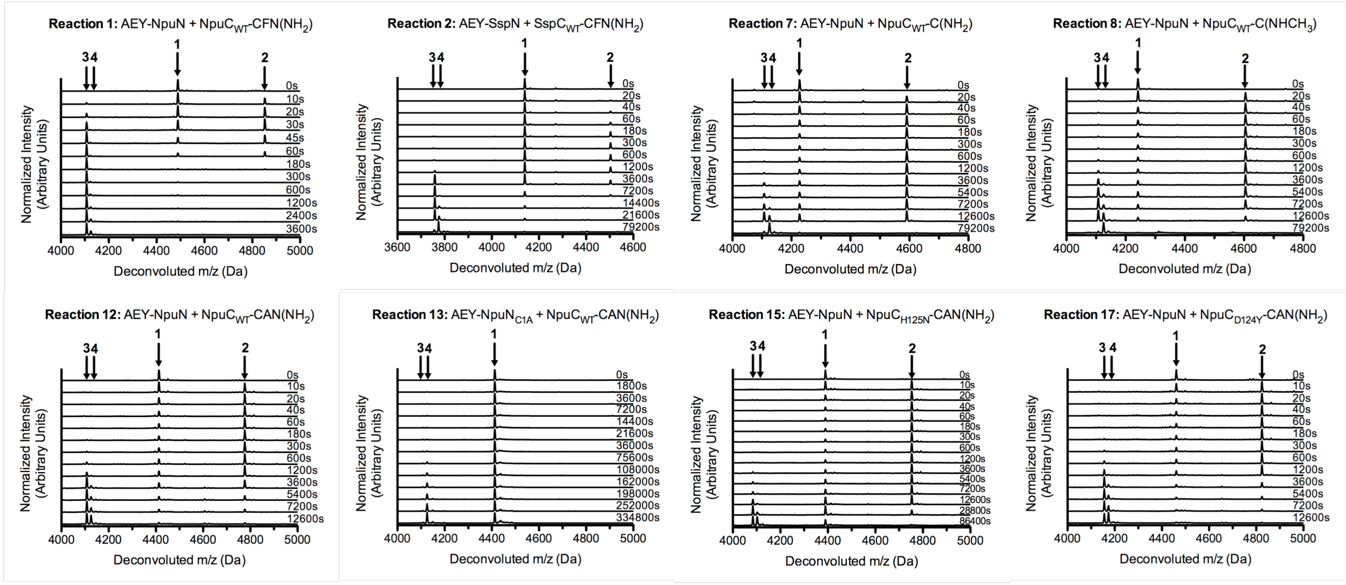

Figure S10. Representative deconvoluted ESI-MS spectra from splicing assays. Reactions 2, 7, 8, 12, 13, 15, and 17 were analyzed by ESI-MS. Reaction 1 was also analyzed this way for comparison to the RP-HPLC assay.

Table S3. Expected and observed masses of intermediates and products in splicing assays.

Molecule	Species	Relevant Rxns	Expected Mass (Da)	Observed Mass (Da)
NpuC$_{WT}$-C^{AEY}FN(NH$_2$)	2	1	4849.47	4849.7
SspC$_{WT}$-C^{AEY}FN(NH$_2$) [a]	2	2	4498.39	4498.3
NpuC$_{N137A}$-C^{AEY}FN(NH$_2$)	2	4	4806.46	4806.5
NpuC$_{WT}$-C^{AEY}(OH)	2	5	4589.34	4589.2
NpuC$_{WT}$-C^{AEY}(OCH$_3$)	2	6	4603.35	4603.4
NpuC$_{WT}$-C^{AEY}(NH$_2$)	2	7	4588.35	4588.2
NpuC$_{WT}$-C^{AEY}(NHCH$_3$)	2	8	4602.37	4602.4
NpuC$_{WT}$-C^{AEY}F(OCH$_3$)	2	9	4750.42	4750.4
NpuC$_{WT}$-C^{AEY}F(NH$_2$)	2	10	4735.42	4735.5
NpuC$_{WT}$-C^{AEY}FA(NH$_2$)	2	11	4806.46	4806.5
NpuC$_{WT}$-C^{AEY}AN(NH$_2$)	2	12	4773.43	4773.5
NpuC$_{H125N}$-C^{AEY}FN(NH$_2$)	2	14	4826.45	4826.4
NpuC$_{H125N}$-C^{AEY}AN(NH$_2$)	2	15	4750.42	4750.4
NpuC$_{D124Y}$-C^{AEY}FN(NH$_2$)	2	16	4897.50	4897.5
NpuC$_{D124Y}$-C^{AEY}AN(NH$_2$)	2	17	4821.47	4821.5
NpuC$_{WT}$(succinimide)	3	1, 3, 6-13	4105.18	4105.3
SspC$_{WT}$(succinimide) [a]	3	2	3754.10	3754.1
NpuC$_{H125N}$(succinimide)	3	14, 15	4082.16	4082.1
NpuC$_{D124Y}$(succinimide)	3	16, 17	4153.21	4153.2
NpuC$_{WT}$(OH)	4	1, 3, 6-13	4123.19	4123.2
SspC$_{WT}$(OH) [a]	4	2	3772.11	3772.1
NpuC$_{H125N}$(OH)	4	14, 15	4100.17	4100.1
NpuC$_{D124Y}$(OH)	4	16, 17	4171.22	4171.3
AEY-CFN(NH$_2$) [b]	5	1, 2, 14, 16	745.30	745.3
AEY-C(OCH$_3$) [b]	5	6	499.19	499.2
AEY-C(NH$_2$) [b]	5	7	484.19	- [c]
AEY-C(NHCH$_3$) [b]	5	8	498.21	498.2
AEY-CF(OCH$_3$) [b]	5	9	646.25	646.3
AEY-CF(NH$_2$) [b]	5	10	631.25	631.3
AEY-CFA(NH$_2$) [b]	5	11	702.29	702.3
AEY-CAN(NH$_2$) [b]	5	12, 15, 17	669.27	669.3
CFN(NH$_2$) [d]	-	3	382.15	382.2
NpuN [e]	-	1, 4-12, 14-17	11856.42	11856.0
SspN [e]	-	2	13967.69	13967.1

[a] All SspC$_{WT}$-based molecules do not contain an N-terminal methionine, whereas NpuC-based molecules do.

