Angewandte Chemie
Eine Zeitschrift der Gesellschaft Deutscher Chemiker

Supporting Information



69451 Weinheim, Germany

# Kinetic Control of One-Pot Trans-Splicing Reactions by Using a Wild-Type and Designed Split Intein**

*Neel H. Shah, Miquel Vila-Perelló,* and Tom W. Muir**

anie_201102909_sm_miscellaneous_information.pdf

**Supporting Information Table of Contents:** **Page #**

## Materials and Methods

### Materials

All buffering salts and isopropyl-β-D-thiogalactopyranoside (IPTG) were purchased from Fisher Scientific (Pittsburgh, PA). β-Mercaptoethanol (BME), DL-dithiothreitol (DTT), benzamide, and nicotinamide-adenine-dinucleotide (NAD) were purchased from Sigma-Aldrich (St. Louis, MO). Tris(2-carboxyethyl)phosphine hydrochloride was purchased from Thermo Scientific (Rockford, IL). 6-Biotin-17-nicotinamide-adenine-dinucleotide (biotinylated NAD), 10x activated DNA, and human PARP1 were purchased from Trevigen (Gaithersberg, MD). Complete protease inhibitor tablets were purchased from Roche Diagnostics (Mannheim, Germany).

The QuikChange XL II site directed mutagenesis kit was from Agilent (La Jolla, CA). Nickel-nitriloacetic acid (Ni-NTA) resin and the KOD DNA polymerase kit were from Novagen (Gibbstown, NJ). T4 DNA ligase and all restriction enzymes were from New England Biolabs (Ipswich, MA). The Phusion High-Fidelity PCR kit was purchased from Thermo Scientific (Rockford, IL). 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. OverExpress C43(DE3) competent cells were purchased from Lucigen (Middleton, WI). Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). All plasmids used in this study were sequenced by GENEWIZ (South Plainfield, NJ) to verify the correct DNA sequence.

Criterion XT Bis-Tris gels (10% and 12%), Immun-blot PVDF membrane (0.2μm), and α-Rabbit-HRP secondary antibody were purchased from Bio-Rad (Hercules, CA). 20x MES-SDS running buffer was purchased from Boston Bioproducts (Ashland, MA). SYPRO Ruby stain and 4x LDS loading dye were purchased from Invitrogen (Carlsbad, CA). Mouse HA.11 monoclonal

antibody (α-HA) was purchased from Covance (Princeton, NJ). PARP1 (H-300) rabbit polyclonal antibody (α-PARP1-N) and PARP1 (F-2) mouse monoclonal antibody (α-PARP1-C) were purchased from Santa Cruz Biotechnology (Santa Cruz, CA). IRDye 800CW goat anti-Mouse IgG, IRDye 700CW goat anti-Mouse IgG, and IRDye 800CW goat anti-Rabbit IgG secondary antibodies were purchased from LI-COR Biotechnology (Lincoln, NE). ECL Plus reagent and streptavidin-horseradish peroxidase (HRP) conjugate (RPN1231) (Streptavidin-HRP) were from GE Health Life Sciences (Piscataway, NJ).

**Equipment**

Preparative and analytical size-exclusion chromatography was carried out on an ÄKTA FPLC system from GE Healthcare. Preparative FPLC was carried out on a HiLoad 16/60 Superdex 75 or HiLoad 16/60 Superdex 200 column. Analytical FPLC was carried out on a Superdex 75 10/300 column. For all runs, proteins were eluted over 1.3 column volumes of buffer (preparative flow rate: 1 mL/min, analytical flow rate: 0.5 mL/min). Analytical RP-HPLC was performed on a Hewlett-Packard 1100 series instrument equipped with a C18 Vydac column (5 μm, 4.6 x 150 mm) at a flow rate of 1 mL/min. All runs used 0.1 % TFA (trifluoroacetic acid) in water (solvent A) and 0.1% TFA 90 % acetonitrile in water (solvent B). For all runs, a two-minute isocratic period in solvent A was followed by a 30-minute linear gradient from 0-73% solvent B in solvent A. Electrospray ionization mass spectrometric analysis (ESI-MS) was performed on a Sciex-API-100 single quadrupole spectrometer using positive ionization. Matrix-assisted laser desorption/ionization time of flight mass spectrometry (MALDI-TOF MS) was carried out on an ABI Voyager DE STR Mass Spectrometer. *In vivo* intein activity assays were carried out on a VersaMax tunable microplate reader from Molecular Devices. Fluorescence spectroscopy was carried out on a HORIBA Jobin Yvon FluroLog-3 spectrofluorometer. SYPRO Ruby-stained gels

were scanned using a Bio-Rad VersaDoc scanner and quantified using ImageJ, developed at the National Institutes of Health. Western blots were imaged and quantified on a LI-COR Odyssey Infrared Imager.

**Bioinformatic Sequence Analysis of Split and Intact Mini-Inteins**

All mini-intein sequences (116 total) were retrieved from the InBase based on the criteria that they were explicitly annotated as having no Block C, D, E, or H motifs (the homing endonuclease motifs).[1] Sequences longer than 200 residues (14 sequences) were removed to avoid alignment problems due to large insertions. The remaining sequences were aligned with a Hidden Markov Model (HMM) from the SUPERFAMILY database (model # 0036550)[2] using HMMER 3.0.[3, 4] Several mis-aligned sequences (25%) were removed from the resulting alignment, and the remaining sequence alignment was manually optimized based on known intein sequence motifs.[1] This optimized alignment was used to build a new HMM, which in turn was used to align all 116 mini-intein sequences extracted from the InBase. The resulting alignment was manually optimized, and this optimized alignment (24 split inteins and 76 intact inteins) was used for all bioinformatic analyses.

To analyze the isoelectric points (pI) of N- and C-intein sequences, the sequence alignment was split into two alignments: one alignment contained all positions that aligned with the N-terminal methionine of the C-inteins through the C-terminal asparagine; the second alignment contained all sequence positions before the C-intein region (Figure S1). The isoelectric point of each sequence fragment was estimated using the pI/MW tool on the ExPASY server.[5] To generate sequence logos, these split alignments were further modified, removing any positions that did not align with the $Npu_{WT}$ intein. The modified alignments were separated based on naturally split and intact inteins and represented as sequence logos using WebLogo.[6] See Figures 1a and 1c in the main text and Figures S1-S3.

**Intein-Activity Coupled Kanamycin Resistance Assays**

The aminoglycoside phosphotransferase ($Kan^R$) and $Npu_{WT}$ gene fragments were cloned into a pBluescript KS (+) vector between KpnI and SacI restriction sites as previously described to generate the construct depicted at the top of Figure S4.[7] The resulting plasmid was modified by PCR insertion of an N-terminal myc-tag onto the $Kan^R$-N gene using the QuikChange kit with primers **1** and **2**. This wild-type plasmid (myc-KanR-NpuDnaE-split) was further modified by QuikChange PCR mutagenesis with primers **3-36** to generate all mutants used for *in vivo* splicing assays (Tables S1 and S2). Plasmids with multiple mutations were introduced by iterative mutagenesis steps with different primer pairs. When possible, multiple mutations were incorporated using a single primer pair spanning several mutation sites.

Intein activity-coupled kanamycin resistance ($Kan^R$) assays (Figure S4) were conducted in 96-well plate format as previously described.[7] Typically, plasmids were transformed into sub-cloning efficiency DH5α cells by heat shock, and the transformed cells were grown for 16-18 hours at 37 °C in Luria-Bertani (LB) media with 100 μg/mL of ampicillin (LB/amp). The over-night cultures were diluted 250-fold into LB/amp solutions containing 8 different kanamycin concentrations. The cells were grown at 30 °C on a 96-well plate, monitoring optical density at 595 nm every 5 minutes for 24 hours while shaking. The endpoint of this growth curve (typically in the stationary phase) was plotted as a function of kanamycin concentration to visualize the dose-response relationship (e.g. the $Npu_{WT}$ vs. $Npu_{MUT}$ curves in Figure 2b) and fitted to a variable-slope dose-response equation to determine IC50 values (Figure S5).

$$OD_{Obs} = OD_{Min} + \frac{\left(OD_{Max} - OD_{Min}\right)}{1 + 10^{\left[\left(\log IC_{50} - \log[Kan]\right)\cdot HillSlope\right]}}$$

**Cloning and Protein Isolation for *In Vitro* Binding, Splicing, and Competition Assays**
(Intein sequences in each protein are shown in bold. Mutated residues are underlined.)

*$His_6$-Ub-$NpuN_{WT}$ (referred to as Ub-$NpuN_{WT}$ in the main text)*

MHHHHHHGGMQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQELIFAGKQLE
DGRTLSDYNIQKESTLHLVLRLRGGGGGGKFAEY**CLSYETEILTVEYGLLPIGKIVEKRI**
**ECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDE**
**IFERELDLMRVDNLPN**

The plasmid pMR-Ub-NpuN(WT) was constructed as previously described.[7] *E. coli* BL21 cells transformed with pMR-Ub-NpuN(WT) were grown in 2 L of LB medium containing ampicillin (100 μg/mL) 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, the cell pellet was resuspended in cold lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, 2 mM BME, pH 8.0) supplemented with Complete protein inhibitor tablets. Cells were lysed by passage through a French press, and the soluble fraction was recovered by centrifugation. 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 2.5 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.

The cleanest fractions containing the desired product were combined and treated with 50 mM DTT for 15 minutes at room temperature to reduce any inter- and intra-molecular disulfide bonds. The reduced sample was loaded onto a Superdex S75 column for preparative size exclusion chromatography and eluted in splicing buffer (100 mM phosphate, 150 mM NaCl, 1mM DTT, 1mM EDTA, pH 7.2). FPLC fractions were analyzed by SDS-PAGE, and the purest fractions (Figure S7)

were pooled and analyzed by analytical FPLC, analytical HPLC, and mass spectrometry (Figures S8 and S9, Table S5).

*$His_6$-Ub-$NpuN_{WT}$-C1A*

MHHHHHHGGMQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQELIFAGKQLE
DGRTLSDYNIQKESTLHLVLRLRGGGGGGKFAEY**ALSYETEILTVEYGLLPIGKIVEKRI**
**ECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDE**
**IFERELDLMRVDNLPN**

The plasmid pMR-Ub-NpuN(WT)-C1A was constructed by PCR mutagenesis of the Cys1 codon in the pMR-Ub-NpuN(WT) plasmid using a QuikChange kit with primers **37** and **38** (Table S3). The $His_6$-Ub-$NpuN_{WT}$-C1A protein was expressed and purified identically as $His_6$-Ub-$NpuN_{WT}$. The purest fractions from preparative FPLC (Figure S7) were pooled and analyzed by analytical HPLC and mass spectrometry (Figures S8 and Table S5).

*$His_6$-Ub-$NpuN_{MUT}$ (referred to as Ub-$NpuN_{MUT}$ in the main text)*

MHHHHHHGGMQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQELIFAGKQLE
DGRTLSDYNIQKESTLHLVLRLRGGGGGGKFAEY**CLSYETKILTVEYGLLPIGKIVEKRI**
**ECTVYSVDNNGNIYTQPVAQWHDRGKQKVFEYCLKDGSLIRATKDHKFMTVDGQMLPIKE**
**IFRRKLDLMRVDNLPN**

The plasmid pMR-Ub-NpuN(13) was constructed by PCR mutagenesis of the E7, E52, E54, E61, D85, E89, and E91 codons in the pMR-Ub-NpuN(WT) plasmid using a QuikChange kit with primers **3/4**, **9/10**, **17/18**, and **21/22** (Table S1). The $His_6$-Ub-$NpuN_{MUT}$ protein was expressed and purified identically as $His_6$-Ub-$NpuN_{WT}$. The purest fractions from preparative FPLC (Figure S7) were pooled and analyzed by analytical FPLC, analytical HPLC, and mass spectrometry (Figures S8 and S9, Table S5).

*$His_6$-Ub-$NpuN_{MUT}$-C1A*

MHHHHHHGGMQIFVKTLTGKTITLEVESSDTIDNVKSKIQDKEGIPPDQQELIFAGKQLE
DGRTLSDYNIQKESTLHLVLRLRGGGGGGKFAEY**ALSYETKILTVEYGLLPIGKIVEKRI**
**ECTVYSVDNNGNIYTQPVAQWHDRGKQKVFEYCLKDGSLIRATKDHKFMTVDGQMLPIKE**
**IFRRKLDLMRVDNLPN**

The plasmid pMR-Ub-NpuN(13)-C1A was constructed by PCR mutagenesis of the Cys1 codon in the pMR-Ub-NpuN(13) plasmid using a QuikChange kit with primers **37** and **38** (Table S3). The $His_6$-Ub-$NpuN_{MUT}$-C1A protein was expressed and purified identically as $His_6$-Ub-$NpuN_{WT}$. The purest fractions from preparative FPLC (Figure S7) were pooled and analyzed by analytical HPLC and mass spectrometry (Figures S8 and Table S5).

*$His_6$-MBP-$NpuN_{WT}$ (Note that the N-terminal methionine is removed in vivo.)*

MGSSHHHHHHSQDPKIEEGKLVIWINGDKGYNGLAEVGKKFEKDTGIKVTVEHPDKLEEK
FPQVAATGDGPDIIFWAHDRFGGYAQSGLLAEITPDKAFQDKLYPFTWDAVRYNGKLIAY
PIAVEALSLIYNKDLLPNPPKTWEEIPALDKELKAKGKSALMFNLQEPYFTWPLIAADGG
YAFKYENGKYDIKDVGVDNAGAKAGLTFLVDLIKNKHMNADTDYSIAEAAFNKGETAMTI
NGPWAWSNIDTSKVNYGVTVLPTFKGQPSKPFVGVLSAGINAASPNKELAKEFLENYLLT
DEGLEAVNKDKPLGAVALKSYEEELAKDPRIAATMENAQKGEIMPNIPQMSAFWYAVRTA
VINAASGRQTVDEALKDAQTNGGGGKFAEY**CLSYETEILTVEYGLLPIGKIVEKRIECTV**
**YSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQMLPIDEIFER**
**ELDLMRVDNLPN**

The plasmid pMR-MBP-NpuN(WT) was constructed in multiple steps. First, the MBP gene was amplified from the commercial pRK793 plasmid using KOD polymerase with primers **45** and **46** to contain a 5' NdeI restriction site and 3' overlapping sequence with the NpuN gene. The $NpuN_{WT}$ gene was analogously amplified from the myc-KanR-NpuDnaE-split plasmid with primers **47** and **48** to contain a 5' overlapping sequence with the MBP gene and a 3' SacII restriction site. The PCR products were mixed and fused/amplified using KOD polymerase and primers **45** and **48** (see Table S3 for sequence of primers **45-48**). The resulting MBP-$NpuN_{WT}$ gene fusion was digested with NdeI and SacII then ligated into a digested pMR-Ub-NpuN(WT) plasmid with T4 DNA ligase. BL21(DE3) cells were transformed with the resulting plasmid and grown in 2 L of LB

media 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. The cells were lysed and the desired protein was enriched over Ni-NTA resin identically as for $His_6$-Ub-$NpuN_{WT}$.

The cleanest fractions containing the desired product were combined and treated with 50 mM DTT for 15 minutes at room temperature to reduce any inter- and intra-molecular disulfide bonds. The reduced sample was loaded onto a Superdex S200 column for preparative size exclusion chromatography and eluted in splicing buffer (100 mM phosphate, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.2). FPLC fractions were analyzed by SDS-PAGE, and the purest fractions (Figure S7) were pooled and analyzed by analytical FPLC, analytical HPLC, and mass spectrometry (Figures S8 and S9, Table S5).

*$NpuC_{WT}$-SUMO-HA (referred to as $NpuC_{WT}$-SUMO in the main text. The sequence given is pre-cleavage by TEV protease to remove the N-terminal $His_6$-tag.)*

MGSSHHHHHHGENLYFQ|G**IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN**CFNSGLV
PRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAK
RQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGYPYDVPDYA

The plasmid pET-NpuC(WT)-SUMO was generated in three steps by overlap-extension PCR.[8] First, the $NpuC_{WT}$ gene was amplified from the myc-KanR-NpuDnaE-split plasmid by Phusion polymerase with primers **39** and **40** to contain 5' and 3' sequences that overlap with regions of a modified pET-SUMO plasmid (Table S3). In a second step, the $NpuC_{WT}$ gene was inserted into the pET-SUMO plasmid, upstream of SUMO, using Phusion polymerase, then the template plasmid was globally digested using DpnI. The resulting product was used to transform DH5α cells, amplified, and isolated. In the third step, a C-terminal HA tag was inserted using a QuikChange kit with primers **41** and **42** (Table S3).

