Supporting information for

Live-cell protein engineering with an ultra-short split intein

Antony J. Burton[#,1], Michael Haugbro[#,1], Eva Parisi[1] & Tom W. Muir[*,1]

[1] Frick Chemistry Laboratory, Princeton, NJ 08544, USA.

# these authors contributed equally to this work

*to whom correspondence should be addressed: muir@princeton.edu

www.pnas.org/cgi/doi/10.1073/pnas.2003613117

Table of Contents

SI Figures

SI Tables

Appendices

1. General materials and methods

Dimethylformamide (DMF), dichloromethane (DCM), diallyl N,N-diisopropylphosphoramidite, tert-butyl hydroperoxide, 5-(methylthio)-1H-tetrazole, Boc-Gly-OH, and Fmoc-Ser-OH were obtained from Sigma-Aldrich (St. Louis, MO). Fmoc amino acids were obtained from Matrix Innovations (Quebec City, Canada) or Novabiochem (Darmstadt, Germany). Fmoc-Lys(Biotin)-OH was obtained from CreoSalus (Louisville, KY). Rink Amide and Trityl ChemMatrix resins were obtained from Biotage (Charlotte, NC). Trifluoroacetic acid (TFA) was obtained from Halocarbon (North Augusta, SC). (7-azabenzotriazol-1-yloxy)tripyrrolidinophosphonium hexafluorophosphate (PyAOP) was obtained from Oakwood (Estill, SC). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) and isopropyl-β-thiogalactopyranoside (IPTG) were obtained from Gold Biotechnology (St. Louis, MO). All other chemicals, unless otherwise indicated, were obtained from Sigma-Aldrich (St. Louis, MO).

Oligonucleotide primers were obtained from Integrated DNA Technologies (Coralville, IA) or Sigma-Aldrich (St. Louis, MO). PrimeSTAR HS DNA Polymerase was purchased from Takara Bio (Kusatsu, Japan). PfuUltra II Fusion HS DNA Polymerase was obtained from Agilent (Santa Clara, CA). Gibson Assembly Master Mix was obtained from New England Biolabs (Ipswich, MA). DNA purification was carried out with kits obtained from Qiagen (Valencia, CA). Plasmid sequencing was performed by GENEWIZ (South Plainfield, NJ). One Shot Bl21 (DE3) chemically competent *E. Coli* cells and Hoechst 33342 solution were obtained from Invitrogen (Carlsbad, CA). *E. coli* B834 (DE3) cells were a kind gift from the lab of Prof. Fred Hughson, Princeton University. SelenoMethionine Medium Complete kit was obtained from Molecular Dimensions, Inc. (Holland, OH). Nickel-nitrilotriacetic acid (Ni-NTA) resin was obtained from Thermo Fisher Scientific (Waltham, MA). MES-SDS running buffer was obtained from Boston Bioproducts (Ashland, MA). Criterion Cassettes, Acrylamide, Immun-blot PVDF membrane (0.2 μm), APS, and TEMED were obtained from Bio-Rad (Hercules, CA).

Dulbecco's Modified Eagle Medium (DMEM) with 4.5g/L glucose, 110 mg/L pyruvate, Dulbecco's phosphate-buffered saline (DPBS) without calcium, magnesium, or phenol red, L-Glutamine (200 mM), Penicillin-Streptomycin (5,000 U/mL), Trypsin-EDTA (0.05%) with phenol red, Opti-MEM Reduced Serum Medium, Falcon Standard Tissue Culture Dishes, and Lipofectamine 2000 Transfection Reagent were obtained from Thermo Fisher Scientific (Waltham, MA). Bovine Serum Albumin (BSA) was obtained from GoldBio (St. Louis, MO). Fetal Bovine Serum (Heat Inactivated) was obtained from Atlanta Biologicals (Flowery Branch, GA). EZQuant Cell Quantifying Kit was obtained from Alstem (Richmond, CA). Poly-L-Lysine Solution was obtained from Sigma-Aldrich (St. Louis, MO). Paraformaldehyde (20% solution) was obtained

from electron microscopy sciences (Hatfield, PA). 35 mm glass-bottom dishes were obtained from MatTek (Ashland, MA). Streptavidin-Alexa Fluor 488 was obtained from BioLegend (San Diego, CA). Electroporation was conducted using a BTX ECM 830 Square Wave Electroporation System (Holliston, MA). 2 mm Gap Sterile Electroporation Cuvettes were obtained from Universal Medical (Oldsmar, FL). HEK293T cells (ATCC CRL-3216) were obtained from ATCC (Manassas, VA).

2. Peptide synthesis

All peptides used in this study were synthesized by standard Fmoc solid-phase peptide synthesis methods on a CEM Liberty Blue microwave-assisted peptide synthesizer. Peptides were synthesized on a 0.1 mmol scale on H-Rink Amide resin (ChemMatrix) using DIC-Oxyma activation. Fmoc deprotection was affected with 20% v/v piperidine in DMF with 0.1 M HOBt. Where necessary, alloc-protected lysine side chains were deprotected on resin using 0.1 eq. $Pd(PPh_3)_4$ and 5 eq. 1,3-dimethylbarbituric acid in DCM for 3 x 20 min with nitrogen bubbling. The resin was then washed with DCM, DMF and 5% w/v sodium diethyldithiocarbamate in DMF to remove excess Pd. 5,6-Carboxyfluorescein (3 eq.) was coupled to orthogonally deprotected lysine side chains or the peptide N-terminus using DIC-HOAt (2.95 eq. each) activation in DMF (2 x 4 h treatments). Fmoc-Lys(Biotin)-OH (3 eq.) was coupled using PyAOP (2.95 eq.) activation with DIEA (6 eq.) in NMP for 2 x 4 h. Fmoc-Lys(Me,Boc)-OH and Fmoc-Lys(Ac)-OH (2.5 eq.) were coupled using PyAOP (2.4 eq.) activation with DIEA (5 eq.) in DMF for 1 x 4 h. Desthiobiotin (5 eq.) was coupled to the peptide N-terminus using PyAOP (4.95 eq.) activation with DIEA (10 eq.) in DMF for 1 x 4 h.

Side-chain deprotection and cleavage from the resin was affected by addition of a 92.5:2.5:2.5:2.5 v/v solution of TFA:TIPS:ethanedithiol:H_2O for 70 minutes. The cleavage solution was reduced to <5 ml volume under a positive flow of N_2, and the crude peptide was precipitated using cold diethyl ether. The crude product was isolated by refrigerated centrifugation, dissolved in 50:50 v/v H_2O:MeCN with 0.1% TFA, and lyophilized.

Phosphorylated desthiobiotin-HP1$\alpha_{1\text{-}18}$-(S13ph)-Vid^N was prepared according to a protocol adapted from previously published methods (1). HP1$\alpha_{1\text{-}18}$-SG-VidaL^N was synthesized via standard Fmoc methods as described above, with the exception of the incorporation of unprotected Fmoc-Ser-OH at position 13 and coupling of desthiobiotin to the deprotected peptide N-terminus. The resin was then treated with 43 eq. 5-(methylthio)-1H-tetrazole and 10 eq. diallyl N,N-diisopropylphosphoramidite in DMF at rt for 2 x 2 h treatments. The resin was then washed with DMF and treated with 40 eq. tert-butyl hydroperoxide in DMF at rt for 1 h. The resin was washed with DMF and DCM and cleaved as described previously.

3. HPLC

Semi-preparative scale reversed-phase high-pressure liquid chromatography (RP-HPLC) was performed on Agilent 1200 series instruments equipped with a Waters XBridge BEH C18 column (5 μm, 10 x 250 mm) at a flow rate of 4 ml min^{-1}. The mobile phase comprised 0.1% v/v trifluoroacetic acid in water (solvent A) and 90% acetonitrile, 0.1% TFA in water (solvent B). Crude peptides were purified on a 0-60% B gradient at 50 °C over 45 min. Analytical scale (RP-HPLC) was performed on a Vydac 218ms C18 column (5 μM particle size, 4.6 x 150 mm dimension) at a flow rate of 1 ml min^{-1}. Peptides were analyzed on a 0-70% B gradient over 30 min. Pure fractions were identified by analytical RP-HPLC and ESI-MS, and pooled and lyophilized.

4. Mass spectrometry

Electrospray ionization mass spectrometry (ESI-MS) was performed on Bruker Daltonics micrOTOF-Q II mass spectrometer. Masses quoted are for the monoisotopic mass as the singly protonated species.

5. Cloning of DNA plasmids

Sequences for the constructs used in this study are provided in ***Appendix 1***.

Bacterial expression

$VidaL^C$ constructs for *E. coli* expression were cloned into pET vectors containing H6-SUMO expression tags by Gibson Assembly (NEB) following the manufacturer's instructions. An *E. coli* codon-optimized G-block containing VidaL intein was purchased from IDT. Where necessary, point mutations were introduced by QuikChange protocols, following the manufacturer's instructions.

Mammalian expression

$VidaL^C$ constructs for expression in mammalian cell lines were cloned into CMV vectors by Gibson Assembly following the manufacturer's instructions. The plasmid containing Alk2 (pCMV5-Alk2) was a gift from Jeff Wrana (Addgene 11741).

Kanamycin resistance assay

A pTXB1 plasmid containing $KanR^N$-Vid^N and Vid^C-$KanR^C$ in different reading frames separated by an internal ribosomal binding site was constructed in two steps. $VidaL^C$ was inserted by Gibson Assembly, and secondly $VidaL^N$ was inserted by inverse PCR using primers optimized for bacterial expression. Each of the extein point mutations for the -2, -1 and +2 positions, along with a C1A negative control mutation, were introduced by QuikChange protocols following the manufacturer's instructions.

6. Protein expression and purification

Expression of VidaL fusion proteins *E. coli* BL21 (DE3) cells were transformed with the appropriate H6-SUMO-Vid^{C} fusion construct and grown at 37 °C to an $OD_{600\ nm} \approx 0.6$ in LB broth with 50 μg mL^{-1} kanamycin. The temperature was lowered to 18 °C and expression was induced by the addition of IPTG (0.5 mM final concentration). After 18 h, the cells were harvested by centrifugation (4000 *g*, 15 min) and stored at -80 °C.

E. coli pellets were resuspended in lysis buffer (50 mM phosphate, 500 mM NaCl, 5 mM imidazole, 1 mM PMSF, 2 mM DTT, 10% v/v glycerol, pH 7.8) and lysed by sonication on ice (30% amplitude, 20 sec on / 30 sec off, total time = 200 sec on). The soluble fraction was isolated by centrifugation (30,000 *g*, 35 min, 4 °C) and incubated for 1 h at 4 °C with Ni^{2+}-NTA resin (2 mL per liter; pre-incubated with lysis buffer). The beads were washed with lysis buffer (50 mL), lysis buffer + 15 mM imidazole (50 mL) and eluted with lysis buffer + 295 mM imidazole (12 mL). The elution fraction was treated with Ulp1 protease (to cleave the H6-SUMO purification tag) and simultaneously dialyzed into lysis buffer overnight at 4 °C. The protein was then added to Ni^{2+}-NTA resin (2 mL per liter; pre-incubated with lysis buffer) and incubated for 1 h at 4 °C. The flow-through was collected and the beads washed with 5 ml lysis buffer to isolate the target protein.

The crude protein was concentrated to <2 mL and further purified by size-exclusion chromatography (SEC). SEC was performed on an AKTA FPLC system (GE Healthcare) equipped with an S75 16/60 Superdex column using a degassed buffer containing 100 mM phosphate, 150 mM NaCl, 1 mM EDTA, 1 mM DTT, 10% v/v glycerol, at pH 7.2. Pure fractions were identified by SDS-PAGE (10% bis-tris gel), and confirmed by ESI-MS. The purified protein was flash frozen in liquid N_2 and stored in single-use aliquots at -80 °C.

To obtain the selenomethionine-substituted VidaL construct for crystallographic studies, the plasmid encoding full-length VidaL with C-A and N-A mutations was transformed into E. coli B834 (DE3) cells. The SelenoMethionine Medium Complete kit was used according to the manufacturer's protocol. Briefly, a starter culture in LB was inoculated in L-methionine-containing minimal medium and allowed grown overnight at 37 °C. The cells were then pelleted by centrifugation and washed with sterile PBS 3 times. The cells were resuspended and inoculated into L-selenomethionine-containing minimal medium. The remaining steps of expression and purification were carried out as described above.

Expression of histone proteins *E. coli* BL21 (DE3) cells were transformed with plasmids encoding wild-type histones or macroH2A1.$1_{1\text{-}119}$-Vid^{N}-H6 and grown at 37 °C to an $OD_{600\ nm}$ ≈ 0.6 in LB broth with appropriate antibiotics. Expression was induced by the addition of IPTG (0.5 mM final concentration) and after 3 h, the cells were harvested by centrifugation (4000 *g*, 15 min) and stored at -80 °C.

E. coli pellets were resuspended in lysis buffer (50 mM tris, 300 mM NaCl, 1 mM EDTA, 5 mM β-mercaptoethanol, and 1 mM PMSF, pH 7.5) and the insoluble lysate was collected by centrifugation (30,000 *g*, 35 min, 4 °C). The pellet was washed twice with lysis buffer + 1% v/v Triton X-100 and centrifuged as before. The pellet was washed again with lysis buffer and centrifuged as before. Histone proteins were extracted from the pellet in a buffer containing 50 mM tris, 300 mM NaCl, 6 M guanidinium chloride, 5 mM β-mercaptoethanol, pH 7.5) at room temperature for 2 h, followed by centrifugation. For macroH2A1.$1_{1\text{-}119}$-Vid^{N}-H6, the supernatant was added to Ni^{2+}-NTA resin (2 mL per liter; pre-incubated with extraction buffer) and incubated for 1 h at 4 °C. The beads were washed with extraction buffer (50 mL), extraction buffer + 15 mM imidazole (50 mL) and eluted with extraction buffer + 295 mM imidazole (12 mL). The isolated histones were filtered through a 0.45 μM filter and purified by preparative scale RP-HPLC. Pure fractions were identified by ESI-MS and analytical RP-HPLC, pooled, and lyophilized.

