

Structure of the Branched Intermediate in Protein Splicing

Supporting Information Appendix

Zhihua Liu^{a,1}, Silvia Frutos^{b,1,2}, Matthew J. Bick^{c,1,3}, Miquel Vila-Perelló^a, Galia T. Debelouchina^a, Seth A. Darst^c and Tom W. Muir^{a,4}

^a Frick Laboratory, Department of Chemistry, Princeton University, Princeton, NJ 08544

^b The Laboratory of Synthetic Protein Chemistry, The Rockefeller University, New York City, NY 10065

^c The Laboratory of Molecular Biophysics, The Rockefeller University, New York City, NY 10065

¹ Z.L., S.F., and M.J.B. contributed equally to this work

² Present address: CIBER-BBN de Bioingeniería, Biomateriales y Nanomedicina, Instituto de Sanidad Carlos III, Madrid 28029, Spain

³ Present address: Molecular Engineering and Sciences, University of Washington, Seattle, WA 98195

⁴ Corresponding author: muir@princeton.edu, Tel: 609-258-5778

Supporting Information Appendix Table of Contents

General Materials	3
General Equipment	4
Peptide Synthesis	5
Cloning of plasmids for protein expression	9
Protein expression and purification.....	10
Buffers used for Ni-NTA affinity purification.....	10
Generation of constructs 1-5 by expressed protein ligation	12
Crystallization, structure determination and refinement of construct 1.....	12
Branch resolution kinetic analysis.	13
Mxe GyrA(S179C) protein splicing kinetic analysis.	14
C-extein cleavage kinetic analysis in constructs 4 and 5.	14
Mxe GyrA(C1S, S179C) C-extein cleavage and N-extein cleavage analysis	15
Nuclear magnetic resonance (NMR) spectroscopy.....	15
Supplemental note on the synthesis of bis-ester containing branched peptide	16
Table S1. X-ray data collection statistics	17
Figure S1: 2IFol – lFcl map of Mxe GyrA branched intermediate active site.	18
Figure S2: Overlay of the terminal Asn and the penultimate His side-chain conformations among reported intein crystal structures.....	19
Figure S3: Splicing activity of Mxe GyrA wt and Mxe GyrA(S179C).....	20
Figure S4: ¹ H- ¹⁵ N HSQC spectra of refolded construct 2 (red) and 3(black).....	21
Figure S5: Kinetic analysis of intein-succinimide formation in branched constructs 2 and 3.	22
Figure S6: Kinetic analysis of the protein splicing of wt Mxe GyrA.....	23
Figure S7: Analysis of C-extein cleavage and N-extein cleavage activities of Mxe GyrA intein (C1S, S179C).	24
Figure S8: Kinetic analysis of C-extein cleavage (succinimide formation) in linear Mxe GyrA constructs 4 and 5.	25
Figure S9: Synthesis of the GyrA peptide used for preparing construct 1.....	26
Figure S10: Synthesis of the GyrA peptide used for preparing construct 2.....	27
Figure S11: Synthesis of the GyrA peptide used for preparing construct 3.....	28
Figure S12: Synthesis of the GyrA peptide used for preparing construct 4.....	29
Figure S13: Synthesis of the GyrA peptide used for preparing construct 5.....	30
References:	31

General Materials

All buffering salts, media for bacteria culture, isopropyl- β -D-thiogalactopyranoside (IPTG), and *N,N*-diisopropylethylamine (DIPEA) were purchased from Fisher Scientific (Pittsburgh, PA). Sodium 2-mercaptoethanesulfonate (MESNa), ethanedithiol (EDT), Coomassie brilliant blue, *N,N*-dimethylformamide (DMF), triisopropylsilane (TIPS), 4-mercaptophenyl acetic acid (MPAA), tetrakis(triphenylphosphine) palladium(0) Pd(PPh₃)₄, Phenylsilane, were purchased from Sigma-Aldrich (St. Louis, MO). Tris(2-carboxyethyl)phosphine hydrochloride (TCEP) was purchased from Thermo Scientific (Rockford, IL). *N*^α-Fmoc protected amino acids and *N*^α-Boc protected amino acids were purchased from Novabiochem (Läufelfingen, Switzerland) or ATGC Bioproducts (Wilmington, MA). L- α -hydroxyisovaleric acid (*hVal*) was purchased from Bachem (Torrance, CA). Piperidine was purchased from Alfa Aesar (Ward Hill, MA). 2-(1H-Benzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate (HBTU) and 1-Hydroxybenzotriazole hydrate (HOBt) were purchased from AnaSpec (Fremont, Ca). Trifluoroacetic acid (TFA) was purchased from Halocarbon (North Augusta, SC). Restriction protease factor Xa (FXa) and complete protease inhibitor tablets were purchased from Roche Diagnostics (Mannheim, Germany). Nickel-nitrilotriacetic acid (Ni-NTA) resin was from Novagen (Gibbstown, NJ). ¹⁵NH₄Cl was purchased from Cambridge Isotope Laboratories (Andover, MA). The QuikChange XL II site directed mutagenesis kit was from Agilent (La Jolla, CA). Sub-cloning efficiency DH5 α cells and One Shot BL21 (DE3) chemically competent *E. coli* were purchased from Invitrogen (Carlsbad, CA) and used to generate “in-house” high-competency cell lines. Oligonucleotides were purchased from Integrated DNA Technologies (Coralville, IA). All plasmids used in this study were sequenced by GENEWIZ. Criterion XT Bis-Tris gels (12%), and Bradford reagent dye concentrate were purchased from Bio-Rad (Hercules, CA). 20x MES-SDS running buffer was purchased from Boston Bioproducts (Ashland, MA).

General Equipment

Size-exclusion chromatography (SEC) was carried out on ÄKTA FPLC or Purifier systems from GE Healthcare (Piscataway, NJ). Preparative SEC was carried out on a Superdex 75 10/300 column (GE Healthcare). For all runs, proteins were eluted over 1.35 CV of buffer (flow rate: 0.5 mL/min). Both semi-preparative and analytical RP-HPLC were performed on Agilent 1100 and 1200 series instruments. Analytical RP-HPLC was carried out on a C18 Vydac column (5 μm , 4.6 x 150 mm) or a C4 Vydac column (5 μm , 4.6 x 150 mm) at a flow rate of 1 mL/min. Semi-preparative RP-HPLC was carried out on a C18 Vydac 218TP152010 column (15-20 μm ; 10 x 250 mm) or a C4 Vydac 214TP510 column (5 μm ; 10 x 250 mm) at a flow rate of 3.5 mL/min. All runs used 0.1 % TFA (trifluoroacetic acid) in water (solvent A) and 90 % acetonitrile in water with 0.1% TFA (solvent B). For analytical and semi-preparative RP-HPLC runs, a two minutes isocratic period in initial conditions was followed by a 30 minutes linear gradient with increasing solvent B concentration. Electrospray ionization mass spectrometric analysis (ESI-MS) was performed on a Bruker Daltonics MicroTOF-Q II mass spectrometer, or on a Sciex-API-100 single quadrupole spectrometer. External calibration was performed immediately before data acquisition using ESI-L Low Concentration Tuning Mix (Agilent). Samples for ESI-MS analysis were prepared by collecting the entire peak during RP-HPLC analysis, followed by a 10-fold dilution in MS buffer (50% acetonitrile, 0.1% formic acid in H₂O) and directly infusing this solution into the mass spectrometer. The acquired intensity vs m/z data was deconvoluted using the Maximum Entropy algorithm (Spectrum Square Associates, Ithaca, NY) to give the observed average molecular weight. Calculated average molecular weights were obtained using Protein Identification and Analysis Tools on the ExPASy Server (<http://web.expasy.org/protparam/>). Cells were lysed using an S-450D Branson Digital Sonifier or a French Press. Coomassie-stained gels were imaged on a LI-COR Odyssey Infrared Imager or a GE ImageQuant LAS 4010 imager. NMR experiments were carried out at a Bruker Avance 800 spectrometer equipped with a TCI CryoProbe.

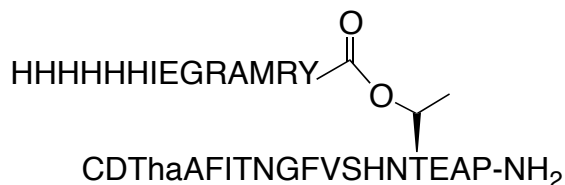
Peptide Synthesis

Peptides were synthesized using either Fmoc SPPS (solid phase peptide synthesis) on Rink-amide ChemMatrix resin (0.5 mmol/g, PCAS BioMatrix Quebec, Canada) or Boc SPPS on MBHA resin (0.5 mmol/g, Creosalus Louisville, KY).

Fmoc SPPS was carried on a Liberty Synthesizer (CEM, Matthews, NC) unless otherwise stated. Chain assembly was carried out with HBTU (4.9 eq.) and HOBt (4.9 eq.) activation using a 5-fold excess of standard N^α -Fmoc protected amino acid over the resin in DMF (dimethylformamide) with DIEA (N,N -diisopropylethylamine, 8 eq.). The Fmoc protecting group was removed with 20% piperidine in DMF. Peptides were cleaved from the resin using 94% TFA, 1% triisopropylsilane (TIS), 2.5% ethanedithiol, and 2.5% H₂O (cleavage cocktail). Crude peptide products were precipitated and washed with cold Et₂O, dissolved in solvent A with a minimal amount of solvent B and then purified by RP-HPLC and characterized by ESI-MS.

Boc SPPS was performed manually in a reaction glass vessel with a frit. Chain assembly was carried out with HBTU (5 eq.) using a 5.2-fold excess of standard N^α -Boc protected amino acid over the resin in DMF with DIEA (8 eq.). The Boc protecting group was removed with 100% TFA (1 minute x 2). Peptides were cleaved from the resin using 10 mL of anhydrous HF with 400 μ L *p*-cresol at 0 °C for 1hr. After cleavage, crude peptide products were precipitated and washed with cold Et₂O, dissolved in solvent A with a minimal amount of solvent B and then purified by RP-HPLC and characterized by ESI-MS.

