Supplementary Information

Assembling Branched and Macrocyclic Peptides on Proteins

Sven Ullrich,[‡,a] Santhanalaxmi Kumaresan,[‡,a] Marina G. Rahman,[a] Bishvanwesha Panda,[a] Richard Morewood,[a] Christoph Nitsche*[,a]

[a] Research School of Chemistry, Australian National University, Canberra, ACT 2601, Australia.

[*] Corresponding author: christoph.nitsche@anu.edu.au

[‡] These authors share first-authorship.

Abbreviations

DMSO	dimethyl sulfoxide
ESI	electrospray ionisation
FDCP	4-fluoro-2,6-dicyanopyridine
FPLC	fast protein liquid chromatography
HEPES	4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid
HPLC	high performance liquid chromatography
HRMS	high resolution mass spectrometry
LB	lysogeny broth
LC	liquid chromatography
MES	2-(*N*-morpholino)ethanesulfonic acid
MS	mass spectrometry
NMR	nuclear magnetic resonance
OD	optical density
PAGE	polyacrylamide gel electrophoresis
PDB	protein data bank
SDS	sodium dodecyl sulfate
SMILES	simplified molecular input line entry system
SOC	super optimal condition
SPPS	solid phase peptide synthesis
TCEP	3,3′,3″-phosphanetriyltripropanoic acid
TLC	thin layer chromatography
UV	ultraviolet light

Materials

Solvents (ChemSupply, Australia; Sigma-Aldrich, USA) and reagents (Sigma-Aldrich; USA, AK Scientific, USA; Ambeed, USA; Thermo Fisher, USA) were used as received without any additional purification. Flash column chromatography was carried out using silica gel 60 with 230–400 mesh (ChemSupply, Australia) and an Isolera One purification system (Biotage, Sweden). Analytical TLC was conducted with silica gel 60 plates (0.2 mm) with aluminium backing and fluorescent coating (Millipore, USA) and evaluated using UV fluorescence (λ_{max} = 254 nm). Ultrapure water was generated in-house using a milli-Q purification system (Millipore, USA). Precast gels (Thermo Fisher, USA) were used for 10%-SDS PAGE analysis. Peptides were synthesised by SPPS and purified by HPLC as previously described.[1, 2]

NMR analysis

NMR spectra were recorded on an Avance III 400 instrument (Bruker, USA). Residual solvent peaks were used as internal reference.[3] Spectra were phase- and baseline-corrected, and visualised in MestReNova 9 (Mestre Lab, Spain).

Intact protein MS

Intact protein MS was conducted with an Orbitrap Elite Hybrid Ion Trap-Orbitrap mass spectrometer (Thermo Fisher, USA) coupled with an UltiMate 3000 UHPLC (Thermo Fisher, USA) using an Agilent ZORBAX SB-C3 RRHT threaded column (Agilent, USA). Intact mass spectra were deconvoluted using Xcalibur 3 (Thermo Fisher, USA) using the default 'Xtract Settings' in 'Generated Masses Mode, M' from 600 Da to 4000 Da. Spectra were visualised in GraphPad Prism 10 (Dotmatics, USA). Expected masses were predicted by enviPat Web (Eawag, Switzerland).[4] Percent abundance within DCP-tagging reactions was evaluated using a Python script co-written by ChatGPT 4 (OpenAI, USA), which is provided in the appendix.

HRMS analysis

The identity of peptides was confirmed using high-resolution ESI+ mass spectrometry using an Orbitrap Elite mass spectrometer (Thermo Fisher, USA) or a Synapt G2-Si QuanTOF mass spectrometer (Waters, USA). Anticipated mass-to-charge ratios were calculated using enviPat Web (Eawag, Switzerland).[4]

Protein expression and purification

The protein expression and purification protocol was adapted and slightly modified from the literature.[5] A pET-29b(+) vector containing an insert between NdeI and XhoI restriction sites encoding for the respective His_6-tagged protein of interest was obtained commercially (Twist Bioscience, USA). The plasmid (1 µL, 100 ng/µL) was transformed into electrocompetent *Escherichia coli* BL21 (DE3) cells (100 µL), and the cells were allowed to recover in SOC medium (1 h, 37 °C) before being spread (50 µL) on LB agar plates containing 50 mg/L kanamycin. The plate was incubated overnight at 37 °C, after which three colonies were inoculated into 10 mL LB medium supplemented with 50 mg/L kanamycin and grown overnight at 37 °C on a shaker (200 rpm). The overnight culture was transferred into 1 L of LB medium containing 50 mg/L kanamycin and the cells were grown at 37 °C on a shaker until an OD_{600} value of 0.60–0.80 was reached, after which the cells were allowed to reach room temperature. IPTG (1 mL, 1 M) was added to the cells to a final concentration of 1 mM to initiate recombinant protein expression and the culture was incubated on a shaker (200 rpm) at room temperature for 16–20 h. The cells were harvested by centrifugation at 5000 × *g* for 20 min at 4 °C and resuspended in buffer A (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 5 mM imidazole). Cell lysis was performed on ice using sonication with the Omni-Ruptor 4000 (Omni International, USA) for 2 × 6 min, 60% power and 60% pulse length. The lysates were centrifuged for 1 h at 16,000 × *g* at 4 °C. The supernatant was purified using FPLC with the ÄKTA pure 25 chromatography system (Cytiva, USA) equipped with a 5 mL HisTrap HP column (Cytiva, USA). The column was washed with buffer A and the recombinant protein was eluted with buffer B (50 mM Tris-HCl, pH 7.5, 300 mM NaCl, 300 mM imidazole). The buffer of the eluted fraction was exchanged to buffer C (50 mM Tris-HCl, pH 7.5, 150 mM NaCl, 1 mM TCEP) with a HiPrep Desalting 26/10 Column (Cytiva, USA) and the recombinant protein was concentrated using an Amicon ultrafiltration centrifugal tube (Millipore, USA) with the appropriate molecular weight cut-off. Protein concentrations were determined by measuring A_{280} with a NanoDrop OneC (Thermo Fisher, USA) using extinction coefficients determined by Expasy ProtParam.[6] The recombinant proteins were analysed by 10%-SDS PAGE and protein mass spectrometry. Aliquots were snap-frozen in liquid nitrogen and kept at −80 °C. Protein mass spectrometry was conducted with an Orbitrap Fusion Tribrid mass spectrometer (Thermo Fisher, USA) coupled with an UltiMate S4 3000 UHPLC (Thermo Fisher, USA) using the Agilent ZORBAX SB-C3 Rapid Resolution HT Threaded Column (Agilent, USA).