To express $NpuC_{WT}$-SUMO-HA, *E. coli* BL21 cells transformed with pET-NpuC(WT)-SUMO were grown in 2 L of LB medium containing kanamycin (50 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. The cells were lysed and the desired protein was enriched over Ni-NTA resin identically as for $His_6$-Ub-$NpuN_{WT}$. The purest fractions were pooled and dialyzed into TEV cleavage buffer (20 mM Tris, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, pH 8.0) then treated with TEV protease overnight at 4°C. The cleavage was confirmed by SDS-PAGE, then the reaction solution was diluted two-fold in wash buffer 1 (50 mM phosphate, 300 mM NaCl, 20 mM imidazole, 2 mM BME, pH 8.0) and incubated with Ni-NTA resin at 4°C for 30 min. The flow-through and two 1 CV washes with wash buffer 1 were collected and pooled. The sample was loaded onto a Superdex S75 column for preparative size exclusion chromatography and eluted in splicing buffer (100 mM phosphate, 150 mM NaCl, 1mM DTT, 1mM EDTA, pH 7.2). FPLC fractions were analyzed by SDS-PAGE, and the purest fractions (Figure S7) were pooled and analyzed by analytical FPLC, analytical HPLC, and mass spectrometry (Figures S8 and S9, Table S5).

*$NpuC_{MUT}$-SUMO-HA (referred to as $NpuC_{MUT}$-SUMO in the main text. The sequence given is pre-cleavage by TEV protease to remove the N-terminal $His_6$-tag.)*

MGSSHHHHHHGENLYFQ|G**IEIATEKYLGEQNVYDIGVERDHNFALENGFIASN**CFNSGLV
PRGSASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAK
RQGKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIGGYPYDVPDYA

The plasmid pET-NpuC(13)-SUMO was generated in two steps from pET-NpuC(WT)-SUMO by PCR mutagenesis with the QuikChange kit using primers **43** and **44** (Table S3) in the first step and primers **29** and **30** (Table S1) in the second step. The resulting plasmid was used to express and purify $NpuC_{MUT}$-SUMO-HA identically as $NpuC_{WT}$-SUMO-HA. The purest fractions from preparative FPLC (Figure S7) were pooled and analyzed by analytical FPLC, analytical HPLC, and mass spectrometry (Figures S8 and S9, Table S5).

*$NpuC_{WT}$-eGFP-HA (The sequence given is pre-cleavage by TEV protease to remove the N-terminal $His_6$-tag.)*

MGSSHHHHHHGENLYFQ|G**IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN**CFNSGLV
PRGSASMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVK
FEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIED
GSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGM
DELYKYPYDVPDYA

The plasmid pET-NpuC(WT)-eGFP was constructed in multiple steps. First, the $NpuC_{WT}$ gene was amplified from the pET-NpuC(WT)-SUMO plasmid using KOD polymerase with primers **49** and **50** to contain a 5' NdeI restriction site and 3' overlapping sequence with the eGFP gene. The eGFP gene was analogously amplified from an eGFP containing plasmid with primers **51** and **52** to contain a 5' overlapping sequence with the $NpuC_{WT}$ gene and a 3' HA tag and HindIII restriction site. The PCR products were mixed and fused/amplified using KOD polymerase and primers **49** and **52** (see Table S3 for sequence of primers **49-52**). The resulting $NpuC_{WT}$-eGFP gene fusion was digested with NdeI and HindIII then ligated into a digested pET-NpuC(WT)-SUMO plasmid with T4 DNA ligase.

*E. coli* BL21 cells transformed with the plasmid were grown in 2 L of LB medium containing ampicillin (100 μg/mL) 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, the cell pellet was resuspended in cold lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, 2 mM BME, pH 8.0) supplemented with Complete protein inhibitor tablets. The cells were lysed and the desired protein was enriched over Ni-NTA resin identically as for $His_6$-Ub-$NpuN_{WT}$. The purest fractions were pooled, dialyzed into cleavage buffer, proteolyzed by TEV protease, and FPLC-purified identically as $NpuC_{WT}$-SUMO-HA. The purest fractions from preparative FPLC (Figure S7) were pooled and analyzed by analytical FPLC, analytical HPLC, and mass spectrometry (Figures S8 and S9, Table S5).

**Cloning and Protein Isolation for Three-Piece Ligations**
(Intein sequences in each protein are shown in bold. Mutated residues are underlined.)

*Construction of "Empty" Split Intein Vectors for Three-Piece Ligations*

Initially, three vectors were generated with the following general organization: N-plasmid → $His_6$-TEV-MCS-$NpuN_{MUT}$, M-plasmid → $His_6$-$NpuC_{MUT}$-MCS-$NpuN_{WT}$, and C-plasmid → $His_6$-$NpuC_{WT}$-MCS (where TEV is a TEV protease cleavage site and MCS is the multiple cloning site from the parent plasmid). All three vectors were created from orthogonal empty Duet vectors from Novagen (N: pACYCDuet-1, M: pCDFDuet-1, and C: pETDuet-1) using overlap-extension PCR cloning.[8] The pACYCDuet-1 vector was previously modified to contain a TEV site after its native $His_6$-tag upstream of the MCS. To generate the N-plasmid, first the $NpuN_{MUT}$ gene was amplified from pMR-Ub-NpuN(13) using Phusion polymerase with primers **53** and **54** (Table S4). The resulting PCR product was used to insert $NpuN_{MUT}$ after the MCS in the modified pACYCDuet-1 vector using Phusion polymerase to give the desired N-plasmid. The C-plasmid was constructed analogously by amplifying $NpuC_{WT}$ from pET-NpuC(WT)-SUMO using primers **55** and **57** (Table S4) and inserting the gene into pETDuet-1. To construct the M-plasmid, the $NpuC_{MUT}$ and $NpuN_{WT}$ genes were simultaneously amplified/fused in one pot from pET-NpuC(13)-SUMO and pMR-Ub-NpuN(WT), respectively, using Phusion polymerase and primers **53**, **54**, **56**, and **57** (Table S4). The resulting PCR product, $NpuC_{MUT}$-MCS-$NpuN_{WT}$, was inserted into pCDFDuet-1 using Phusion polymerase to generate the M-plasmid.

*SH3-$NpuN_{MUT}$-$His_6$ (referred to as SH3-$NpuN_{MUT}$ in the main text.)*

MEAIAKYDFKATADDELSFKRGDILKVLNEECDQNWYKAELNGKDGFIPKNYIEM**CLSYE TKILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGKQKVFEYCLKDGSL IRATKDHKFMTVDGQMLPIKEIFRRKLDLMRVDNLPN**GSSHHHHHHSG

The plasmid pACYC-Grb2-N-Int-His was generated as follows. First, the gene for the N-terminal SH3 domain of Grb2 (amino acids 1-55) was amplified from a plasmid containing the

human Grb2 gene using Phusion polymerase with primers **58** and **59**. The PCR product was then inserted into the empty N-plasmid using Phusion polymerase. Next, a C-terminal $His_6$-tag was inserted after the $NpuN_{MUT}$ gene using a QuikChange kit with primers **66** and **67**. Finally the N-terminal $His_6$-TEV gene was deleted using a QuikChange kit with primers **60** and **61** to generate the desired plasmid (see Table S4 for all primer sequences). The SH3-$NpuN_{MUT}$-$His_6$ protein was expressed and purified identically as $His_6$-Ub-$NpuN_{WT}$ substituting ampicillin with chloramphenicol (50 μg/mL) in the bacterial growth medium and without DTT in the FPLC buffer. Note that lysis and enrichment over Ni-NTA resin were carried out with BME in the buffers. After preparative FPLC, fractions were analyzed by SDS-PAGE, and the purest fractions (Figure S10) were pooled and analyzed by mass spectrometry (Table S5).

*$His_6$-$NpuC_{MUT}$-GB1-$NpuN_{WT}$-$His_6$ (referred to as $NpuC_{MUT}$-GB1-$NpuN_{WT}$ in the main text. Note that the N-terminal methionine is removed in vivo.)*

MGSSHHHHHHSGG**IEIATEKYLGEQNVYDIGVERDHNFALENGFIASN**CFNYKLILNGKT
LKGETTTEAVDAATAEKVFKQYANDNGVDGEWTYDDATKTFTVTEY**CLSYETEILTVEYG**
**LLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKF**
**MTVDGQMLPIDEIFERELDLMRVDNLPN**GSSHHHHHHSG

The pCDFDuet-GB1-M-2His-Y plasmid was constructed as follows. First, the GB1 gene was amplified from another plasmid using the Phusion polymerase with primers **62** and **63**. This gene was then inserted into the empty M-plasmid using Phusion polymerase. Next, a C-terminal $His_6$-tag was inserted after $NpuN_{WT}$ using a QuikChange kit with primers **66** and **67**. Finally, a single tyrosine residue was inserted between the C-terminus of GB1 and the N-terminus of $NpuN_{WT}$ to mimic the endogenous N-extein sequence. This was done using a QuikChange kit with primers **64** and **65** (See Table S4 for all primer sequences).

To express $His_6$-$NpuC_{MUT}$-GB1-$NpuN_{WT}$-$His_6$, *E. coli* BL21 cells transformed with pCDFDuet-GB1-M-2His-Y were grown in 2 L of LB medium containing streptomycin (50 μg/mL) at 37 °C until $OD_{600}$ = 0.6. Expression was induced by addition of 0.5 mM IPTG for 3 hours at 37 °C. After

harvesting the cells by centrifugation, the cell pellet was resuspended in cold lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, no BME, pH 8.0) supplemented with Complete protein inhibitor tablets. Cells were lysed by passage through a French press, and the soluble fraction was recovered by centrifugation. 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 CV of lysis buffer (no BME), 5 CV of wash buffer 1 (lysis buffer with 20 mM imidazole and no BME), and 2.5 CV of wash buffer 2 (lysis buffer with 50 mM imidazole and no BME). The protein was eluted with elution buffer (lysis buffer with 250 mM imidazole and no BME) in four 1.5 CV elution fractions. The wash and elution fractions were analyzed by SDS-PAGE.

The cleanest fractions containing the desired product were combined, loaded onto a Superdex S75 column for preparative size exclusion chromatography and eluted in splicing buffer (100 mM phosphate, 150 mM NaCl, no DTT, 1mM EDTA, pH 7.2). FPLC fractions were analyzed by SDS-PAGE, and the purest fractions (Figure S10) were pooled and analyzed by mass spectrometry (Table S5). Note that thiols were omitted from the entire M-fragment purification to prevent unexpected thiol-induced N- and C-intein cleavage reactions in *cis* that were not observed in *trans* during our splicing assays.

*$His_6$-TEV-$NpuC_{WT}$-eGFP-HA (referred to as $NpuC_{WT}$-eGFP in the main text. Note that the N-terminal methionine is removed in vivo.)*

MGSSHHHHHHGENLYFQG**IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN**CFNSGLV
PRGSASMVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEGDATYGKLTLKFICTTGKL
PVPWPTLVTTLTYGVQCFSRYPDHMKQHDFFKSAMPEGYVQERTIFFKDDGNYKTRAEVK
FEGDTLVNRIELKGIDFKEDGNILGHKLEYNYNSHNVYIMADKQKNGIKVNFKIRHNIED
GSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSALSKDPNEKRDHMVLLEFVTAAGITLGM
DELYKYPYDVPDYA

The $His_6$-TEV-$NpuC_{WT}$-eGFP-HA protein was expressed from the same plasmid as $NpuC_{WT}$-eGFP-HA, pET-NpuC(WT)-eGFP. The protein was over-expressed and enriched over Ni-NTA resin

as indicated above. The cleanest fractions containing the desired product were combined without TEV cleavage, loaded onto a Superdex S75 column for preparative size exclusion chromatography and eluted in splicing buffer (100 mM phosphate, 150 mM NaCl, no DTT, 1mM EDTA, pH 7.2). FPLC fractions were analyzed by SDS-PAGE, and the purest fractions (Figure S10) were pooled and analyzed by mass spectrometry (Table S5).

*PARP1(1-363)-NpuN$_{MUT}$-His$_6$ (referred to as PARP1-N in the main text.)*

MAESSDKLYRVEYAKSGRASCKKCSESIPKDSLRMAIMVQSPMFDGKVPHWYHFSCFWKV
GHSIRHPDVEVDGFSELRWDDQQKVKKTAEAGGVTGKGQDGIGSKAEKTLGDFAAEYAKS
NRSTCKGCMEKIEKGQVRLSKKMVDPEKPQLGMIDRWYHPGCFVKNREELGFRPEYSASQ
LKGFSLLATEDKEALKKQLPGVKSEGKRKGDEVDGVDEVAKKKSKKEKDKDSKLEKALKA
QNDLIWNIKDELKKVCSTNDLKELLIFNKQQVPSGESAILDRVADGMVFGALLPCEECSG
QLVFKSDAYYCTGDVTAWTKCMVKTQTPNRKEWVTPKEFREISYLKKLKVKKQDRIFPPE
TSA**CLSYETKILTVEYGLLPIGKIVEKRIECTVYSVDNNGNIYTQPVAQWHDRGKQKVFE**
**YCLKDGSLIRATKDHKFMTVDGQMLPIKEIFRRKLDLMRVDNLPN**GSSHHHHHHSG

The plasmid pACYCDuet-PARP1-N-Int-His was generated as follows. First, the gene for the N-terminal region of PARP1 was amplified from a plasmid containing the full-length human PARP1[9] gene using Phusion polymerase with primers **68** and **69**. The PCR product was then inserted into the empty N-plasmid using Phusion polymerase. Next, a C-terminal His$_6$-tag was inserted after the NpuN$_{MUT}$ gene using a QuikChange kit with primers **66** and **67**. Finally the N-terminal His$_6$-TEV gene was deleted using a QuikChange kit with primers **70** and **71** to generate the desired plasmid (see Table S4 for all primer sequences).

To express PARP1-N, BL21(DE3) cells transformed with this plasmid were grown in 2 L of LB with chloramphenicol (50 μg/mL) at 37 °C until $OD_{600}$ = 0.6. Expression was induced by addition of 0.5 mM IPTG for 3 hours at 37 °C. After harvesting the cells by centrifugation, the cell pellet was resuspended in cold lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, no BME, pH 8.0) supplemented with Complete protein inhibitor tablets. Cells were lysed by passage through a French press, and the soluble fraction was recovered by centrifugation. The soluble

fraction was mixed with 1 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 CV of lysis buffer, 10 CV of wash buffer 1 (lysis buffer with 20 mM imidazole and no BME), and 2 x 2.5 CV of wash buffer 2 (lysis buffer with 50 mM imidazole and no BME). The protein was eluted with elution buffer (lysis buffer with 250 mM imidazole and no BME) in four 1.5 CV elution fractions. The wash and elution fractions were analyzed by SDS-PAGE. The cleanest elution fractions were used for three-piece ligations with no further preparations (Figure S11).

*$His_6$-$NpuC_{MUT}$-PARP1(*S364C-655*)-$NpuN_{WT}$-$His_6$ (referred to as PARP1-M in the main text.)*

MGSSHHHHHHSGG**IEIATEKYLGEQNVYDIGVERDHNFALENGFIASN**CVAATPPPSTAS
APAAVNSSASADKPLSNMKILTLGKLSRNKDEVKAMIEKLGGKLTGTANKASLCISTKKE
VEKMNKKMEEVKEANIRVVSEDFLQDVSASTKSLQELFLAHILSPWGAEVKAEPVEVVAP
RGKSGAALSKKSKGQVKEEGINKSEKRMKLTLKGGAAVDPDSGLEHSAHVLEKGGKVFSA
TLGLVDIVKGTNSYYKLQLLEDDKENRYWIFRSWGRVGTVIGSNKLEQMPSKEDAIEHFM
KLYEEKTGNAWHSKNFTKYPKKFYPLEIDYGQDEEAVKKL**CLSYETEILTVEYGLLPIGK
IVEKRIECTVYSVDNNGNIYTQPVAQWHDRGEQEVFEYCLEDGSLIRATKDHKFMTVDGQ
MLPIDEIFERELDLMRVDNLPN**GSSHHHHHHSG

The pCDFDuet-PARP1-M-2His plasmid was constructed as follows. First, the gene for the middle region of PARP1 was amplified from a plasmid containing the full-length human PARP1 gene using Phusion polymerase with primers **72** and **73**. This gene was then inserted into the empty M-plasmid using Phusion polymerase. Next, a C-terminal $His_6$-tag was inserted after $NpuN_{WT}$ using a QuikChange kit with primers **66** and **67** (See Table S4 for all primer sequences).

To express PARP1-M, C43(DE3) cells transformed with this plasmid were grown in 2 L of LB with streptomycin (50 μg/mL) at 37 °C until $OD_{600}$ = 0.6. Expression was induced by addition of 0.5 mM IPTG for 3 hours at 37 °C. After harvesting the cells by centrifugation, the cell pellet was resuspended in cold lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, no BME, pH 8.0) supplemented with Complete protein inhibitor tablets. Cells were lysed by passage through a

French press, and the soluble fraction was recovered by centrifugation. The soluble fraction was mixed with 2 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 CV of lysis buffer, 10 CV of wash buffer 1 (lysis buffer with 20 mM imidazole and no BME), and 2x 2.5 CV of wash buffer 2 (lysis buffer with 50 mM imidazole and no BME). The protein was eluted with in eight 1.5 CV elution fractions over a step-wise gradient of imidazole concentrations (2x 100 mM, 2x 200 mM, 2x 300 mM, and 2x 500 mM imidazole in lysis buffer with no BME). The wash and elution fractions were analyzed by SDS-PAGE. The cleanest elution fraction was used for three-piece ligations with no further preparations (Figure S11). Note that thiols were omitted from the PARP1-M isolation procecure to prevent unexpected thiol-induced N- and C-intein cleavage reactions in *cis* that were not observed in *trans* during our splicing assays.