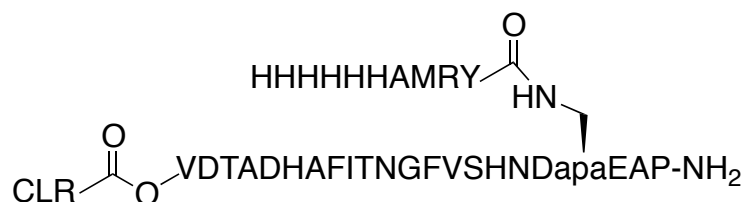
Synthesis of the GyrA peptide used in the semi-synthesis of construct 1:



The synthetic route employed is summarized in Figure S9. The linear chain Cys¹⁸⁵-Pro⁺⁴ (residue

numbers refer to GyrA intein + C-extein) was assembled using Fmoc SPPS as described above. His187 was replaced by Fmoc- β -2-thienyl-Ala-OH (abbreviated as Tha). The Thr at the +1 position was introduced without side-chain protection to allow elongation of the branched chain. Any premature, undesired Thr side-chain acylation during the linear chain assembly was cleaved with ethanolamine in 5% water/DMF (2 x 30 min). Fmoc-Tyr(Trt)-OH (10 eq.) was coupled to the β -hydroxyl of the Thr side-chain using the symmetric anhydride method with DIC (5 eq.) and DMAP (dimethylaminopyridine, 0.1 eq.) for 3 hr. This was followed by an overnight coupling using Fmoc-Tyr(Trt)-OH (4 eq.) activated with MSNT (1-(2-mesi-tylenesulfonyl)-3-nitro-1H-1,2,4-triazole, 4 eq.) and NMI (N-methylimidazole, 12 eq.) in DCM/DMF (50:50). All Fmoc deprotections during elongation of the branched peptide were performed with 2% DBU (1,8-diazabicyclo [5.4.0] undec-7-ene) to prevent ester cleavage by piperidine. All subsequent residues were manually coupled using DIC/HOBt activation for 1 hr each: 5 eq. of amino acid, 5 eq. of DIC and 5 eq. of HOBt. Following cleavage, the crude peptide was purified by RP-HPLC (semiprep C₁₈ column, gradient of 20-40% solvent B) using the method described above. The purity and the molecular weight were confirmed by analytical RP-HPLC and ESI-MS (Figure S9). The isolated yield after synthesis and purification was ~15%.

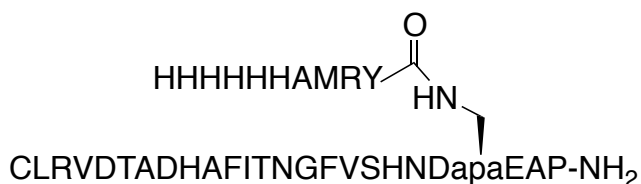
Synthesis of the peptide used in the semi-synthesis of construct 2:



The synthetic route employed is summarized in Figure S10. Residues *h*Val¹⁸²-Pro⁺⁴ were assembled using Fmoc SPPS as described above, except Fmoc-Dapa(Alloc)-OH was used in the synthesis. After completion of the linear peptide *h*Val¹⁸²-Pro⁺⁴, Alloc was selectively removed by treatment with Pd(PPh₃)₄ (0.1 eq.), phenylsilane (25 eq.) in dry DCM for 2 x 30 min. The branched

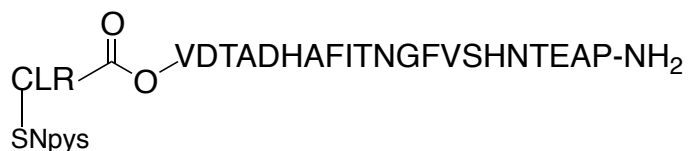
sequence was assembled off the Dapa side-chain using Fmoc SPPS methods employing Boc-His(Trt)-OH in the final cycle. Undesired *h*Val acylation was then reversed by treatment with ethanolamine in 5% water/DMF (2 x 30 min). Fmoc-Arg(pbf)-OH (10 eq.) was coupled manually to the α -hydroxyl group of *h*Val using the symmetric anhydride method with DIC and (5 eq.), DMAP (0.1 eq.) for 3 hr, followed by a second overnight coupling using Fmoc-Arg(pbf)-OH (5 eq.), MSNT (5 eq.) and NMI (15 eq.). The Fmoc group was deprotected with 2% DBU when an ester is present in peptide backbone. Fmoc-Leu-OH and Boc-Cys(Trt)-OH were manually coupled using DIC/HOBt activation for 15 min each: 5 eq. of amino acid, 5 eq. of DIC and 5 eq. of HOBt. Following cleavage, the crude peptide was purified by RP-HPLC (semiprep C₁₈ column, gradient of 22.5-32.5% solvent B) using the method described above. The purity and the molecular weight were confirmed by analytical RP-HPLC and ESI-MS (Figure S10). The isolated yield after synthesis and purification was ~8 %.

Synthesis of the peptide used in the semi-synthesis of construct 3:



The synthetic route employed is summarized in Figure S11. Residues Cys¹⁷⁹-Pro⁺⁴ were assembled using Fmoc SPPS employing Fmoc-Dapa(Alloc)-OH and Boc-Cys(Trt)-OH in the synthesis. At this point, Alloc was selectively removed by treatment with Pd(PPh₃)₄ (0.1 eq.), phenylsilane (25 eq.) in dry DCM for 2 x 30 min. The branched sequence was assembled off the Dapa side-chain using Fmoc SPPS methods. Following cleavage, the crude peptide was purified by RP-HPLC (semiprep C₁₈ column, gradient of 22.5-32.5% solvent B) using the method described above. The purity and the molecular weight were confirmed by analytical RP-HPLC and ESI-MS (Figure S11). The isolated yield after synthesis and purification was ~10%.

Synthesis of the peptide used in the semi-synthesis of construct 4:



The synthetic route employed is summarized in Figure S12. Residues *hVal*¹⁸²-Pro⁺⁴ were assembled using Boc SPPS methods as described above. Boc-Arg(Tos)-OH (10 eq.) was coupled manually to the α -hydroxyl group of *hVal* side chain using the symmetric anhydride method with DIC and (5 eq.), DMAP (dimethylaminopyridine, 0.1 eq.) for 3 hr x 2, followed by a second overnight coupling using Boc-Arg(Tos)-OH (5 eq.), MSNT (5 eq.) and NMI (15 eq.). Boc-Leu-OH and Boc-Cys(SNpys)-OH were manually coupled using DIC/HOBt activation for 15 min each: 5 eq. of amino acid, 5 eq. of DIC and 5 eq. of HOBt. Following cleavage, the crude peptide was purified by RP-HPLC (semiprep C₁₈ column, gradient of 22.5-32.5% solvent B) using the method described above. The purity and the molecular weight were confirmed by analytical RP-HPLC and ESI-MS (Figure S12). The isolated yield after synthesis and purification was ~ 2 %.

Synthesis of the peptide used in the semi-synthesis of construct 5:



The synthetic route employed is summarized in Figure S13. The peptide was prepared using Fmoc SPPS methods as described above, except Fmoc-Cys(Trt)-OH was used. Following cleavage, the crude peptide was purified by RP-HPLC (semiprep C₁₈ column, gradient of 22.5-32.5% solvent B) using the method described above. The purity and the molecular weight were confirmed by analytical RP-HPLC and ESI-MS (Figure S13). The isolated yield after synthesis and purification was ~15 %.

Cloning of plasmids for protein expression

The DNA encoding the *Mxe* GyrA intein (residues 1-184, C1A, T184A) was PCR amplified from the pTXB1 vector to incorporate an additional N-terminal His₆-tag and the Factor Xa recognition site. This PCR product was re-cloned back into the NdeI and SapI restrictions sites of the pTXB1 vector, which contain a mutated *Mxe* GyrA intein and the chitin-binding domain. The C1A mutation was then introduced using Quick-Change mutagenesis kit (Stratagene) to generate the final fusion protein (His₆-FXa-GyrA(C1A, T184A, 1-184)-GyrA-CBD) used to prepare branched constructs.

His₆-FXa-GyrA(C1A, T184A, 1-184)-GyrA-CBD:

MASSRVDGGRHHHHHGMASMTGGQQMGATLYDDDDKDRWGSGHIEGRAITGDALV
ALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGPNVLADRLFHSGEHPVYTVRTVE
GLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGK
PEFAPTTYTVGVPGLVRFLEAHHRDPDAQIADELTDGRFYAKVASVTDAGVQPV
YSLRVDACITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGPNVLADRLFHS
GEHPVYTVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSV
DCAGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQIADELTDGRFYAKVASVT
DAGVQPVYSLRVDADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAYTAGQL
VTYNGKTYKCLQPHTSLAGWEPSNVPALWQLQ

GyrA(C1A, 1-184) sequence is indicated in bold and the FXa recognition site is underlined.

The GyrA(1-178, C1A)-GyrA-CBD plasmids were derived from the GyrA(1-184)-GyrA-CBD pTXB1 plasmid mentioned above. The gene encoding S179-T184 was deleted using Quick-Change mutagenesis kit (Stratagene). The resulting plasmids encoded for the following sequences:

His₆-FXa-GyrA(C1A, 1-178)-GyrA-CBD:

MASSRVDGGRHHHHHGMASMTGGQQMGATLYDDDDKDRWGSGHIEGRAITGDALV
ALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGPNVLADRLFHSGEHPVYTVRTVE
GLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFARGK
PEFAPTTYTVGVPGLVRFLEAHHRDPDAQIADELTDGRFYAKVASVTDAGVQPV
YCITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGPNVLADRLFHSGEHPVY
TVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAGFA
RGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQIADELTDGRFYAKVASVTDAGVQP
VYSLRVDADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAYTAGQLVTYNGK
TYKCLQPHTSLAGWEPSNVPALWQLQ

GyrA(C1A, 1-178) sequence is indicated in bold and the FXa recognition site is underlined.

His₆-FXa-AMRY-GyrA(C1S, 1-178)-GyrA-CBD:

MASSRVDGGRHHHHHGMASMTGGQQMGATLYDDDDKDRWGSGHIGRAMRYSIT

**GDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHSGEHPVYT
 VRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSVDCAG
 FARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQIADELTDGRFYAKVASVTDA
 GVQPVY CITGDALVALPEGESVRIADIVPGARPNSDNAIDLKVLDRHGNPVLADRLFHS
 GEHPVYTVRTVEGLRVTGTANHPLLCLVDVAGVPTLLWKLIDEIKPGDYAVIQRSAFSV
 DCAGFARGKPEFAPTTYTVGVPGLVRFLEAHHRDPDAQIADELTDGRFYAKVASVT
 DAGVQPVYSLRVDTADHAFITNGFVSHATGLTGLNSGLTTNPGVSAWQVNTAYTAGQL
 VTYNGKTYKCLQPHTSLAGWEPSNPALWQLQ**

AMRY-GyrA(C1A, 1-178) sequence is indicated in bold and the FXa recognition site is underlined.