[b] For all reactions, even if RP-HPLC analysis did not yield sufficient separation between species **1**-**4** for quantification, a peak corresponding to the spliced product, **5**, was collected and identified by ESI-MS. The expected and observed monoisotopic masses for these species correspond to the $[M+H]^+$ ion.

[c] While an RP-HPLC peak consistent with the spliced product for reaction 8 was observed (based on rate of appearance and retention time), mass spectrometric data could not be obtained to unambiguously identify this peak.

[d] This molecule corresponds to the cleaved C-extein in control reaction 3. The expected and observed monoisotopic masses for this molecule correspond to the $[M+H]^+$ ion.

[e] While the starting material AEY-IntN and the excised IntN species could not be separated by RP-HPLC, their combined peak could be collected and analyzed as a mixture by ESI-MS to confirm the presence of extein-free NpuN or SspN. The expected and observed masses for these molecules are the average masses.

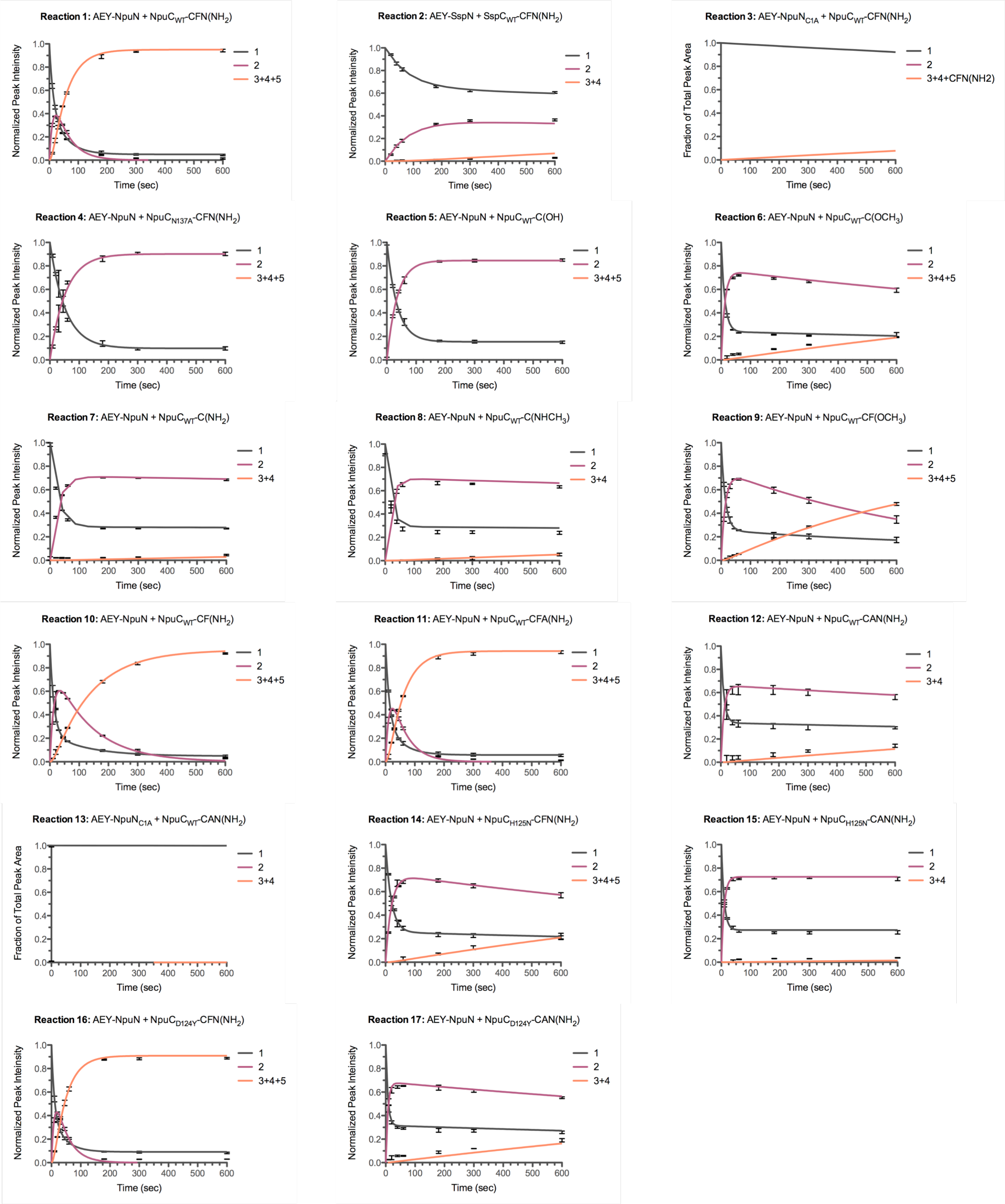

Figure S11. Reaction progress curves for splicing assays: the first ten minutes. Error bars represent the standard deviation from three or four independent measurements. The solid lines are reaction progress curves generated by plotting the relevant rate equations (see methods section above) with the average best-fit values for all fitted kinetic parameters (Table 1, main text).