*$His_6$-$NpuC_{WT}$-PARP1(*T656C-1015*) (referred to as PARP1-C in the main text.)*

MGSSHHHHHHSGG**IKIATRKYLGKQNVYDIGVERDHNFALKNGFIASN**CVNPGTKSKLPK
PVQDLIKMIFDVESMKKAMVEYEIDLQKMPLGKLSKRQIQAAYSILSEVQQAVSQGSSDS
QILDLSNRFYTLIPHDFGMKKPPLLNNADSVQAKAEMLDNLLDIEVAYSLLRGGSDDSSK
DPIDVNYEKLKTDIKVVDRDSEEAEIIRKYVKNTHATTHNAYDLEVIDIFKIEREGECQR
YKPFKQLHNRRLLWHGSRTTNFAGILSQGLRIAPPEAPVTGYMFGKGIYFADMVSKSANY
CHTSQGDPIGLILLGEVALGNMYELKHASHISKLPKGKHSVKGLGKTTPDPSANISLDGV
DVPLGTGISSGVNDTSLLYNEYIVYDIAQVNLKYLLKLKFNFKTSLW

The pETDuet-PARP1-C plasmid was constructed as follows. First, the gene for the C-terminal region of PARP1 was amplified from a plasmid containing the full-length human PARP1 gene using Phusion polymerase with primers **74** and **75** (Table S4). Then, this gene was then inserted into the empty C-plasmid using Phusion polymerase. To express PARP1-C, BL21(DE3) cells transformed with this plasmid were grown in 1 L of LB with ampicillin (100 μg/mL) at 37 °C until $OD_{600}$ = 0.6. Expression was induced by addition of 0.5 mM IPTG for 3 hours at 37°C. After harvesting the cells by centrifugation, the cell pellet was resuspended in cold lysis buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, no BME, pH 8.0) supplemented with Complete protein

inhibitor tablets. Cells were lysed by passage through a French press, and the soluble fraction was recovered by centrifugation. The soluble fraction was mixed with 2 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 CV of lysis buffer, 5 CV of wash buffer 1 (lysis buffer with 20 mM imidazole and no BME), and 2.5 CV of wash buffer 2 (lysis buffer with 50 mM imidazole and no BME). The protein was eluted with elution buffer (lysis buffer with 250 mM imidazole and no BME) in four 1.5 CV elution fractions. The wash and elution fractions were analyzed by SDS-PAGE. The cleanest elution fraction was used for three-piece ligations with no further preparations (Figure S11).

**Fluorescence Binding Assays**

All fluorescence emission spectra were recorded at 30 °C with $\lambda_{ex}$ = 295 nm, an excitation slit width of 3 nm, $\lambda_{em}$ = 310-440 nm, an emission slit width of 7 nm, an integration time of 0.5 sec, and 0.5 nm wavelength increments. Measurements were carried out in a quartz Starna Fluorometer cell with a 1 mm x 10 mm sample chamber cross-section (maximum volume 850 μL). Samples were excited on the 1 mm face and emission was recorded from the 10 mm face. All purified protein stocks and fluorescence samples were prepared in filtered splicing buffer (100 mM phosphate, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.2). Protein concentrations were quantified multiple times by UV absorbance at 280 nm and by the Bradford assay. Both methods gave consistent and similar results (within 10%), and an average value of all measurements was used. Prior to every titration, proteins and blank solutions were incubated with 1 mM TCEP (added from a pH-neutralized 100 mM stock solution) at 30 °C for a minimum of 10 minutes. At each titration point, C-intein solution was added to an N-intein solution in the cuvette, and the

mixture gently mixed by pipetting then allowed to equilibrate for 2 minutes prior to measuring emission spectra.

Equilibrium binding curves were measured by monitoring the decrease in intrinsic tryptophan fluorescence of Ub-NpuN in the presence of increasing concentrations of NpuC-SUMO. (Figure S12. Note that W47 of NpuN is the only tryptophan residue in the entire protein mixture.) Before each binding curve titration, the background fluorescence of NpuC-SUMO was measured over the entire concentration range used and found to be reproducible and linear. The slope of this line was used to subtract out the contribution of NpuC-SUMO to the total fluorescence observed during titrations. For the $NpuN_{WT}$ + $NpuC_{WT}$ titration, $NpuC_{WT}$-SUMO was incrementally added from a 10 μM stock to 600 μL of 50 nM Ub-$NpuN_{WT}$. For the $NpuN_{MUT}$ + $NpuC_{WT}$ titration, $NpuC_{WT}$-SUMO was incrementally added from a 10 μM stock to 600 μL of 300 nM Ub-$NpuN_{MUT}$. For the $NpuN_{WT}$ + $NpuC_{MUT}$ titration, $NpuC_{MUT}$-SUMO was incrementally added from a 30 μM stock to 600 μL of 300 nM Ub-$NpuN_{WT}$. For the $NpuN_{MUT}$ + $NpuC_{MUT}$ titration, $NpuC_{MUT}$-SUMO was incrementally added from a 30 μM stock to 600 μL of 200 nM Ub-$NpuN_{MUT}$.

To determine binding affinities, data were processed as follows: First, emission intensity at 345 nm was adjusted by subtracting the C-intein fluorescence. After background correction, the intensity was corrected for dilution, then all corrected intensities were normalized as a fraction of the initial fluorescence intensity when [NpuC] = 0 nM. Corrected and normalized data from three binding curves were globally fit to the standard quadratic binding equation using GraphPad Prism (Table 1 and Figure S13):

$$F_{obs} = F_{NpuN} + \left(F_{Sat} - F_{NpuN}\right)\frac{\left(\left[NpuC\right] + K_d + \left[NpuN\right]\right) - \sqrt{\left(\left[NpuC\right] + K_d + \left[NpuN\right]\right)^2 - 4\left[NpuN\right]\left[NpuC\right]}}{2\left[NpuN\right]}$$

where [NpuC] is the C-intein concentration in nM (x-values), $F_{obs}$ is the corrected and normalized fluorescence intensity at 345 nm (y-values), $F_{sat}$ is the normalized fluorescence intensity at

saturation (allowed to vary), $F_{NpuN}$ is the normalized intrinsic NpuN fluorescence (allowed to vary, with best fit values always ~1.0), $K_d$ is the binding affinity in nM (allowed to vary), and [NpuN] is the protein concentration in nM (fixed at the appropriate concentration used for each titration).

**Splicing Assays**

For all pair-wise splicing assays, protein concentrations were quantified multiple times by UV abosrbance at 280 nm and by the Bradford assay. Both methods gave consistent and similar results (usually within 10%), and the average value of all measurements were used. For a typical assay, individual protein stock solutions of Ub-NpuN and NpuC-SUMO constructs were prepared in filtered splicing buffer (100 mM phosphate, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.2) at 2x the final concentration (e.g. 2.0 μM stock solution for a 1.0 μM reaction). 1 mM TCEP was added (from a pH-neutralized 100 mM stock solution) to each protein solution, and the proteins were incubated at 30 °C for 5 min. To initiate a reaction, pairs of N- and C-inteins were mixed at equal volumes (i.e. equimolar ratios). Small aliquots of the reaction solution were removed at the desired time points and quenched in a 2x, 4x, or 6x concentrated gel loading dye to afford a final quencher solution with 40 mM Tris (~pH 7.0), 10% (v/v) glycerol, 1% (w/v) SDS, 0.02% (w/v) bromophenol blue, and 2% (v/v) BME. Aliquots of starting materials and time points were loaded onto Bis-Tris gels and run in MES-SDS running buffer. Gels were stained using SYPRO Ruby stain and imaged on a VersaDoc scanner (Figure S14).

To quantify reaction rates, each lane of a gel was analyzed using ImageJ, and the intensity of each band in a lane was expressed as a fractional intensity of the total band intensity (which remained relatively constant between lanes). These normalized intensities were plotted as a function of time and fit to second-order rate equations:

For reactant depletion: $$Y = S \cdot \left( \frac{1}{1 + \left( [Intein]_0 \cdot k_{rxn} \cdot t \right)} \right) + Z$$

For product formation: $$Y = Y_{\max} \cdot \left( 1 - \frac{1}{1 + \left( [Intein]_0 \cdot k_{rxn} \cdot t \right)} \right)$$

$Y$ is the fractional intensity of a species, $t$ is time in minutes, $[Intein]_0$ is the initial intein concentration (fixed based on reaction conditions), S is a scaling factor for reactant depletion (allowed to vary), Z indicates the fraction of product remaining at the reaction endpoint (allowed to vary), $Y_{max}$ is a scaling factor for product formation, and $k_{rxn}$ is the second-order rate constant for the splicing reaction (allowed to vary). Initial half-lives were calculated from the best-fit value for the second-order rate constant:

$$t_{1/2} = \frac{1}{[Intein]_0 \cdot k_{rxn}}$$

For all reactions, individual product formation data sets were normalized based on scaling factors, and 3-4 normalized data sets from different reactions were globally fit to the same equation given above (Table 1 and Figure S15). It is noteworthy that for all reactions, a first-order and second-order fit gave similar half-lives (initial half-lives for second-order fits), however regression analysis with a second-order rate equation consistently gave better fits. Furthermore, the rate of product (Ub-SUMO) and NpuN formation were consistent with the rate of Ub-NpuN and NpuC-SUMO depletion. For all intein fragment pairs that showed splicing activity, reactions plateaued at roughly 60-70% completion and showed no detectable levels of side reactions.

**Competition Assays**

For all competition splicing assays, protein concentrations were quantified multiple times by UV abosrbance at 280 nm and by the Bradford assay. Both methods gave consistent and similar results (within 10%), and the average value of all measurements were used. All competition and control reactions were carried out with all four components at 0.5 μM. For a typical assay, 2.0 μM

protein stock solutions of MBP-NpuN$_{WT}$, Ub-NpuN$_{WT}$, Ub-NpuN$_{MUT}$, NpuC$_{WT}$-eGFP, NpuC$_{WT}$-SUMO, and NpuC$_{MUT}$-SUMO were prepared in filtered splicing buffer (100 mM phosphate, 150 mM NaCl, 1 mM DTT, 1 mM EDTA, pH 7.2). For the 500 mM NaCl reactions, protein stocks were prepared by mixing high-salt splicing buffer (100 mM phosphate, 4.5 M NaCl, 1 mM DTT, 1 mM EDTA, pH 7.2) with regular splicing buffer and the desired protein at appropriate ratios to achieve a final NaCl concentration of 500 mM. 1mM TCEP was added (from a pH-neutralized 100 mM stock solution) to each protein solution, and the proteins were incubated at 30 °C (25 °C or 37 °C for alternate temperature assays) for 5 min. Before a reaction, equal volumes of the N-inteins (MBP-NpuN$_{WT}$ + Ub-NpuN$_{WT/MUT}$) were combined in one tube, and equal volumes of the C-inteins (NpuC$_{WT}$-eGFP + NpuC$_{WT/MUT}$-SUMO) were combined in another tube.

Reactions were initiated by combining equal volumes of N- and C-intein mixtures to afford a final concentration of 0.5 μM for each of the four pieces in the reaction. Control and competition reactions were run in parallel. Small aliquots of the reaction solution were removed at the desired time points and quenched in a 4x concentrated gel loading dye to afford a final quencher solution with 40 mM Tris (~pH 7.0), 10% (v/v) glycerol, 1% (w/v) SDS, 0.02% (w/v) bromophenol blue, and 2% (v/v) BME. Aliquots of starting materials and time points were loaded onto Bis-Tris gels and run in MES-SDS running buffer. The resolved proteins were transferred from the gel onto PVDF membrane in Towbin buffer (25 mM Tris base, 192 mM glycine, pH 8.3, 15 % (v/v) methanol) at 100 V for 90 minutes. Immuno-blotting was performed using an α-HA antibody from Covance (1:10000 dilution) and fluorescent goat α-mouse IRDye 800 secondary antibody from Li-Cor (1:15000 dilution). See Figure S16 for Western blot examples.

Western blot detection and quantification was performed using the Odyssey Infrared Imaging System and the Odyssey v 2.1.10 analysis software from Li-COR. For each blot, the intensity of every band was quantified. To control for variable loading, the total intensity of all

bands in each lane was normalized to match the total intensity of all bands in the last time point lane. The individual intensity of all four products at each time point was then expressed as a fraction of the total product intensity at the last time point. These fractions were plotted as a function of time in stacked bar graphs to depict reaction progress and the extent of formation of each product (Figure 3b and Figure S17).

**Model Three-Piece Ligations**

For all model three-piece ligations, protein concentrations were quantified by the Bradford assay. Prior to each reaction, the starting materials, which were purified in a splicing buffer with no reducing agents (100 mM phosphate, 150 mM NaCl, 1 mM EDTA, pH 7.2), were treated with 5 mM TCEP for 30 minutes at 30 °C. Note that thiols were not present in any starting materials nor in the reaction buffer to prevent the cleavage reactions mentioned in the M-fragment purification protocol above. As an initial test, SH3-NpuN$_{MUT}$ (N), NpuC$_{MUT}$-GB1-NpuN$_{WT}$ (M), and NpuC$_{WT}$-eGFP (C) were mixed in pair-wise combinations alongside a three-piece reaction. Proteins were diluted in splicing buffer and mixed to achieve an N : M : C ratio of roughly 1 eq : 1.2 eq :1 eq (1 eq = 1.75 μM) for the three-piece reaction, and one component was replaced with additional splicing buffer for each of the two-piece reactions. Small aliquots of each reaction solution were removed at the desired time points and quenched in a 4x concentrated gel loading dye to afford a final quencher solution with 40mM Tris (~pH 7.0), 10% (v/v) glycerol, 1% (w/v) SDS, 0.02% (w/v) bromophenol blue, and 2% (v/v) BME. Aliquots of starting materials and time points were loaded side-by-side on a single Bis-Tris gel and run in MES-SDS running buffer. The gel was Coomassie-stained to visualize the resolved proteins (Figure S18).

A larger three-piece reaction (3 mL) was carried out with a more extensive time course using identical starting material ratios and concentrations (Figure S19a). After the reaction, the

remaining reaction solution was diluted two-fold with a binding buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, no BME, pH 8.0) and flowed through a 2 mL Ni-NTA column to trap unreacted starting materials, intermediates, and spliced intein fragments. The column was washed with 6 mL of binding buffer. The flow-through and washes were combined and concentrated 10-fold to ~1 mL. Aliquots of the concentrated solution were desalted by RP-HPLC and analyzed by MALDI-TOF MS to confirm the full-length spliced product (Table S5).

In parallel, a sequential one-pot three-piece ligation was set up to determine if the multi-step reaction could be driven to completion. First, proteins were pre-treated with 5 mM TCEP as in the previous reactions. Next, the M and C fragments were mixed at a 1.2 eq to 1.0 eq ratio and allowed to react at 30°C. After 4 hours, 1.4 eq of the N fragment were added, and the reaction was continued overnight. Time points were quenched in SDS-containing gel loading dye as in the previous reactions then run on a Bis-Tris gel in MES-SDS running buffer. The resolved proteins were visualized by Coomassie staining (Figure S19b).

**PARP1 Three-Piece Ligations**

For all PARP1 ligations, PARP1 fragments were enriched over Ni-NTA resin (see above), and elution fractions from the nickel columns of all three proteins were analyzed on the same Coomassie-stained Bis-Tris gel. Relative concentrations were estimated based on the intensity of product bands in the purest fractions, and these estimates were used to guide the reaction set-up. All reactions were set up in elution buffer (50 mM phosphate, 300 mM NaCl, 250 mM imidazole, no BME, pH 8.0) to achieve a starting material ratio of N : M : C close to 1 eq : 1.5 eq : 1 eq with minimal dilution. Four small-scale reactions (400 μL) were set up in parallel to compare two-piece and three-piece reactions, similarly as with the model system. Starting materials and elution buffer were pre-incubated at 30 °C (without TCEP) for 10 minutes, then the reagents were mixed

together with 1 mM TCEP to initiate the reactions. Aliquots of each reaction solution were removed at the desired time points and quenched in a 4x concentrated gel loading dye to afford a final quencher solution with 40 mM Tris (~pH 7.0), 10% (v/v) glycerol, 1% (w/v) SDS, 0.02% (w/v) bromophenol blue, and 2% (v/v) BME. Aliquots of starting materials and time points were loaded side-by-side on a single Bis-Tris gel and run in MES-SDS running buffer. The resolved proteins were transferred from the gel onto PVDF membrane in CAPS transfer buffer (10 mM *N*-cyclohexyl-3-aminopropanesulfonic acid, 10% (v/v) methanol, pH 10.5) at 100 V for 60 minutes. Immuno-blotting was performed by simultaneously using α-PARP1-N and α-PARP1-C antibodies from Santa Cruz Biotechnology (each at 1:500 dilution) followed by fluorescent goat α-mouse IRDye 700 and goat α-rabbit IRDye 800 secondary antibodies from Li-Cor (each at 1:15000 dilution). The blot was visualized on the Li-Cor infrared imaging system (Figure S20).