Protein expression and purification.

Buffers used for Ni-NTA affinity purification

Buffer	Composition
Lysis	50 mM phosphate, 300 mM NaCl, 5 mM imidazole, pH 8.0
Isolation	50 mM phosphate, 300 mM NaCl, 5 mM imidazole, 1% Triton, pH 8.0
Resuspension	50 mM phosphate, 300 mM NaCl, 5 mM imidazole, 7 M urea, pH 8.0
Wash	50 mM phosphate, 300 mM NaCl, 20 mM imidazole, 7 M urea, pH 8.0
Elution	50 mM phosphate, 300 mM NaCl, 250 mM imidazole, 7 M urea, pH 8.0

Over-expression and purification of His6-FXa-GyrA(C1A, T184A, 1-178)-GyrA-CBD, His6-FXa-GyrA(C1A, 1-178)-GyrA-CBD, His6-FXa-AMRY-GyrA(C1A, 1-178)-GyrA-CBD, Flag-AMRY-GyrA-TEAP-His6, Flag-AMRY-GyrA(S179C)-TEAP-His6, Flag-AMRY-GyrA(C1S, S179C)-TEAP-His6

E. coli BL21(DE3) cells transformed with the desired plasmid were grown in 1 L of LB containing 100 µg/mL of ampicillin at 37°C until OD₆₀₀ = 0.6. Protein expression was induced by addition of 0.5 mM IPTG for 3 hours at 37 °C. After harvesting the cells by centrifugation (10,500 rcf, 30 min), the cell pellets were transferred to 50 mL conical tubes with 5 mL of Lysis Buffer and stored at -80°C. The cell pellets were resuspended by adding an additional 15 mL of Lysis Buffer supplemented with Complete protease inhibitor cocktail. Cells were lysed by sonication (35% amplitude, 8x 15 second pulses separated by 60 seconds on ice). The soluble fraction was discarded after centrifugation (35,000 rcf, 30 min) and the insoluble fraction was washed with Isolation Buffer. The resulting pellet from 1 L LB media was extracted with 20 mL of Resuspension Buffer overnight at 4 °C. The mixture was

centrifuged at 35,000 ref for 30 minutes, and the supernatant was mixed with 2 mL of Ni-NTA resin, previously equilibrated with Resuspension Buffer, 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) x 2 of Resuspension Buffer, 5 CV x 2 of Wash Buffer. The protein was eluted with Ni Column Elution Buffer in four 1.5 CV elution fractions.

Generation and purification of GyrA(C1A, T184A, 1-184)-MES, GyrA(C1A, 1-178)-MES, AMRY-GyrA(C1A, 1-178)-GyrA-MES

Ni-affinity purified fusion protein was diluted to 1 mg/mL with a buffer containing 100 mM sodium phosphate (Pi), 150 mM NaCl, 7 M urea, pH 8.0. The protein was subsequently refolded by dialyzing against the same buffer but containing 4 M urea, 2 M urea, and 0 M urea, respectively. The dialysis was carried out at 4 °C for 2 hr each. A thiolysis cocktail consisting of 100 mM MESNa, 1 mM TCEP was added after the final dialysis, and the pH was adjusted to 8.0. The reaction was allowed to proceed for 24-48 hr under argon. The reaction mixture was passed through Ni-NTA resin, previously equilibrated with a buffer containing 100 mM Pi, 150 mM NaCl, 100 mM MESNa, pH 8.0. The resin was washed with 2 x 5 CV of the same buffer but containing 1 M NaCl, and then washed with 2 x 5 CV of the same buffer containing 1% Tween 20. The protein α -thioester and GyrA-CBD were eluted with 6 M Gn HCl, 100 mM MESNa and 0.2 M AcOH in four 1.5 CV elution fractions. Protein α -thioester was purified by RP-HPLC over a semi-preparative C₄ column using a gradient of 42.5-55% buffer B in 30 min. Pure α -thioester products were pooled, lyophilized and then refolded by a quick dialysis from a buffer containing 50 mM HEPES, 100 mM NaCl, 100 mM MESNa, 6 M urea, pH 7.0 against 50 mM HEPES, 100 mM NaCl, 100 mM MESNa, pH 7.0 at 4 °C. The N-terminal His₆-tag was proteolytically removed with Factor Xa (1U enzyme/500 mg protein for 24 h at 4 °C). The final protein α -thioester was purified by RP-HPLC with a semi-preparative C₄ column using a gradient of 42.5-60% buffer B in 30 min. Pure α -thioester products were pooled and lyophilized. ~5 mg of pure α -thioester-containing

products were obtained from 1L of *E. coli* culture.

Generation of constructs 1-5 by expressed protein ligation

Expressed protein ligation (EPL) was performed by mixing the appropriate synthetic peptide (3 eq.) and protein α -thioester (1 eq, 1 mM) in freshly prepared EPL buffer (100 mM Pi, 100 mM NaCl, 6M Gn HCl, 200 mM MESNa, 50 mM MPAA, 30 mM TCEP, pH 7.9). The reaction progress was followed by analytical RP-HPLC and ESI-MS. All reactions were complete within 2 days. For construct **1**, semi-synthetic proteins were separated from unreacted precursors on a Ni-NTA affinity column. The resin was washed with buffer (100 mM Tris-HCl pH 7.5, 6M Urea, 100 mM NaCl and 1 mM TCEP), the protein eluted from the column in 8 x 0.5 mL fractions of buffer supplemented with 500 mM imidazole. The EPL products were purified by semi-preparative scale RP-HPLC on a C₄ column using a gradient of 42.5 - 55% solvent B in 30 min. Pure protein was lyophilized, and was refolded by step-wise dialysis from a buffer containing 50 mM Tris-HCl pH 7.5, 100 mM NaCl and 6 M urea against 50 mM Tris-HCl pH 7.5, 100 mM NaCl at 4 °C. The His₆-tag from the branch was proteolytically removed with Factor Xa (1U enzyme/500 mg protein for 6 h at 25 °C). The final construct **1** was purified by semi-preparative scale RP-HPLC on C₄ column using a gradient of 42.5 - 55% buffer B in 30min. For construct **2-5**, semi-synthetic proteins were directly purified by semi-preparative scale RP-HPLC on a C₄ column using a gradient of 42.5 - 55% solvent B in 30 min. For all constructs, pure semi-synthetic protein was pooled and lyophilized, and analyzed by analytical RP-HPLC and ESI-MS (yield ~ 40%, Figure 2d).

Crystallization, structure determination and refinement of construct 1

Pure lyophilized construct **1** was refolded at 4 °C by step-wise dialysis from a buffer containing, 50 mM Tris-HCl pH 7.6, 100 mM NaCl, 1 mM TCEP and 6 M urea against the same buffer but containing 4 M urea, 2 M urea and 0 M urea, respectively. Then the sample was concentrated to 10 mg/mL and purified over a Superdex 75 column using 50 mM Tris-HCl pH 7.5, 100 mM NaCl, 1 mM

TCEP as running buffer. The protein eluted as a monomer and was concentrated to a final concentration of 22 mg/mL before crystallization. Protein concentrations were estimated by measuring the absorbance at 280 nm. Crystals of construct **1** were obtained by the hanging drop, vapor-diffusion method at 4 °C by mixing 1 µL protein solution containing 50 mM Tris, 100 mM NaCl, 1 mM TCEP, pH 7.5 with 1 µL of the reservoir solution containing 100 mM sodium cacodylate pH 6.5, 200 mM magnesium acetate and 20% (w/v) PEG 8000. Crystals were cryoprotected by soaking in reservoir solution plus 20% glycerol for 5 minutes prior to flash freezing in liquid nitrogen. Crystals of construct **1** grew to maximum size of approximately 100 x 100 x 100 µm, and displayed numerous cracks throughout. In order to isolate a single lattice during the diffraction experiment, it was necessary to collect the data on a microfocus beamline using a 5 µm aperture at the Advanced Photon Source Northeastern Collaborative Access Team beamline 24ID-E. X-ray diffraction data were integrated and scaled using HKL2000.(1) The structure was solved by molecular replacement using Phaser(2) with the linear *Mxe* GyrA intein structure (PDB code: 1AM2)(3) as a search model. The model was improved by iterative cycles of manual building with Coot(4) and refinement with Phenix(5) against 2.79 Å diffraction data (S.I. Table 1). The final structure contains a nearly complete model of the branched *Mxe* GyrA construct, including the C- and N-exteins, but is missing residues 111-132 and 141-148, which are remote from the active site.

Branch resolution kinetic analysis.

Lyophilized protein constructs **2** and **3** were dissolved in a denaturing buffer (100 mM Pi, 150 mM NaCl, 7 M urea, pH 5.0) to < 1 mg/mL at 4 °C. The protein was subsequently refolded by step-wise dialysis into 100 mM Pi, 150 mM NaCl, pH 5.0, decreasing the urea concentration from 7, to 4, 2 and 0 M. After refolding, succinimide formation was triggered by addition of 1 mM TCEP, and adjusting pH and temperature to 7.5 and 25 °C, respectively. 40 µL aliquots were taken at various time points and the reaction quenched by mixing with 70 µL of 8 M Gn HCl and 4% TFA in water. 100 µL of quenched

sample at each time point was analyzed by RP-HPLC on a C₁₈ analytical column using a gradient of 35-50% solvent B in 30 min at 50 °C. Starting material and product amounts were determined by integration of the 214 nm signal on the RP-HPLC chromatograms (Figure 4a-b).

Mxe GyrA(S179C) protein splicing kinetic analysis.