7. Split-kanamycin resistance assay

Assays coupling intein *trans*-splicing activity to kanamycin resistance in *E. Coli* were performed as previously described (2, 3). A plasmid encoding KanR^{N}-Vid^{N} and Vid^{C}-KanR^{C} containing each of the selected -2, -1 and +2 extein residues (A, E, G, K, L, M, P, Q, S, W) was transformed into DH5α competent cells and 3 mL cultures were grown in LB broth with 100 μg/mL ampicillin at 37 °C for 16 h. The cultures were diluted 8-fold and added to LB media with kanamycin (0, 50, 250, 500, 650, 800, 900, and 1000 μg/mL) at 30 °C for 24 h with shaking. The optical density at 650 nm was measured at 24 h and was plotted as a function of kanamycin concentration. The data was then fit to a dose response curve with variable slope as shown below.

$$OD_{obs} = OD_{min} + \frac{(OD_{max} - OD_{min})}{1 + 10^{[(\log IC_{50} - \log[Kan]) \cdot HillSlope]}}$$

OD_{min} was fixed to background absorbance at 650 nm. The assay was performed in triplicate for each extein position and the IC_{50} values reported as the mean and standard error of the replicates.

8. Fluorescence anisotropy

Equilibrium measurements were performed using 20 nM biotin-Vid^{N}-fluorescein with varying concentrations of splicing-inactivated Vid^{C}-MBP (0 – 2 μM) in splicing buffer with 10% v/v glycerol (100 mM phosphate, 150 mM NaCl, 1 mM EDTA, 10% v/v glycerol, pH 7.2) at 30 °C. High-salt experiments were carried out in splicing buffer with 10% glycerol and 4 M NaCl, and low-salt experiments were carried out in splicing buffer with 10% glycerol and 5 mM NaCl. Samples were diluted from stock solutions, mixed, and transferred to a cuvette of 1 cm path-length. Fluorescence anisotropy was measured immediately on a FluoroMax fluorimeter (Horiba). Data fitting for one-site binding was performed with SigmaPlot 12.0 software.

9. X-ray crystallography

Full length VidaL, with C-A and N-A mutations, was expressed and purified as described above. The purified protein was dialyzed into a buffer containing HEPES (25 mM), NaCl (150 mM) and DTT (1 mM) at pH 7.5 and concentrated to 13 mg mL^{-1}. The protein was subjected to screening using a Phoenix crystallization robot (Art Robbins). Crystal screen, PEG/ion, and index crystallization screens were used (Hampton Research). Crystals were grown at 20 °C using the sitting-drop vapor diffusion method with 0.3 μL of protein mixed with 0.3 μL of screening solution. Crystals were obtained within one week from a buffer containing 0.1 M sodium acetate (pH 4.6) with 2 M ammonium sulfate. The selenomethionine variant was crystallized using identical methods under the same conditions. Crystals were cryoprotected using the screening condition + 50% v/v glycerol before freezing in liquid nitrogen. Diffraction data for the selenomethionine variant were collected at the National Synchrotron Light Source II (Brookhaven National Laboratory), beamline FMX, and phasing was determined using single-wavelength anomalous dispersion (SAD). The data were processed using the XDS package (4). The final structure was obtained after iterative rounds of model building in COOT (5), and refinement with PHENIX Refine (6).

The selenomethionine structure was used to solve the wild-type VidaL structure by molecular replacement using Phaser MR in the CCP4 suite (7), with data collected at the Princeton macromolecular crystallography core facility, using a Rigaku Micromax-007HF X-ray source, and a Dectris Pilatus 3R 300K area detector. The data were processed using the HKL-3000 software package (8). The final structure was obtained after iterative rounds of model building in COOT (5), and refinement with PHENIX Refine (6).

Data collection and refinement statistics for the structures are displayed in ***SI Table 1***.

10. *In vitro* protein *trans*-splicing

VidaL^{N} and VidaL^{C} constructs (8 and 4 μM, respectively) in splicing buffer (100 mM phosphate, 150 mM NaCl, 1 mM EDTA, 10% v/v glycerol, pH 7.2) were treated with 1 mM TCEP for 10 min at 30 °C. The constructs were then mixed (4 and 2 μM final concentration, respectively) and aliquots were removed at selected time points and quenched by addition of 4x SDS loading buffer and boiled for 5 min. The samples were resolved by SDS-PAGE (12% bis-tris gel, 150V, 60 min), and analyzed by Coomassie stain and western blotting with anti-Flag and streptavidin-800. Where appropriate, the temperature and concentration of NaCl in the assay was adjusted. Determination of the rate of C-terminal cleavage was performed as above, but in the absence of added VidaL^{N} and with 2 mM TCEP.

11. Histone dimer and tetramer formation

macroH2A$_{1-119}$-Vid^{N}-H6 and H2B (0.5 mg each) were suspended in 500 μL denaturing buffer (6 M guanidinium chloride, 20 mM tris, 1 mM DTT, pH 7.5) and dialyzed overnight into refolding buffer at 4 °C (10 mM tris, 2 M NaCl, 1 mM EDTA, 1 mM DTT, pH 7.5). Two further 2 h dialysis steps with fresh buffer were then performed. Wild-type H2A-H2B dimers, and H3-H4 tetramers were prepared similarly on the same scale. The complexes were injected onto an S200 increase 10/300 GL gel filtration column, pre-equilibrated with refolding buffer. Fractions containing dimers or tetramers (as identified by SDS-PAGE gel electrophoresis) were concentrated to approx. 20 μM, diluted to 50% v/v glycerol, and stored at -20 °C.

12. macroH2A1.1 semi-synthesis

macroH2A$_{1-119}$-Vid^{N}-H6 and H2B dimers in refolding buffer (4 μM final concentration) were mixed in a 1:1 volume ratio with Vid^{C}-macroH2A1.1$_{120-372}$ in purification buffer (3 μM final concentration), in a total volume of 300 μL. The splicing reaction was left to proceed at room temperature for 3 h. The reaction was injected onto an S200 increase 10/300 GL gel filtration column, pre-equilibrated with refolding buffer. Fractions containing macroH2A1.1-H2B dimers (as identified by SDS-PAGE gel electrophoresis) were concentrated to 8 μM, diluted to 50% v/v glycerol, and stored at -20 °C.

13. Mononucleosome assembly

Mononucleosome assemblies were carried out in a total volume of 50 μL, using Slide-a-lyzer MINI dialysis devices (3.5 kDa cutoff). 147 bp Widom DNA (0.5 μM; 1 eq.; sequence below) was mixed

with wild-type or macroH2A1.1 dimers (2.6 eq.) and wild-type tetramers (1.3 eq.) in refolding buffer. The mixtures were dialyzed against 200 mL nucleosome start buffer at 4 °C for 2 h (10 mM tris, 1.4 M KCl, 0.1 mM EDTA, 1 mM DTT, pH 7.0). 330 mL of nucleosome end buffer (10 mM tris, 10 mM KCl, 0.1 mM EDTA, 1 mM DTT, pH 7.0) was then added at a rate of 1 mL min^{-1}, followed by two one-hour dialyses against nucleosome end buffer. The mixtures were centrifuged at 17,000 *g* for 10 min at 4 °C, and the supernatant removed. Nucleosome concentrations were determined by $A_{260\ nm}$ and was purity confirmed by native gel electrophoresis (5% acrylamide gel followed by SYBR Gold staining) and SDS-PAGE analysis (15% tris gel followed by Coomassie stain).

Widom 601 147 bp DNA:

```
ACAGGATGTATATATCTGACACGTGCCTGGAGACTAGGGAGTAATCCCCTTGGCGGTTA
AAACGCGGGGGACAGCGCGTACGTGCGTTTAAGCGGTGCTAGAGCTGTCTACGACCAAT
TGAGCGGCCTCGGCACCGGGATTCTCCAG
```

14. PARP1 activity assays

PARP1 ADP ribosylation assays were performed in a buffer containing 50 mM tris, 20 mM NaCl, 2 mM $MgCl_2$, and 2 mM DTT, at pH 7.5. The final mononucleosome (macroH2A1.1 and wild type) concentration was 200 nM. PARP1 was added to a final assay concentration of 100 nM. Biotin-NAD+ was added to a final concentration of 25 μM. The reactions were left to proceed at 30 °C for 30 min, before the addition of 4x SDS-PAGE loading buffer and boiling to quench the reaction. Samples were resolved on a 12% bis-tris SDS-PAGE gel and western blotting was performed with streptavidin-800 detection. Anti-H3 antibody was used as a loading control for the nucleosome-containing lanes.

15. Mammalian cell culture

HEK 293T cells were cultured in DMEM supplemented with FBS and penicillin/streptomycin in an incubator at 37 °C and 5% CO_2.

16. Cell viability experiments

HEK 293T cells were transfected with the appropriate constructs using Lipofectamine 2000 following the manufacturer's instructions. After 24 h, 25,000 cells were plated into each well of a 96-well plate. At each time point, 20 μL of EZQuant reagent (diluted three-fold in PBS) was added

and 2 h later the absorbance at 450 nm was recorded. Each viability experiment was performed in triplicate.

17. Protein *trans*-splicing in isolated nuclei

HEK 293T cells were transfected with the appropriate VidaL^{C} construct (FLAG-Vid^{C}-H3$_{19-135}$) using Lipofectamine 2000 following the manufacturer's instructions. After 24-28 h the cells were harvested in ice cold PBS and pelleted by centrifugation at 1200 *g* for 5 min at 4 °C. Intact nuclei were isolated by resuspending the cells in hypotonic lysis buffer (10 mM Tris base, 15 mM NaCl, 1.5 mM $MgCl_2$, and Roche Complete protease inhibitor; pH 7.6; 1 mL per 1 x 10^7 cells) and centrifuging at 400 *g* for 5 min at 4 °C. The nuclei were further purified from cellular contaminants by resuspension in lysis buffer (1 mL per 1 x 10^7 cells) and homogenization with 10 strokes of a loose pestle Dounce homogenizer. The homogenized nuclei were again pelleted by centrifugation at 400 *g* for 5 min at 4 °C. Purified nuclei were resuspended in reaction buffer (20 mM HEPES, 1.5 mM $MgCl_2$, 150 mM KCl, 1 mM DTT, 1 mg/mL BSA, Roche Complete protease inhibitor; pH 7.6; 500 μL per 1 x 10^7 cells), the appropriate VidaL^{N} construct (0.5 or 1 μM final concentration) was added, and the reaction was incubated at 37 °C. At selected time points, aliquots were removed from the reaction mixture and immediately quenched by addition of 80 mM iodoacetamide, and flash frozen in liquid N_2. After thawing the frozen nuclei, chromatin was isolated as described below, and samples were resolved by SDS-PAGE (12% bis-tris gel, 150V, 60 min) and analyzed by western blotting.

18. In-cell delivery of VidaL^{N} by electroporation

HEK 293T cells were transfected with the appropriate VidaL^{C} construct using Lipofectamine 2000 following the manufacturer's instructions. After 26 h, the cells were trypsinized and resuspended in cold DMEM (10 mL per 10 cm plate). The cells were pelleted by centrifugation at 1300 rpm for 5 min, resuspended and washed with cold PBS (2 x 10 mL) and isolated by centrifugation. Finally, the cells were resuspended in PBS at a density of 1 x 10^7 cells per mL. Each electroporation cuvette was filled with 400 μL of cells, with 6 μM of the appropriate VidaL^{N} peptide. Electroporation was performed on a BTX-830 electroporator (Harvard biosciences), using 5 x 0.1 msec pulses at 415 V, with a 1.1 sec delay time. The cells were re-plated into poly-lysine coated plates or microscopy dishes containing DMEM at 37 °C with no antibiotics. After 16 h the cells were harvested in PBS supplemented with iodoacetamide (80 mM final concentration), and recovered by centrifugation at 1200 *g* for 5 min at 4 °C.

Workup for isolation of the chromatin fraction The cells were lysed through the addition of lysis buffer 1 (50 mM HEPES, 140 mM NaCl, 1 mM EDTA, 10% v/v glycerol, 0.5% v/v NP-40, 0.25% v/v triton-X-100, pH 7.5, protease inhibitor, 80 mM iodoacetamide; 300 ul per 1 x 10^7 cells) for 10 min on ice. After centrifugation at 1200 *g* for 5 min at 4 °C the insoluble fraction was washed with lysis buffer 2 (10 mM tris, 200 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, pH 8.0, protease inhibitor, 80 mM iodoacetamide) and centrifuged. Finally, the chromatin fraction was suspended in lysis buffer 3 (10 mM tris, 100 mM NaCl, 1 mM EDTA, 0.5 mM EGTA, 0.1% w/v sodium deoxycholate, 0.5% w/v *N*-lauroylsarcosine, pH 8.0, protease inhibitor, 80 mM iodoacetamide) and sheared by sonication (2 x 5 sec pulses, 25% amplitude). 1% v/v Triton X-100, 100 mM DTT and 1 x SDS-PAGE loading buffer were added, and the samples boiled for 20 min. The samples were resolved by SDS-PAGE (12% bis-tris gel, 150V, 60 min) and analyzed by western blotting.

Biotin immunoprecipitation of spliced product from the chromatin fraction The chromatin fraction was suspended in lysis buffer 3 (as above) and sheared to approximately mono- and di-nucleosomes by 12 cycles of a probe sonicator (25% amplitude, 15 sec on, 45 sec off). 0.5% v/v triton-X100 was added, and the sample centrifuged at 10 k*g* for 10 min at 4 °C. The supernatant (≈300 μL) was diluted with 700 μL binding buffer (25 mM tris, 150 mM NaCl, 1 mM DTT, 0.5% v/v NP-40 alternative), and added to pre-equilibrated magnetic streptavidin beads. The samples were rotated at rt for 1 h. The supernatant was removed, and the beads were sequentially washed with binding buffer (1x), 0.5% w/v SDS in PBS (2x), 1 M NaCl in PBS (2x) and tris-buffered saline (1x). 1x SDS-PAGE loading buffer with 2 mM DTT was added, and the beads were boiled for 20 min to liberate the immunoprecipitated proteins. The samples were resolved by SDS-PAGE (12% bis-tris gel) and analyzed by western blotting.

Workup for whole-cell isolation The cells were lysed through the addition of 120 μL hypotonic lysis buffer (10 mM tris, 10 mM KCl, pH 7.6, protease inhibitor, 80 mM iodoacetamide) and incubated on ice for 10 min. The suspension was sheared by sonication (2 x 5 sec pulses, 25% amplitude). 1% v/v Triton X-100, 100 mM DTT and 1 x SDS-PAGE loading buffer were added, and the samples boiled for 20 min. The samples were resolved by SDS-PAGE (12% bis-tris gel, 150V, 60 min) and analyzed by western blotting.