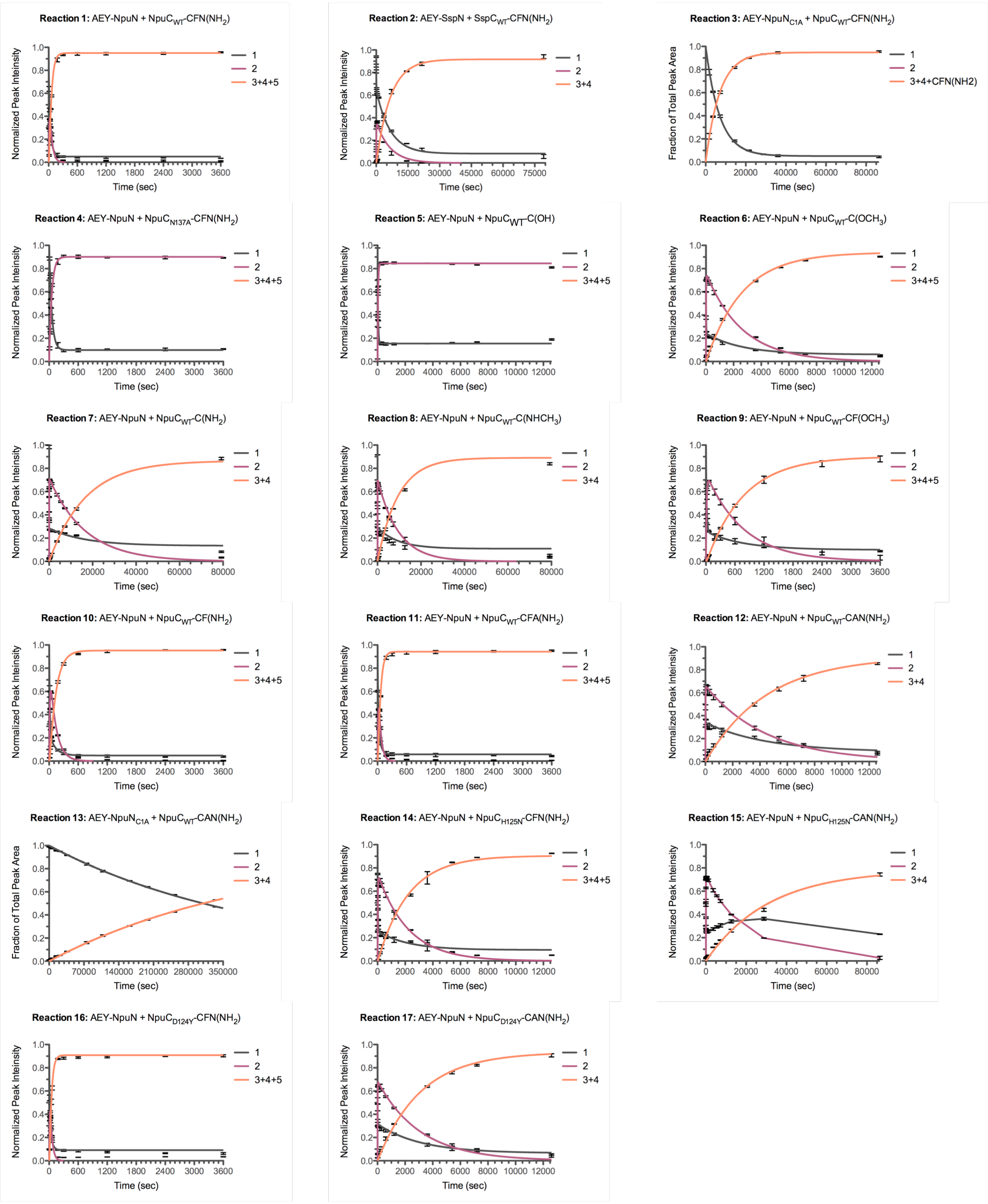

Figure S12. Reaction progress curves for splicing assays: the entire reaction. Error bars represent the standard deviation from three or four independent measurements. The solid lines are reaction progress curves generated by plotting the relevant rate equations (see methods section above) with the average best-fit values for all fitted kinetic parameters (Table 1, main text). Note that for reaction 13, the lines for species **1** (black) and **2** (purple) are not fitted curves (see methods section above for details).