Table S1. Recombinant proteins used in this study. Within the amino acid sequences, His_6-tags and engineered TEV cleavage sites are shown in grey, and mutations are shown in red.

Name	Abbreviation(s)	Mutation(s)	Yield[a]	Amino acid sequence
Extra-superfolder green fluorescent protein[7, 8]	esGFP, GFP, GFP-1C	C48S, C64L, C70M, K238C	~ 60 mg	MGSSHHHHHHENLYFQGMSKGEELFTGVVPILVELDGDVN GHKFSVRGEGEGDATNGKITLKLISTTGKLPVPWPTLVTT LGYGVQMFARYPDHMKRHDFFKSAMPEGYVQERTISFKDD GTFKTRAEVKFEGDTIVNRIKLKGIDFKEDGNILGHKLEY NFNSHKVYITADKQKNGIKANFKIRHNVEDGSVQLADHYQ QNTPIGDGPVRLPDNHYLSTQSVILEDPNEKRDHMVLHEF VTAAGITHGMDELYC
Linked Zika virus NS2B-NS3 protease construct[9, 10]	gZiPro, ZiPro, ZiP ZiP-1C	V36C, C80S, C143S	~ 45 mg	MGSSHHHHHHENLYFQSGKSVDMYIERAGDITWEKDAEVT GNSPRLDVALDESGDFSLVEDDGPPMREGGGGSGGGGSGA LWDVPAPKEVKKGETTDGVYRVMTRRLLGSTQCGVGVMQE GVFHTMWHVTKGSALRSGEGRLDPYWGDVKQDLVSYSGPW KLDAAWDGHSEVQLLAVPPGERARNIQTLPGIFKTKDGDI GAVALDYPAGTSGSPILDKSGRVIGLYGNGVVIKNGSYVS AITQGRREEETPVE
Ubiquitin[11, 12]	Ubq, Ub Ubq-1C	S20C	~ 55 mg	MGSSHHHHHHENLYFQSMQIFVKTLTGKTITLEVEPCDTI ENVKAKIQDKEGIPPDQQRLIFAGKQLEDGRTLSDYNIQK ESTLHLVLRLRGG
B1 domain of protein G[13, 14]	GB1 GB1-1C	I6C	~ 60 mg	MTYKLCLNGKTLKGETTTEAVDAATAEKVFKQYANDNGVD GEWTYDDATKTFTVTEHHHHHH

SpyCatcher 3[15, 16]	SpyC3 SC3	-	~ 80 mg	MGSSHHHHHHENLYFQSVTTLSGLSGEQGPSGDMTTEEDS ATHIKFSKRDEDGRELAGATMELRDSSGKTISTWISDGHV KDFYLYPGKYTFVETAAPDGYEVATPIEFTVNEDGQVTVD GEATEGDAHT

[a] = yield obtained from 1 L culture of *E. coli* BL21 under the outlined conditions.

Table S2. Deconvoluted intact protein mass spectrometry data for proteins used in this study.

Protein	Molecular formula[a]	Calculated mass[b]	Observed mass
GFP-1C	$C_{1287}H_{1981}N_{359}O_{384}S_9$[c]	28911.44	28912.57
ZiP-1C	$C_{1195}H_{1854}N_{342}O_{381}S_8$	27360.44	27362.57
Ubq-1C	$C_{473}H_{757}N_{137}O_{143}S_4$	10777.52	10778.56
GB1-1C	$C_{311}H_{468}N_{86}O_{102}S_3$	7138.33	7139.36
SC3	$C_{618}H_{927}N_{169}O_{217}S_4$	14321.58	14323.71

[a] = calculated with additional methionine residue.

[b] = mass calculated to have highest abundance using enviPat.

[c] = calculated for protein with matured chromophore.