A large-scale three-piece ligation of PARP1 (6.5 mL) was set up identically as the small-scale reactions and allowed to proceed for 21 hours. After the allotted time, the reaction mixture was dialyzed into binding buffer (50 mM phosphate, 300 mM NaCl, 5 mM imidazole, no BME, pH 8.0) then incubated with 2 mL of Ni-NTA resin for 30 min at 4 °C to trap unreacted starting materials, intermediates, and spliced intein fragments. The column was washed with 6 mL of binding buffer, then the flow-through and wash fractions were combined and concentrated ~10-fold. The concentrated sample was loaded onto an S200 size exclusion column and eluted in PARP1 buffer (20mM Tris HCl, 200mM NaCl, 1mM DTT, pH 8.0). The reaction and purification progress was analyzed by Western blotting against PARP1-N and PARP1-C (Figure 4c). The purified PARP1 was run along-side commercial PARP1 and BSA standards on a gel, and the gel was Coomassie stained to assess purity and estimate protein concentration (Figure S21). In addition to the clarified three-piece ligation reaction, 0.5 mL of PARP1-C post Ni-NTA column

(Figure S11) was also purified by size exclusion chromatography in PARP1 buffer. Purification of three-piece-ligated PARP1 from 3L of bacterial culture yielded roughly 10μg of pure PARP1.

**MS analysis of Three-Piece Ligated PARP1**

Mass spectrometric analysis of PARP1 was conducted by Dr. J. Fernandez at The Rockefeller University Proteomics Resource Center. Briefly, a Coomassie-stained gel band containing purified three-piece ligated PARP1 was cut out of a gel, treated with iodoacetamide to alkylate cysteine residues, and proteolyzed by trypsin. The tryptic protein fragments were extracted from the gel and analyzed by LC/MS/MS. Observed masses were searched against a database of mammalian protein sequences using MASCOT.[10] The observed peptide ions gave 57% sequence coverage of wild-type human PARP1 (Figure S22a). In addition, ions corresponding to tryptic peptides spanning the ligation junction were observed (Figures S22b and S22d) and selected for fragmentation by tandem MS/MS. The resulting fragment ions verified the amino acid sequences at both splice junctions (Figures S22c and S22e).

**PARP1 Automodification Assays**

For all auto-ADP-ribosylation reactions, 10 ng of PARP1 obtained through three-piece ligation were diluted into PARP1 assay buffer (50 mM Tris, 10 mM $MgCl_2$, 1.25 mM DTT, pH 8.0) and incubated with 5 μM biotinylated NAD at 25°C for one hour. To demonstrate that auto-ADP-ribosylation of semi-synthetic PARP1 was enzymatic, could be stimulated by DNA and histones, and could be inhibited by benzamide, as reported for the native enzyme,[11] reactions were set up in the presence or absence of activated DNA (from Trevigen as 10x stock), 4 μg of histone octamers,[9] and 10 mM benzamide. Reactions were quenched by addition of 4X LDS loading dye (Invitrogen) and immediately separated on a 10% Bis-Tris gel with MES-SDS running buffer.

Proteins were transferred onto a PVDF membrane using a modified Towbin buffer (25 mM Tris base, 192 mM glycine, pH 8.3, 0.08 % (w/v) SDS, 10 % (v/v) MeOH) at 100 V for 30 min. The membrane probed with streptavidin-HRP (1:5000 dilution). The membrane was incubated with ECL reagent and visualized following exposure to film (Figure 4d, Figure S23a). PARP1 loading in each reaction was confirmed by analysis of reaction samples (at time 0) by Western blotting. Briefly, samples were separated on a 10% Bis-Tris gel, transferred into a PVDF membrane using CAPS buffer at 100V for 60 min. Immuno-blotting was performed using the α-PARP1-N antibody followed by α-rabbit-HRP and ECL Plus (Figure 4d). In order to demonstrate that semi-synthetic PARP1 activity was comparable to that of the native enzyme, an analogous set of reactions was set up using a commercial source of PARP1 and analyzed as described above (Figure S23b).

**Histone Poly(ADP-ribosylation) Assays**

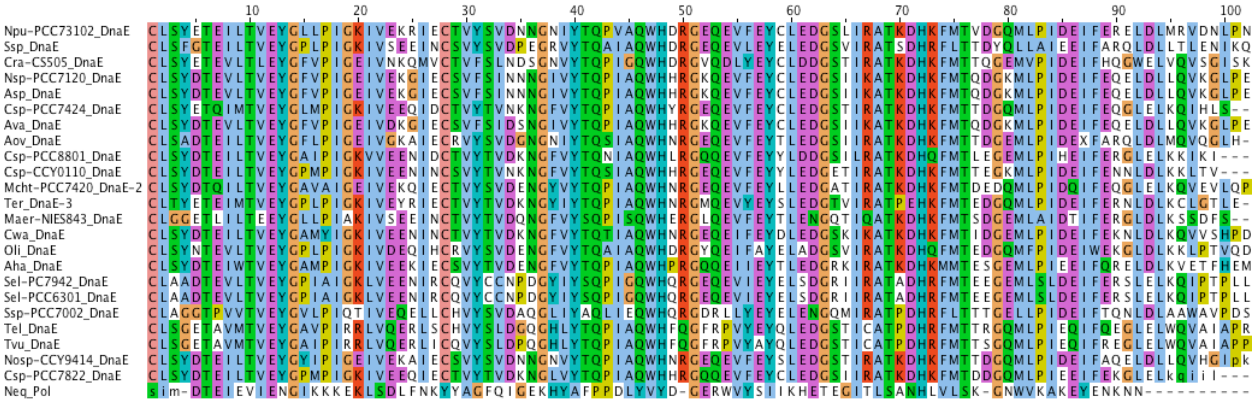
For all histone modification assays, H2A-H2B histone dimers containing 2 μg of H2B[9] were diluted into PARP1 assay buffer and treated with 10 ng of PARP1, 25 μM biotinylated NAD, and 175 μM NAD for one hour at 25°C. To demonstrate that PARP1 activity was stimulated by DNA and inhibited by benzamide reactions were carried out in the presence or absence of activated DNA and 10 mM benzamide. Reactions were quenched by addition of 4X LDS loading dye (Invitrogen) and immediately separated on a 10% Bis-Tris gel with MES-SDS running buffer. Proteins were transferred onto a PVDF membrane and analyzed identically as for the automodification assays (Figure S24a). To demonstrate that histone poly(ADP-ribosylation) by our semi-synthetic PARP1 was not solely due to the presence of the catalytic domain but required the rest of the protein, we set up another group of reactions using purified PARP1-C (Figure S24b). Coomassie staining of membranes was used as a loading control for histones.

## References


[1] F. B. Perler, *Nucleic Acids Res.* **2002**, *30*, 383.
[2] J. Gough, K. Karplus, R. Hughey, C. Chothia, *J. Mol. Biol.* **2001**, *313*, 903.
[3] A. Krogh, M. Brown, I. S. Mian, K. Sjölander, D. Haussler, *J. Mol. Biol.* **1994**, *235*, 1501.
[4] S. R. Eddy, *Bioinformatics* **1998**, *14*, 755.
[5] E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel, A. Bairoch, *The Proteomics Protocols Handbook* **2005**, 571.
[6] G. E. Crooks, G. Hon, J.-M. Chandonia, S. E. Brenner, *Genome Res.* **2004**, *14*, 1188.
[7] S. W. Lockless, T. W. Muir, *Proc. Natl. Acad. Sci. USA* **2009**, *106*, 10999.
[8] A. V. Bryksin, I. Matsumura, *BioTechniques* **2010**, *48*, 463.
[9] P. M. Moyle, T. W. Muir, *J. Am. Chem. Soc.* **2010**, *132*, 15878.
[10] D. N. Perkins, D. J. Pappin, D. M. Creasy, J. S. Cottrell, *Electrophoresis* **1999**, *20*, 3551.
[11] E. Kun, E. Kirsten, J. Mendeleyev, C. P. Ordahl, *Biochemistry* **2004**, *43*, 210.


## Figures and Tables

N-Inteins:




C-Inteins:

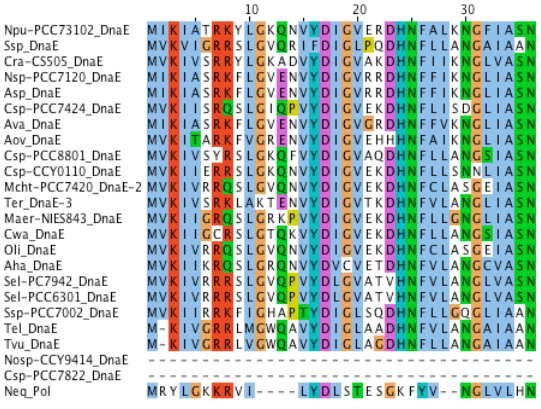



**Figure S1.** Sequence alignment of 24 naturally split inteins from the InBase. Sequences include 23 cyanobacterial DnaE inteins and the Pol intein from *N. equitans*. Note that two C-intein sequences were not available from the InBase.

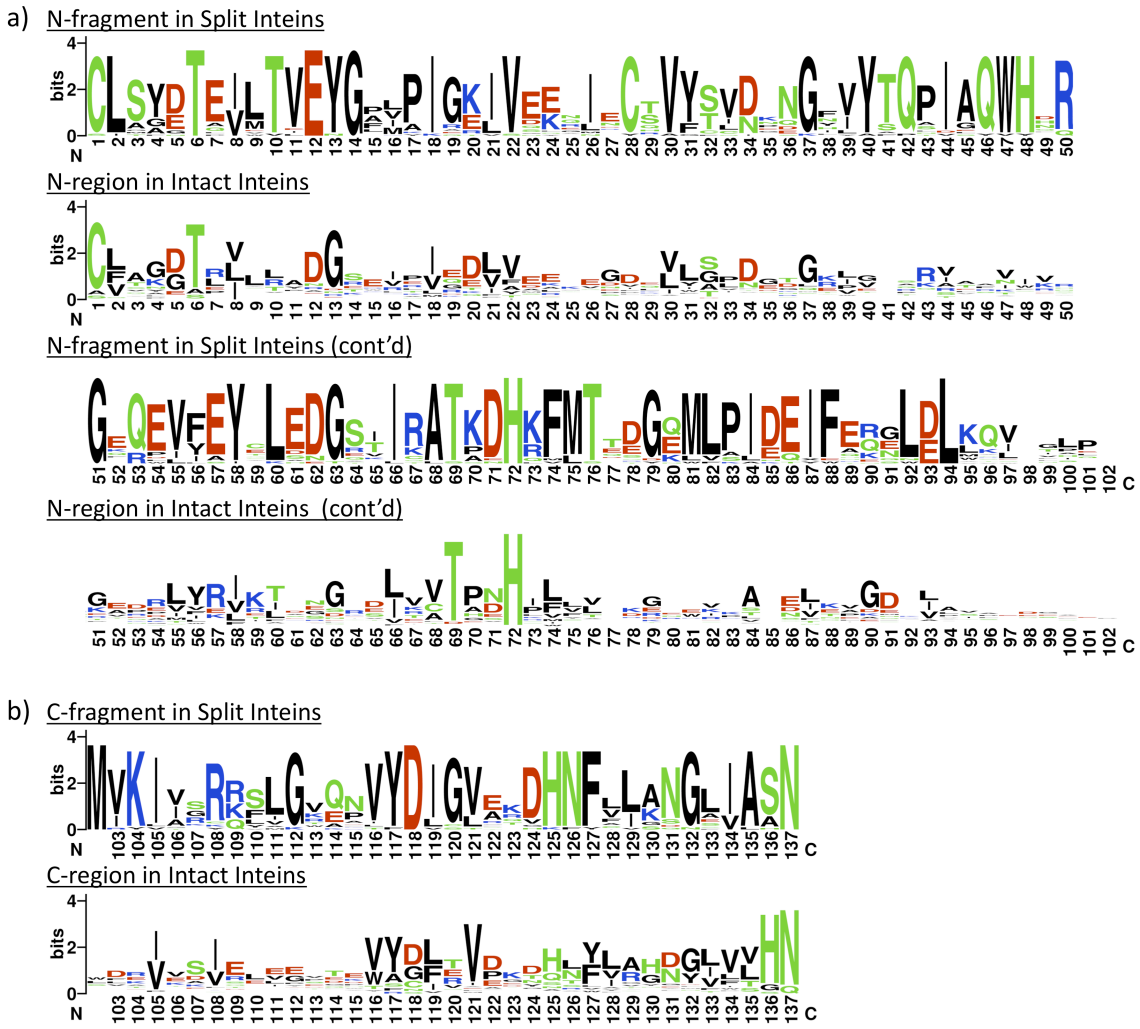



**Figure S2.** Sequence logos of naturally split and intact mini inteins. a) Sequence positions that align with the N-intein of $Npu_{WT}$. b) Sequence positions that align with the C-intein of $Npu_{WT}$.

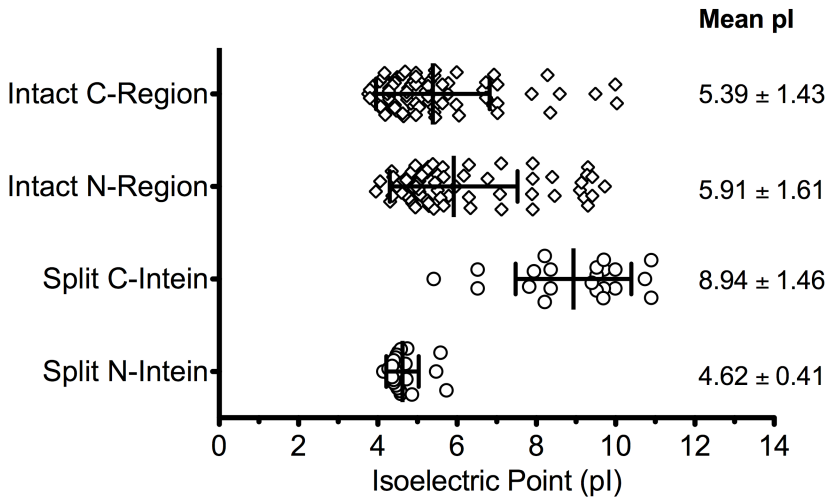



**Figure S3.** Charge distribution in split and intact mini-inteins.

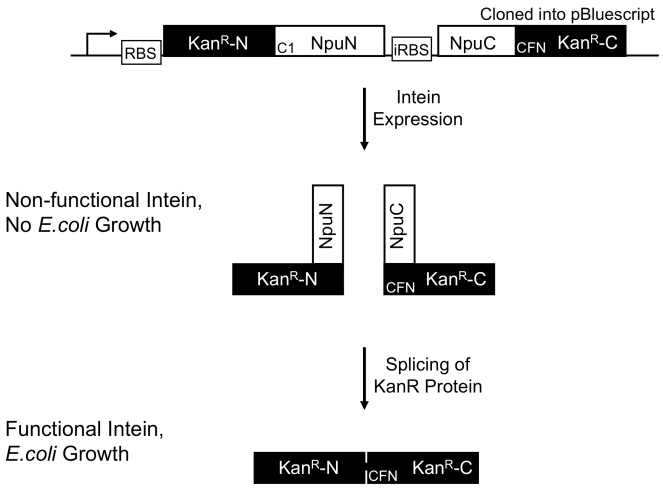



**Figure S4.** Scheme depicting the intein activity-coupled kanamycin resistance (Kan$^R$) assay.

**Table S1.** Sequences of reverse-complement primer pairs used to generate Kan$^R$ assay mutant constructs. Mutated codons are indicated in bold on the sense-strand primer.