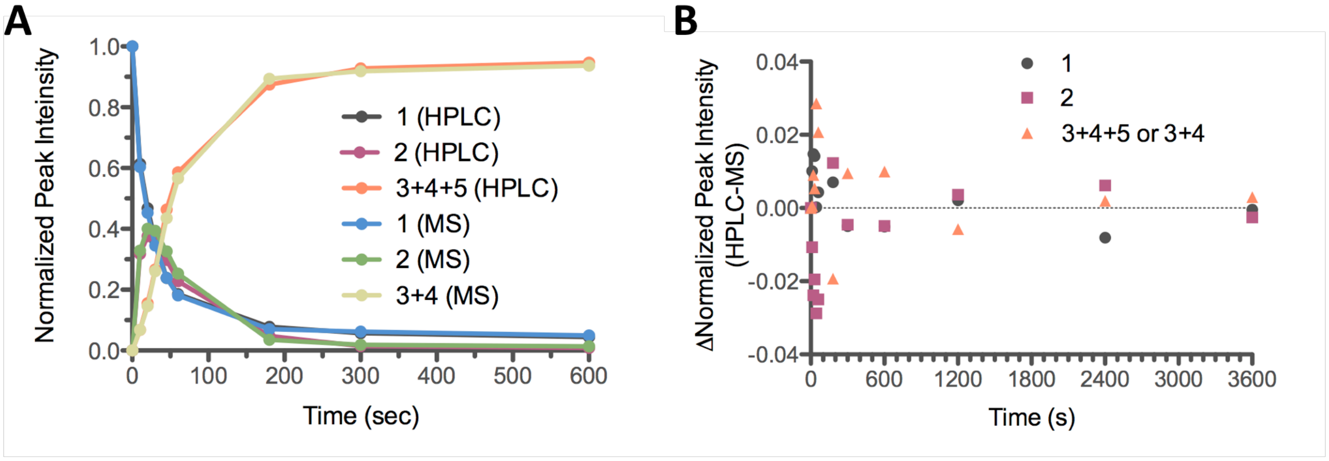

Figure S13. Comparison of RP-HPLC and ESI-MS splicing assays for Reaction 1. A. Analysis and quantification of the same samples from one iteration of reaction 1 by RP-HPLC (first three reaction curves) and ESI-MS (second three reaction curves). **B.** Difference between the normalized peak intensities for the RP-HPLC and ESI-MS reactions for each species in the kinetic model. Note that the maximal error is roughly 3% and is only that large for the first few time points.

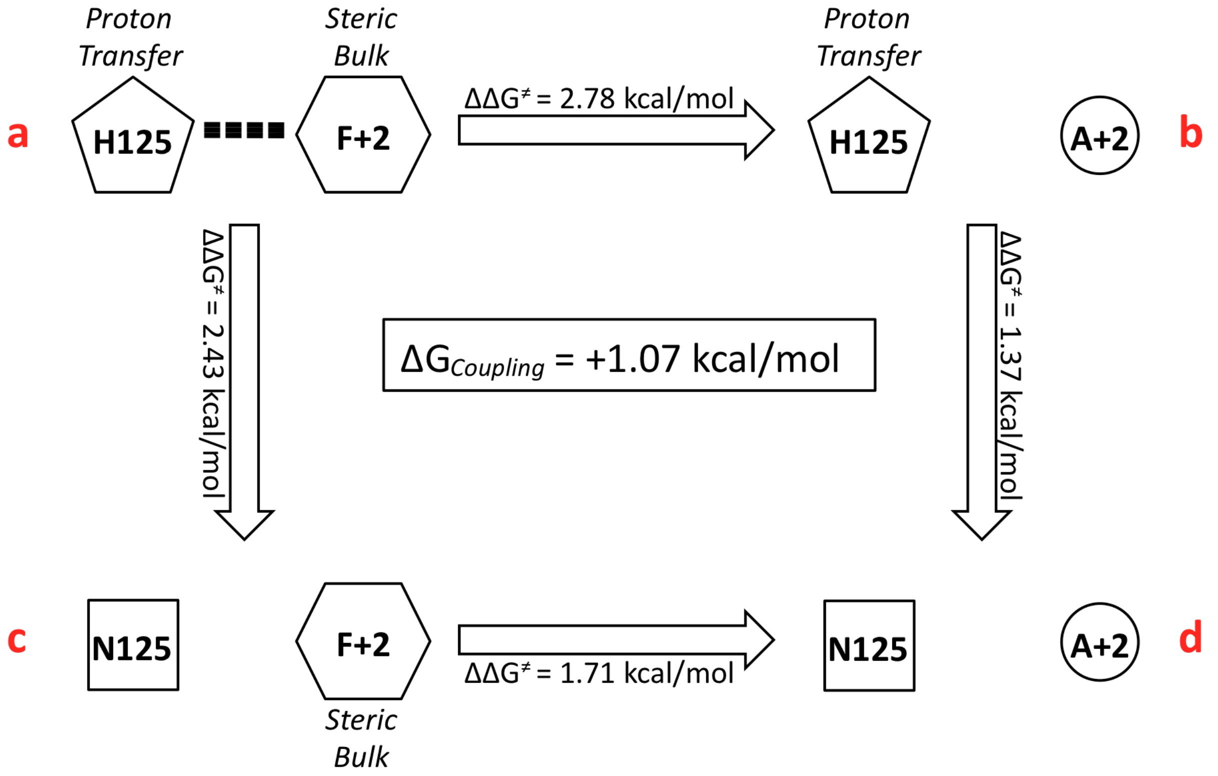

Figure S14. Mutant cycle analysis of the His$_{125}$-Phe$_{+2}$ interaction's contribution to splicing. The coupling free energy indicates a loss of favorable interaction energy for BI resolution upon mutation of His$_{125}$ and Phe$_{+2}$.