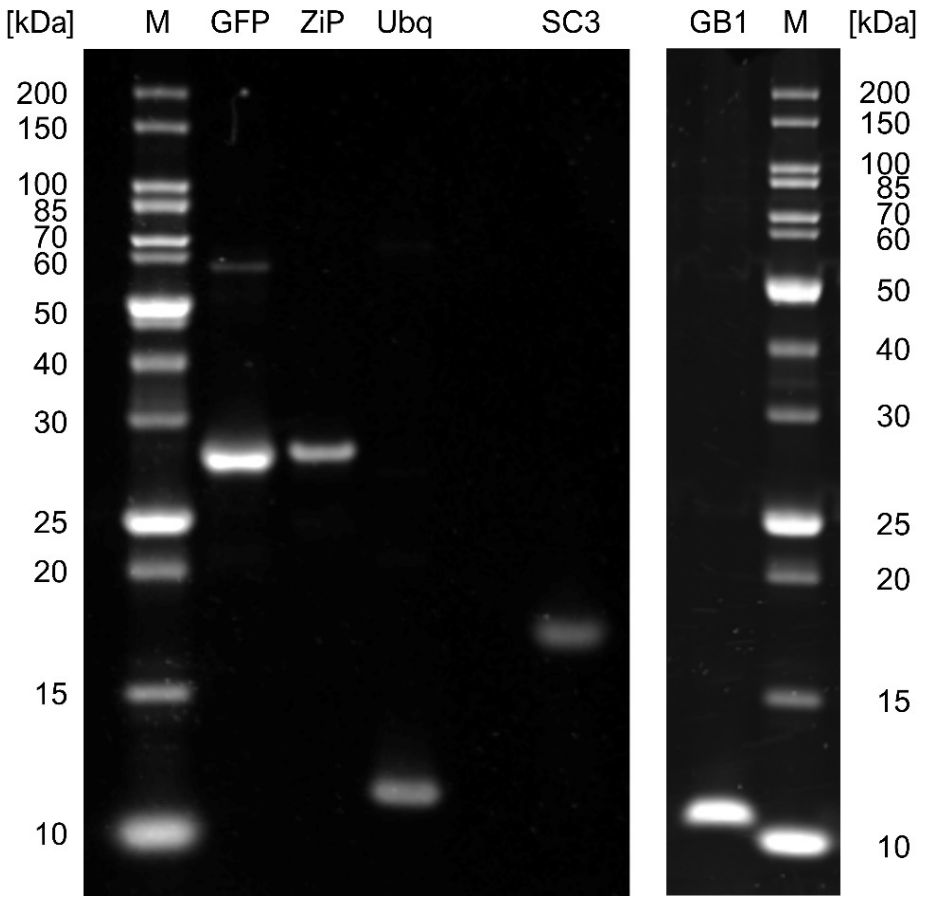


Figure S1. Coomassie-stained SDS-PAGE gel analysis of Ni-NTA FPLC-purified, unmodified protein samples. M = molecular weight marker (P7717S; New England Biolabs, USA). Two sections from separate gels are presented and the uncropped gel images are shown in Figure S21 and Figure S22, respectively.

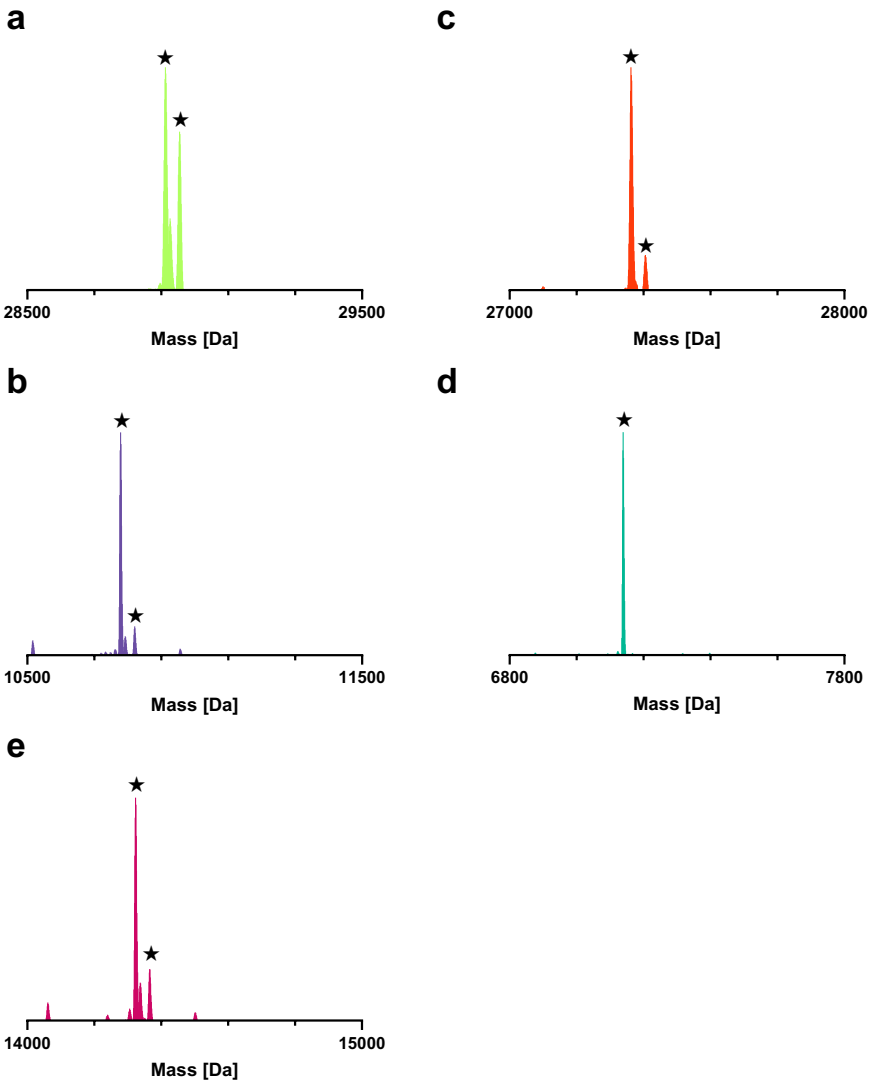


Figure S2. Deconvoluted intact protein mass spectra of proteins used in this study. (a) GFP-1C (green), (b) ZiP-1C (red), (c) Ubq-1C (purple), (d) GB1-1C (blue), (e) SC3 (pink). The protein peaks are marked by a star (★; including a +42 Da species, likely resulting from post-translational acetylation,[17] which was observed in several expression batches).

Table S3. Sequences of peptides used in this study.

Compound	N-terminus	Sequence[a]	C-terminus
P1	H- (amine)	CSHPQFC	-NH_2 (amide)
P2	H- (amine)	CAYTNCG	-NH_2 (amide)
P3	H- (amine)	CGKRKSCF	-NH_2 (amide)
P4	H- (amine)	CSDEVCW	-NH_2 (amide)
P5	H- (amine)	CHYLC	-NH_2 (amide)
P6	H- (amine)	CGSGYGSGC	-NH_2 (amide)
P7	H- (amine)	cPΞSyCAK	-NH_2 (amide)
P8	H- (amine)	XRKKRX	-NH_2 (amide)

[a] Ξ = L-propargylglycine, X = L-2-amino-4-(2-amino-3-mercaptopropanamido)butanoic acid.