Ni-column purified *Mxe* GyrA protein constructs Flag-AMRY-GyrA-TEAP-His6 and Flag-AMRY-GyrA(S179C)-TEAP-His6 were refolded as described above. After refolding, 1 mM TCEP was added, and pH and temperature was adjusted to 7.5 and 25 °C, respectively, to initiate the reaction. During the reaction, 20 µL aliquots of the reaction solution were removed at desired time points and quenched in 20 µL of 2x concentrated SDS gel loading dye on ice. Samples were boiled for 5 minutes then centrifuged at 17,000 rcf for 1 minute. Aliquots of starting materials and time points (15 µL) were loaded onto 12% acrylamide Bis-Tris gels and run in MES-SDS running buffer. The gels were Coomassie-stained, and imaged using the Li-Cor Odyssey scanner (Figures S3 and S6). Densitometry analysis was used to quantify the coomassie stained gel.

C-extein cleavage kinetic analysis in constructs 4 and 5.

Lyophilized protein constructs **4** and **5** were refolded as described above. After refolding, the reaction was triggered by addition of 1 mM TCEP, and adjusting pH and temperature to 7.5 and 25 °C, respectively. 40 µL samples were taken at various each time points and the reaction quenched by mixing them with 70 µL of 8 M Gn HCl and 4% TFA in water. 100 uL of quenched sample at each time point was quickly desalted by RP-HPLC with a C₁₈ analytical column using a gradient of 0-20% solvent B in 5 min followed by 100% solvent B for 5 min at 25 °C. Protein products elute with 100% solvent B as a single peak. Protein products were subsequently analyzed and identified by ESI-MS, and starting material and product amounts were determined by integration of ESI-MS chromatograms (Figure 4c-d).

For construct **4** and **5** C-extein cleavage rate calculation (Figure S8), the starting material concentrations were calculated using equation:

$$\text{StartingMaterials\%} = \frac{\text{AMRYGyr}^{\text{A}}\text{TEAP}}{\text{AMRYGyr}^{\text{A}}\text{TEAP} + \text{Gyr}^{\text{A}}\text{Suc} + \text{AMRYGyr}^{\text{A}}\text{Suc}} \times 100\%$$

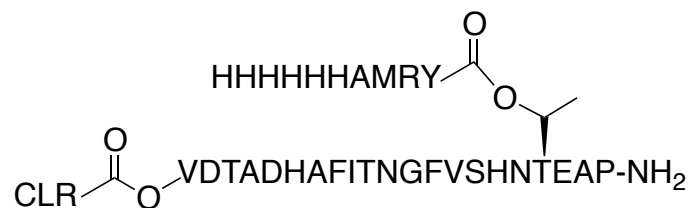
Mxe GyrA(C1S, S179C) C-extein cleavage and N-extein cleavage analysis

Ni-column purified *Mxe* GyrA protein construct Flag-AMRY-GyrA(C1S, S179C)-TEAP-His6 was refolded as described above. After dialysis, 1 mM TCEP was added, and pH and temperature was adjusted to 7.5 and 25 °C, respectively, to initiate the reaction. A typical reaction volume was 500 µL and was carried out in an Eppendorf tube. At each time point, 40 uL of reaction solution was quenched by mixing with 70 uL of 8 M Gn HCl and 4% TFA. 100 uL of quenched sample at each time point was quickly analyzed by RP-HPLC with a C₁₈ analytical column using a gradient of 0-73% solvent B in 30 min at 25 °C. Reaction product peaks were subsequently collected and their identities were confirmed by mass spectrometry (Figure S7). No spliced product was detected under these conditions.

Nuclear magnetic resonance (NMR) spectroscopy.

Lyophilized protein constructs **2** and **3** were dissolved in a denaturing buffer (50 mM Pi, 100 mM NaCl, 7 M urea, pH 5.0) to < 1 mg/mL at 4 °C. The protein was subsequently refolded by step-wise dialysis into 50 mM Pi, 100 mM NaCl, pH 5.0, decreasing the urea concentration from 7, to 4, 2 and 0 M. After dialysis, protein constructs were purified by SEC as described above. Pure fractions were pooled, and concentrated. Two-dimensional ¹H-¹⁵N HSQC (heteronuclear single-quantum correlation spectra) experiments were performed on an Avance III spectrometer operating at a ¹H Larmor frequency of 800 MHz and equipped with a TCI cryoprobe optimized for ¹H detection (Bruker BioSpin Corporation, Billerica, MA). Spectra were acquired at 10 °C, with acquisition times of 184 ms (4096 points) and 25 ms (220 points) in the direct and indirect dimensions, respectively. 320 scans per point were acquired for each sample. Data were processed with the NMRPipe software(6) and visualized in Sparky.(7) (Figure S4)

Supplemental note on the synthesis of bis-ester containing branched peptide



Considerable effort was made to prepare the bis-ester containing branched peptide shown above, but without success. Various SPPS strategies were employed, in each case leading to the generation of complex crude mixtures from which the desired peptide could not be isolated. Interestingly, we observed acylation of the Asn198 side-chain (presumably through an imide linkage) during Boc-SPPS, but only in the presence of the acylated Thr+1 side-chain (i.e. ester branch) – again no such reactivity was observed in the absence of Thr side-chain acylation, or with a protected Asn side-chain (as was the case for the Fmoc-synthesis of Dapa containing peptides).

Table S1. X-ray data collection statistics

Data collection	
Space group	$P4_12_12$
Unit-cell parameters (Å, °)	$a = b = 58.746, c = 111.762,$ $\alpha = \beta = \gamma = 90.0$
Temperature (K)	100
Matthews coefficient (Å ³ Da ⁻¹)	2.56
Solvent content (%)	52.04
Resolution (Å)	50.00 – 2.78
Wavelength (Å)	0.97923
R_{merge} (%)	0.059 (0.188)
$\langle I/\sigma(I) \rangle$	39.6 (9.7)
Completeness (%)	99.7 (96.4)
Reflections (observed)	46397
Reflections (unique)	5310
Redundancy	8.7 (7.2)
Refinement	
Resolution (Å)	40.489 – 2.786
Reflections	5236
R_{work}/R_{free} (%)	22.98 / 26.17
No. of monomers in asymmetric unit	1
No. of non-H atoms	
Total	1342
Protein	1313
Ligand	1
Water	29
B factors (Å ²)	
Overall	35.94
Protein	36.13
Ligand	24.27
Water	27.85
R.m.s. deviations	
Bond lengths (Å)	0.003
Bond angles (°)	0.770
Dihedral angles (°)	14.204
<i>MolProbity</i> validation	
Clashscore (all-atom contacts)	10.01 (97 th percentile)
Poor rotamers (%)	10
Ramachandran outliers (%)	0.00
Ramachandran favored (%)	94.61
C^β deviations > 0.25 Å	0
<i>MolProbity</i> score	2.59 (89 th percentile)
Residues with bad bonds (%)	0.00
Residues with bad angles (%)	0.23 (Gly85, Pro134)

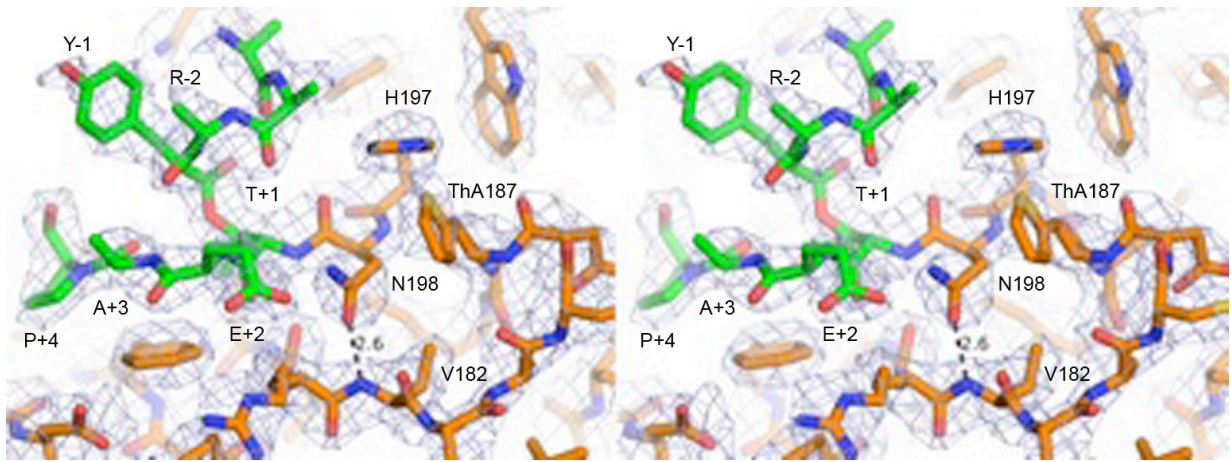


Figure S1: 2IFol – 1Fcl map of Mxe GyrA branched intermediate active site.

Stereo view of the final 2.79-Å-resolution 2IFol – 1Fcl map, contoured at 1σ , of the GyrA intein splicing junction. Key residues are labeled. The C- and N-terminal exteins are colored green.

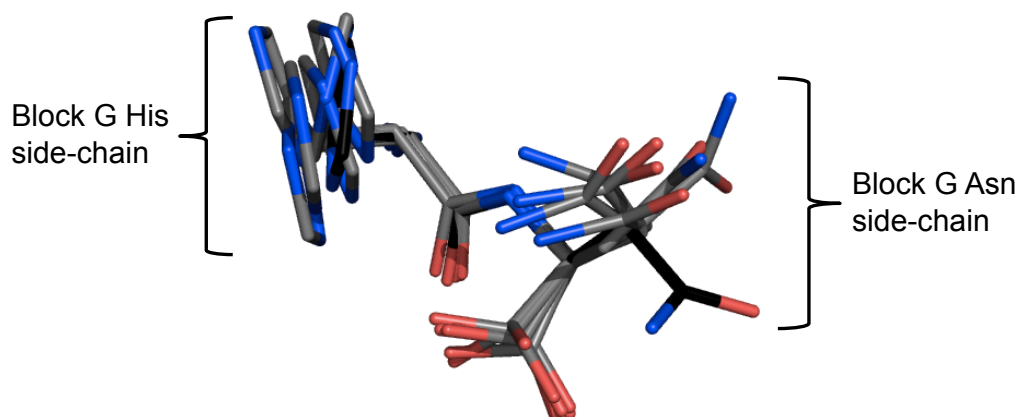


Figure S2: Overlay of the terminal Asn and the penultimate His side-chain conformations among reported intein crystal structures.