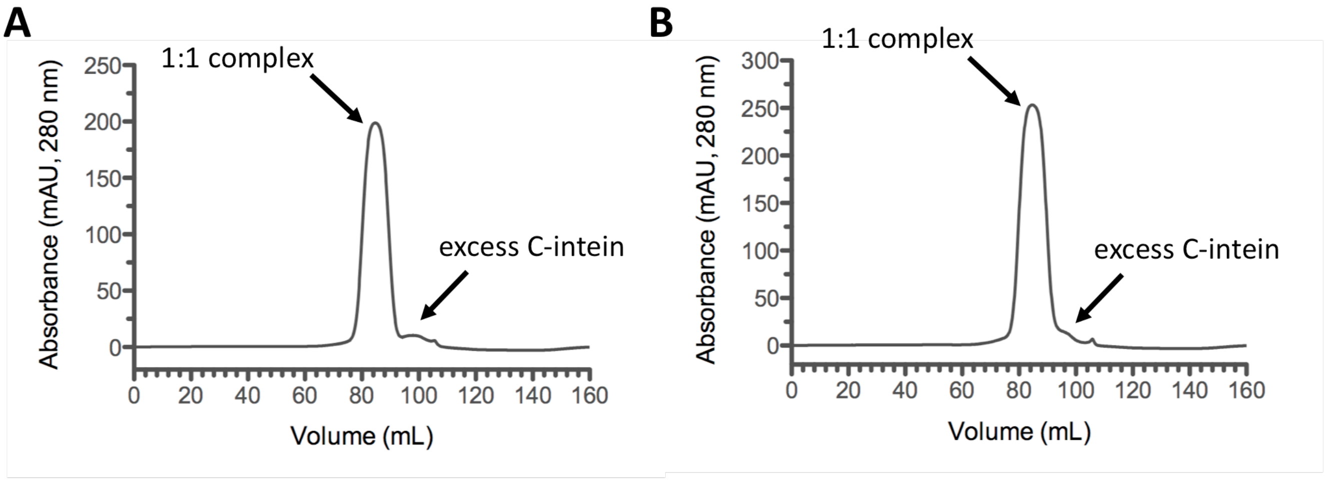

Figure S15. Preparative size-exclusion chromatography of complexes for NMR. Unlabeled AEY-$\mathrm{NpuN}_{\mathrm{C1A}}$ was mixed with slight excess of **A.** $\mathrm{NpuC}_{\mathrm{N137A}}$(*$^{13}C$,$^{15}N$*)-CFN($NH_2$) or **B.** $\mathrm{NpuC}_{\mathrm{N137A}}$(*$^{13}C$,$^{15}N$*)-CAN($NH_2$) and purified on a Superdex 75 16/60 preparative gel filtration column (CV = 125 mL). The sample was eluted with 1.35 CV of NMR buffer (25 mM sodium phosphates, 100 mM NaCl, 1 mM DTT, pH 6.5).

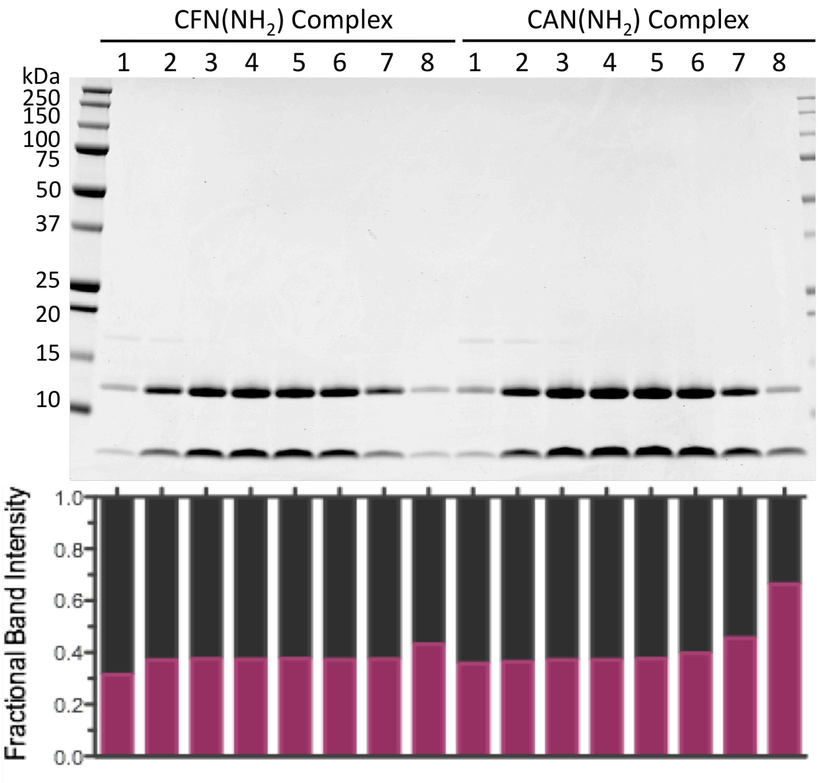

Figure S16. SDS-PAGE analysis of complexes for NMR. Eight fractions, equally spaced throughout the entire major size-exclusion peak shown in Figure S15, were analyzed by SDS-PAGE. For the CFN(NH_2) complex, fractions 2-7 had a roughly constant ratio of N- to C-intein (black and magenta, respectively). For the CAN(NH_2) complex, fractions 1-6 had a roughly constant ratio of N- to C-intein. These fractions were pooled for NMR analysis.