19. LCMS/MS confirmation of spliced product formation

PTS was used to install the dually modified H3K4me1-H3K27ac tail to FLAG-VidaL^C-H3-$_{35\text{-}135}$-GFP, and the spliced product was isolated by biotin immunoprecipitation as described above. The spliced product was resolved by SDS-PAGE and stained by Coomassie. The gel band was excised and subjected to in-gel reduction, alkylation, and overnight trypsin digestion (9). Samples

were dried in a SpeedVac (Thermo Fisher) and resuspended in 21 μL of 0.1% formic acid (pH 3.0). 2 μL was injected using an Easy-nLC 1200 UPLC system onto a 45 cm x 75 μm (inner diameter) nano-capillary column packed with 1.9 μm C18-AQ (Dr. Maisch, Germany) mated to a metal emitter in-line with an Orbitrap Fusion Lumos (Thermo Scientific). The column temperature was 45 °C and a one-hour gradient method was run at a flow rate of 300 nL min^{-1}. The mass spectrometer was operated in data dependent mode with a 120,000 resolution MS1 scan (positive mode, profile data type, AGC 4 $x10^5$, Max IT 50 ms, 375-1500 m/z) in the Orbitrap, followed by HCD fragmentation in the ion trap with 35% collision energy. A dynamic exclusion list was invoked to exclude previously fragmented peptides for 60 s and maximum cycle time of 3 s was used. Peptides were isolated for fragmentation using a 1.2 Da window in the quadrupole. The ion trap was operated in Rapid mode with AGC 1 $x10^4$, maximum IT of 360 ms, and a minimum of 5000 ions.

Raw files were searched using Sequest HT algorithms (10) within the Proteome Discoverer 2.2 suite (Thermo Fisher). 10 ppm MS1 and 0.4 Da MS2 mass tolerances were specified. Carbamidomethylation of cysteine was specified as a fixed modification. Oxidation of methionine, acetylation of protein N-termini, conversion of glutamine to pyroglutamate, and deamidation of asparagine were specified as dynamic modifications. Trypsin digestion with a maximum of two missed cleavages was allowed. The Uniprot human proteome reference database was used (proteome ID UP000005640), and the modified Histone H3.1 protein was added to the database.

Scaffold (version 4.8.4, Proteome Software Inc.) was used to validate MS/MS based peptide and protein identifications. Peptide identifications were accepted if they could be established at greater than 90.0% probability by the Scaffold Local FDR algorithm. Protein identifications were accepted if they could be established at greater than 99% probability and contained at least 2 identified peptides. Protein probabilities were assigned by the Protein Prophet algorithm (11). Proteins that contained similar peptides and could not be differentiated based on MS/MS analysis alone were grouped to satisfy the principles of parsimony.

20. Confocal microscopy

After electroporation, approximately 800,000 cells were plated onto poly-lysine-coated 35 mm glass-bottom microscopy dishes with DMEM in the absence of antibiotics. At selected time points, the media was removed, and the cells were washed 3x with PBS before addition of Hoechst stain diluted 1:10,000 in PBS. Confocal microscopy was performed at 40x magnification using a

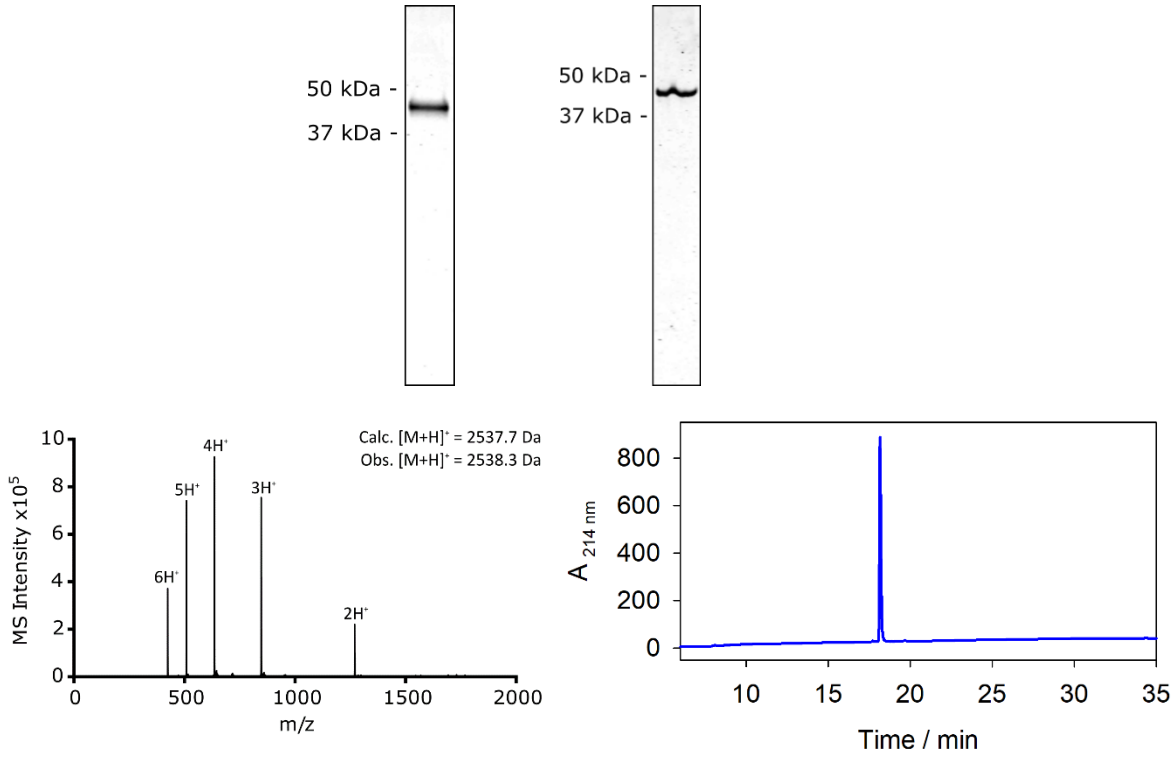
Nikon A1/HD25 microscope (Nikon Instruments, Inc., Melville, NY). Images shown are representative of the multiple cross-sectional images taken during each session.

21. Immunofluorescence microscopy

After electroporation, approximately 800,000 cells were plated onto poly-lysine-coated 35 mm glass-bottom microscopy dishes with DMEM in the absence of antibiotics. At selected time points, the media was removed, and the cells were washed with PBS. Cells were then fixed with 400 μL of 4% paraformaldehyde in PBS for 20 min at 37 °C. Cells were washed 3x with PBS and permeabilized with 400 μL of 0.1% triton X-100 in PBS for 20 min at RT. Cells were washed with PBS and blocked with 400 μL of 2% BSA in PBS for 20 min at RT. Cells were washed 3x with PBS, and incubated with 400 μL of Streptavidin-Alexa Fluor 488 diluted 1:500 and Hoechst diluted 1:10,000 in PBS. Confocal microscopy was performed at 40x magnification using a Nikon A1/HD25 microscope (Nikon Instruments, Inc., Melville, NY). Images shown are representative of the multiple cross-sectional images taken during each session.

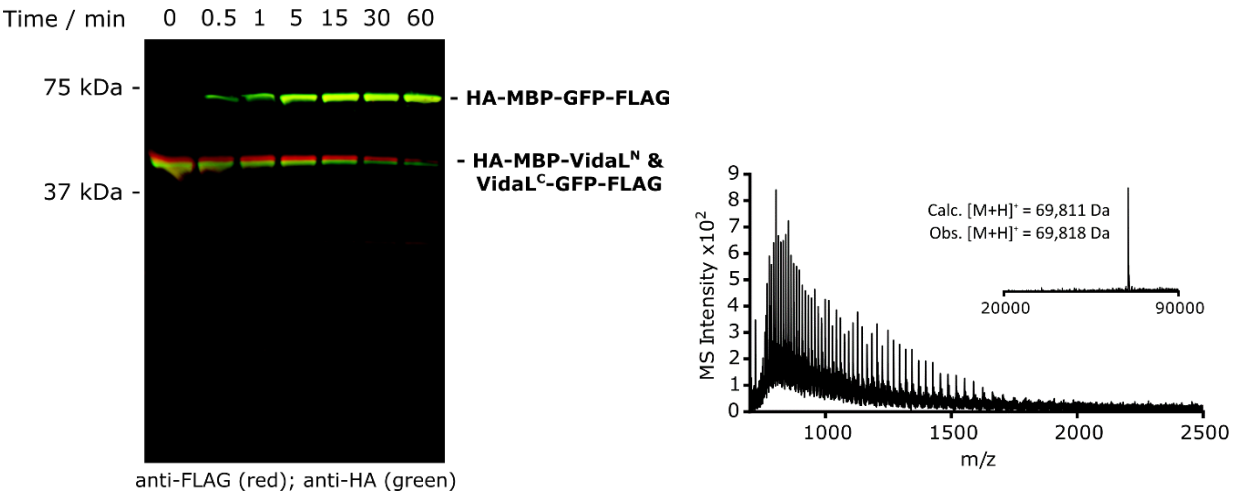
22. Western blotting

SDS-PAGE gels (12% bis-tris) were transferred to PVDF membranes and blocked with 3% w/v BSA in TBS-T (25 mM tris, 150 mM NaCl, 0.1% v/v Tween-100, pH 7.6) at rt for 1 h. Appropriate primary antibodies (***SI Table 2***) were added to the membrane for 1 h at rt (anti-FLAG) or overnight at 4 °C (other epitopes). Membranes were washed three times for 5 min each with TBS-T, and appropriate secondary antibodies were added for 1 h at rt in TBS-T (***SI Table 2***). Membranes were imaged using a LI-COR Odyssey Infrared Imager.

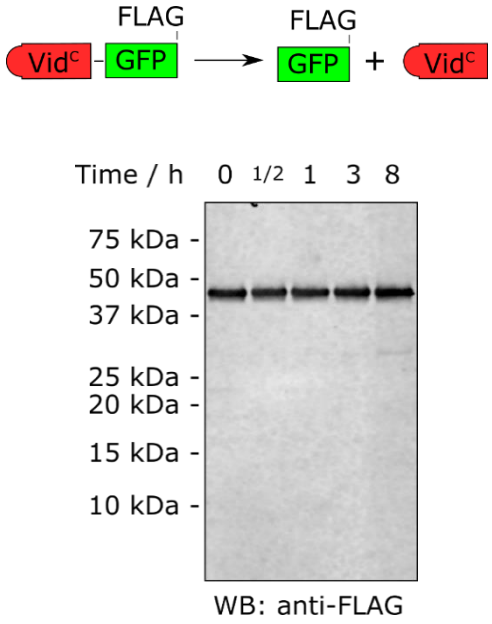


H_2N-K(biotin)KESGCLPKEAVVQIRLTKKG-NH_2

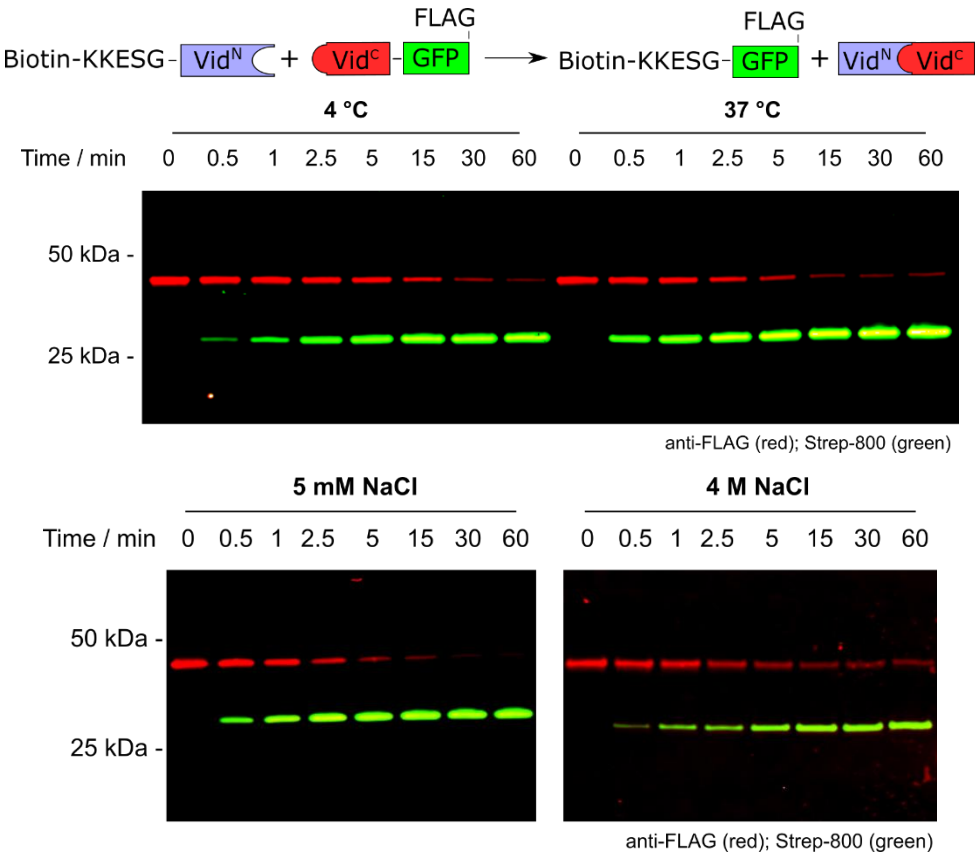
SI Figure 1. Characterization of VidaL constructs for *in vitro* studies.
Top: Coomassie-stained SDS-PAGE gels for Vid^C-GFP-FLAG (left) and HA-MBP-Vid^N (right). Bottom: Characterization of biotin-ESG-Vid^N. ESI-MS (left) and RP-HPLC trace (right).