Overlay was performed based on C α of the last three residues of the intein using Pymol. All oxygen atoms are colored in red, and nitrogen atoms are colored in blue. Carbon atoms in *Mxe* GyrA branched intermediate are colored in black, and carbon atoms in other intein structures (PDB code 1AM2, 4E2U, 1DQ3, 1UM2, 2CW8, 2IN9, 4E2T)(8-12) colored in grey. The terminal Asn side-chain of the *Mxe* GyrA branched intermediate adopts a unique χ_1 rotamer conformation.

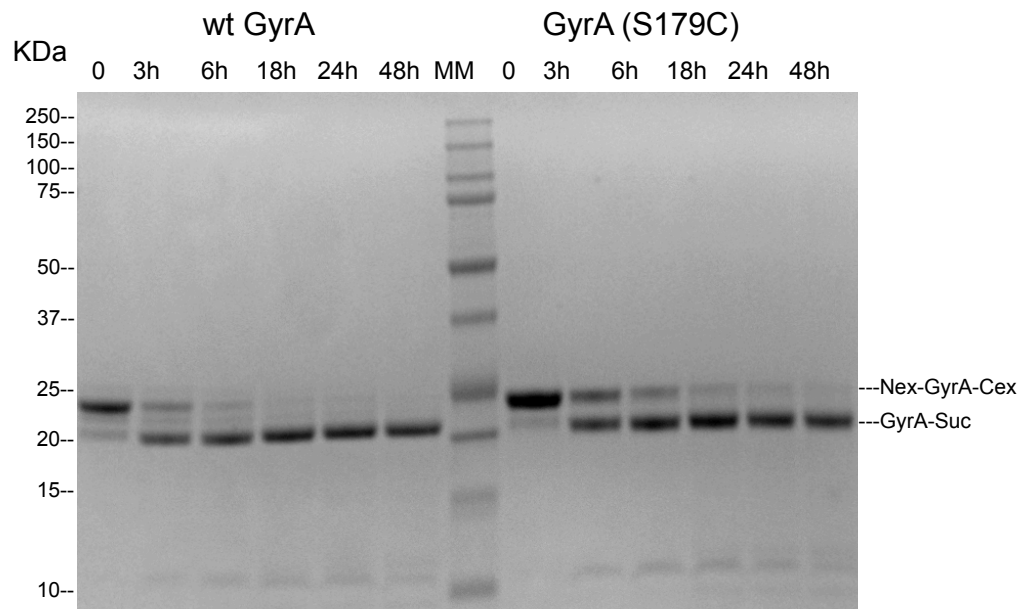


Figure S3: Splicing activity of Mxe GyrA wt and Mxe GyrA(S179C).

SDS-PAGE analysis of protein splicing activity of wt GyrA and GyrA S179C mutant. Gel is stained by coomassie and imaged by ImageQuant LAS 4010. Nex refers to the N-extein (Flag-AMRY), Cex refers to the C-extein (TEAP-His6), GyrA-Suc refers to the GyrA succinimide product.

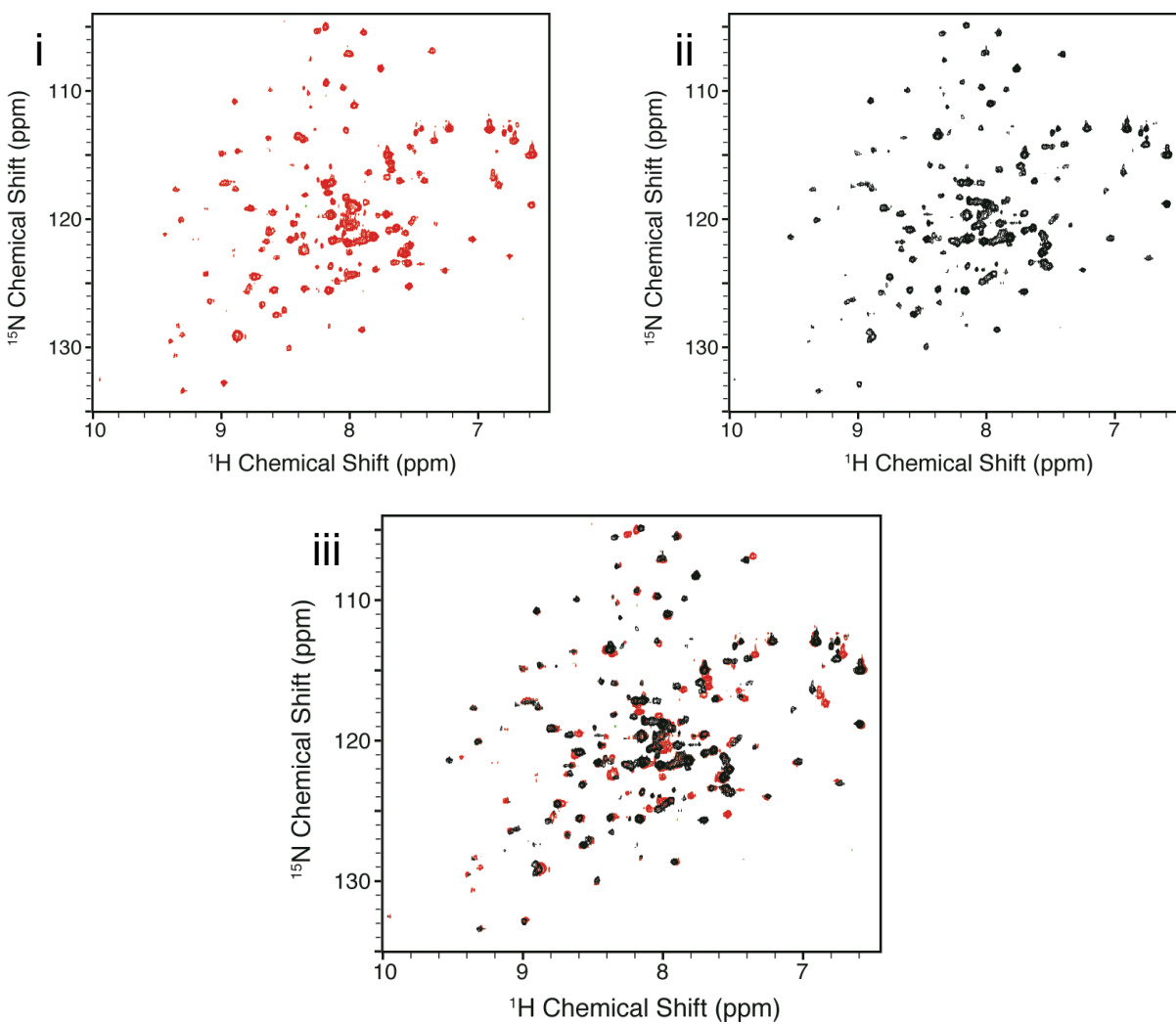
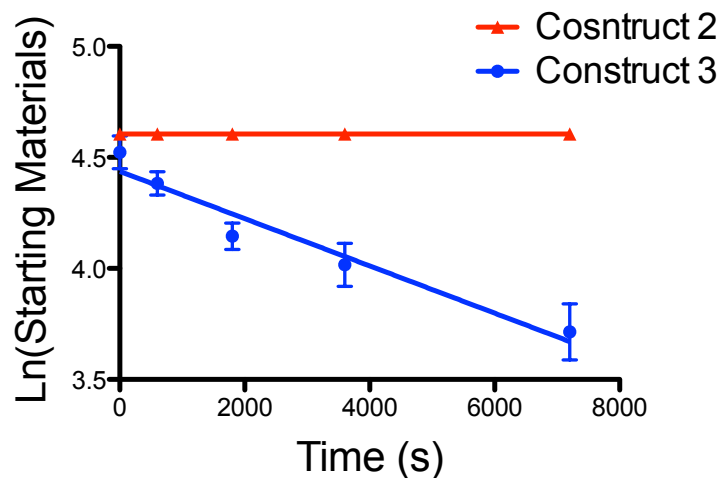


Figure S4: ^1H - ^{15}N HSQC spectra of refolded construct 2 (red) and 3 (black).

(i) ^1H - ^{15}N HSQC spectra of refolded construct 2. (ii) ^1H - ^{15}N HSQC spectra of refolded construct 3. (iii) Overlay of ^1H - ^{15}N HSQC spectra of refolded construct 2 (red) and construct 3 (black).

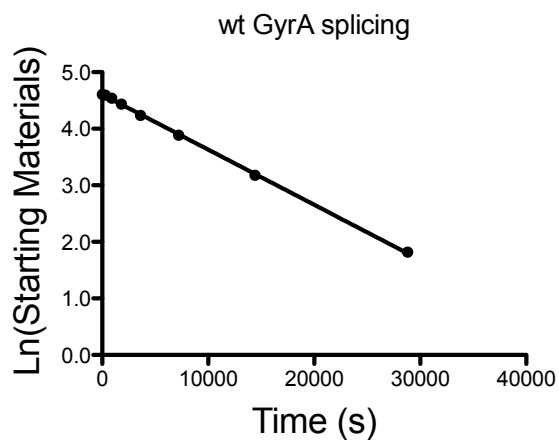
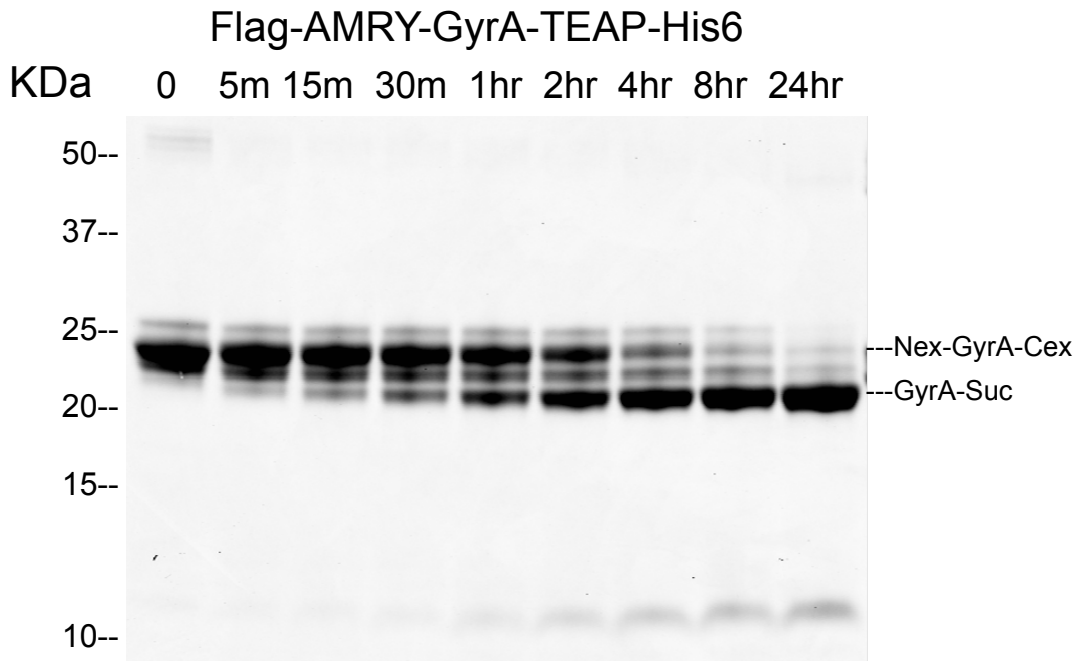


First-order rate constants of succinimide formation reaction

Construct	k_{obs} (s^{-1})	$T_{1/2}$
2	Inactive	--
3	$1.1 (\pm 0.1) \times 10^{-4}$	~ 2 hr

Figure S5: Kinetic analysis of intein-succinimide formation in branched constructs 2 and 3.