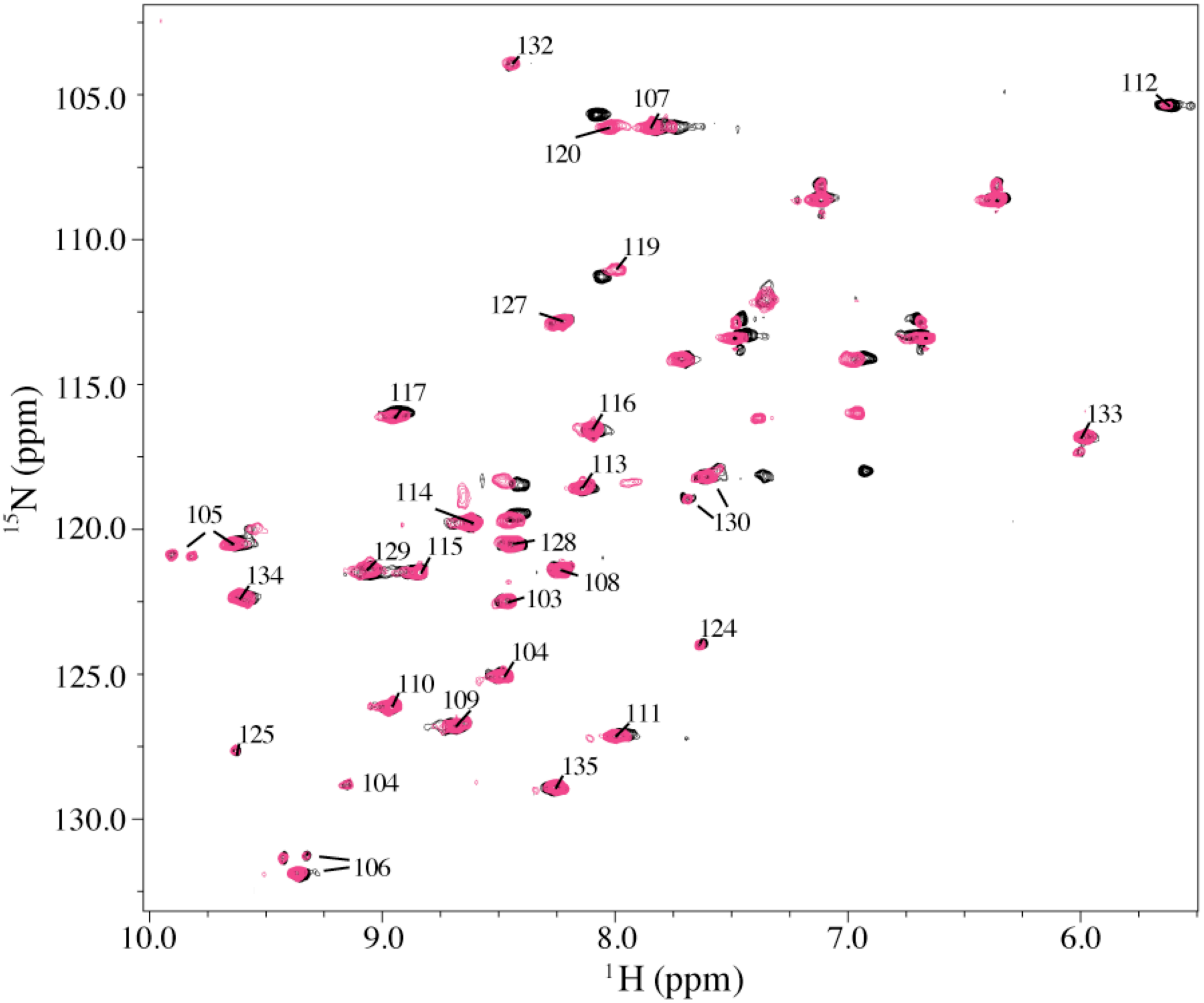

Figure S17. C-extein dependence in ^{1}H-^{15}N HSQC spectra of labeled Npu$_{C}$ in complex with NpuN. C-intein resonances for the complex bearing a CFN(NH$_2$) C-extein are shown in black, and C-intein resonances for the complex bearing a CAN(NH$_2$) C-extein are shown in red. Cross-peaks assigned to specific backbone amide N-H groups are labeled with that residue number. Note that residues K$_{104}$, I$_{105}$, A$_{106}$, and K$_{130}$ show multiple distinct resonances per residue. This was also true but less dramatic for T$_{107}$, R$_{108}$, K$_{109}$, and Y$_{110}$. The backbone N-H resonances for the N-terminal methionine, along with residues D$_{118}$, V$_{121}$, E$_{122}$, R$_{123,}$ N$_{126}$, N$_{131}$, S$_{136}$, and A$_{137}$ could not be assigned.