LCMS analysis

Peptides were analysed using analytical reverse phase LCMS 1260/6120 (Agilent, USA) using UV detection and ESI+ spectrometry using a Poroshell 120, EC-C_{18} 3.0 mm × 50 mm, 2.7 µm column (Agilent, USA) held at 30 °C. Ultrapure water (solvent 1) and LCMS-grade acetonitrile (solvent 2), both with 0.1% formic acid additive were used.

Method A This method used a flow rate of 0.3 mL/min and started at 5% solvent 2 in solvent 1. From minute 1, the proportion of solvent 2 was increased gradually up to 90% at minute 11. This percentage was held until minute 18.

Method B This method used a flow rate of 1 mL/min and started at 5% solvent 2 in solvent 1. From minute 2, the proportion of solvent 2 was increased gradually up to 90% at minute 22. The percentage was held until minute 24. The proportion of solvent 2 was decreased gradually down to 5% at minute 25. This percentage was held until minute 30.

Method C This method used a flow rate of 0.3 mL/min and started at 5% solvent 2 in solvent 1. From minute 2, the proportion of solvent 2 was increased gradually up to 60% at minute 12. The percentage was held until minute 12.5.

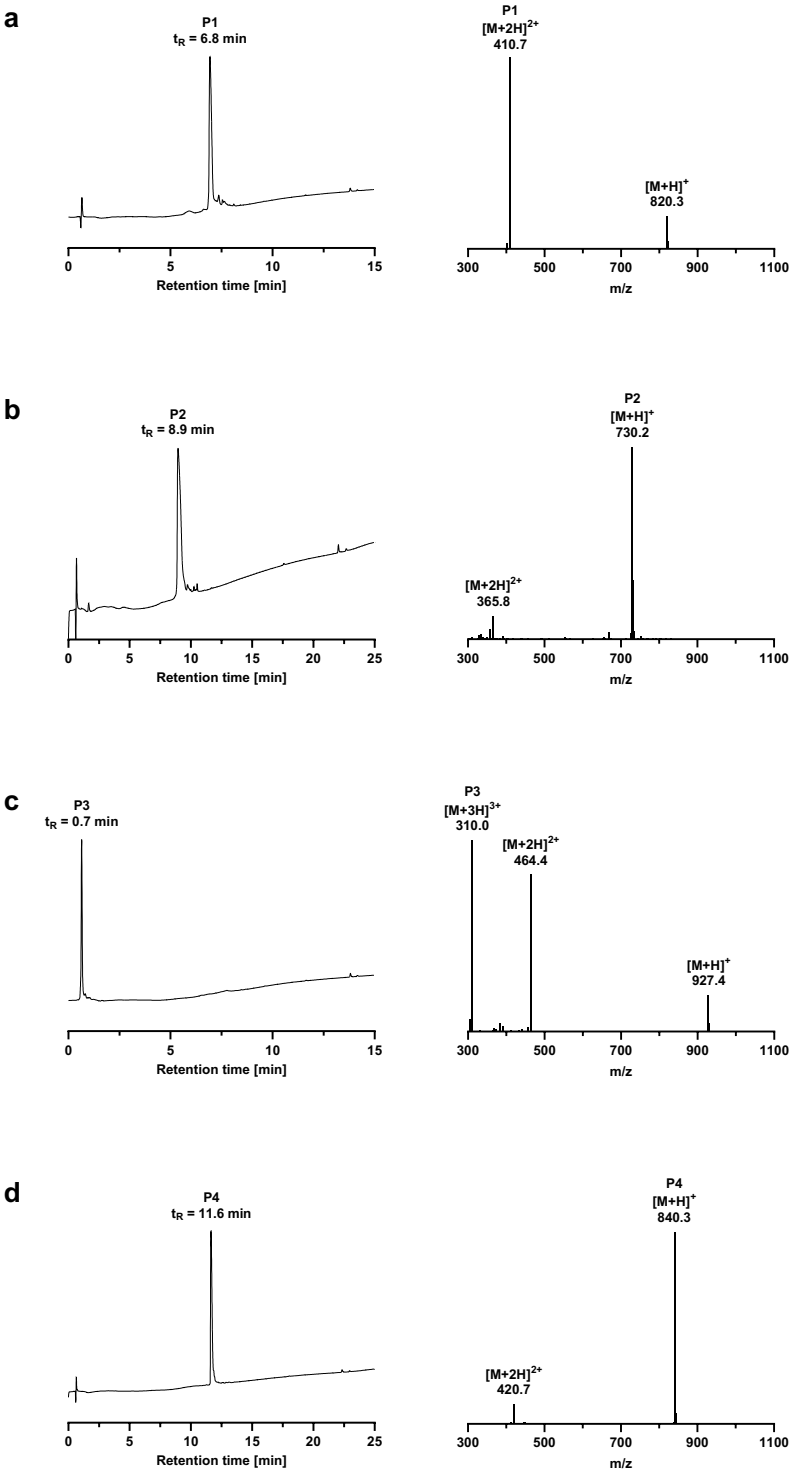