Rate of intein-succinimide formation rate was quantified by integration of the product peak in the RP-HPLC chromatogram. The data from the time course was fitted to a first order rate function ($\ln[A] = \ln[A]_0 - kt$) to obtain the observed rate constants. Error bars represent the standard deviation from three independent reactions.



First-order rate constants of protein splicing in *Mxe* GyrA wt

Construct	k_{obs} (s^{-1})	$T_{1/2}$
<i>Mxe</i> GyrA wt	1.0×10^{-4}	~ 2 hr

Figure S6: Kinetic analysis of the protein splicing of wt *Mxe* GyrA.

SDS-PAGE analysis of splicing activity of the wild type *Mxe* GyrA intein. Gel is stained by coomassie brilliant blue and imaged on a LI-COR Odyssey Infrared Imager. Nex refers to the N-extein, Cex refers to the C-extein, GyrA-Suc refers to the GyrA intein-succinimide. Densitometry analysis was used to quantify the coomassie stained gel, the data from the time course fit to a first order rate function to obtain the observed rate constant.

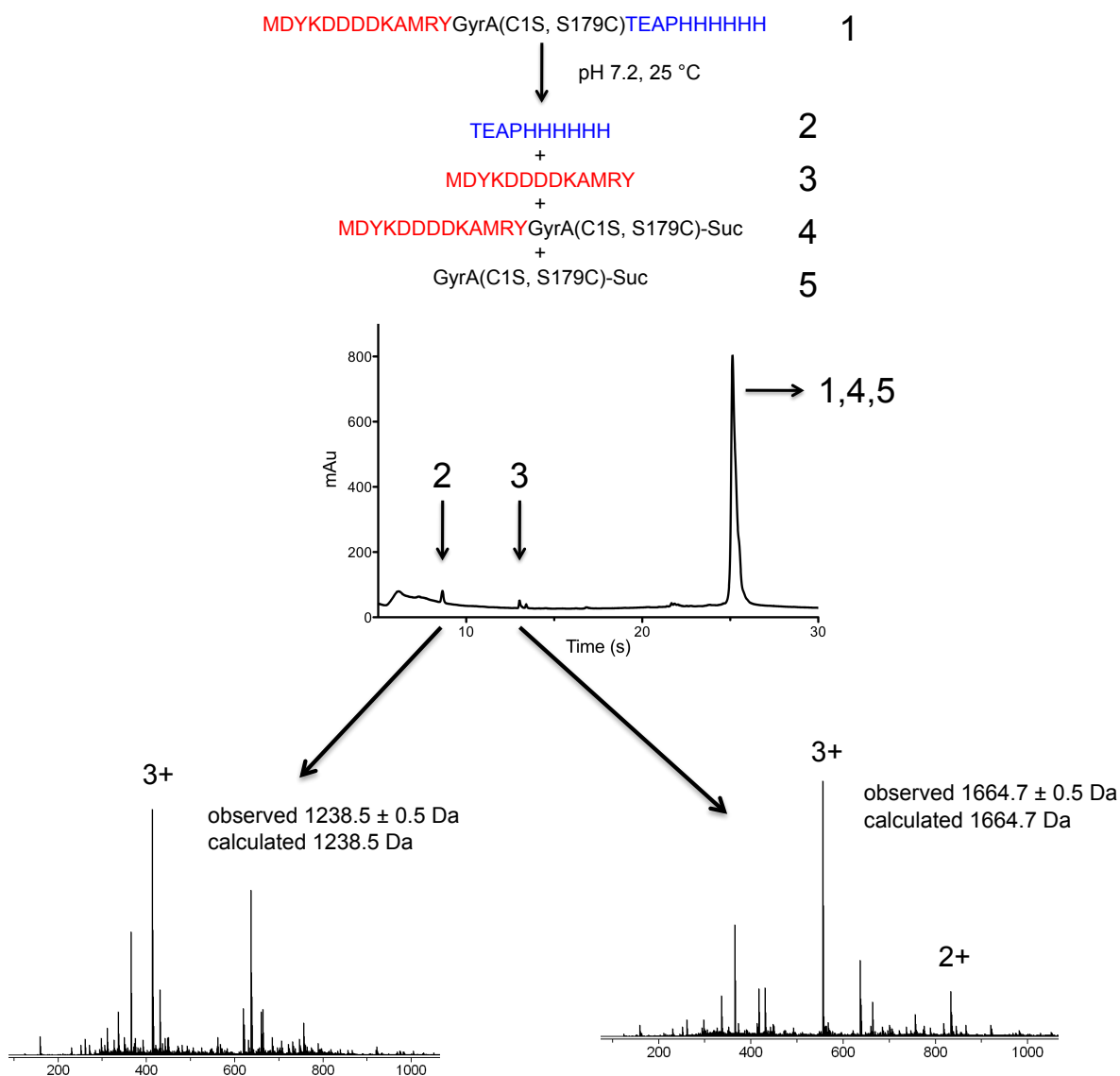
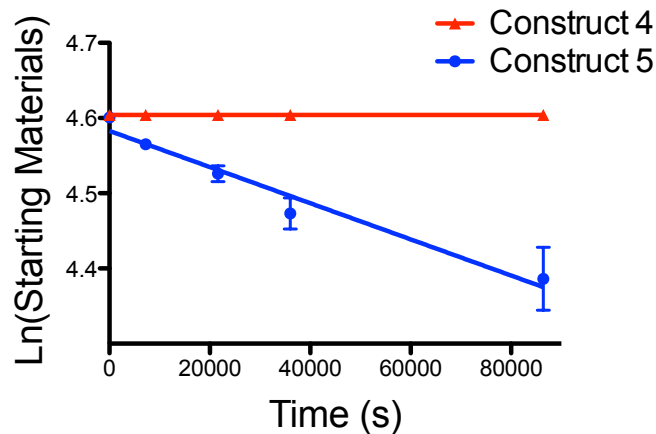


Figure S7: Analysis of C-extein cleavage and N-extein cleavage activities of *Mxe* GyrA intein (C1S, S179C).

At the 12 hr time point, C-extein cleavage product (2) and N-extein cleavage product (3) were identified by RP-HPLC (C_{18} column, 0-73% solvent B over 30 min, flow rate 1 mL/min, 214 nm detection) and ESI-MS analysis. No spliced product was observed.



First-order rate constants of
C-extein cleavage reaction

Construct	K_{obs} (s^{-1})	$T_{1/2}$
4	Inactive	--
5	$2.4 (\pm 0.2) \times 10^{-6}$	~ 80 hr

Figure S8: Kinetic analysis of C-extein cleavage (succinimide formation) in linear Mxe GyrA constructs 4 and 5.

C-extein cleavage rate was quantified by integration of the ESI-MS chromatograms. The data from the time course was fit to a first order rate function ($\ln[A] = \ln [A]_0 - kt$) to obtain the observed rate constants. Error bars represent the standard deviation from three independent reactions.

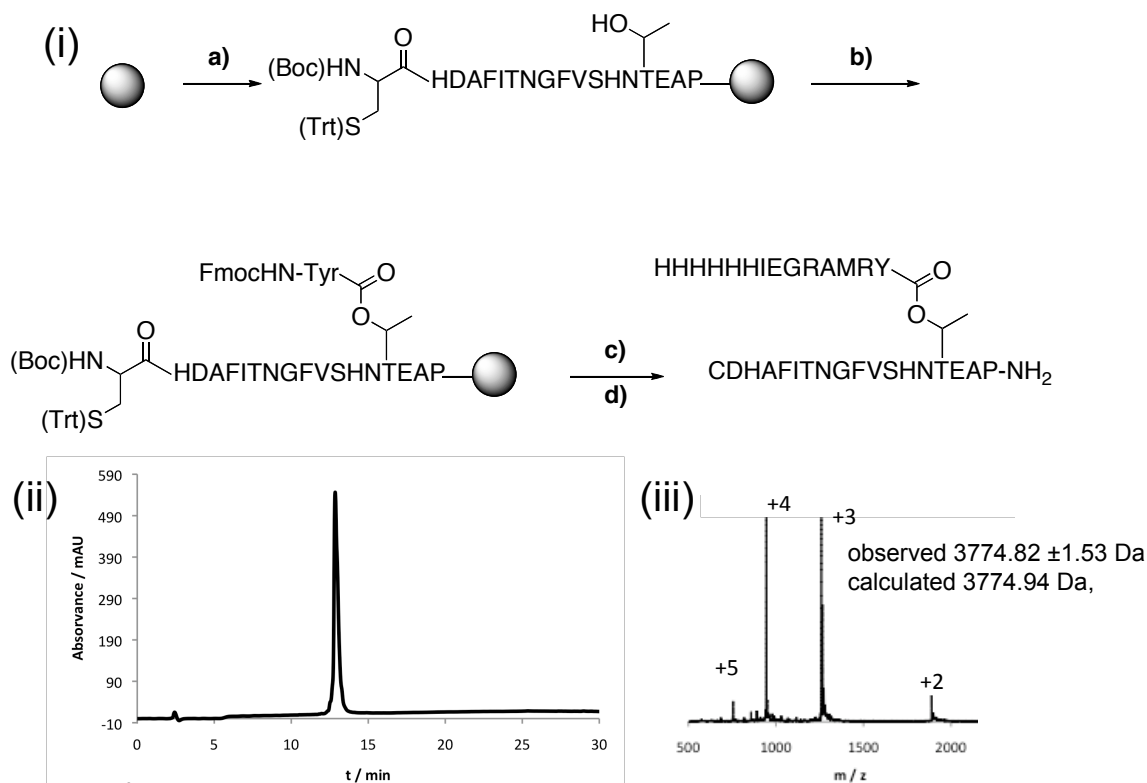


Figure S9: Synthesis of the GyrA peptide used for preparing construct 1.