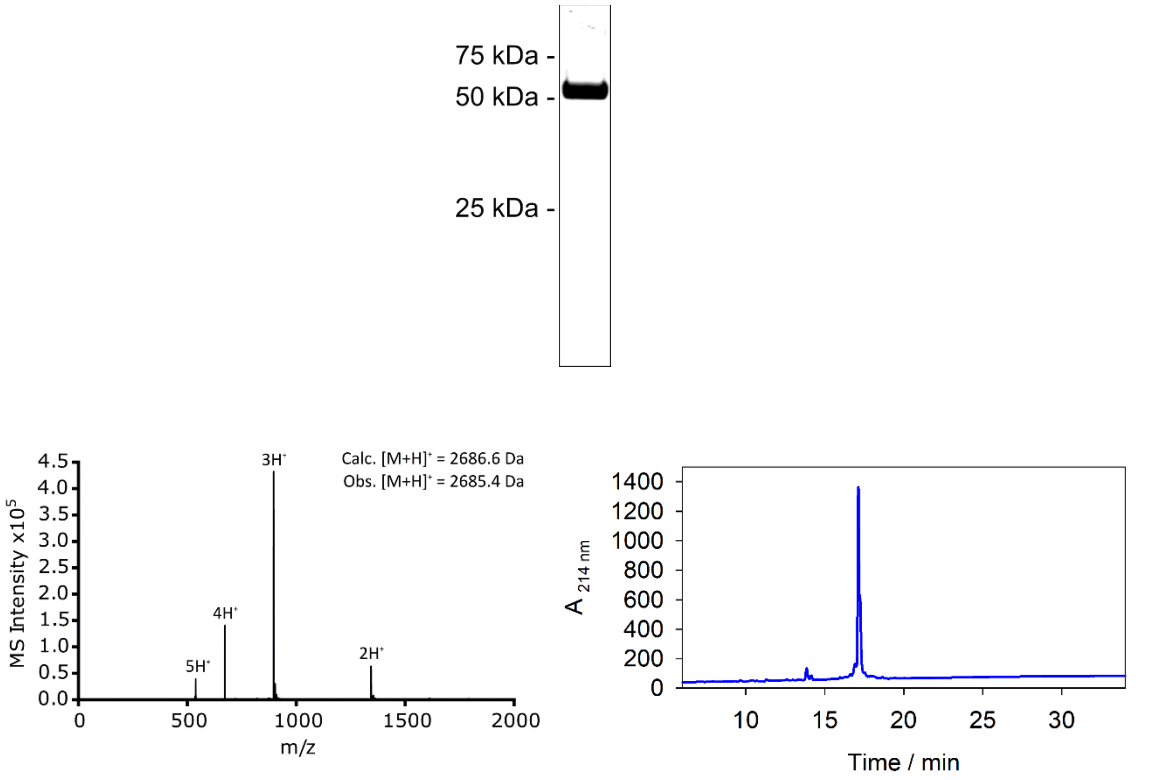
SI Figure 2. Protein *trans*-splicing with MBP and GFP exteins.
Left: Western blot of a PTS time course for the reaction between HA-MBP-Vid^{N} and Vid^{C}-GFP-FLAG. Dually tagged spliced product is observed at ≈70 kDa. Right: ESI-MS of HA-MBP-GFP-FLAG spliced product (inset = deconvoluted).



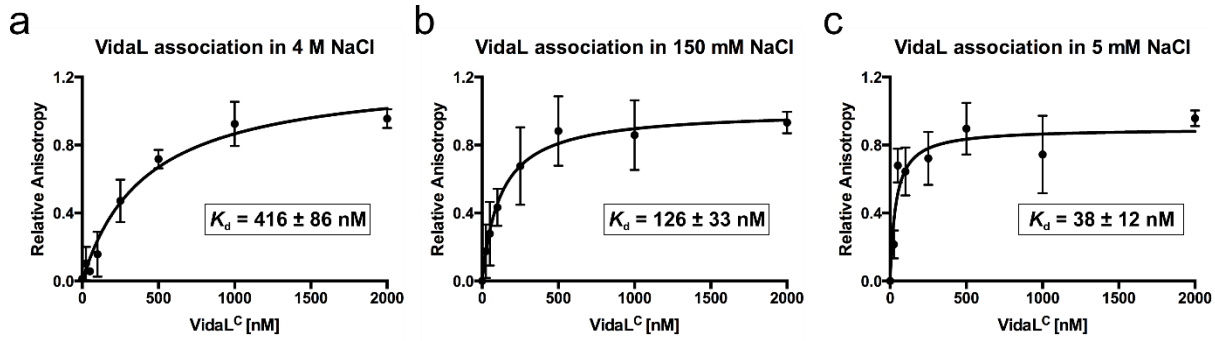
SI Figure 3. C-terminal cleavage of VidaL^{C}-GFP-FLAG.
anti-FLAG western blot visualizing the rate of C-terminal cleavage of Vid^{C}-GFP-FLAG in the presence of 2 mM TCEP. <2% C-terminal cleavage (observed at ≈30 kDa) is observed after 8 h at 30 °C.



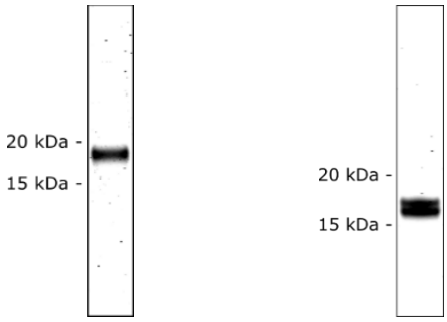
SI Figure 4. Temperature and NaCl dependence of splicing.
Top: Western blots (anti-FLAG (red) and Strep-800 (green)) visualizing Vid^C-GFP-FLAG starting material and biotin-GFP-FLAG spliced product for PTS carried out at 4 and 37 °C. Bottom: Western blots for PTS carried out in the presence of 5 mM and 4 M NaCl.



H_2N-ESGCLPKEAVVQIRLTKKGSGK(fluorescein)-NH_2



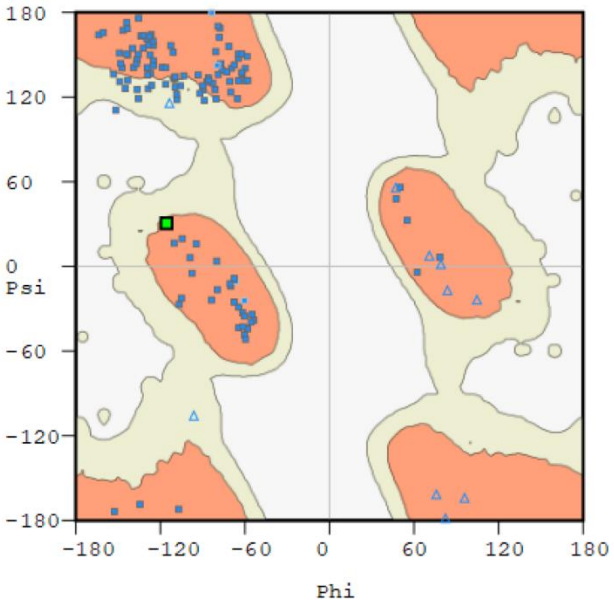
SI Figure 5. Association of VidaL^N and VidaL^C by fluorescence anisotropy.
Top: Coomassie-stained SDS-PAGE gel for Vid^C-MBP carrying a N123A and S_{+1}A double mutation at the splice junction. Middle: Characterization of Vid^N-fluorescein peptide; ESI-MS (left) and RP-HPLC trace (right). Bottom: Binding between Vid^C-MBP carrying a N123A and S_{+1}A double mutation and VidaL^N-fluorescein as measured by fluorescence anisotropy in 4M (a), 150 mM (b), and 5 mM (c) NaCl in splicing buffer. n=3, errors = S.E.M.



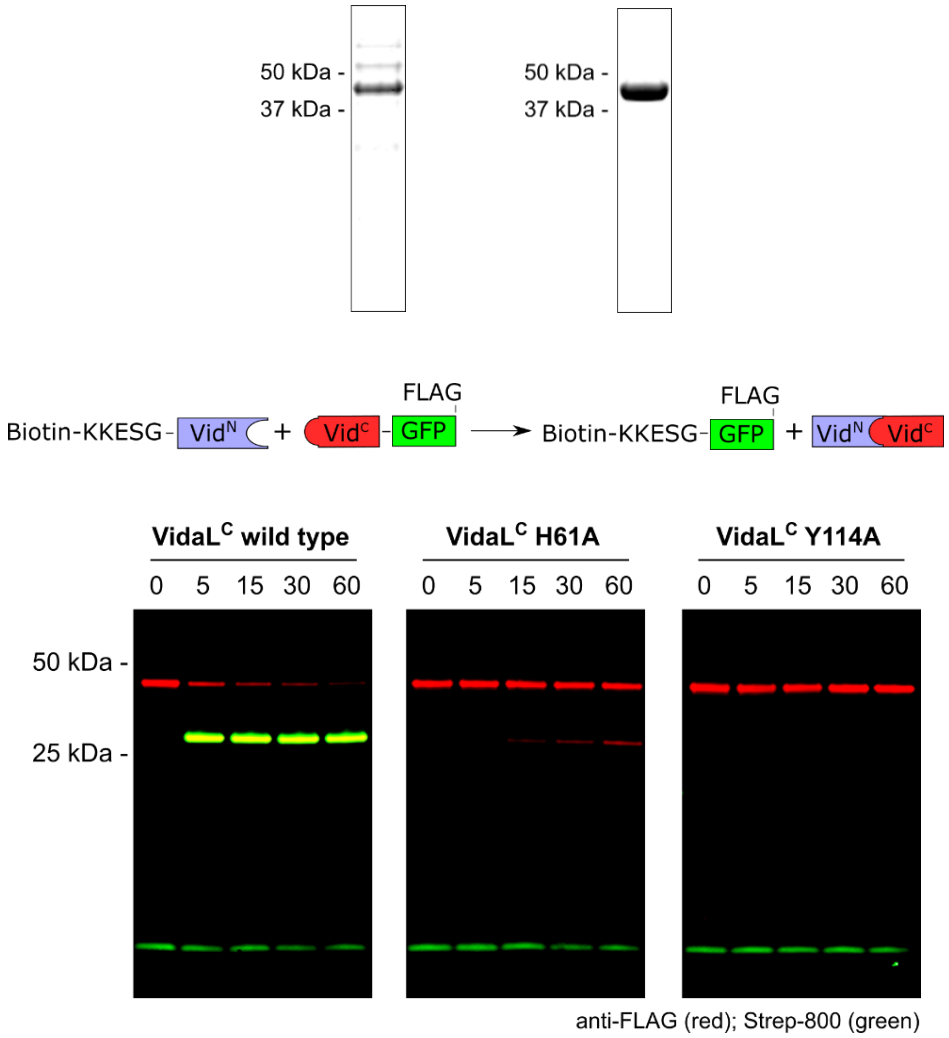
SI Figure 6. Expression of fused VidaL constructs for crystallization.
Left: Coomassie-stained SDS-PAGE gel for fused VidaL intein carrying C1A and N123A mutations. Right: Coomassie-stained SDS-PAGE gel for SeMet variant of this protein.

SI Table 1. Crystallization data collection and refinement statistics.
Highest-resolution shell shown in parentheses. R_{free} represents the R-factor calculated from 5% of the reflections not used during refinement.

	VidaL	**VidaL (SeMet variant)**
PDB identifier	**6vgv**	**6vgw**
Data Collection		
Wavelength (Å)	1.5418	0.9793
Space group	C2	C2
Cell dimensions		
a, b, c (Å)	100.1, 40.8, 36.2	101.5, 40.6, 36.0
β (°)	103.5	103.6
Resolution (Å)	26.95-1.65 (1.75-1.65)	49.33-1.51 (1.58-1.51)
R_{pim}	0.023 (0.185)	0.029 (0.250)
$I / \sigma I$	4.0 (3.5)	15.4 (2.7)
CC ½ (%)	99.5 (91.9)	99.9 (83.8)
Completeness (%)	98.9 (92.6)	97.2 (88.8)
Anomalous completeness (%)	-	96.2 (81.9)
Redundancy	5.9 (2.5)	10.3 (9.4)
Anomalous redundancy	-	5.1 (5.0)
Wilson B Factor (Å^2)	16.6	16.6
Refinement		
Resolution (Å)	1.65	1.51
Reflections / unique	104971 / 18525	223210 / 21755
$\boldsymbol{R_{work}}$ **/** $\boldsymbol{R_{free}}$	**0.163 / 0.209**	**0.169 / 0.190**
No. atoms / ions		
Protein	1159	1182
Other	150	123
Average B-factor (Å^2)	21.7	20.9
R.M.S deviations		
Bond lengths (Å)	0.005	0.01
Bond angles (°)	0.80	0.88
Ramachandran (%)		
Favored	99.3	96.7
Allowed	0.7	3.3

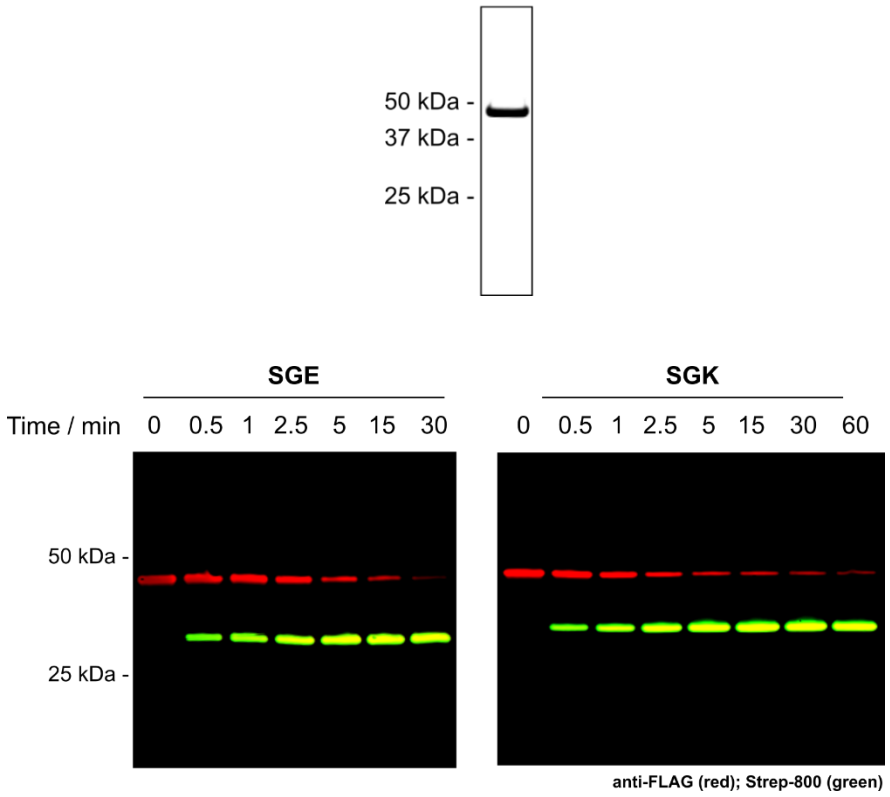


SI Figure 7. Ramachandran plot generated from the X-ray crystal structure of VidaL. Gly-1 is highlighted, residing in α-helical Ramachandran space. Plot generated in Coot (5).