Figure S3. *(continued)*

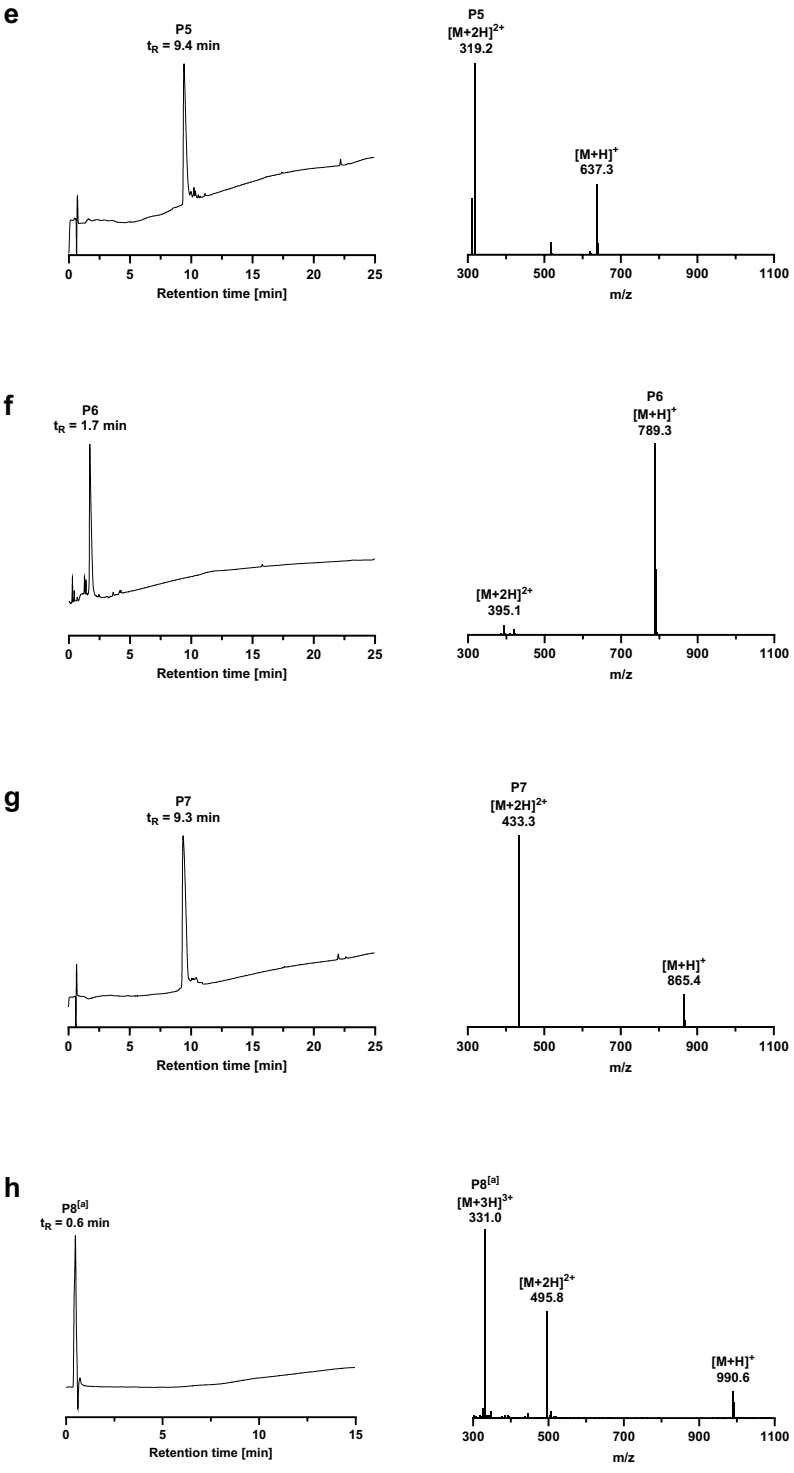


Figure S3. LCMS analysis of peptides **P1–P8** showing their LC trace and ESI+ spectrum. (a) **P1** (method A), (b) **P2** (method B), (c) **P3** (method A), (d) **P4** (method B), (e) **P5** (method B), (f) **P6** (method B), (g) **P7** (method B), (h) **P8** (method C). [a] = observed as cyclic disulfide.

UHPLC analysis

Peptides **P3** and **P8** were additionally analysed using a Nexera UHPLC (Shimadzu, Japan) equipped with a UV detector and an Alltima HP C_{18}-AQ, 2.1 mm × 150 mm, 5 µm column (VWR, USA) held at 40 °C. Ultrapure water (solvent 1) and LCMS-grade acetonitrile (solvent 2), both with 0.1% trifluoroacetic acid additive were used.

Method D This method used a flow rate of 1 mL/min and started at 100% solvent 1. From minute 2, the proportion of solvent 2 was increased gradually up to 50% at minute 22. This percentage was held until minute 24. From minute 24 to minute 25, 100% solvent 1 were used.

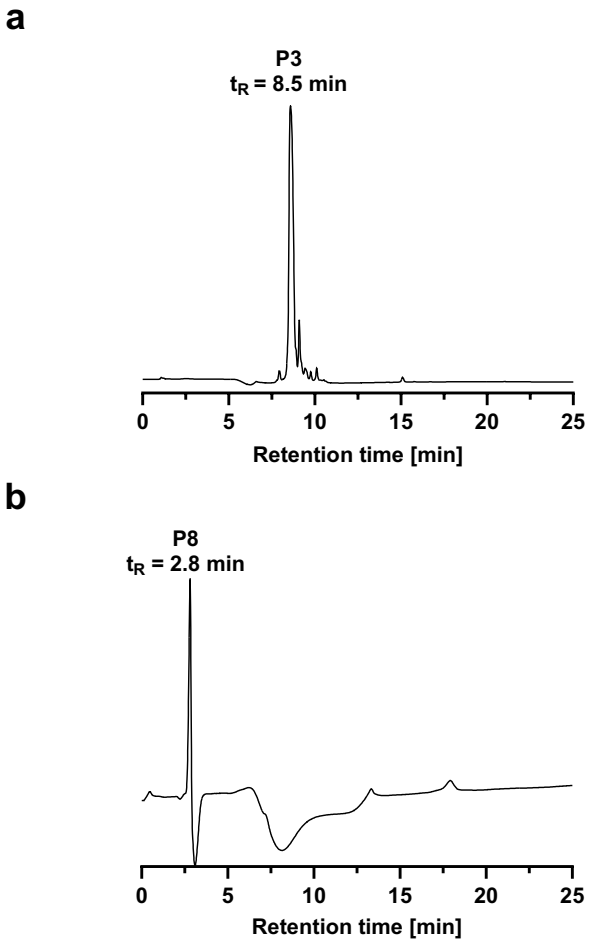


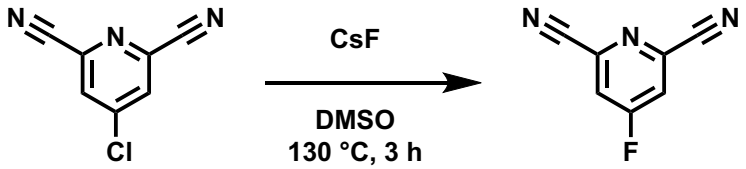
Figure S4. UHPLC traces of peptides (a) **P3** (method D) and (b) **P8** (method D).