(i) Scheme of synthetic route. Reaction conditions: a) Standard Fmoc-SPPS. b) Ethanolamine in 5% water/DMF for 2 x 30 min, then 4.0 eq. Fmoc-Tyr(*t*Bu)-OH, 2.0 eq. DIC, 0.1 eq. DMAP for 2 x 30 min, followed by 4 eq. MSNT, 12eq. NMI in DMF for 16 hr. c) Fmoc-SPPS using 5 eq. of amino acid, 5 eq. of HOBt and 5 eq. of DIC and 2% of DBU in DMF to remove the Fmoc. d) 1% TIPS, 2.5% ethanedithiol, 2.5% water and 94% TFA, 1.5 h, 25 °C. (ii) RP-HPLC analysis of purified branched peptide (C_{18} column, gradient of 20-40% solvent B, 214 nm detection). (iii) ESI-MS analysis of the branched GyrA peptide **1**.

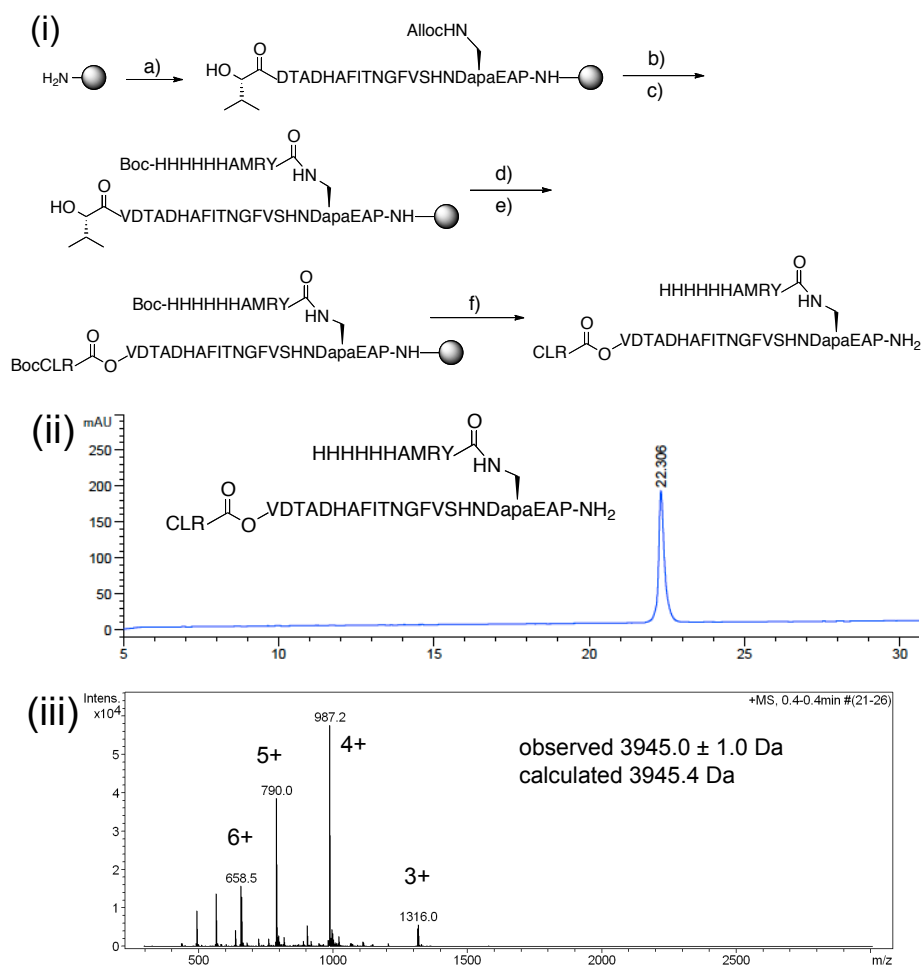


Figure S10: Synthesis of the GyrA peptide used for preparing construct 2.

(i) Scheme of synthetic route. Reaction conditions: a) Standard Fmoc-SPPS. b) 0.1 eq. Pd(PPh₃)₄, 25 eq. PhSiH₃, in DCM 30 min. c) Standard Fmoc-SPPS. d) Ethanolamine in 5% water/DMF for 2 x 30 min, then 5 eq. Fmoc-Arg(Pbf)-OH, 2.5 eq. DIC, 0.1 eq. DMAP for 3hr, followed by 4 eq. MSNT, 12eq. NMI in DMF for 16 hr. e) Standard Fmoc-SPPS, but 2% DBU was used to deprotect Fmoc in each cycle. f) 1% TIPS, 2.5% ethanedithiol, 2.5% water, 94% TFA, 1.5 hr at 25 °C. (ii) RP-HPLC analysis of purified branched peptide (C₁₈ column, 10-40% solvent B over 30 min, flow rate 1 mL/min, 214 nm detection). (iii) ESI-MS analysis of the branched ester GyrA peptide 2.

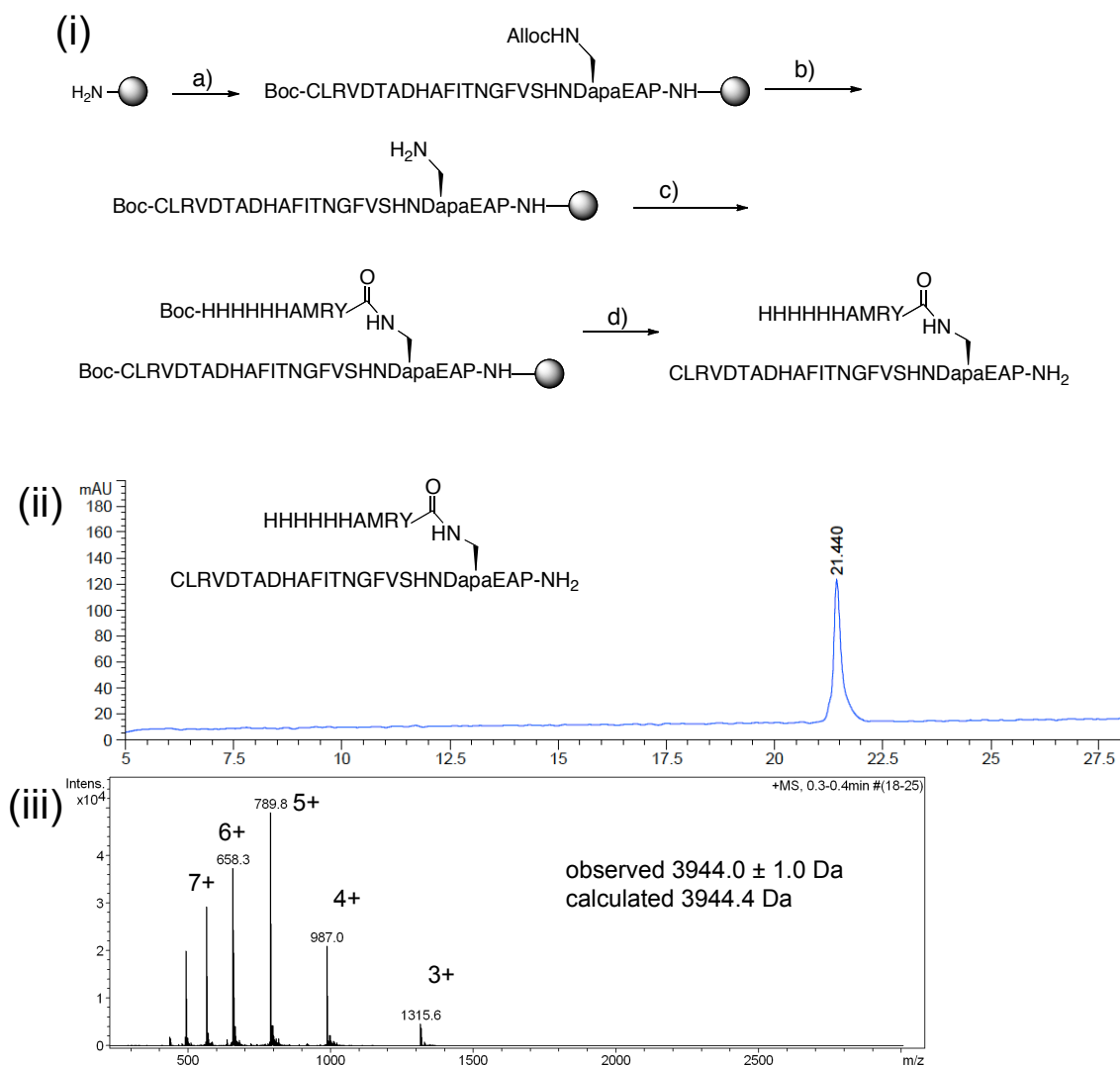


Figure S11: Synthesis of the GyrA peptide used for preparing construct 3.

(i) Scheme of synthetic route. Reaction conditions: a) standard SPPS using Fmoc chemistry. b) 0.1 eq. $\text{Pd}(\text{PPh}_3)_4$, 25 eq. PhSiH_3 , in DCM 30 min. c) standard SPPS using Fmoc chemistry. d) 1% TIPS, 2.5% ethanedithiol, 2.5% water, 94% TFA, 1.5 hr at 25 °C. (ii) RP-HPLC analysis of purified branched peptide (C_{18} column, 10-40% solvent B over 30 min, flow rate 1 mL/min, 214 nm detection). (iii) ESI-MS analysis of the branched GyrA peptide **3**.