| ID # | Primer Sequence 5' → 3' | Mutation |
|---|---|---|
| **1** | TACAAGGGGTGTTATG**GAACAGAAACTGATCAGCGAAGAAGATCTG**AGCCATATTCAACG | myc insertion |
| **2** | CGTTGAATATGGCTCAGATCTTCTTCGCTGATCAGTTTCTGTTCCATAACACCCCTTGTA | myc insertion |
| **3** | TTAAGCTATGAAACG**AAA**ATATTGACAGTAGAA | E7K |
| **4** | TTCTACTGTCAATATTTTCGTTTCATAGCTTAA | E7K |
| **5** | CAATGGCACGATCGCGGA**AAG**CAAGAGGTGTTTGAGTAT | E52K |
| **6** | ATACTCAAACACCTCTTGCTTTCCGCGATCGTGCCATTG | E52K |
| **7** | CACGATCGCGGAGAACAA**AAG**GTGTTTGAGTATTGTTTG | E54K |
| **8** | CAAACAATACTCAAACACCTTTTGTTCTCCGCGATCGTG | E54K |
| **9** | GTGTTTGAGTATTGTTTG**AAG**GATGGTTCATTGATTCGG | E61K |
| **10** | CCGAATCAATGAACCATCCTTCAAACAATACTCAAACAC | E61K |
| **11** | GGTCAAATGTTGCCAATT**AAG**GAAATATTTGAACGTG | D85K |
| **12** | CACGTTCAAATATTTCCTTAATTGGCAACATTTGACC | D85K |
| **13** | CCAATTGATGAAATATTT**CGT**CGTGAATTGGATTTGATG | E89R |
| **14** | CATCAAATCCAATTCACGACGAAATATTTCATCAATTGG | E89R |
| **15** | GATGAAATATTTGAACGT**AAG**TTGGATTTGATGCGGGTT | E91K |
| **16** | AACCCGCATCAAATCCAACTTACGTTCAAATATTTCATC | E91K |
| **17** | CAATGGCACGATCGCGGA**AAG**CAA**AAG**GTGTTTGAGTAT | E52K, E54K |
| **18** | ATACTCAAACACCTTTTGCTTTCCGCGATCGTGCCATTG | E52K, E54K |
| **19** | GGTCAAATGTTGCCAATT**AAG**GAAATATTT**CGT**CGTG | D85K, E89R |
| **20** | CACGACGAAATATTTCCTTAATTGGCAACATTTGACC | D85K, E89R |
| **21** | CAAATGTTGCCAATT**AAG**GAAATATTT**CGT**CGT**AAG**TTGGATTTGATGCGG | D85K, E89R, E91K |
| **22** | CCGCATCAAATCCAACTTACGACGAAATATTTCCTTAATTGGCAACATTTG | D85K, E89R, E91K |
| **23** | ACAAGGGGTGTTATGATC**GAA**ATAGCCACACGTAAATAT | K104E |
| **24** | ATATTTACGTGTGGCTATTTCGATCATAACACCCCTTGT | K104E |
| **25** | ATGATCAAAATAGCCACA**GAA**AAATATTTAGGCAAACAA | R108E |
| **26** | TTGTTTGCCTAAATATTTTTCTGTGGCTATTTTGATCAT | R108E |
| **27** | ACACGTAAATATTTAGGC**GAA**CAAAATGTCTATGACATT | K113E |
| **28** | AATGTCATAGACATTTTGTTCGCCTAAATATTTACGTGT | K113E |

| | | |
|---|---|---|
| **29** | CATAATTTTGCACTC**GAA**AATGGCTTCATAGCT | K130E |
| **30** | AGCTATGAAGCCATTTTCGAGTGCAAAATTATG | K130E |
| **31** | AGGGGTGTTATGATC**GAA**ATAGCCACA**GAA**AAATATTTAGGCAAA | K104E, R108E |
| **32** | TTTGCCTAAATATTTTTCTGTGGCTATTTCGATCATAACACCCCT | K104E, R108E |
| **33** | ATCAAAATAGCCACA**GAA**AAATATTTAGGC**GAA**CAAAATGTCTATGAC | R108E, K113E |
| **34** | GTCATAGACATTTTGTTCGCCTAAATATTTTTCTGTGGCTATTTTGAT | R108E, K113E |
| **35** | AGGGGTGTTATGATC**GAA**ATAGCCACA**GAA**AAATATTTAGGC**GAA**CAAAATGTCTATGAC | K104E, R108E, K113E |
| **36** | GTCATAGACATTTTGTTCGCCTAAATATTTTTCTGTGGCTATTTCGATCATAACACCCCT | K104E, R108E, K113E |

**Table S2.** N- and C-intein mutation combinations analyzed *in vivo*. For each intein, three variants were tested: an N-intein mutant, a C-intein mutant, and a charge-swapped intein.

| | **Ion Cluster 1** | | | **Ion Cluster 2** | | | **Ion Cluster 3** | | | **Ion Cluster 4** | |
|---|---|---|---|---|---|---|---|---|---|---|---|
| | **N-Intein** | | **C-intein** | **N-Intein** | | **C-intein** | **N-Intein** | | **C-intein** | **N-Intein** | **C-intein** |
| **Intein** | E52K | E54K | K113E | D85K | E89R | R108E | E61K | E91K | K104E | E7K | K130E |
| **3a** | X | | X | | | | | | | | |
| **3b** | | X | X | | | | | | | | |
| **3c** | X | X | X | | | | | | | | |
| **4a** | | | | X | | X | | | | | |
| **4b** | | | | | X | X | | | | | |
| **4c** | | | | X | X | X | | | | | |
| **5a** | | | | | | | X | | X | | |
| **5b** | | | | | | | | X | X | | |
| **5c** | | | | | | | X | X | X | | |
| **6** | X | X | X | X | X | X | X | X | X | | |
| **12** | | | | | | | | | | X | X |
| **13** [a] | X | X | X | X | X | X | X | X | X | X | X |
| **14** | X | X | X | X | X | X | | | | | |
| **15** | X | X | X | | | | X | X | X | | |
| **16** | | | | X | X | X | X | X | X | | |
| **26** | X | X | X | X | X | X | | | | X | X |
| **27** | X | X | X | | | | X | X | X | X | X |
| **28** | | | | X | X | X | X | X | X | X | X |

a) Note that intein **13** (shown in red) is equivalent to $Npu_{MUT}$ described in the main text.

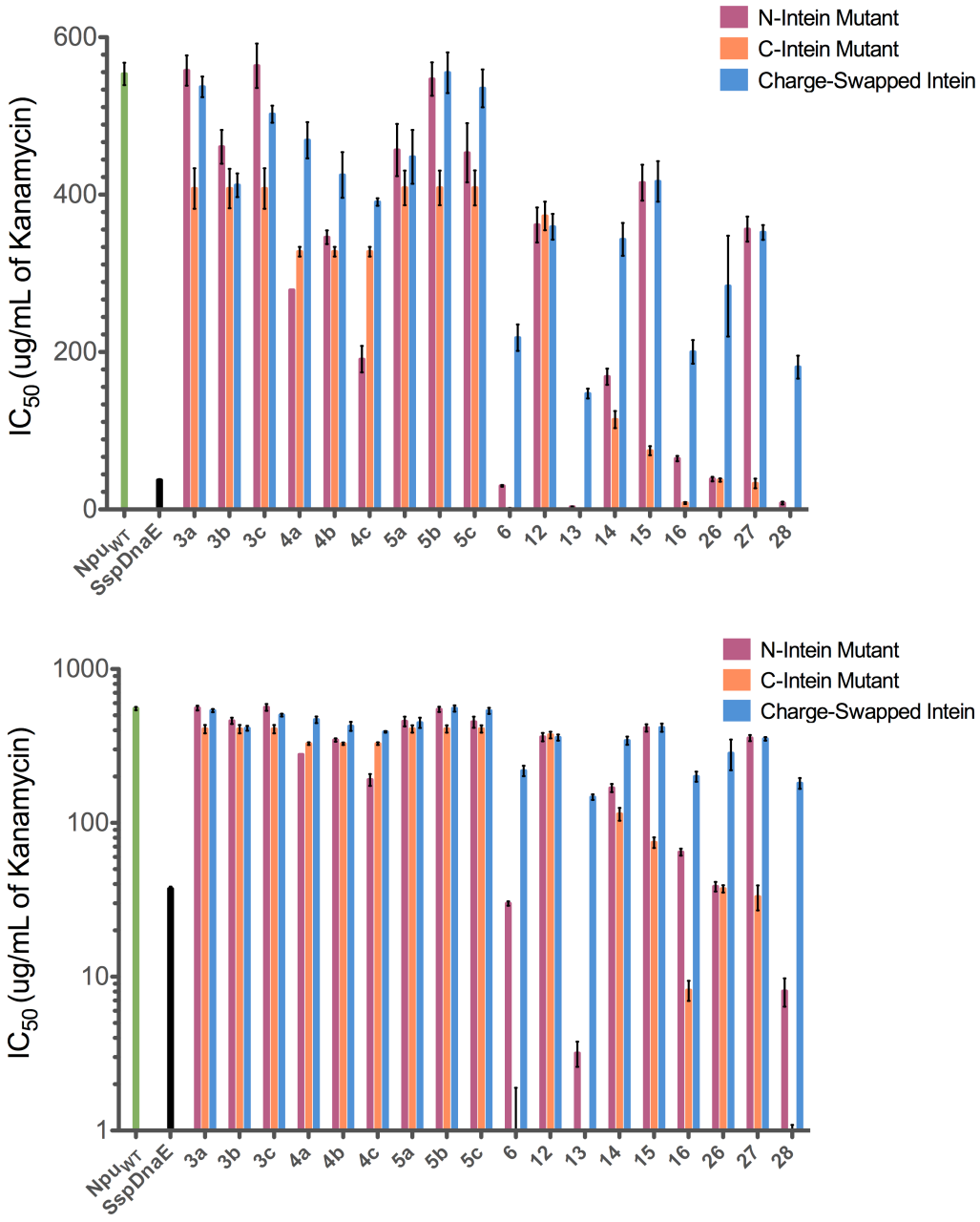



**Figure S5.** Kanamycin resistance IC50 values for all inteins analyzed *in vivo* (shown on a linear scale, above, and log scale, below). IC50 values were determined from a best fit of three to five dose response curves to a standard dose-response equation with variable slope. Error bars represent the standard error of the fit.

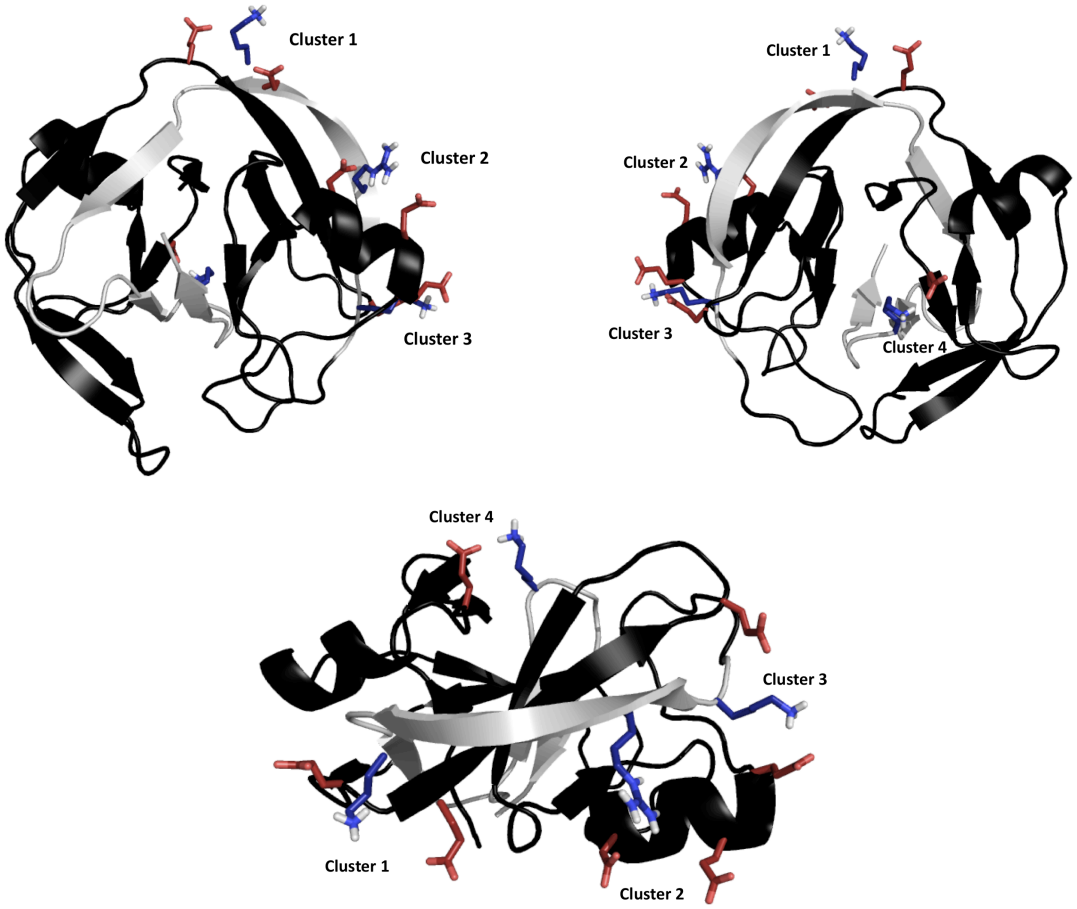



**Figure S6**. Ion clusters mutated in $Npu_{MUT}$ (intein 13 in Table S2 and Figure S3) shown on the NMR structure of $Npu_{WT}$ (PDB code 2KEQ).

**Table S3.** Sequences of primers used to generate constructs for *in vitro* assays.

| ID # | Primer Sequence 5' → 3' |
|---|---|
| **37** | AAGTTTGCGGAATAT**GCG**TTAAGCTATGAAACG |
| **38** | CGTTTCATAGCTTAACGCATATTCCGCAAACTT |
| **39** | ATCATCATCATCATCACGGC**GAGAACCTGTATTTTCAGGGC**ATCAAAATAGCCACACGTA |
| **40** | GCCGCGCGGCACCAGGCCGCTGTTGAAGCAATTAGAAGCTAT |
| **41** | GAGAACAGATTGGTGGT**TATCCGTATGATGTGCCGGATTATGCG**TGAAGACAAGCTTAGG |
| **42** | CCTAAGCTTGTCTTCACGCATAATCCGGCACATCATACGGATAACCACCAATCTGTTCTC |
| **43** | TATTTTCAGGGCATC**GAA**ATAGCCACA**GAA**AAATATTTAGGC**GAA**CAAAATGTCTATGAC |
| **44** | GTCATAGACATTTTGTTCGCCTAAATATTTTTCTGTGGCTATTTCGATGCCCTGAAAATA |
| **45** | GGAGATATACATATGGGCAGCAGCCATCACCATCATCACCACAGCCAGGATCCG |
| **46** | CCGCAAACTTGCCGCCGCCACCATTAGTCTGCGCGTCTTTCAGGGC |
| **47** | GCCCTGAAAGACGCGCAGACTAATGGTGGCGGCGGCAAGTTTGCGG |
| **48** | CGAAATAAAATCCGCGGTAAGCTTTAATTCGGCAAATTATC |
| **49** | TAAGAAGGAGATATACATATGGGCAGCAGCCATCATC |
| **50** | CAGCTCCTCGCCCTTGCTCACCATGCTAGCGCTGCCGCGCGGCACCAGGCC |
| **51** | GGCCTGGTGCCGCGCGGCAGCGCTAGCATGGTGAGCAAGGGCGAGGAGCTG |
| **52** | CCTAAGCTTGTCTTCACGCATAATCCGGCACATCATACGGATACTTGTACAGCTCGTCC |

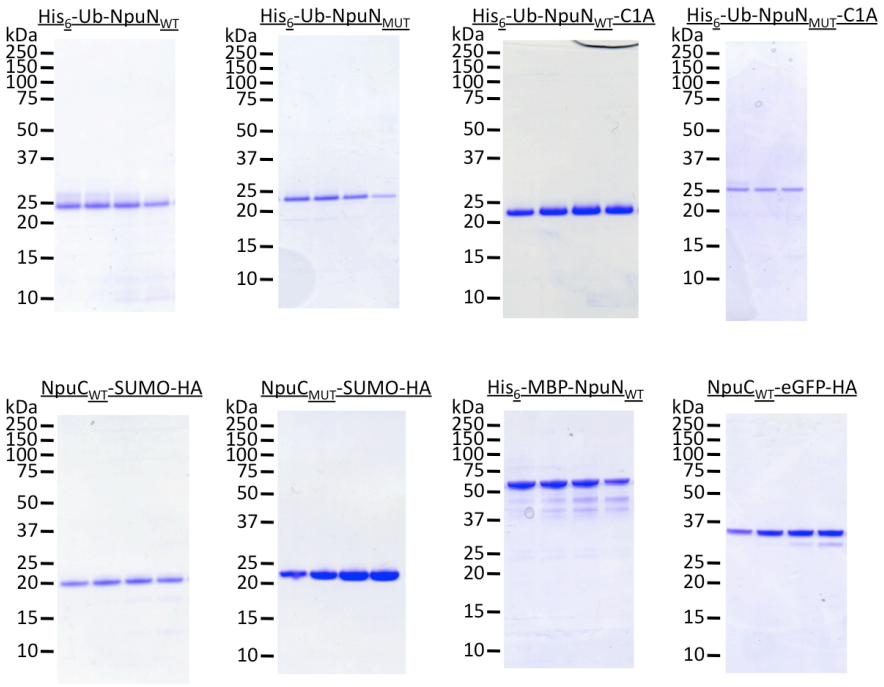



**Figure S7.** Pure size exclusion fractions of proteins used for splicing, binding, and competition assays, analyzed by SDS-PAGE and Coomassie staining.