Table S4. HRMS analysis of peptides used in this study.

Compound	Molecular formula	Ion	Calculated m/z	Observed m/z
P1	$C_{34}H_{49}N_{11}O_9S_2$	$[M+Na]^+$	842.3048	842.3066
P2	$C_{28}H_{43}N_9O_{10}S_2$	$[M+Na]^+$	752.2467	752.2457
P3	$C_{38}H_{66}N_{14}O_9S_2$	$[M+2H]^{2+}$	464.2362	464.2383
P4	$C_{34}H_{49}N_9O_{12}S_2$	$[M+H]^+$	840.3015	840.2997
P5	$C_{27}H_{40}N_8O_6S_2$	$[M+Na]^+$	659.2404	659.2395
P6	$C_{29}H_{44}N_{10}O_{12}S_2$	$[M+Na]^+$	811.2474	811.2441
P7	$C_{37}H_{56}N_{10}O_{10}S_2$	$[M+H]^+$	865.3695	865.3672
P8	$C_{38}H_{75}N_{19}O_8S_2$[a]	$[M+H]^+$	990.5560	990.5652

[a] = calculated with disulfide bridge.

Synthesis of FDCP



4-Fluoro-2,6-dicyanopyridine (FDCP) was synthesised following a protocol similar to our previous work, matching the reported characterisation data.[18] Caesium fluoride (1.30 g, 8.6 mmol) was dried under vacuum at 130 °C for 2 h in an oven-dried round bottom flask. After allowing the mixture to cool, 4-chloro-2,6-dicyanopyridine (1.08 g, 6.6 mmol) in dry DMSO was added under a positive pressure of dry nitrogen. The reaction mixture was stirred at 130 °C for 3 h, with progress monitored by TLC. Upon cooling to room temperature, the reaction was quenched with ice-cold ultrapure water and extracted with CH_2Cl_2. The organic layer was washed, dried over Na_2SO_4 and concentrated under reduced pressure to give crude FDCP. Further purification by flash column chromatography (100% toluene) yielded FDCP (0.78 g, 5.3 mmol, 80%) as a white crystalline solid.

^{1}H NMR (400 MHz, MeOD): $\delta = 8.14$ (d, $J = 7.9$ Hz, 2H). ^{13}C NMR (100 MHz, MeOD): $\delta = 170.4$ (d, $J = 268$ Hz, 1C), 138.2 (d, $J = 10$ Hz, 2C), 121.7 (d, $J = 21$ Hz, 2C), 116.1 (d, $J = 4.5$ Hz, 2C). ^{19}F NMR (376 MHz, MeOD): $\delta = -96.5$ (t, $J = 7.8$ Hz, 1F).