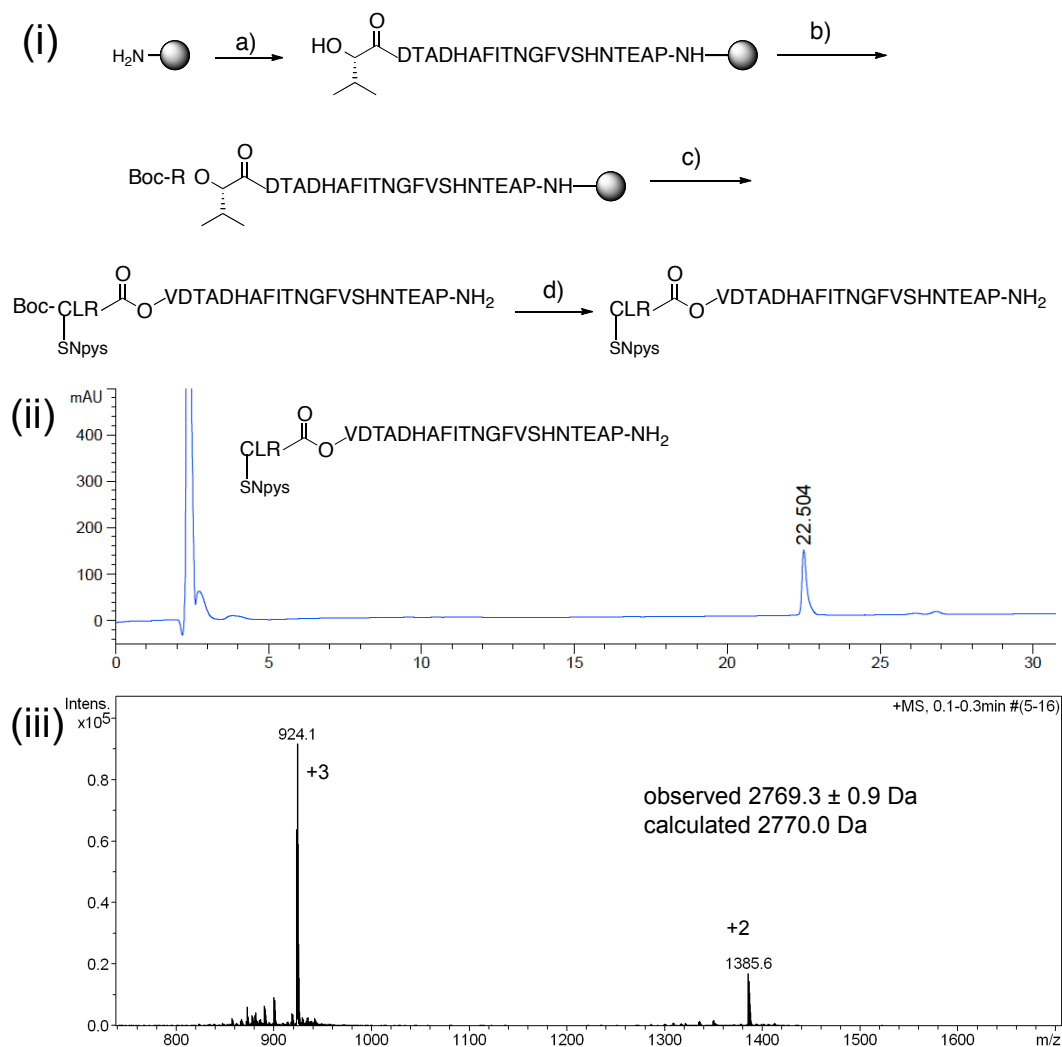


Figure S12: Synthesis of the GyrA peptide used for preparing construct 4.

(i) Scheme of synthetic route. Reaction conditions: a) Standard Boc-SPPS. b) Ethanolamine in 5% water/DMF for 2 x 30 min, then 10 eq. Boc-Arg(Tos)-OH, 5 eq. DIC, 0.1 eq. DMAP for 3hr, followed by 4 eq. MSNT, 12eq. NMI in DMF for 16 hr. This entire process was repeated for better esterification yield. c) Standard Fmoc-SPPS, but 2% DBU was used to deprotect Fmoc in each cycle. d) 1% TIPS, 2.5% ethanedithiol, 2.5% water and 94% TFA, 1.5 hr at 25 °C. (ii) RP-HPLC analysis of purified branched peptide (C_{18} column, 10-40% solvent B over 30 min, flow rate 1 mL/min, 214 nm detection). (iii) ESI-MS analysis of the linear ester GyrA peptide 4.

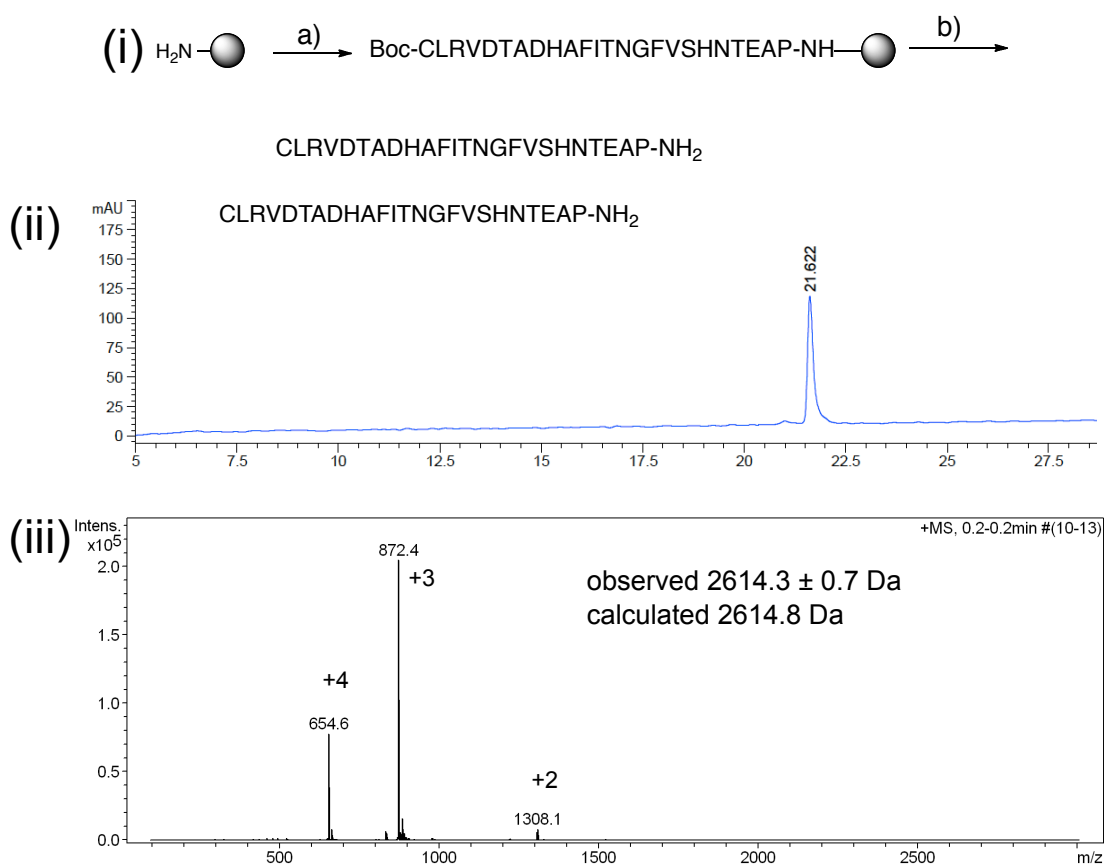


Figure S13: Synthesis of the GyrA peptide used for preparing construct 5.

(i) Scheme of synthetic route. Reaction conditions: a) Standard Fmoc-SPPS. b) 1% TIPS, 2.5% ethanedithiol, 2.5% water and 94% TFA, 1.5 hr at 25 °C. (ii) RP-HPLC analysis of purified branched peptide (C₁₈ column, 10-40% solvent B over 30 min, flow rate 1 mL/min, 214 nm detection). (iii) ESI-MS analysis of the linear GyrA peptide.

References:

1. Otwinowski Z & Minor W (1997) Processing of X-ray diffraction data collected in oscillation mode. *Methods Enzymol.*, eds Carter CW & Sweet RM (Academic Press, New York), Vol 276, pp 307-326.
2. McCoy AJ, *et al.* (2007) Phaser crystallographic software. *J. Appl. Crystallogr.* 40(Pt 4):658-674.
3. Klabunde T, Sharma S, Telenti A, Jacobs WR, & Sacchettini JC (1998) Crystal structure of GyrA intein from *Mycobacterium xenopi* reveals structural basis of protein splicing. *Nature Structural Biology* 5(1):31-36.
4. Emsley P & Cowtan K (2004) Coot: model-building tools for molecular graphics. *Acta Crystallogr D Biol Crystallogr* 60(Pt 12 Pt 1):2126-2132.
5. Adams PD, *et al.* (2010) PHENIX: a comprehensive Python-based system for macromolecular structure solution. *Acta Crystallogr D Biol Crystallogr* 66(Pt 2):213-221.
6. Delaglio F, *et al.* (1995) Nmrpipe - a Multidimensional Spectral Processing System Based on Unix Pipes. *J. Biomol. NMR* 6(3):277-293.
7. Goddard TD & Kneller DG (Sparky 3. *University of California, San Francisco.*
8. Ichiyanagi K, Ishino Y, Ariyoshi M, Komori K, & Morikawa K (2000) Crystal structure of an archaeal intein-encoded homing endonuclease PI-PfuI. *J. Mol. Biol.* 300(4):889-901.
9. Mizutani R, Anraku Y, & Satow Y (2004) Protein splicing of yeast VMA1-derived endonuclease via thiazolidine intermediates. *Journal of Synchrotron Radiation* 11:109-112.
10. Matsumura H, *et al.* (2006) Crystal structure of intein homing endonuclease II encoded in DNA polymerase gene from hyperthermophilic Archaeon *Thermococcus kodakaraensis* strain KOD1. *Proteins-Structure Function and Bioinformatics* 63(3):711-715.
11. Van Roey P, *et al.* (2007) Crystallographic and mutational studies of *Mycobacterium tuberculosis* recA mini-inteins suggest a pivotal role for a highly conserved aspartate residue. *J. Mol. Biol.* 367(1):162-173.
12. Oeemig JS, Zhou DW, Kajander T, Wlodawer A, & Iwai H (2012) NMR and Crystal Structures of the *Pyrococcus horikoshii* RadA Intein Guide a Strategy for Engineering a Highly Efficient and Promiscuous Intein. *J. Mol. Biol.* 421(1):85-99.