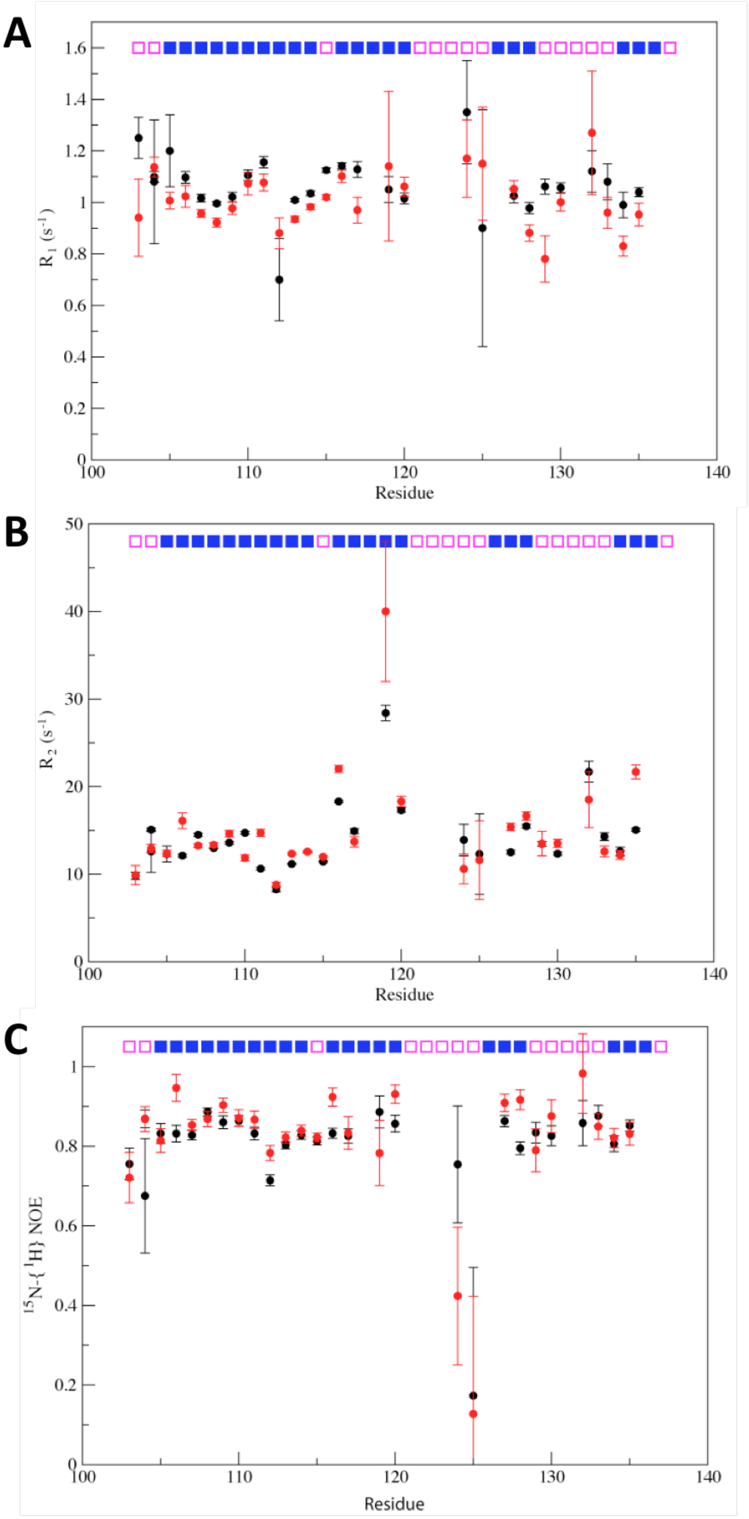

Figure S18. Relaxation data for complexed NpuC with different C-exteins. A. R$_1$, **B.** R$_2$, and **C.** ^{15}N-^{1}H heteronuclear NOE data are shown for the two complexes. Data for the complex bearing a CFN(NH$_2$) C-extein are shown in black, and data for the complex bearing a CAN(NH$_2$) C-extein are shown in red. The secondary structure elements are depicted above the graphs as solid blue rectangles, and the empty pink rectangles represent residues in loop regions.

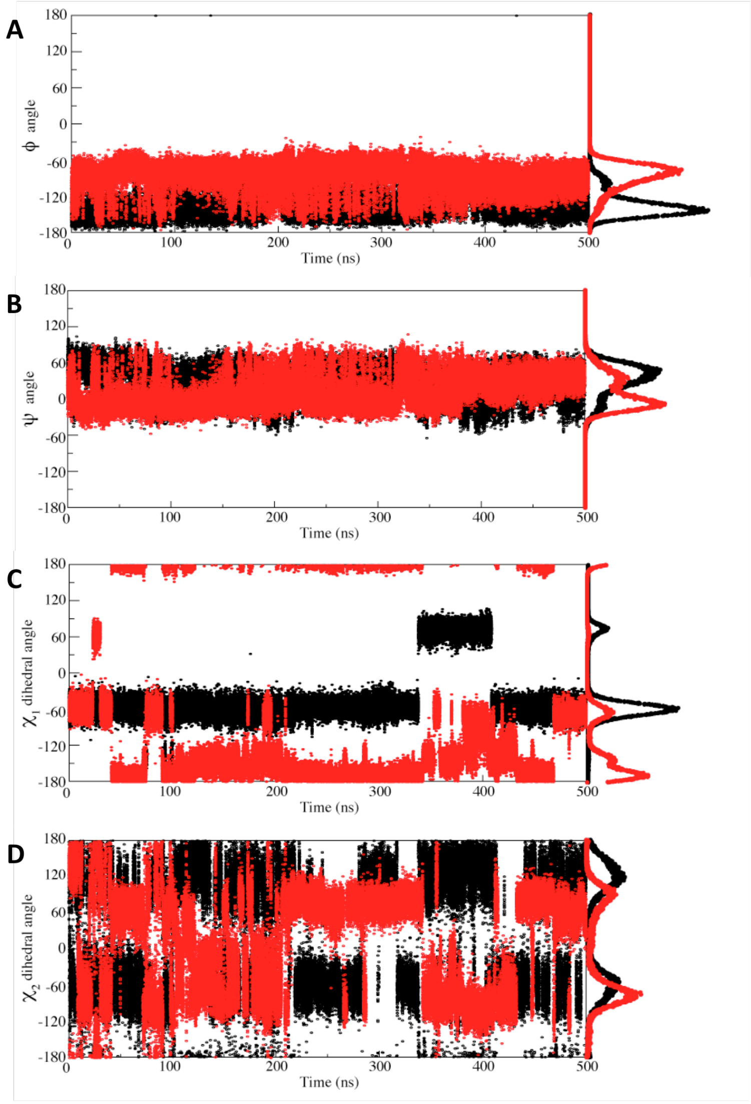

Figure S19. His$_{125}$ ϕ, ψ, χ_1, and χ_2 dihedral angles in the WT CFN and CAN simulations. A. ϕ, B. ψ, C. χ_1, and D. χ_2. Trajectories from the simulation with the CFN(NH$_2$) C-extein are shown in black, and trajectories from the simulation with the CAN(NH$_2$) C-extein are shown in red.