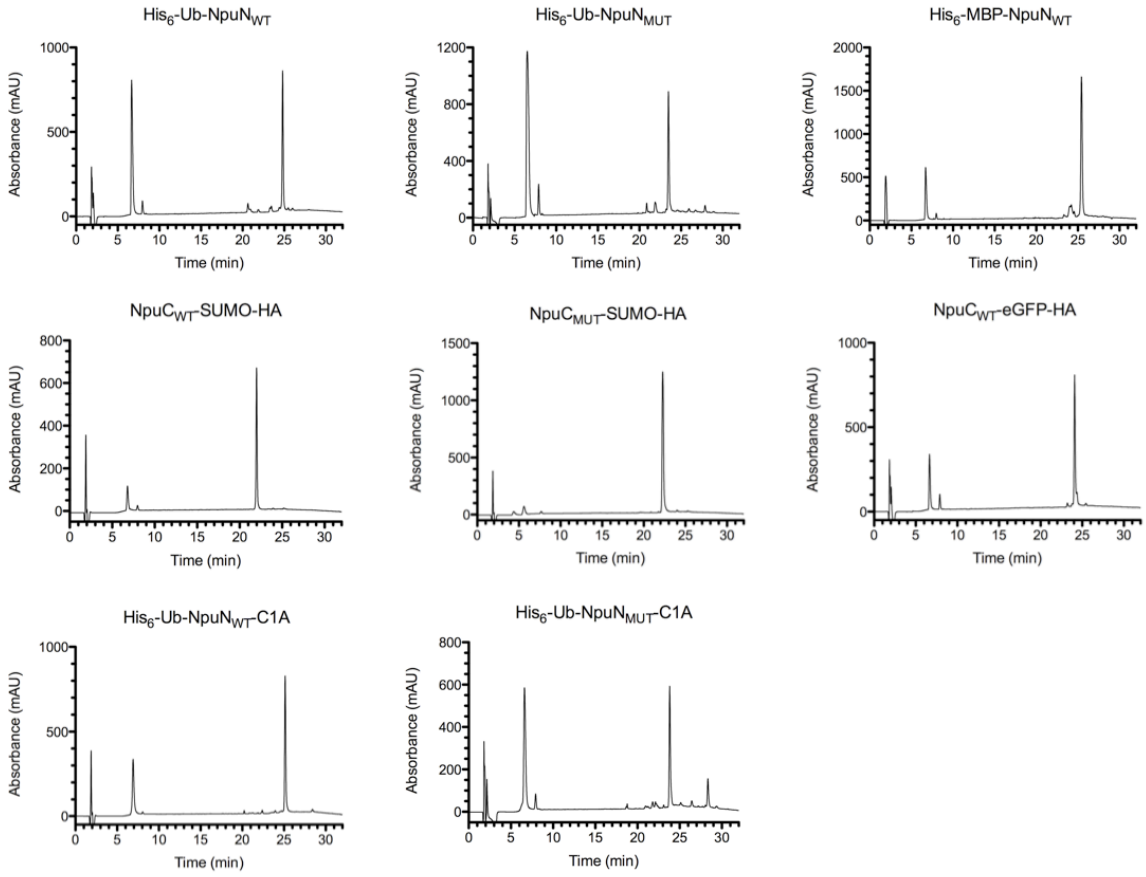



**Figure S8.** Analytical HPLC chromatographs of purified proteins for splicing, binding, and competition assays. Absorbance was monitored at 214 nm. Note that the two peaks near 6.5 min and 8 min correspond to reduced and oxidized DTT in the pure protein samples.

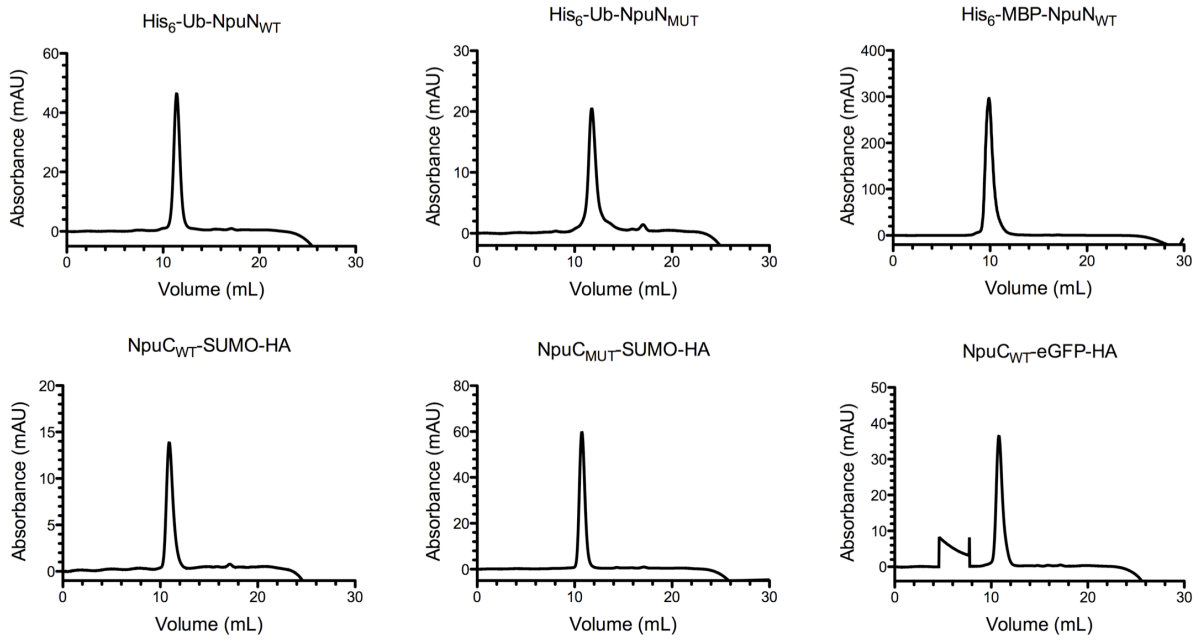



**Figure S9.** Analytical size-exclusion chromatography of purified proteins for splicing and competition assays. Note that the irregular absorbance in the $NpuC_{WT}$-eGFP-HA chromatograph between 4 and 8 minutes is due to a pressure spike. All proteins were run on a 25 mL S75 column.

**Table S4.** Sequences of primers used to generate constructs for three-piece ligations.

| ID # | Primer Sequence 5' → 3' |
|---|---|
| **53** | TCGGCGCGCCTGCAGGTCGACAAGCTTGCGGCCGCATGTTTAAGCTATGAAACG |
| **54** | TGTTCGACTTAAGCATTAATTCGGCAAATTATCAAC |
| **55** | CACCATCATCACCACAGCGGCGGCATCAAAATAGCCACACGT |
| **56** | CACCATCATCACCACAGCGGCGGCATCGAAATAGCCACAGAA |
| **57** | GACCTGCAGGCGCGCCGAGCTCGAATTCGGATCCTGGTTGAAGCAATTAGAAGC |
| **58** | AACCTGTATTTTCAGGGCATGGAAGCCATCGCCAAA |
| **59** | CGTTTCATAGCTTAAACACATTTCTATGTAGTTCTT |
| **60** | TTAATAAGGAGATATACCATGGAAGCCATCGCCAAA |
| **61** | TTTGGCGATGGCTTCCATGGTATATCTCCTTATTAA |
| **62** | GCACTCGAAAATGGCTTCATAGCTTCTAATTGCTTCAACTACAAGCTTATCCTGAACGG |
| **63** | CCATATTCTACTGTCAATATTTCCGTTTCATAGCTTAAACATTCGGTTACCGTGAAGGTTT |
| **64** | ACCTTCACGGTAACCGAATATTGTTTAAGCTATGAAACG |
| **65** | CGTTTCATAGCTTAAACAATATTCGGTTACCGTGAAGGT |
| **66** | GTTGATAATTTGCCGAATGGCAGCAGCCACCATCACCACCATCATAGCGGCTAATGCTTAAGTCGAACAGAA |
| **67** | TTCTGTTCGACTTAAGCATTAGCCGCTATGATGGTGGTGATGGTGGCTGCTGCCATTCGGCAAATTATCAAC |
| **68** | CCTGTATTTTCAGGGCATGGCGGAGTCTTCGG |
| **69** | CGTTTCATAGCTTAAACAGGCGCTGGTTTCTGGG |
| **70** | TTAATAAGGAGATATACCATGGCGGAGTCTTCGGAT |
| **71** | ATCCGAAGACTCCGCCATGGTATATCTCCTTATTAA |
| **72** | GGCTTCATAGCTTCTAATTGCGTGGCGGCCACGCCTCCG |
| **73** | CCGTTTCATAGCTTAAACACAGCTTCTTCACTGCCTC |
| **74** | GGCTTCATAGCTTCTAATTGCGTAAATCCTGGCACC |
| **75** | CTGTTCGACTTAAGCATTACCACAGGGAGGTCTTAAAATTG |

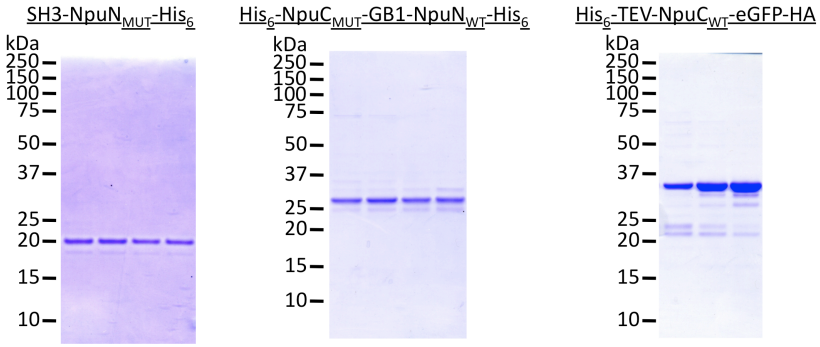



**Figure S10.** Pure size exclusion fractions of proteins used for the model three-piece ligation, analyzed by SDS-PAGE and Coomassie staining.

**Table S5.** Expected and observed masses of proteins.

| **Purified Protein** [a] | **Expected Mass (Da)** | **Observed Mass (Da)** |
|---|---|---|
| His$_6$-Ub-NpuN$_{WT}$ | 22303.1 | 22301.9 |
| His$_6$-Ub-NpuN$_{MUT}$ | 22338.6 | 22337.4 |
| His$_6$-Ub-NpuN$_{WT}$-C1A | 22271.1 | 22272.3 |
| His$_6$-Ub-NpuN$_{MUT}$C1A | 22306.6 | 22307.0 |
| NpuC$_{WT}$-SUMO-HA | 17655.8 | 17656.4 |
| NpuC$_{MUT}$-SUMO-HA | 17631.6 | 17635.3 |
| His$_6$-MBP-NpuN$_{WT}$ | 54509.6 | 54526.0 |
| NpuC$_{WT}$-eGFP-HA | 33317.6 | 33321.8 |
| SH3-NpuN$_{MUT}$-His$_6$ | 19493.5 | 19495.7 |
| His$_6$-NpuC$_{MUT}$-GB1-NpuN$_{WT}$-His$_6$ | 24736.5 | 24739.2 |
| His$_6$-TEV-NpuC$_{WT}$-eGFP-HA | 35224.9 | 35220.2 |
| **Three-Piece Ligated Protein** [b] | **Expected Mass (Da)** | **Observed Mass (Da)** |
| SH3-GB1-eGFP-HA | 42160.7 | 42131.7 |

a) All purified proteins were analyzed by ESI-MS. b) The three-piece ligated protein was analyzed by MALDI-TOF MS.

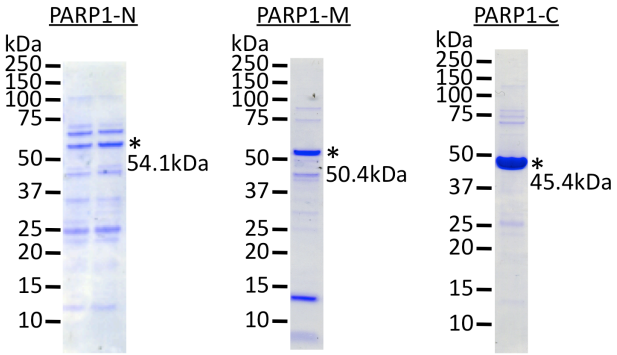



**Figure S11.** Semi-pure nickel column fractions of proteins used for the PARP1 three-piece ligation, analyzed by SDS-PAGE and Coomassie staining. Bands marked by an asterisk correspond to the desired protein product with the given molecular weight.

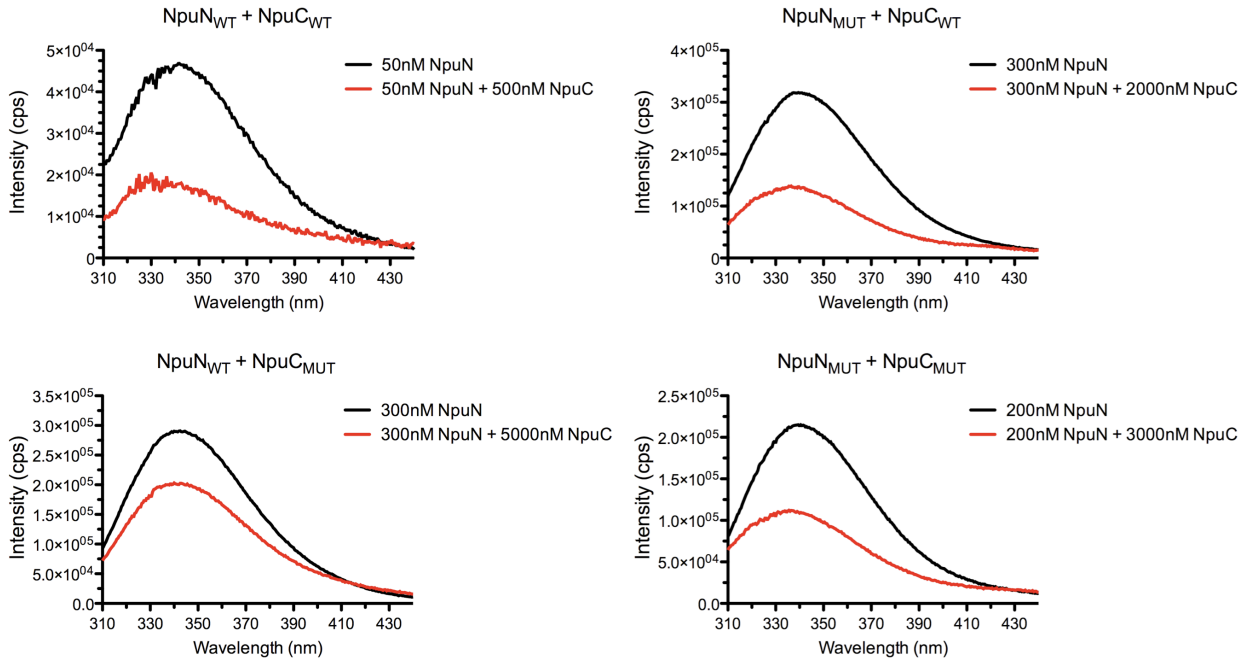



**Figure S12.** Fluorescence emission spectra ($\lambda_{ex}$ = 295nm) of $His_6$-Ub-NpuN-C1A constructs in the absence and presence of NpuC-SUMO-HA constructs. The background fluorescence of NpuC-SUMO-HA is subtracted.