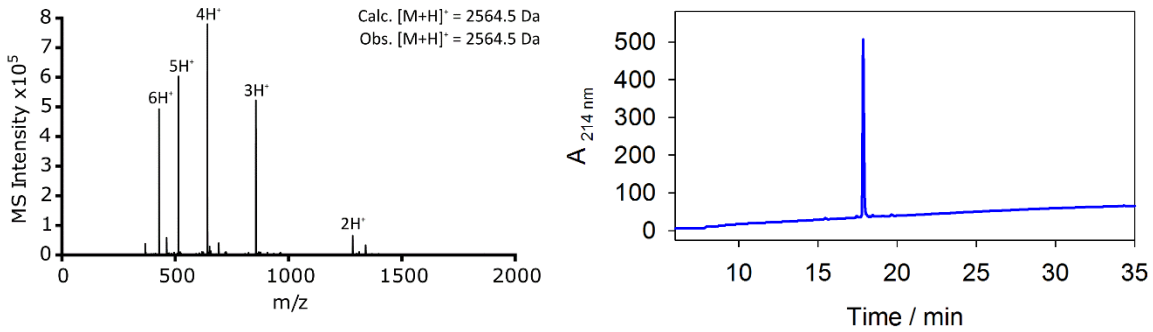


SI Figure 8. Characterization and evaluation of H61A and Y114A mutants of VidaL.
Top: Coomassie-stained SDS-PAGE gel for Vid^{C}-H61A-GFP-FLAG (left) and Y114A (right). Bottom: Western blot (anti-FLAG (red) and Strep-800 (green)) visualizing Vid^{C}-GFP-FLAG starting material and biotin-GFP-FLAG spliced product. Splicing is observed with the wild-type residues, with no splicing observed with the H61A and Y114A mutations. Biotin-ESG-Vid^{N} is observed at ≈7 kDa.

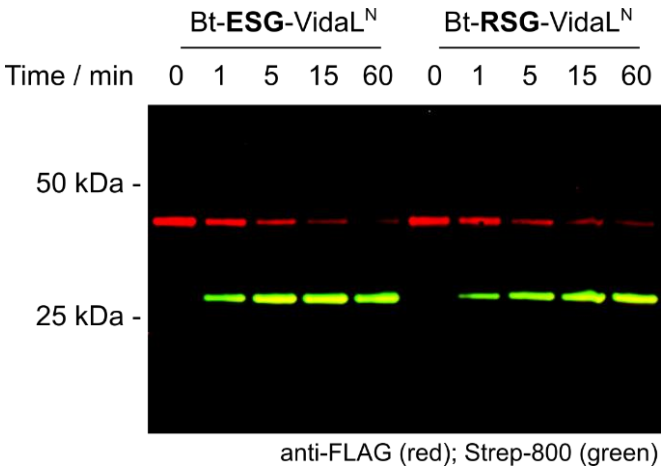
.



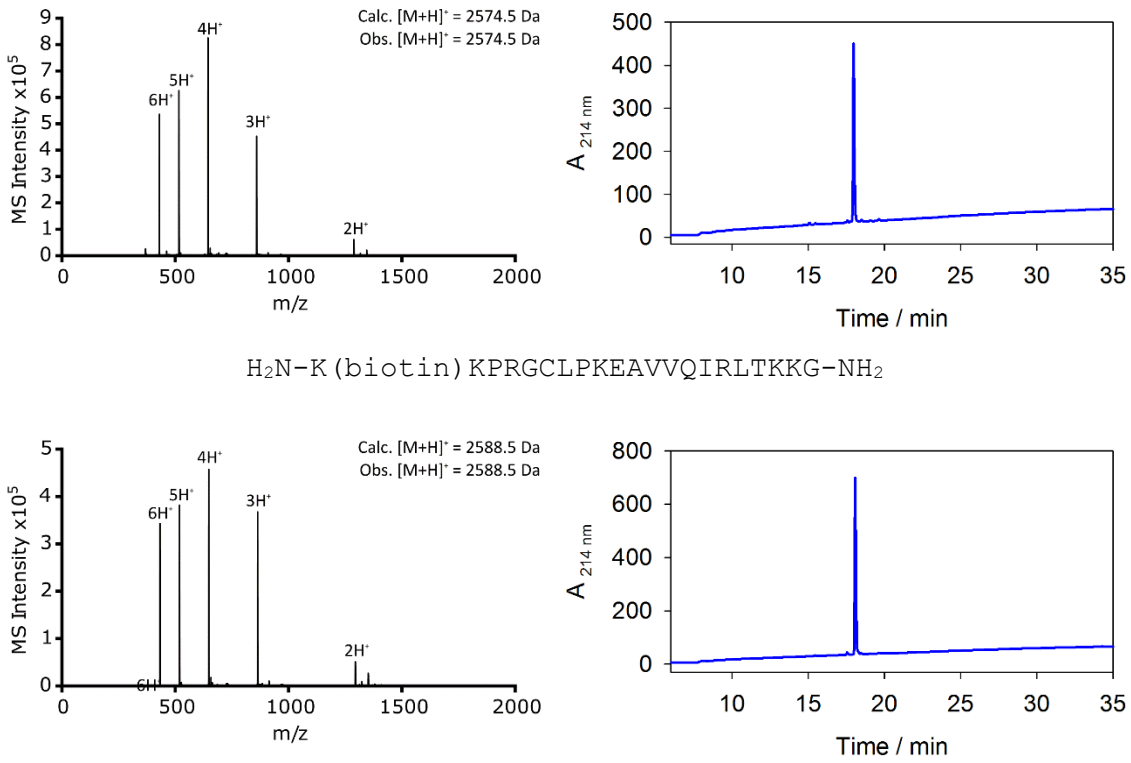
SI Figure 9. Evaluation of +3 extein dependence of VidaL.
Top: Coomassie-stained SDS-PAGE gel for Vid^{C}-$S_{+1}G_{+2}E_{+3}$-GFP-FLAG. Bottom: Western blot (anti-FLAG (red) and Strep-800 (green)) visualizing Vid^{C}-GFP-SGE-FLAG starting material and biotin-GFP-SGE-FLAG spliced product (left). Wild-type VidC-GFP-SGK-FLAG is provided for comparison (right). No difference in splicing rate is observed when mutating the +3 extein residue from K to E.



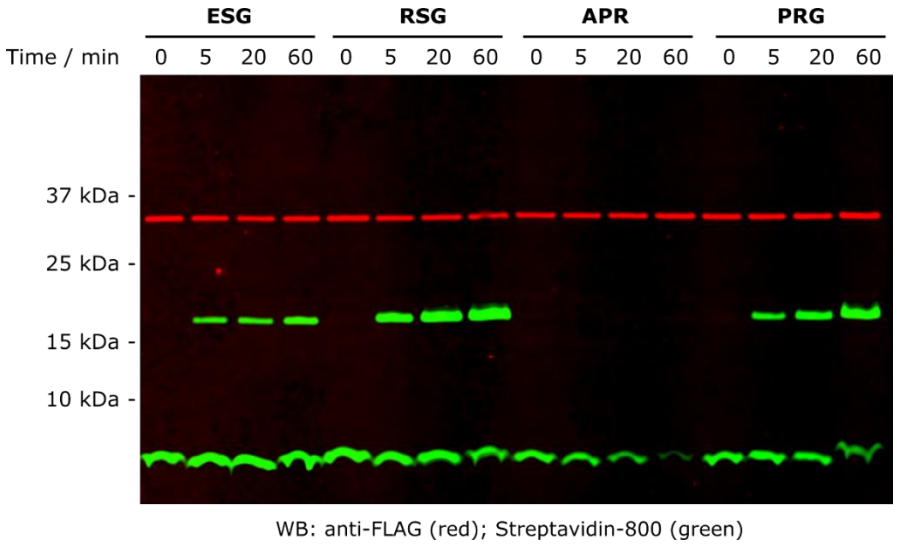
H_2N-K(biotin)KRSGCLPKEAVVQIRLTKKG-NH_2



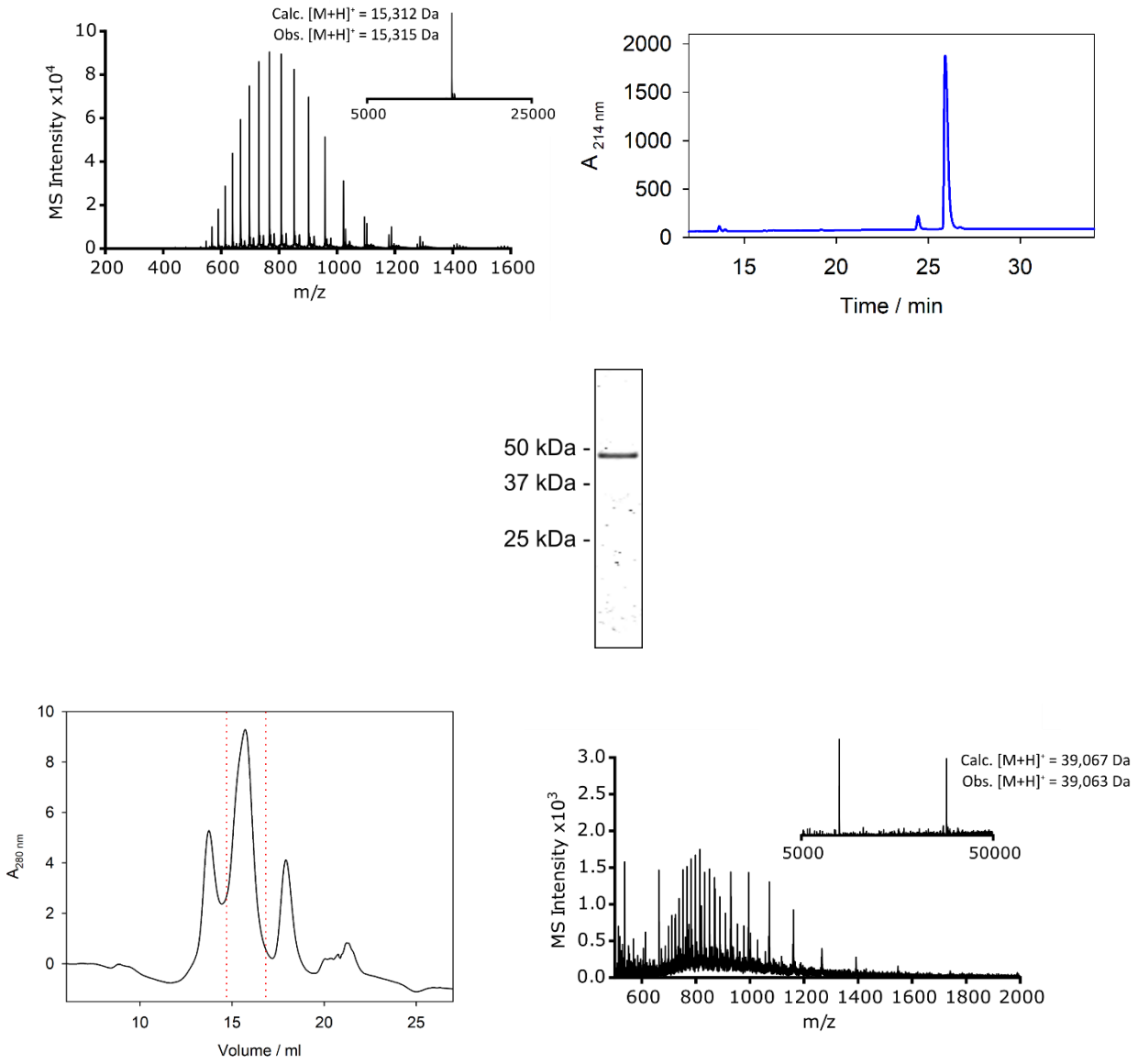
SI Figure 10. Evaluation of -3 extein dependence of VidaL.
Top: Characterization of biotin-R_{-3}SG-Vid^N; ESI-MS (left) and RP-HPLC trace (right). Bottom: Western blot (anti-FLAG (red) and Strep-800 (green)) visualizing Vid^C-GFP-FLAG starting material and biotin-RSG/ESG-GFP-FLAG spliced product. No difference in splicing rate is observed when mutating the -3 extein residue from E to R.



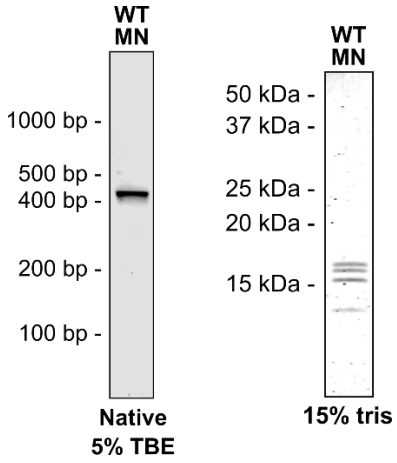
SI Figure 11. Characterization of biotin-PRG-VidaL^N (top) and biotin-APR-VidaL^N (bottom) ESI-MS (left) and RP-HPLC trace (right).



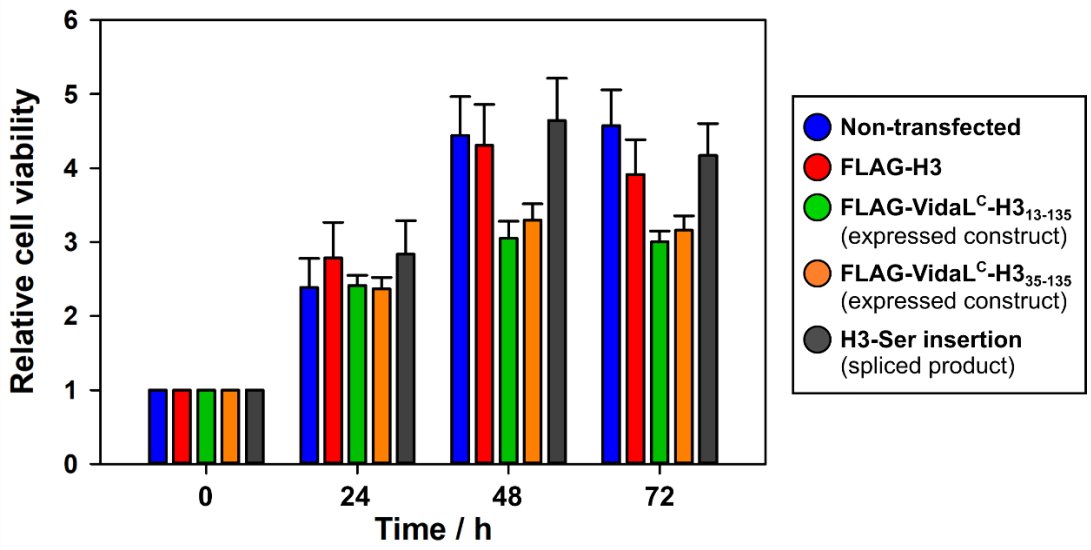
SI Figure 12. In-nucleo extein dependence of VidaL.
Biotin-xyx-Vid^{N} (green; ≈7 kDa), where xyz = ESG, RSG, APR, or PRG, was added to nuclei extracted from cells expressing FLAG-Vid^{C}-H3$_{19-135}$ (red). Spliced product is observed at ≈17 kDa by western blotting against the installed biotin for ESG, RSG, and PRG exteins, but not for APR, confirming a -1 Gly extein dependence in nuclei.