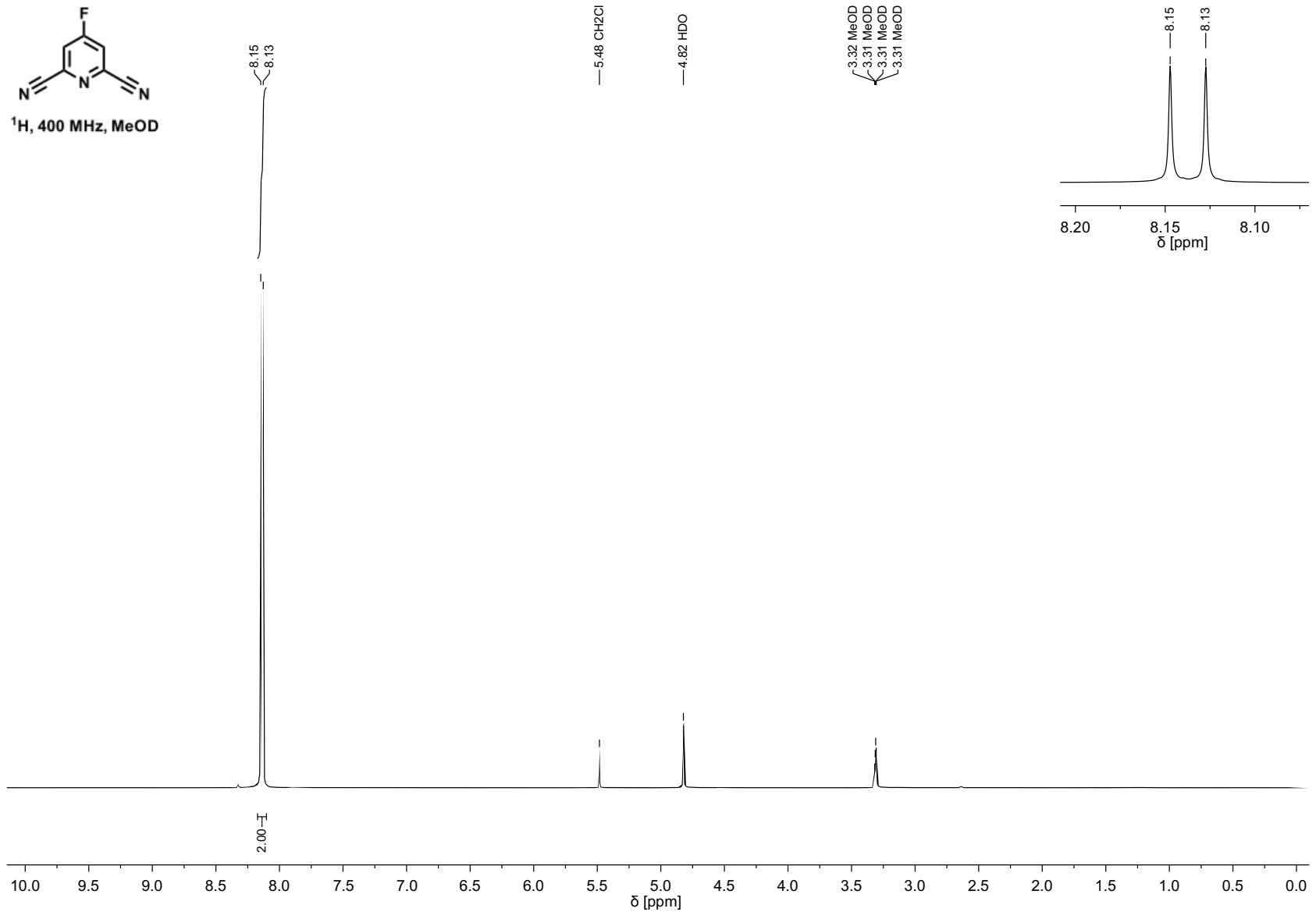


Figure S5. ^{1}H-NMR spectrum (400 MHz, MeOD) of FDCP.

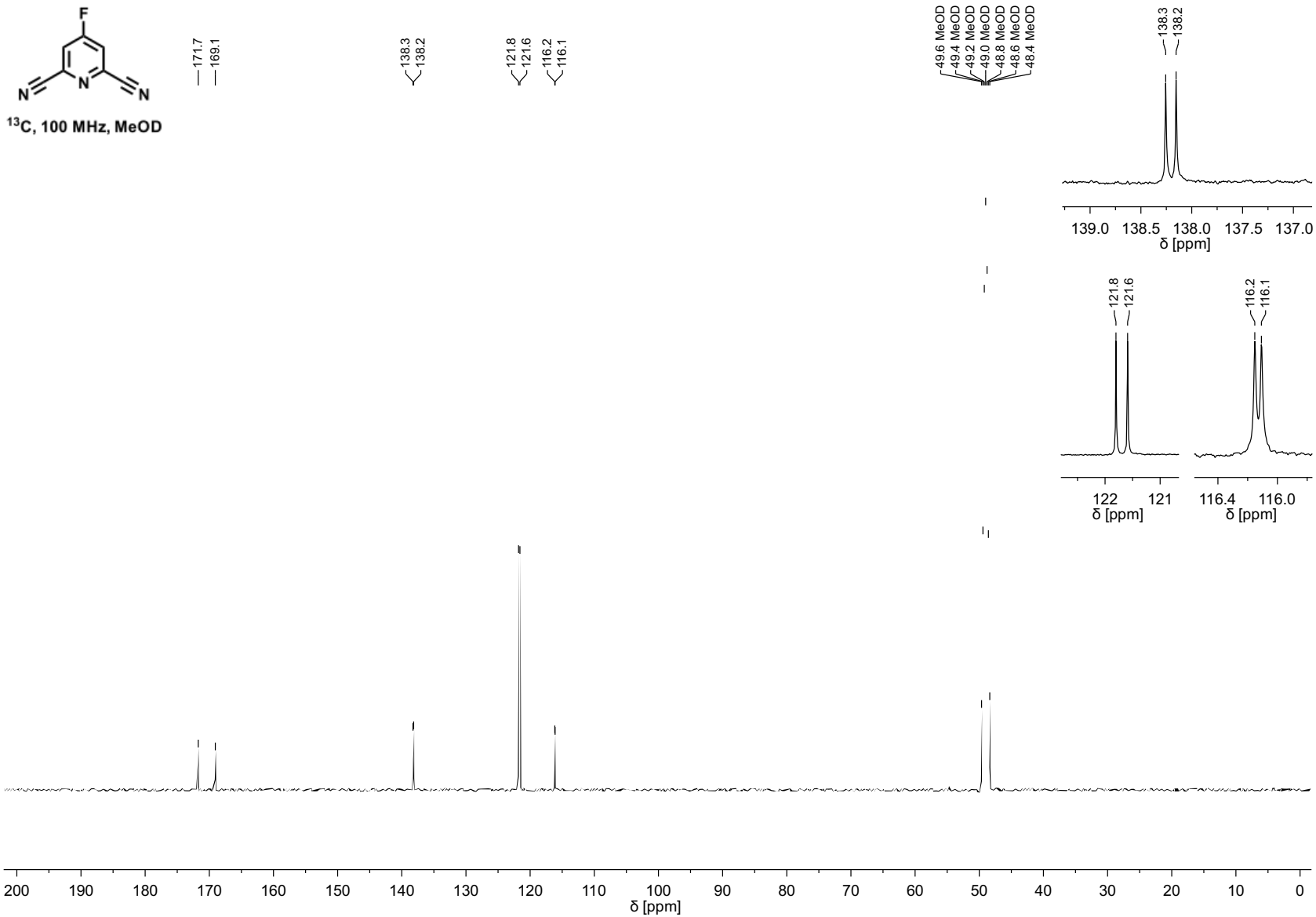


Figure S6. ^{13}C-NMR spectrum (100 MHz, MeOD) of FDCP.