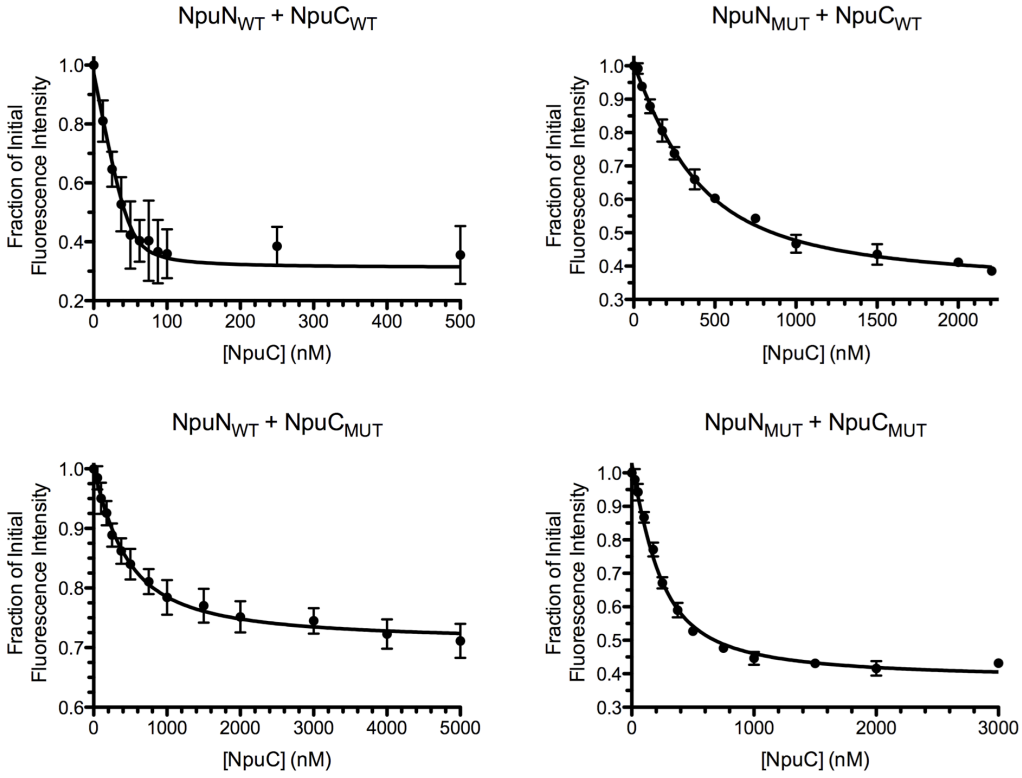



**Figure S13**. Binding curves for $Npu_{WT}$ and $Npu_{MUT}$ fragment combinations.

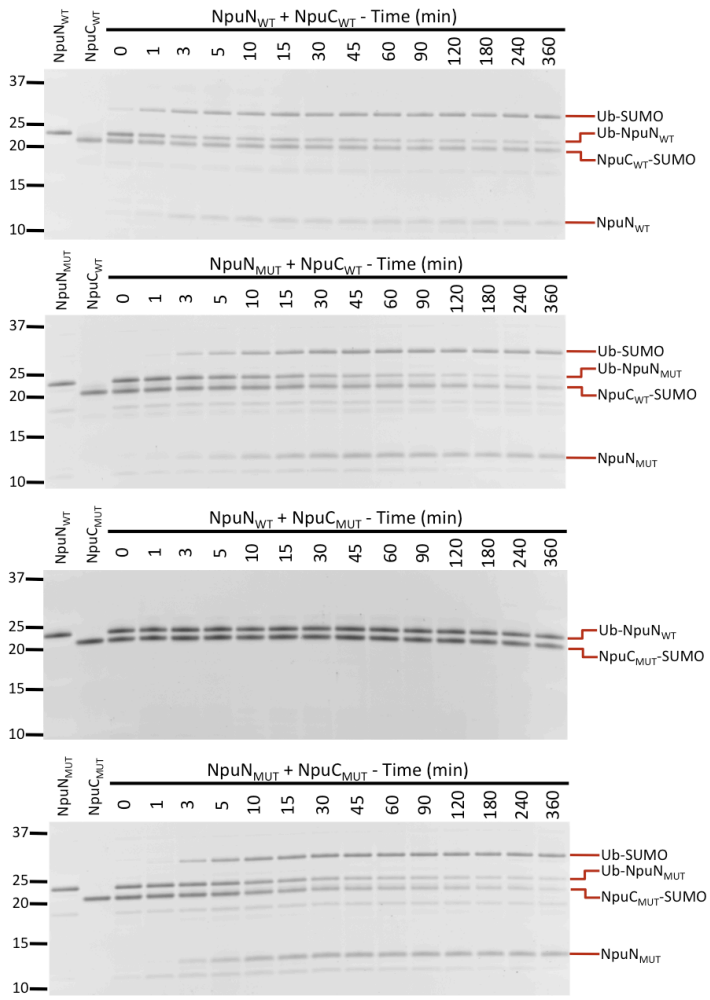



**Figure S14.** SYPRO Ruby-stained gels from 1.0 μM splicing assays, representative of all splicing assays performed.

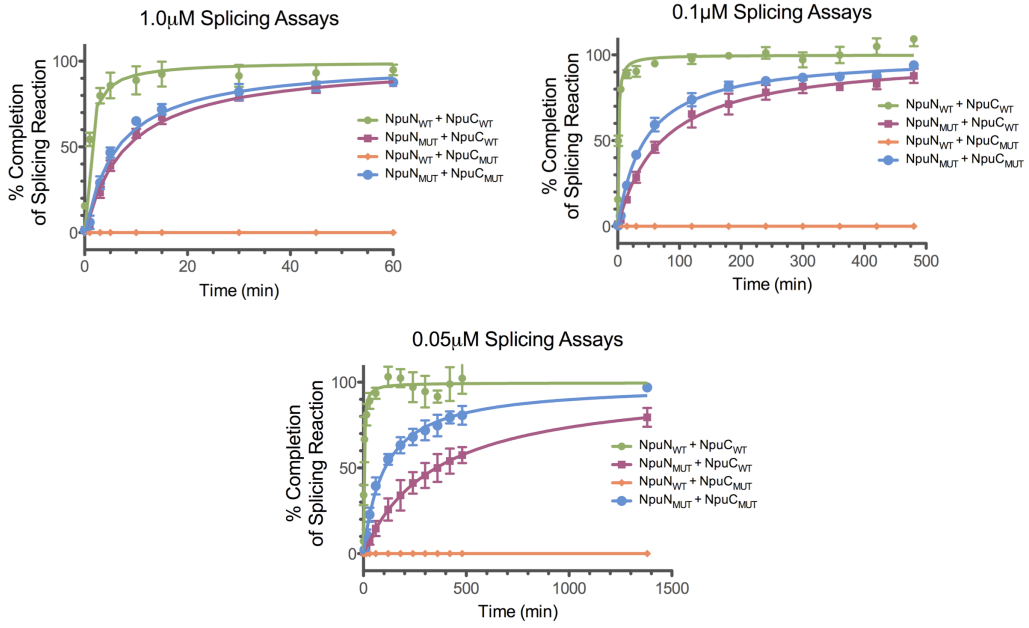



**Figure S15.** Splicing kinetics for Npu$_{WT}$ and Npu$_{MUT}$ fragment combinations.

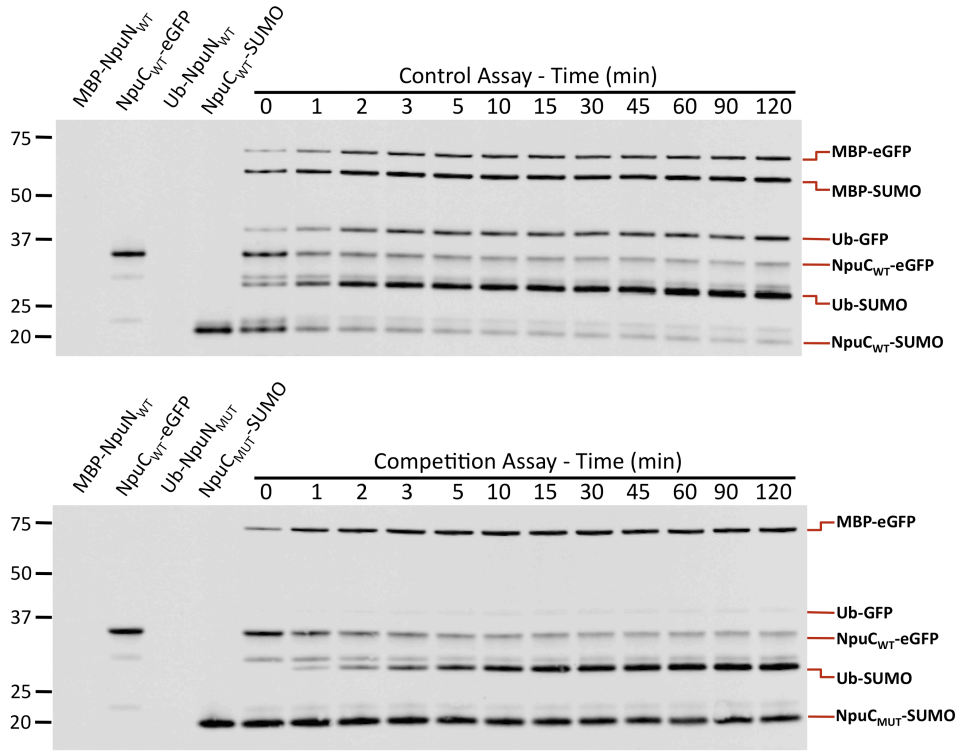



**Figure S16.** Representative Western blots (α-HA) of competition assays at 30 °C and 150 mM NaCl.

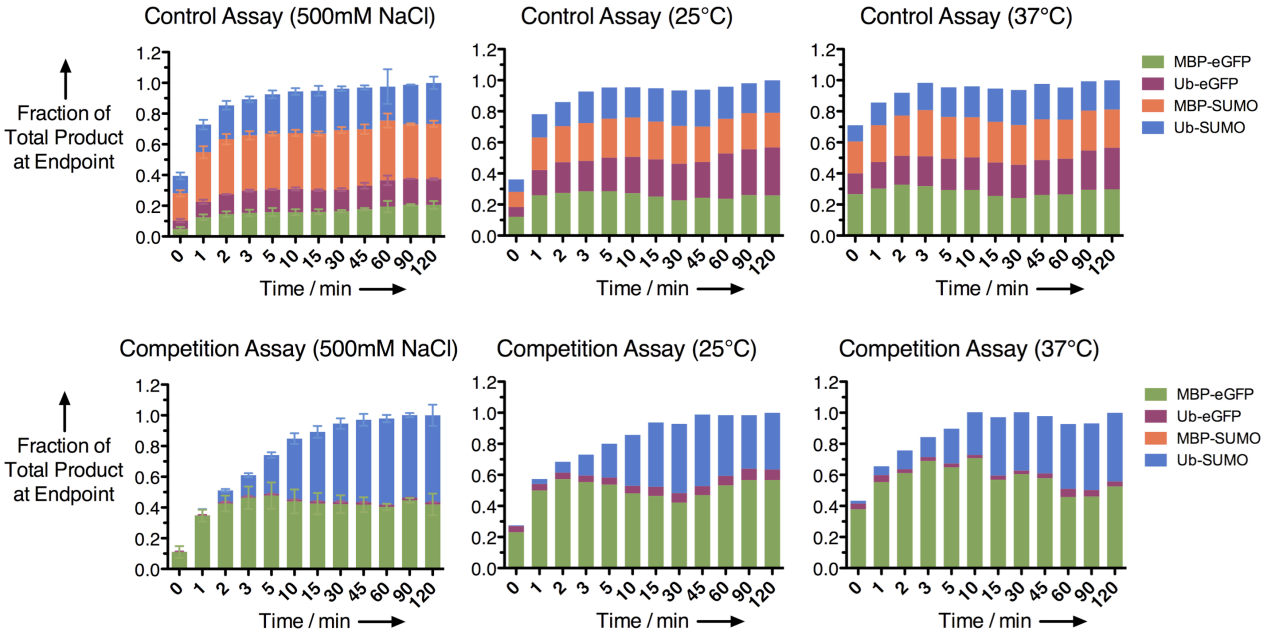



**Figure S17.** Graphs of competition assays at 500 mM NaCl, 25 °C, and 37 °C.

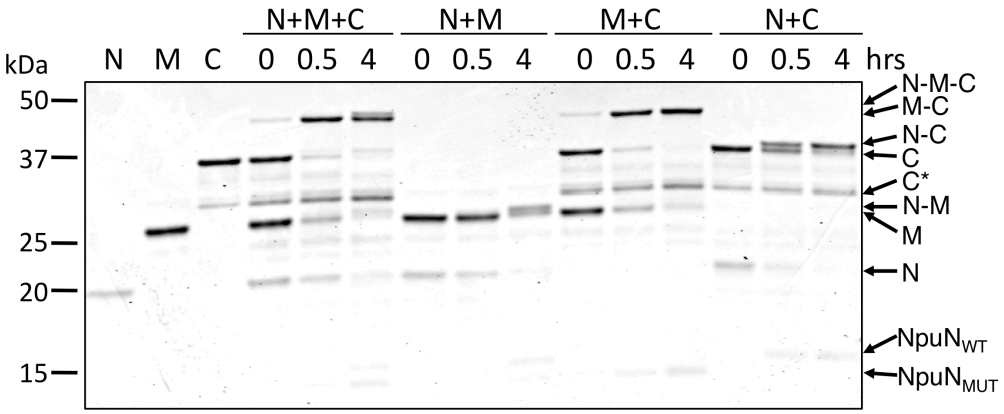



**Figure S18.** Three-piece and pair-wise ligation reactions for the SH3-GB1-eGFP model system. Note that the N-fragment stains poorly in the gel.

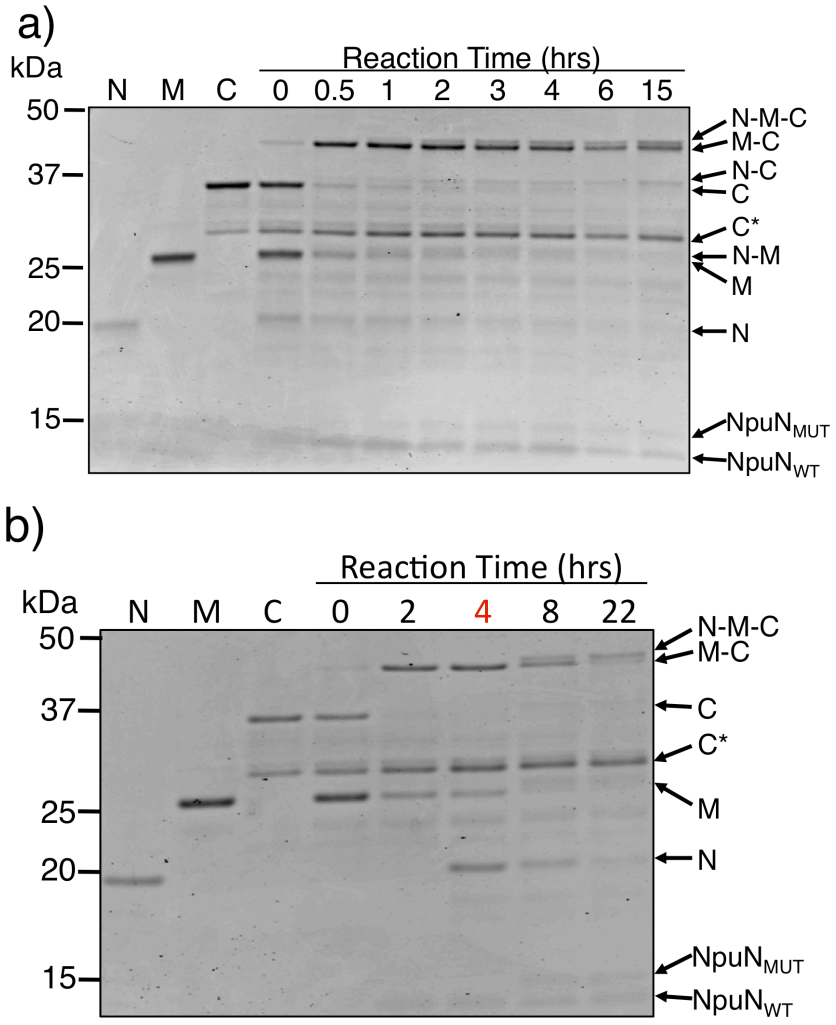



**Figure S19.** Three-piece ligations of model system proteins over an extended time course. a) 15-hour ligation mixing all three components simultaneously with ratios of N:M:C = 1:1.2:1 eq. b) Sequential ligation in which the M- and C-fragments were allowed to react in isolation for 4 hours (M:C = 1.2:1eq), followed by addition of N (1.4eq). Note that the N-fragment stains poorly in the gel.

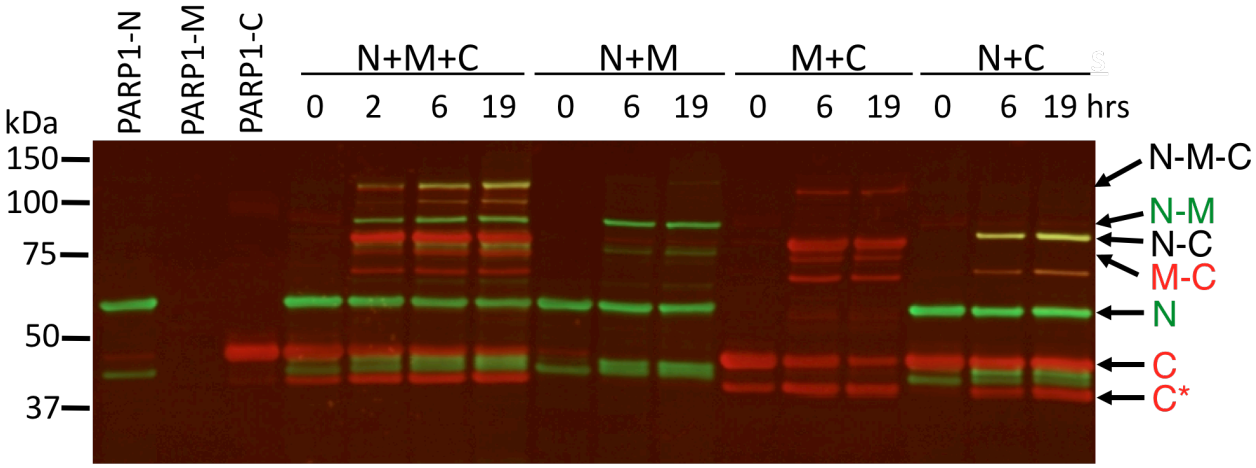



**Figure S20.** Three-piece and pair-wise reactions of PARP1 fragments analyzed by dual-antibody western blotting. Color scheme: green bands (α-PARP1-N), red bands (α-PARP1-C), and yellow bands (α-PARP1-N + α-PARP1-C). C* corresponds to C-extein cleavage of PARP1-C to yield free PARP1 catalytic domain. All other bands correspond to impurities from the fragment preparations or products of reactions with those impurities.