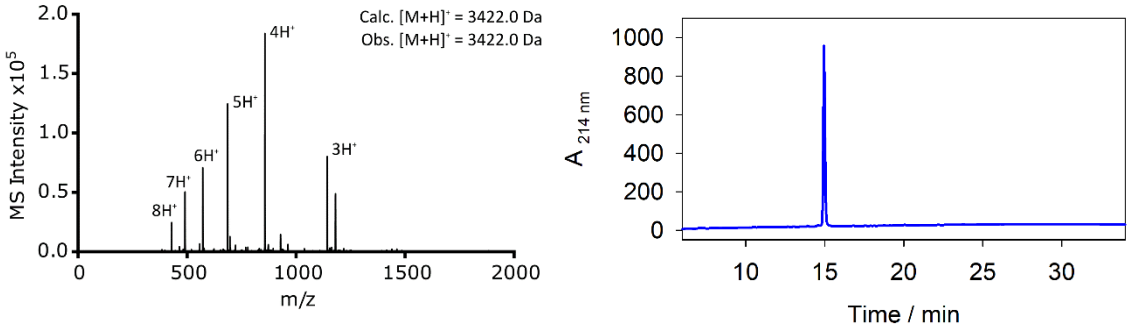
SI Figure 13. Traceless assembly of macroH2A1.1.
Top: Characterization of H2Amacro$_{1\text{-}119}$-Vid^{N}-H6; ESI-MS (left) and RP-HPLC trace (right). Middle: Coomassie-stained SDS-PAGE gel for Vid^{C}-macro1.1$_{120\text{-}372}$. Bottom: FPLC trace for the purification of macroH2A1.1-H2B dimers (left; desired peak is highlighted), mass spectrum of H2Amacro1.1 after PTS (right; inset = deconvoluted spectrum). N.B. Macro-H2A 1.1 co-elutes with wild-type H2B during product isolation by RP-HPLC thereby giving rise to the two charge-state series in the ms spectrum.



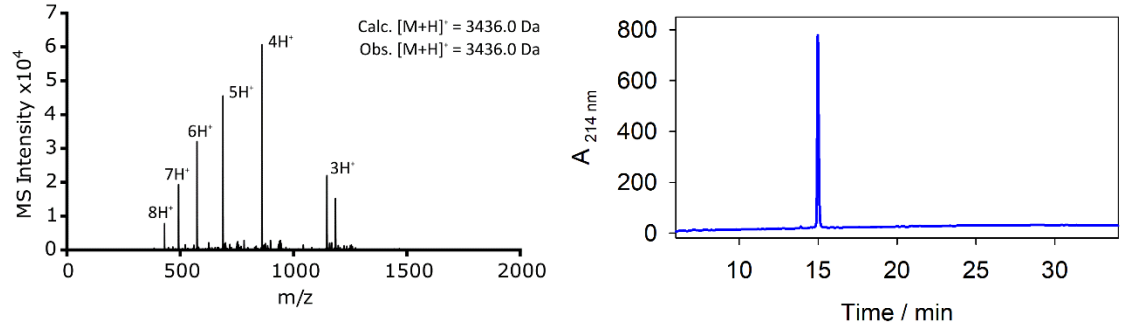
SI Figure 14. Characterization of wild-type mononucleosomes.
Native 5% TBE gel (left) and denaturing SDS-PAGE gel (right) of wild-type mononucleosomes reconstituted on Widom 601 DNA.



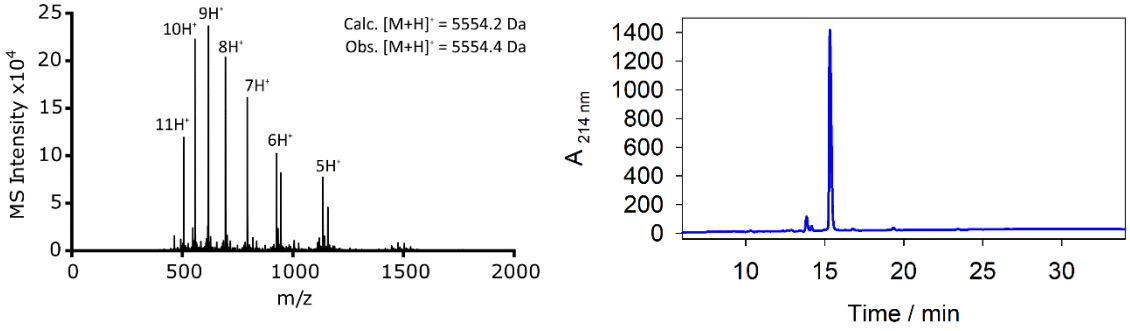
SI Figure 15. Viability of HEK 293T cells transfected with constructs used in this study.
Cell viability, as measured by EZQuant, is not affected for >24 h in the presence of the Vid^{C} fusion proteins (expressed constructs) and the H3 Ser insertion (spliced product). t=0 represents 16 h after the cells were plated into a 96-well plate. Error bars = S.D. (n=3).



H_2N-K(biotin)-ARTKQTARKSTGCLPKEAVVQIRLTKKG-NH_2

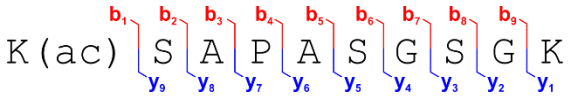


H_2N-K(biotin)-ARTK(me1)QTARKSTGCLPKEAVVQIRLTKKG-NH_2

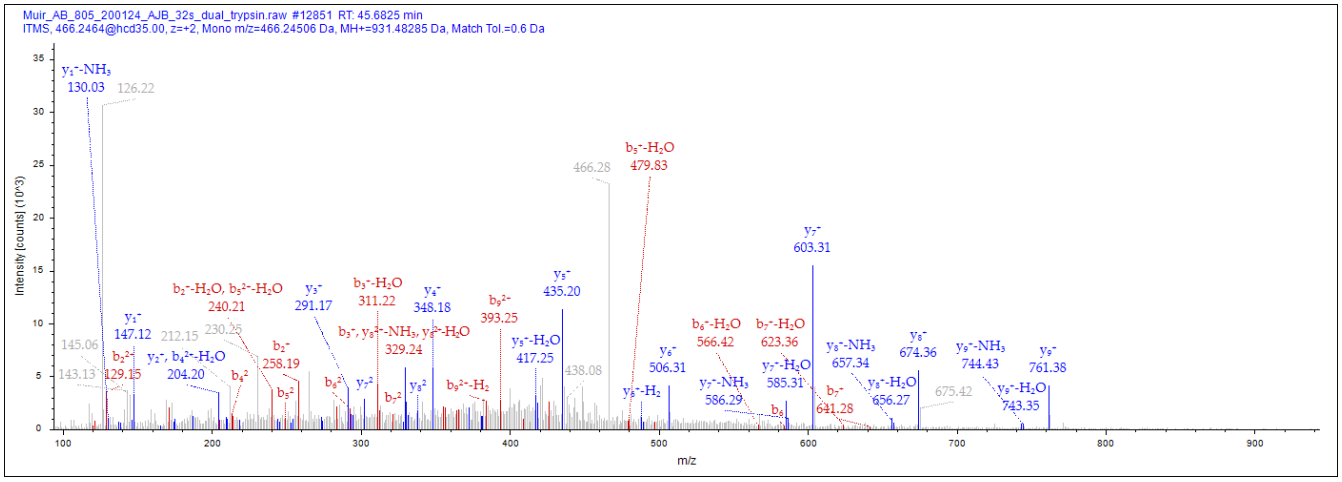


H_2N-K(biotin)-
ARTK(me1)QTARKSTGGKAPRKQLATKAARK(ac)SAPASGCLPKEAVVQIRLTKKG-NH_2

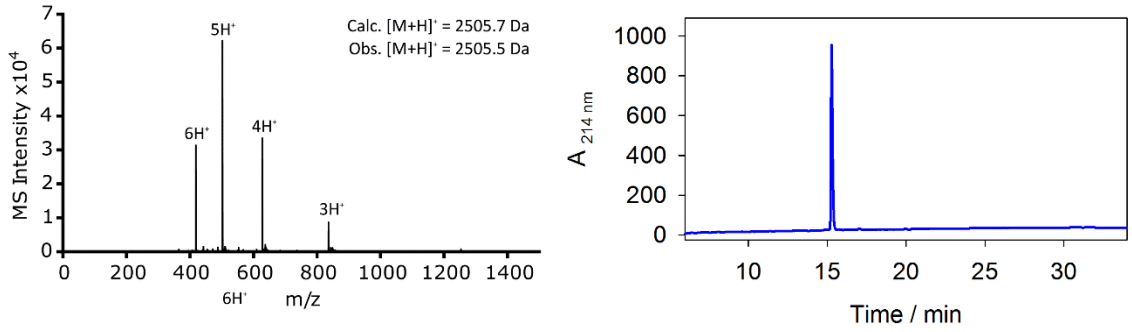
SI Figure 16. Characterization of H3-VidaL^N peptides for live-cell semi-synthesis.
Top: Characterization of biotin-H3$_{1-12}$-Vid^N; ESI-MS (left) and RP-HPLC trace (right). Middle: Characterization of biotin-H3$_{1-12}$(H3K4me1)-Vid^N. Bottom: Characterization of biotin-H3$_{1-31}$(H3K4me1-H3K27ac)-Vid^N.



#1	b^+	b^{2+}	Seq.	y^+	y^{2+}	#2
1	**171.11280**	86.06004	**K-Acetyl**			10
2	**258.14483**	**129.57605**	**S**	**761.37881**	**381.19304**	9
3	**329.18195**	**165.09461**	**A**	**674.34678**	**337.67703**	8
4	**426.23471**	**213.62099**	**P**	**603.30967**	**302.15847**	7
5	**497.27182**	**249.13955**	**A**	**506.25690**	**253.63209**	6
6	**584.30385**	**292.65556**	**S**	**435.21979**	**218.11353**	5
7	**641.32532**	**321.16630**	**G**	**348.18776**	**174.59752**	4
8	728.35734	**364.68231**	**S**	**291.16630**	**146.08679**	3
9	785.37881	**393.19304**	**G**	**204.13427**	**102.57077**	2
10			**K**	**147.11280**	74.06004	1

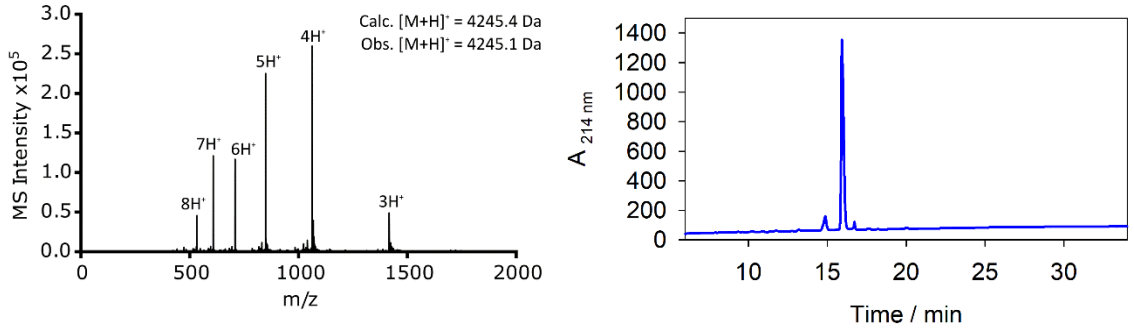


SI Figure 17. LCMS/MS confirmation of H3 spliced product.
Top: Annotated peptide sequence indicating the observed b_n and y_n ions. This peptide contains the unique splice junction, and the installed H3K27ac PTM. Middle: b_n and y_n ion table. Bottom: LCMS/MS spectrum.

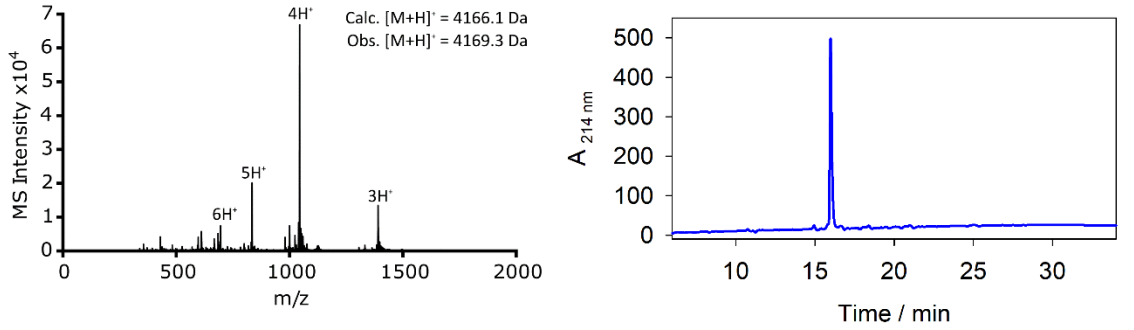


H_2N-K(biotin)KESGALPKEAVVQIRLTKKG-NH_2

SI Figure 18. Characterization of biotin-C1A-Vid^N.
ESI-MS (left) and RP-HPLC trace (right).