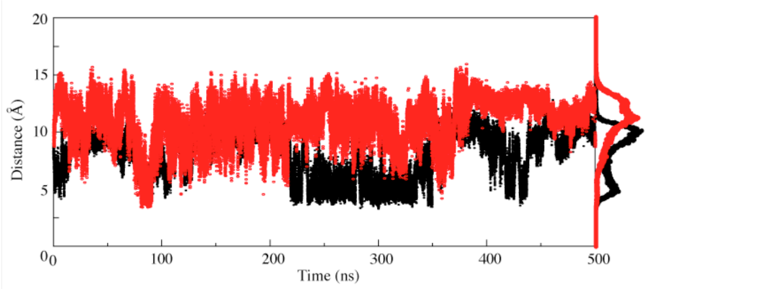

Figure S20. Distance between His$_{125}$ and Phe$_{+2}$/Ala$_{+2}$ Cβ atoms in the WT CFN and CAN simulations. The trajectory from the simulation with the CFN(NH$_2$) C-extein is shown in black, and the trajectory from the simulation with the CAN(NH$_2$) C-extein is shown in red.

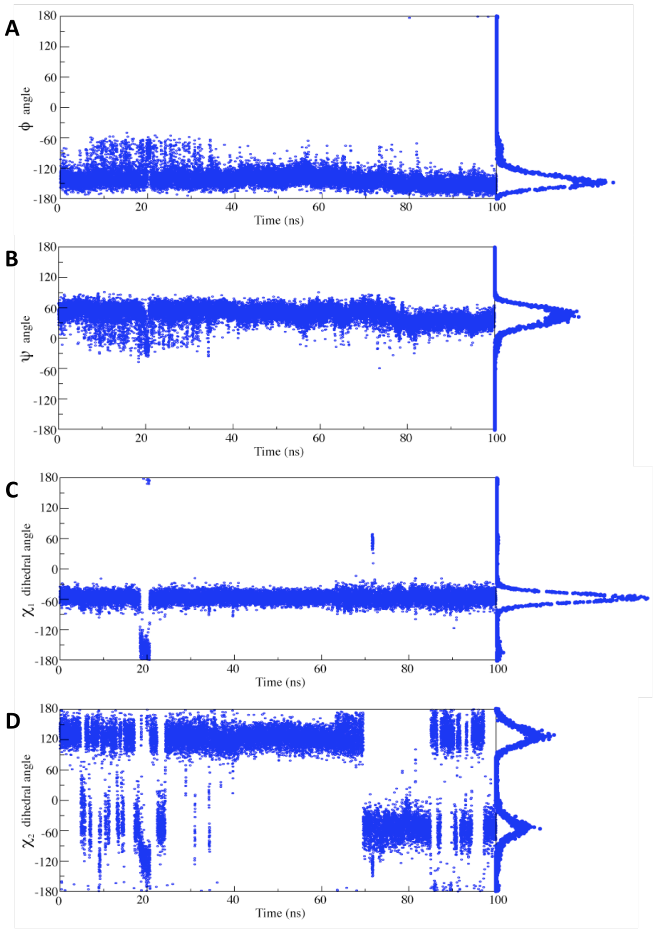

Figure S21. His$_{125}$ ϕ, ψ, χ_1, and χ_2 dihedral angles in the D124Y CAN simulations. A. ϕ, B. ψ, C. χ_1, and D. χ_2.

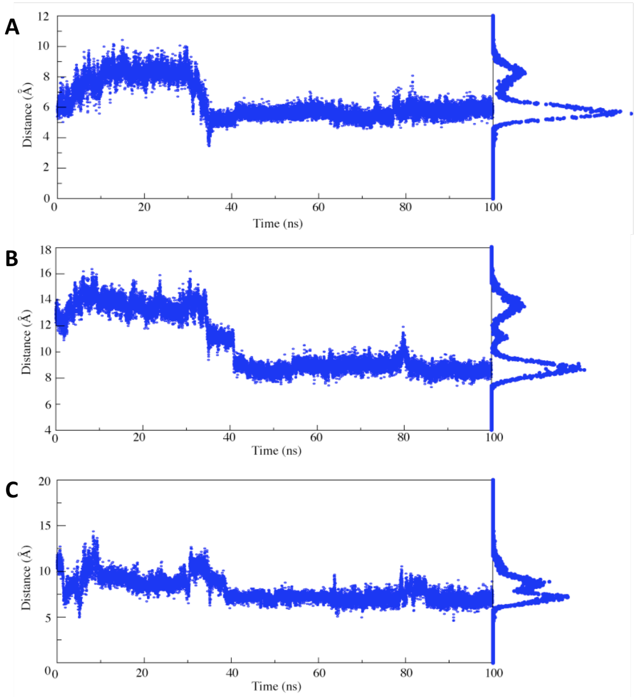

Figure S22. Distances measurements in the D124Y CAN simulation. A. His$_{125}$ Cβ to Asn$_{137}$ Cβ, B. Ile$_{119}$ NH to Cys$_{+1}$ NH, and C. His$_{125}$ Cβ to Ala$_{+2}$ Cβ