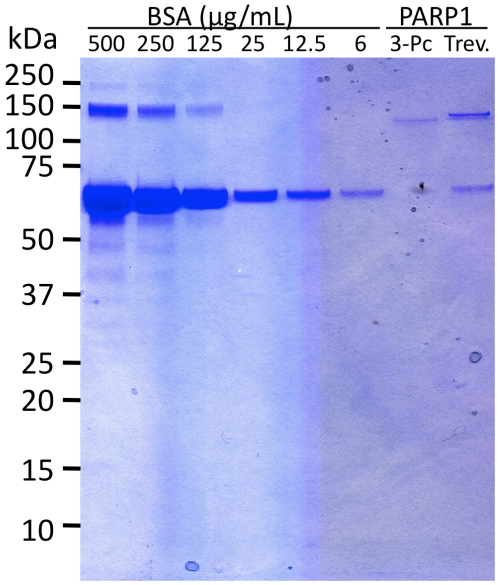
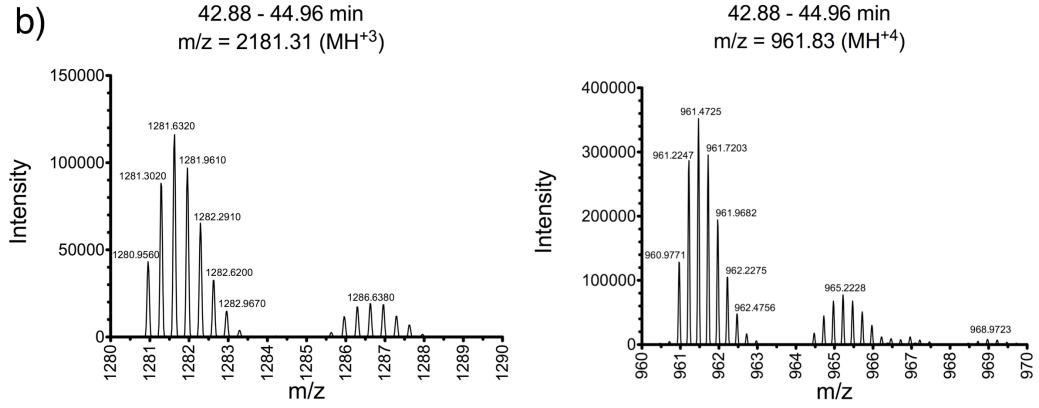



**Figure S21.** Analysis of purified PARP1 by gel electrophoresis and Coomassie staining. The three-piece ligated PARP1 (3-pc.) is shown along-side bovine serum albumin (BSA) concentration standards and a commercially available PARP1 sample from Trevigen (Trev.). Note that the Trevigen PARP1 runs slightly higher than our 3-pc. PARP1 because it is tagged for purification. The commercial sample also contains BSA.

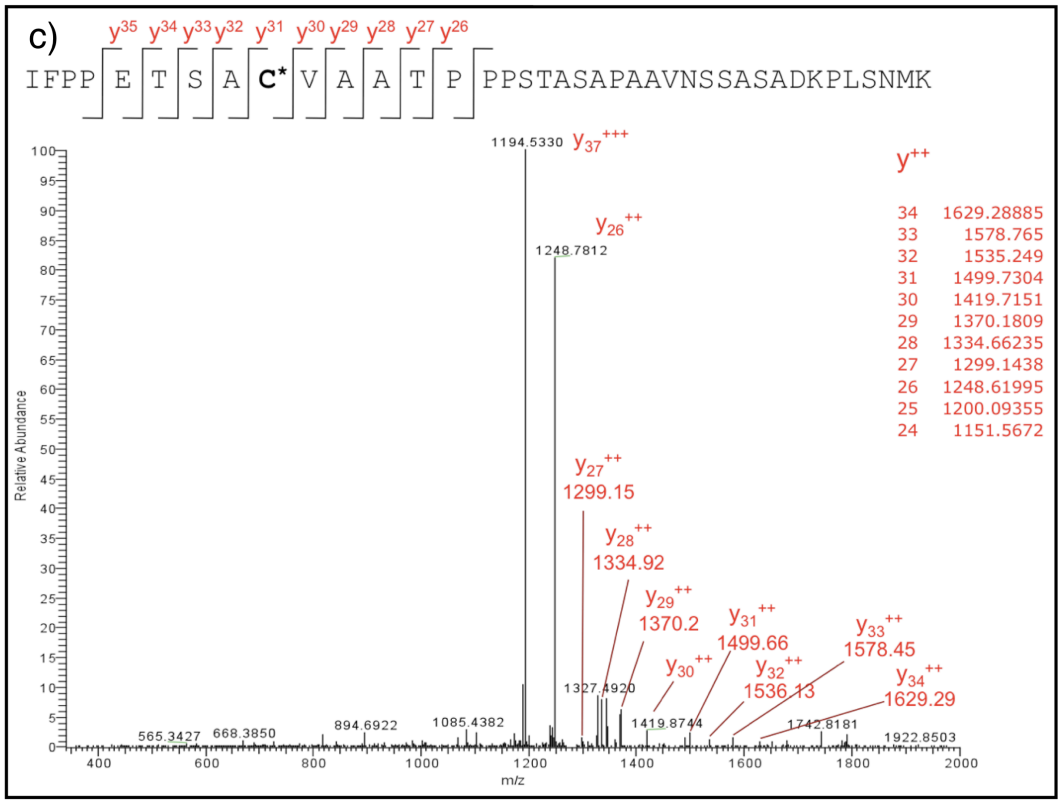
a)

```
  1 MAESSDKLYR VEYAKSGRAS CKKCSESIPK DSLRMAIMVQ SPMFDGKVPH
 51 WYHFSCFWKV GHSIRHPDVE VDGFSELRWD DQQKVKKTAE AGGVTGKGQD
101 GIGSKAEKTL GDFAAEYAKS NRSTCKGCME KIEKGQVRLS KKMVDPEKPQ
151 LGMIDRWYHP GCFVKNREEL GFRPEYSASQ LKGFSLLATE DKEALKKQLP
201 GVKSEGKRKG DEVDGVDEVA KKKSKKEKDK DSKLEKALKA QNDLIWNIKD
251 ELKKVCSTND LKELLIFNKQ QVPSGESAIL DRVADGMVFG ALLPCEECSG
301 QLVFKSDAYY CTGDVTAWTK CMVKTQTPNR KEWVTPKEFR EISYLKKLKV
351 KKQDRIFPPE TSACVAATPP PSTASAPAAV NSSASADKPL SNMKILTLGK
401 LSRNKDEVKA MIEKLGGKLT GTANKASLCI STKKEVEKMN KKMEEVKEAN
451 IRVVSEDFLQ DVSASTKSLQ ELFLAHILSP WGAEVKAEPV EVVAPRGKSG
501 AALSKKSKGQ VKEEGINKSE KRMKLTLKGG AAVDPDSGLE HSAHVLEKGG
551 KVFSATLGLV DIVKGTNSYY KLQLLEDDKE NRYWIFRSWG RVGTVIGSNK
601 LEQMPSKEDA IEHFMKLYEE KTGNAWHSKN FTKYPKKFYP LEIDYGQDEE
651 AVKKLCVNPG TKSKLPKPVQ DLIKMIFDVE SMKKAMVEYE IDLQKMPLGK
701 LSKRQIQAAY SILSEVQQAV SQGSSDSQIL DLSNRFYTLI PHDFGMKKPP
751 LLNNADSVQA KVEMLDNLLD IEVAYSLLRG GSDDSSKDPI DVNYEKLKTD
801 IKVVDRDSEE AEIIRKYVKN THATTHNAYD LEVIDIFKIE REGECQRYKP
851 FKQLHNRRLL WHGSRTTNFA GILSQGLRIA PPEAPVTGYM FGKGIYFADM
901 VSKSANYCHT SQGDPIGLIL LGEVALGNMY ELKHASHISK LPKGKHSVKG
951 LGKTTPDPSA NISLDGVDVP LGTGISSGVN DTSLLYNEYI VYDIAQVNLK
1001 YLLKLKFNFK TSLW
```

b)




c)



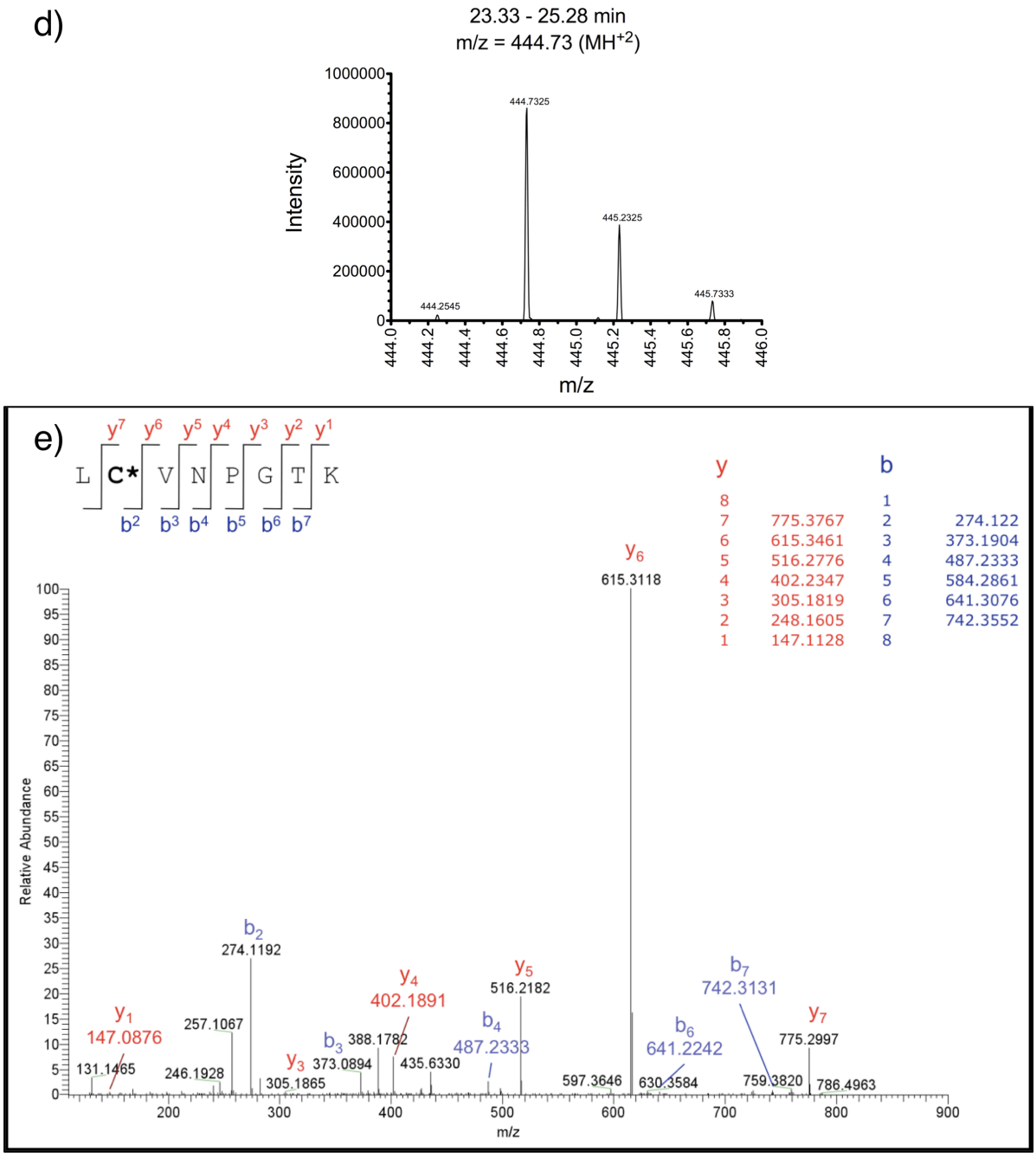



**Figure S22**. MS analysis of three-piece-ligated PARP1. a) Sequence coverage of tryptic peptides identified by MS. Observed peptides are shown in red. Sequences in bold black correspond to tryptic peptides spanning the ligation junctions. b) Parent ions observed for peptide 1, spanning the first ligation site. c) MS/MS analysis of peptide 1, verifying the sequence surrounding the ligation site. d) Parent ion observed for peptide 2, spanning the second ligation site. e) MS/MS analysis of peptide 2, verifying the sequence surrounding the ligation site.

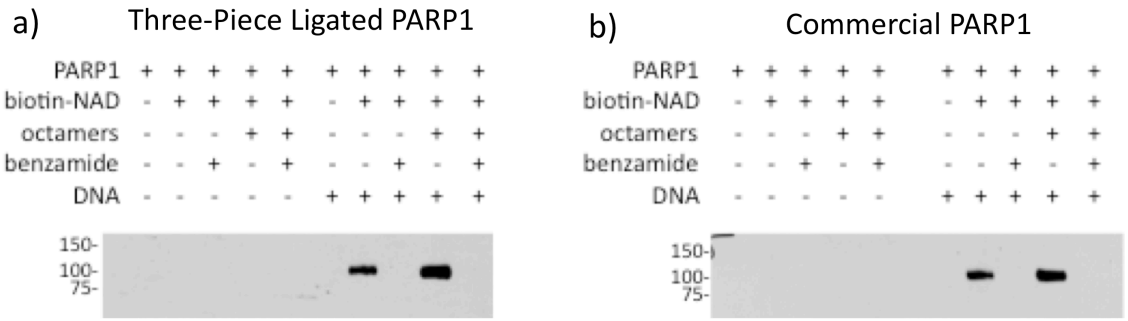



**Figure S23.** Comparison of automodification for commercial and three-piece ligated PARP1 by blotting against biotinylated-NAD with streptavidin-HRP. a) Three-piece ligated PARP1. b) Commercially available PARP1 from Trevigen.

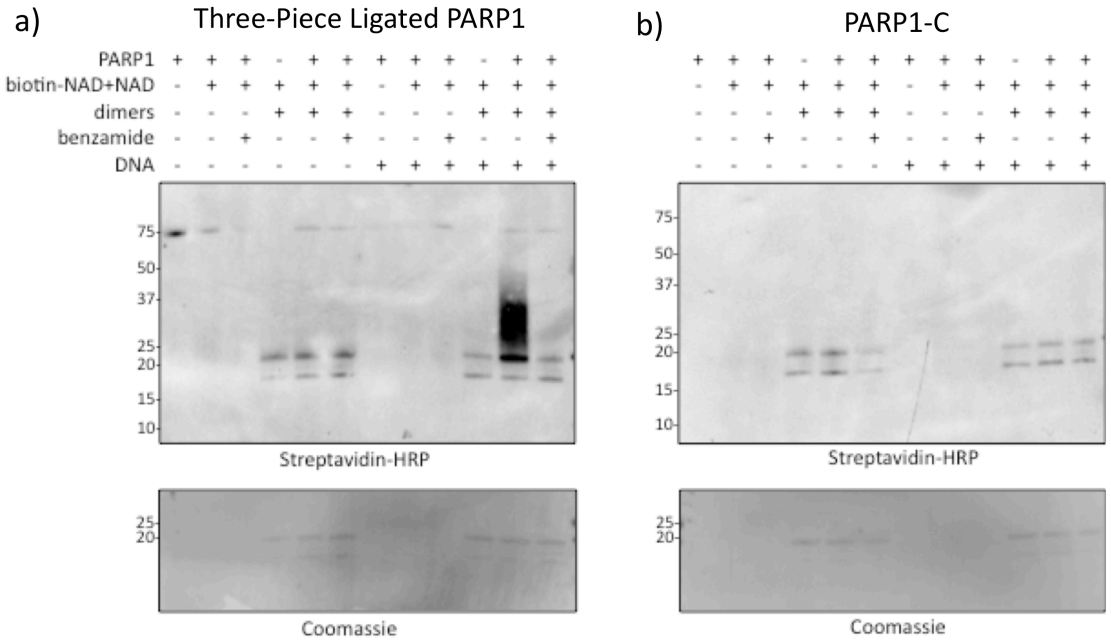



**Figure S24.** The formation of poly-(ADP-ribose) chains (PARylation) on histones by three-piece ligated PARP1. a) PARylation by three-piece ligated PARP1. b) Control showing lack of PARylation with only the PARP1 C-terminal fragment. PARylation is detected by blotting against biotinylated-NAD with streptavidin-HRP. Histone loading is shown in the Coomassie-stained membranes post-blotting.