H_2N-K(biotin)GKKTKRTADSSSSEDEEESGCLPKEAVVQIRLTKKG-NH_2



(desthiobiotin)HN-GKKTKRTADSSSS(ph)EDEEESGCLPKEAVVQIRLTKKG-NH_2

SI Figure 19. Characterization of HP1α-Vidal^N peptides.
Top: Characterization of biotin-HP1$\alpha_{1\text{-}18}$-Vid^N; ESI-MS (left) and RP-HPLC trace (right). Bottom: Characterization of desthiobiotin-HP1$\alpha_{1\text{-}18}$(S13ph)-Vid^N.

SI Table 2: Antibodies used in this study.

Epitope	Vendor (product)	Dilution
anti-FLAG	Sigma Aldrich (M2)	1:1,000 (TBS-T)
anti-H3K4me1	Abcam (8895)	1:1,000 (TBS-T)
anti-H3K27ac	Abcam (4729)	1:1,000 (TBS-T)
anti-H4	Abcam (31830)	1:2,000 (TBS-T)
anti-Mouse	LI-COR IRDye 680 nm	1:10,000 (TBS-T)
Biotin	LI-COR IRDye Streptavidin 800 nm	1:10,000 (TBS-T, 0.01% w/v SDS)
anti-Mouse	Bio-Rad IgG-HRP	1:2,000 (TBS-T)
anti-Rabbit	Bio-Rad IgG-HRP	1:2,000 (TBS-T)
Biotin	Bio-Rad Streptavidin-HRP	1:2,000 (TBS-T)

Appendix 1. Amino acid sequences for the constructs used in this study.

Bacterial expression

a. H6-SUMO-Vid^{C}-GFP-FLAG

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGM
IEEKKVTVQE LRELYLSGEY TIEIDTPDGY QTIGKWFDKG VLSMVRVATA TYETVCAFNH
MIQLADNTWV QACELDVGVD IQTAAGIQPV MLVEDTSDAE CYDFEVMHPN HRYYGDGIVS
HNSGKMVSKG EELFTGVVPI LVELDGDVNG HKFSVSGEGE GDATYGKLTL KFICTTGKLP
VPWPTLVTTL TYGVQCFSRY PDHMKQHDFF KSAMPEGYVQ ERTIFFKDDG NYKTRAEVKF
EGDTLVNRIE LKGIDFKEDG NILGHKLEYN YNSHNVYIMA DKQKNGIKVN FKIRHNIEDG
SVQLADHYQQ NTPIGDGPVL LPDNHYLSTQ SALSKDPNEK RDHMVLLEFV TAAGITLGMD
ELYKDYKDDD DK

b. H6-SUMO-HA-MBP-Vid^{N}

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGY
PYDVPDYAKI EEGKLVIWIN GDKGYNGLAE VGKKFEKDTG IKVTVEHPDK LEEKFPQVAA
TGDGPDIIFW AHDRFGGYAQ SGLLAEITPD KAFQDKLYPF TWDAVRYNGK LIAYPIAVEA
LSLIYNKDLL PNPPKTWEEI PALDKELKAK GKSALMFNLQ EPYFTWPLIA ADGGYAFKYE
NGKYDIKDVG VDNAGAKAGL TFLVDLIKNK HMNADTDYSI AEAAFNKGET AMTINGPWAW
SNIDTSKVNY GVTVLPTFKG QPSKPFVGVL SAGINAASPN KELAKEFLEN YLLTDEGLEA
VNKDKPLGAV ALKSYEEELA KDPRIAATME NAQKGEIMPN IPQMSAFWYA VRTAVINAAS
GRQTVDEAPK DAQTNESGCL PKEAVVQIRL TKKG

c. H6-SUMO-Vid^{C}-(**N123A**)-(**S^{+1}A**)-MBP

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGM
IEEKKVTVQE LRELYLSGEY TIEIDTPDGY QTIGKWFDKG VLSMVRVATA TYETVCAFNH
MIQLADNTWV QACELDVGVD IQTAAGIQPV MLVEDTSDAE CYDFEVMHPN HRYYGDGIVS
H**AA**GKGSASK IEEGKLVIWI NGDKGYNGLA EVGKKFEKDT GIKVTVEHPD KLEEKFPQVA

ATGDGPDIIF WAHDRFGGYA QSGLLAEITP DKAFQDKLYP FTWDAVRYNG KLIAYPIAVE
ALSLIYNKDL LPNPPKTWEE IPALDKELKA KGKSALMFNL QEPYFTWPLI AADGGYAFKY
ENGKYDIKDV GVDNAGAKAG LTFLVDLIKN KHMNADTDYS IAEAAFNKGE TAMTINGPWA
WSNIDTSKVN YGVTVLPTFK GQPSKPFVGV LSAGINAASP NKELAKEFLE NYLLTDEGLE
AVNKDKPLGA VALKSYEEEL AKDPRIAATM ENAQKGEIMP NIPQMSAFWY AVRTAVINAA
SGRQTVDEAP KDAQTN

d. H6-SUMO-Vid^{N}-**C**1**A**-Vid^{C}-**N**123**A** (Selenomethionine highlighted in red)

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGE
SG**A**LPKEAVV QIRLTKKG**M**I EEKKVTVQEL RELYLSGEYT IEIDTPDGYQ TIGKWFDKGV
LSMVRVATAT YETVCAFNH**M** IQLADNTWVQ ACELDVGVDI QTAAGIQPV**M** LVEDTSDAEC
YDFEV**M**HPNH RYYGDGIVSH **A**SGK

e. H6-SUMO-Vid^{C}-(**H61A**)-GFP-FLAG

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGM
IEEKKVTVQE LRELYLSGEY TIEIDTPDGY QTIGKWFDKG VLSMVRVATA TYETVCAFN**A**
MIQLADNTWV QACELDVGVD IQTAAGIQPV MLVEDTSDAE CYDFEVMHPN HRYYGDGIVS
HNSGKMVSKG EELFTGVVPI LVELDGDVNG HKFSVSGEGE GDATYGKLTL KFICTTGKLP
VPWPTLVTTL TYGVQCFSRY PDHMKQHDFF KSAMPEGYVQ ERTIFFKDDG NYKTRAEVKF
EGDTLVNRIE LKGIDFKEDG NILGHKLEYN YNSHNVYIMA DKQKNGIKVN FKIRHNIEDG
SVQLADHYQQ NTPIGDGPVL LPDNHYLSTQ SALSKDPNEK RDHMVLLEFV TAAGITLGMD
ELYKDYKDDD DK

f. H6-SUMO-Vid^{C}-(**Y114A**)-GFP-FLAG

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGM
IEEKKVTVQE LRELYLSGEY TIEIDTPDGY QTIGKWFDKG VLSMVRVATA TYETVCAFNH
MIQLADNTWV QACELDVGVD IQTAAGIQPV MLVEDTSDAE CYDFEVMHPN HR**A**YGDGIVS
HNSGKMVSKG EELFTGVVPI LVELDGDVNG HKFSVSGEGE GDATYGKLTL KFICTTGKLP
VPWPTLVTTL TYGVQCFSRY PDHMKQHDFF KSAMPEGYVQ ERTIFFKDDG NYKTRAEVKF
EGDTLVNRIE LKGIDFKEDG NILGHKLEYN YNSHNVYIMA DKQKNGIKVN FKIRHNIEDG
SVQLADHYQQ NTPIGDGPVL LPDNHYLSTQ SALSKDPNEK RDHMVLLEFV TAAGITLGMD
ELYKDYKDDD DK

g. H6-SUMO-Vid^{C}-SG**E**-GFP-FLAG

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGM
IEEKKVTVQE LRELYLSGEY TIEIDTPDGY QTIGKWFDKG VLSMVRVATA TYETVCAFNH
MIQLADNTWV QACELDVGVD IQTAAGIQPV MLVEDTSDAE CYDFEVMHPN HRYYGDGIVS
HNSG**E**MVSKG EELFTGVVPI LVELDGDVNG HKFSVSGEGE GDATYGKLTL KFICTTGKLP
VPWPTLVTTL TYGVQCFSRY PDHMKQHDFF KSAMPEGYVQ ERTIFFKDDG NYKTRAEVKF
EGDTLVNRIE LKGIDFKEDG NILGHKLEYN YNSHNVYIMA DKQKNGIKVN FKIRHNIEDG
SVQLADHYQQ NTPIGDGPVL LPDNHYLSTQ SALSKDPNEK RDHMVLLEFV TAAGITLGMD
ELYKDYKDDD DK

h. macroH2A1.1$_{1\text{-}119}$-Vid^{N}-H6

MSSRGGKKKS TKTSRSAKAG VIFPVGRMLR YIKKGHPKYR IGVGAPVYMA AVLEYLTAEI
LELAGNAARD NKKGRVTPRH ILLAVANDEE LNQLLKGVTI ASGGVLPNIH PELLAKKRGC
LPKEAVVQIR LTKKGHHHHH H

i. H6-SUMO-Vid^{C}-ΔmacroH2A1.1$_{120-372}$

MGSSHHHHHH GSGLVPRGSA SMSDSEVNQE AKPEVKPEVK PETHINLKVS DGSSEIFFKI
KKTTPLRRLM EAFAKRQGKE MDSLRFLYDG IRIQADQTPE DLDMEDNDII EAHREQIGGM
IEEKKVTVQE LRELYLSGEY TIEIDTPDGY QTIGKWFDKG VLSMVRVATA TYETVCAFNH
MIQLADNTWV QACELDVGVD IQTAAGIQPV MLVEDTSDAE CYDFEVMHPN HRYYGDGIVS
HNSKGKLEAI ITPPPAKKAK SPSQKKPVSK KAGGKKGARK SKKKQGEVSK AASADSTTEG
TPADGFTVLS TKSLFLGQKL QVVQADIASI DSDAVVHPTN TDFYIGGEVG NTLEKKGGKE
FVEAVLELRK KNGPLEVAGA AVSAGHGLPA KFVIHCNSPV WGADKCEELL EKTVKNCLAL
ADDKKLKSIA FPSIGSGRNG FPKQTAAQLI LKAISSYFVS TMSSSIKTVY FVLFDSESIG
IYVQEMAKLD AN

Mammalian expression

a. FLAG-Vid^{C}-H3$_{19-135}$

GDYKDDDDKG MIEEKKVTVQ ELRELYLSGE YTIEIDTPDG YQTIGKWFDK GVLSMVRVAT
ATYETVCAFN HMIQLADNTW VQACELDVGV DIQTAAGIQP VMLVEDTSDA ECYDFEVMHP
NHRYYGDGIV SHNSGKQLAT KAARKSAPAT GGVKKPHRYR PGTVALREIR RYQKSTELLI
RKLPFQRLVR EIAQDFKTDL RFQSSAVMAL QEACEAYLVG LFEDTNLCAI HAKRVTIMPK
DIQLARRIRG ERA

b. FLAG-Vid^{C}-H3$_{13-135}$

GDYKDDDDKG MIEEKKVTVQ ELRELYLSGE YTIEIDTPDG YQTIGKWFDK GVLSMVRVAT
ATYETVCAFN HMIQLADNTW VQACELDVGV DIQTAAGIQP VMLVEDTSDA ECYDFEVMHP
NHRYYGDGIV SHNSGKAPRK QLATKAARKS APATGGVKKP HRYRPGTVAL REIRRYQKST
ELLIRKLPFQ RLVREIAQDF KTDLRFQSSA VMALQEACEA YLVGLFEDTN LCAIHAKRVT
IMPKDIQLAR RIRGERA

c. FLAG-Vid^{C}-H3$_{13-135}$-GFP

GDYKDDDDKG MIEEKKVTVQ ELRELYLSGE YTIEIDTPDG YQTIGKWFDK GVLSMVRVAT
ATYETVCAFN HMIQLADNTW VQACELDVGV DIQTAAGIQP VMLVEDTSDA ECYDFEVMHP
NHRYYGDGIV SHNSGKAPRK QLATKAARKS APATGGVKKP HRYRPGTVAL REIRRYQKST
ELLIRKLPFQ RLVREIAQDF KTDLRFQSSA VMALQEACEA YLVGLFEDTN LCAIHAKRVT
IMPKDIQLAR RIRGERAKGS GMVSKGEELF TGVVPILVEL DGDVNGHKFS VSGEGEGDAT
YGKLTLKFIC TTGKLPVPWP TLVTTLTYGV QCFSRYPDHM KQHDFFKSAM PEGYVQERTI
FFKDDGNYKT RAEVKFEGDT LVNRIELKGI DFKEDGNILG HKLEYNYNSH NVYIMADKQK
NGIKVNFKIR HNIEDGSVQL ADHYQQNTPI GDGPVLLPDN HYLSTQSALS KDPNEKRDHM
VLLEFVTAAG ITLGMDELYK

d. FLAG-H3$_{1-28}$-Vid^{C}-H3$_{35-135}$-GFP

GDYKDDDDKG ARTKQTARKS TGGKAPRKQL ATKAARKSMI EEKKVTVQEL RELYLSGEYT
IEIDTPDGYQ TIGKWFDKGV LSMVRVATAT YETVCAFNHM IQLADNTWVQ ACELDVGVDI
QTAAGIQPVM LVEDTSDAEC YDFEVMHPNH RYYGDGIVSH NSGKVKKPHR YRPGTVALRE
IRRYQKSTEL LIRKLPFQRL VREIAQDFKT DLRFQSSAVM ALQEACEAYL VGLFEDTNLC
AIHAKRVTIM PKDIQLARRI RGERAKGSGM VSKGEELFTG VVPILVELDG DVNGHKFSVS
GEGEGDATYG KLTLKFICTT GKLPVPWPTL VTTLTYGVQC FSRYPDHMKQ HDFFKSAMPE

GYVQERTIFF KDDGNYKTRA EVKFEGDTLV NRIELKGIDF KEDGNILGHK LEYNYNSHNV
YIMADKQKNG IKVNFKIRHN IEDGSVQLAD HYQQNTPIGD GPVLLPDNHY LSTQSALSKD
PNEKRDHMVL LEFVTAAGIT LGMDELYK

e. FLAG-Vid^{C}-HP1α $_{19-135}$

GDYKDDDDKG MIEEKKVTVQ ELRELYLSGE YTIEIDTPDG YQTIGKWFDK GVLSMVRVAT
ATYETVCAFN HMIQLADNTW VQACELDVGV DIQTAAGIQP VMLVEDTSDA ECYDFEVMHP
NHRYYGDGIV SHNSGYVVEK VLDRRVVKGQ VEYLLKWKGF SEEHNTWEPE KNLDCPELIS
EFMKKYKKMK EGENNKPREK SESNKRKSNF SNSADDIKSK KKREQSNDIA RGFERGLEPE
KIIGATDSCG DLMFLMKWKD TDEADLVLAK EANVKCPQIV IAFYEERLTW HAYPEDAENK
EKETAKS

f. FLAG-Alk2-Vid^{C}-ΔHP1α-GFP

GDYKDDDDKG MVDGVMILPV LIMIALPSPS MEDEKPKVNP KLYMCVCEGL SCGNEDHCEG
QQCFSSLSIN DGFHVYQKGC FQVYEQGKMT CKTPPSPGQA VECCQGDWCN RNITAQLPTK
GKSFPGTQNF HLEVGLIILS VVFAVCLLAC LLGVALRKFK RRNQERLNPR DVEYGTIEGL
ITTNVGDSTL ADLLDHSCTS GSGSGLPFLV QRTVARQITL LECVGKGRYG EVWRGSWQGE
NVAVKIFSSR DEKSWFXETE LYNTVMLRHE NILGFIASDM TSRHSSTQLW LITHYHEMGS
LYDYLQLTTL DTVSCLRIVL SIASGLAHLH IEIFGTQGKP AIAHRDLKSK NILVKKNGQC
CIADLGLAVM HSQSTNQLDV GNNPRVGTKR YMAPEVLDET IQVDCFDSYK RVDIWAFGLV
LWEVARRMVS NGIVEDYKPP FYDVVPNDPS FEDMRKVVCV DQQRPNIPNR WFSDPTLTSL
AKLMKECWYQ NPSARLTALR IKKTLTKIDN SLDKLKTDCG SGMIEEKKVT VQELRELYLS
GEYTIEIDTP DGYQTIGKWF DKGVLSMVRV ATATYETVCA FNHMIQLADN TWVQACELDV
GVDIQTAAGI QPVMLVEDTS DAECYDFEVM HPNHRYYGDG IVSHNSGKKT KRTADSSSSE
DEEEYVVEKV LDRRVVKGQV EYLLKWKGFS EEHNTWEPEK NLDCPELISE FMKKYKKMKE
GENNKPREKS ESNKRKSNFS NSADDIKSKK KREQSNDIAR GFERGLEPEK IIGATDSCGD
LMFLMKWKDT DEADLVLAKE ANVKCPQIVI AFYEERLTWH AYPEDAENKE KETAKSGSGM
VSKGEELFTG VVPILVELDG DVNGHKFSVS GEGEGDATYG KLTLKFICTT GKLPVPWPTL
VTTLTYGVQC FSRYPDHMKQ HDFFKSAMPE GYVQERTIFF KDDGNYKTRA EVKFEGDTLV
NRIELKGIDF KEDGNILGHK LEYNYNSHNV YIMADKQKNG IKVNFKIRHN IEDGSVQLAD
HYQQNTPIGD GPVLLPDNHY LSTQSALSKD PNEKRDHMVL LEFVTAAGIT LGMDELYK

Kanamycin resistance assay

a. myc-Kan^{RN}-Vid^{N} (X = -2 position; Z = -1 position)

MEQKLISEED LSHIQRETSC SRPRLNSNMD ADLYGYKWAR DNVGQSGATI YRLYGKPDAP
ELFLKHGKGS VANDVTDEMV RLNWLTEFMP LPTIKHFIRT PDDAWLLTTA IPGKTAFQVL
EEYPDSGENI VDALAVFLRR LHSIPVCNCP FNSDRVFRLA QAQSRMNNGL VDASDFDDER
NGWPVEQVWK EMHKLLPFE**X** **Z**CLPKEAVVQ IRLTKKG

b. Vid^{C} -Kan^{RC} (X = +2 position)

MIEEKKVTVQ ELRELYLSGE YTIEIDTPDG YQTIGKWFDK GVLSMVRVAT ATYETVCAFN
HMIQLADNTW VQACELDVGV DIQTAAGIQP VMLVEDTSDA ECYDFEVMHP NHRYYGDGIV
SHNS**X**KSVVT HGDFSLDNLI FDEGKLIGCI DVGRVGIADR YQDLAILWNC LGEFSPSLQK
RLFQKYGIDN PDMNKLQFHL MLDEFF

Appendix 2. Uncropped western blots

Fig. 1b

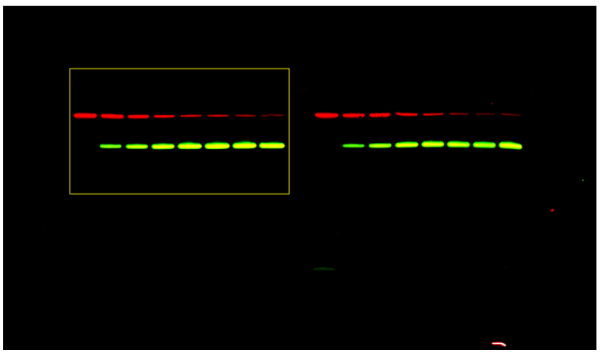


Fig. 3a

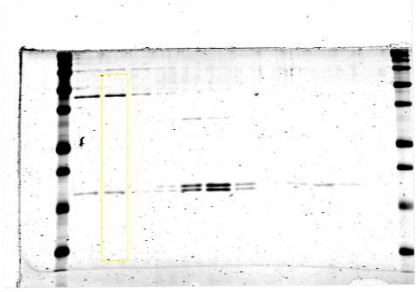
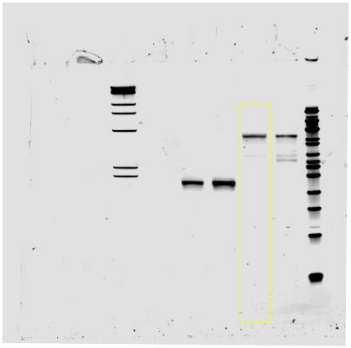


Fig. 3b



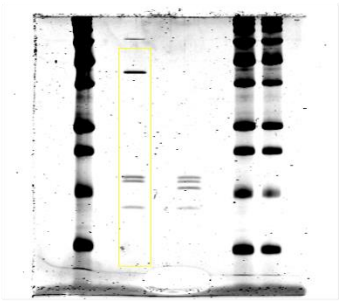


Fig. 3c

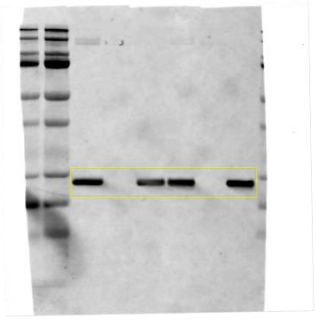


Fig. 4b

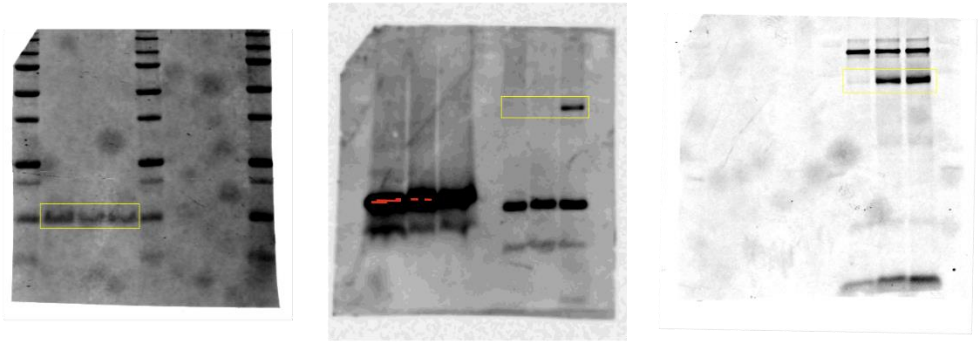


Fig. 4d

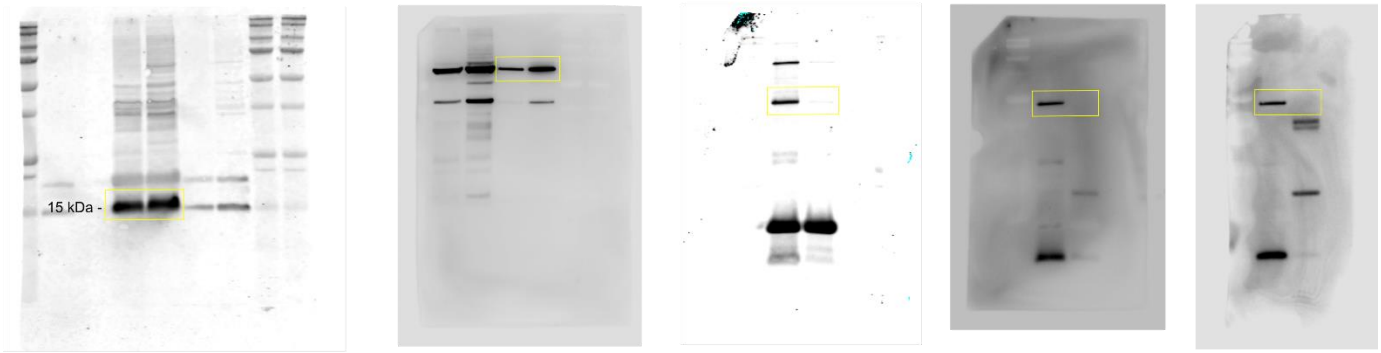


Fig. 5b

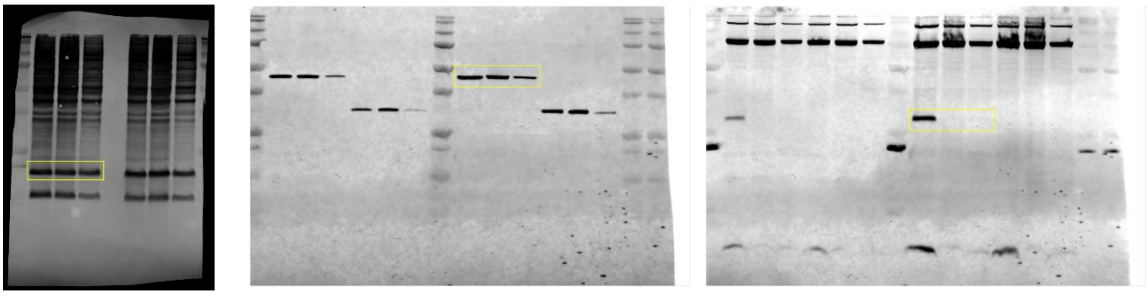
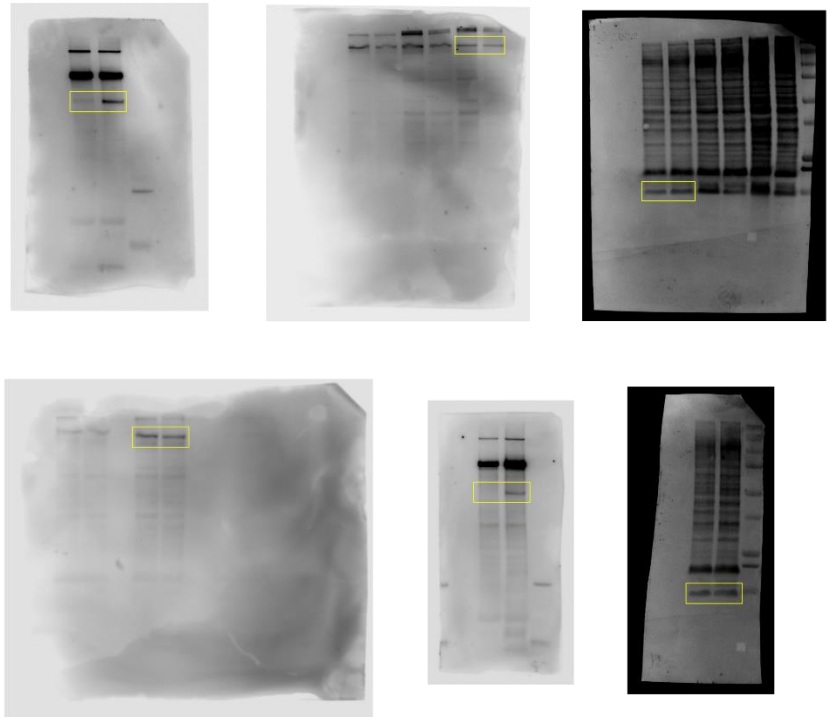


Fig. 5e



References

1. Shapiro G, Swoboda R, & Stauss U (1994) Fmoc solid phase synthesis of serine phosphopeptides via selective protection of serine and on resin phosphorylation. *Tetrahedron Lett.* 35(6):869-872.
2. Lockless SW & Muir TW (2009) Traceless protein splicing utilizing evolved split inteins. *Proc. Natl. Acad. Sci. U.S.A.* 106(27):10999-11004.
3. Stevens AJ*, et al.* (2017) A promiscuous split intein with expanded protein engineering applications. *Proc. Natl. Acad. Sci. U.S.A.* 114(32):8538-8543.
4. Kabsch W (2010) Xds. *Acta Crystallogr. D Biol. Crystallogr.* 66(Pt 2):125-132.
5. Emsley P, Lohkamp B, Scott WG, & Cowtan K (2010) Features and development of Coot. *Acta Crystallogr. D Biol. Crystallogr.* 66(Pt 4):486-501.
6. Liebschner D*, et al.* (2019) Macromolecular structure determination using X-rays, neutrons and electrons: recent developments in Phenix. *Acta Crystallogr. D Struct. Biol.* 75(Pt 10):861-877.
7. Winn MD*, et al.* (2011) Overview of the CCP4 suite and current developments. *Acta Crystallogr. D Biol. Crystallogr.* 67(Pt 4):235-242.
8. Minor W, Cymborowski M, Otwinowski Z, & Chruszcz M (2006) HKL-3000: the integration of data reduction and structure solution--from diffraction images to an initial model in minutes. *Acta Crystallogr. D Biol. Crystallogr.* 62(Pt 8):859-866.
9. Shevchenko A, Tomas H, Havlis J, Olsen JV, & Mann M (2006) In-gel digestion for mass spectrometric characterization of proteins and proteomes. *Nat. Protoc.* 1(6):2856-2860.
10. Eng JK, McCormack AL, & Yates JR, 3rd (1994) An approach to correlate tandem mass spectral data of peptides with amino acid sequences in a protein database. *J. Am. Soc. Mass Spectrom.* 5(11):976-989.
11. Nesvizhskii AI, Keller A, Kolker E, & Aebersold R (2003) A statistical model for identifying proteins by tandem mass spectrometry. *Anal. Chem.* 75(17):4646-4658.