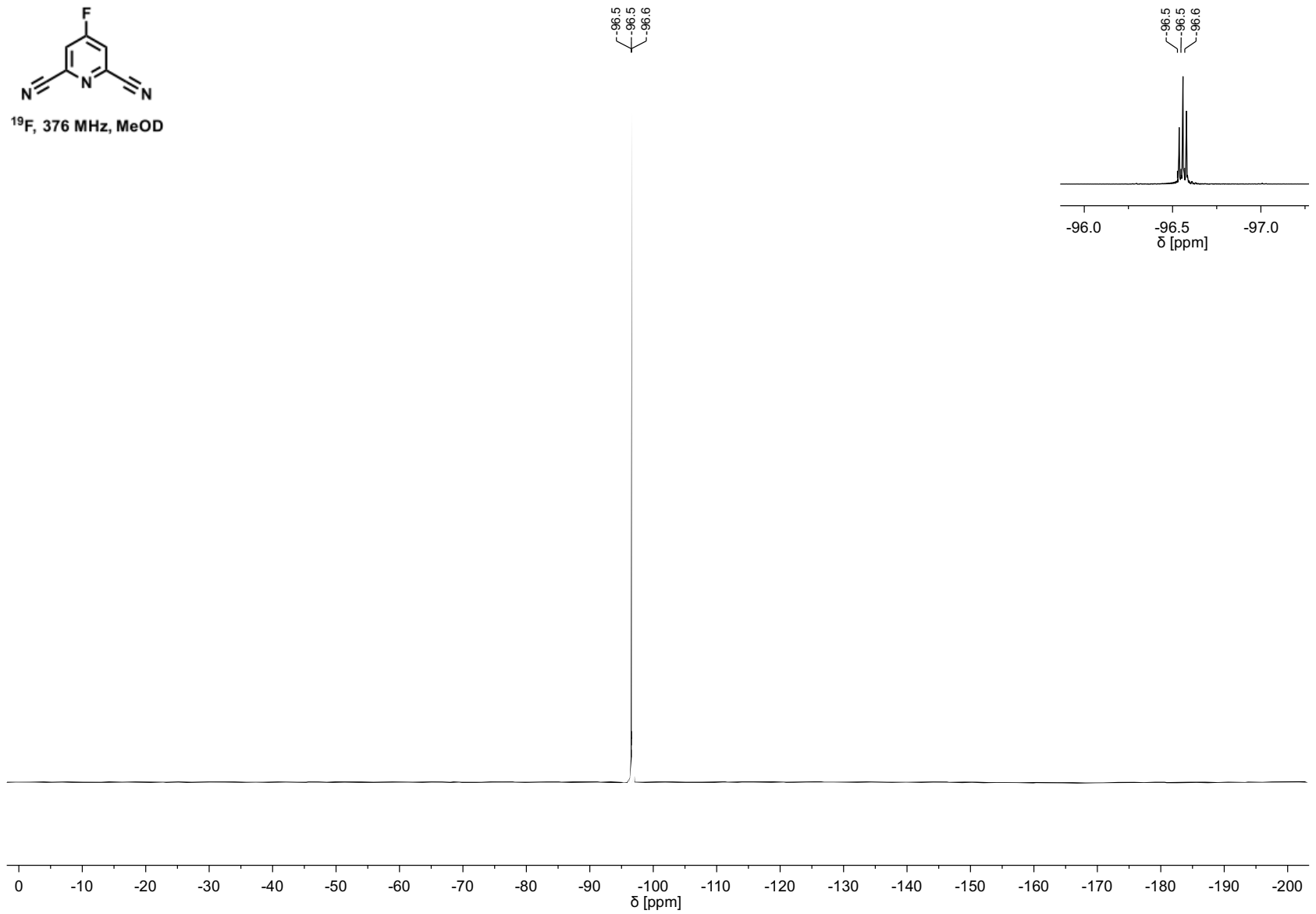


Figure S7. ^{19}F-NMR spectrum (376 MHz, MeOD) of FDCP.

Synthesis of DCP-tagged proteins

Single cysteine protein (GFP-1C, ZiP-1C, Ubq-1C, GB1-1C) solution (150 μM) was exposed to excess of FDCP from a highly concentrated stock (100 mM in DMSO) in aqueous reducing buffer and was kept on a rotisserie before analysis by intact protein mass spectrometry. Various conditions were explored, as outlined in **Table S5**, with the two most effective conditions being 15 eq. of FDCP in 50 mM Tris-HCl pH 7.5, 2 mM TCEP at 25 °C for 24 h or alternatively at 4 °C for 72 h, resulting in DCP-tagged protein. Excess reagent was removed using spin desalting columns (Thermo Fisher, USA) or buffer exchange using centrifugal filters (Millipore, USA) with a molecular weight cut-off suitable for the respective protein.

Synthesis of branched peptide-protein conjugates

DCP-tagged protein (Ubq-DCP, GB1-DCP) solution (60 μM) was exposed to 20 eq. of peptides (P1–P7) containing N-terminal cysteines in aqueous reducing buffer (50 mM Tris-HCl pH 7.5, 2 mM TCEP) at 25 °C for 24 h, leading to the attachment of two peptides through the cyanopyridine-aminothiol reaction. Excess peptide was removed by buffer exchange using centrifugal filters (Millipore, USA) with a molecular weight cut-off suitable for the respective protein.

Synthesis of cyclic peptide-protein conjugates

DCP-tagged protein (Ubq-DCP, GB1-DCP) solution (60 μM) was treated with 30 eq. of a peptide containing two 1,2-aminothiol functionalities (P8) in aqueous reducing buffer (50 mM Tris-HCl pH 7.5, 2 mM TCEP) at 25 °C, resulting in the generation of a cyclic peptide on the protein. Excess peptide was removed by buffer exchange using centrifugal filters (Millipore, USA) with a molecular weight cut-off suitable for the respective protein.

3D-visualisation of protein conjugates

For the generation of the peptide-protein conjugate models, the peptide was drawn in ChemDraw (Revvity, USA), and copied into Chimera (UCSF, USA) as a SMILES string. The structure was minimised using Chimera (UCSF, USA) before attached to a S20C mutant of PDB structure 1UBQ, before the whole conjugate was minimised again. Unless otherwise specified, proteins are represented by their predicted folds as generated by ColabFold[19] and visualised in ChimeraX.[20]

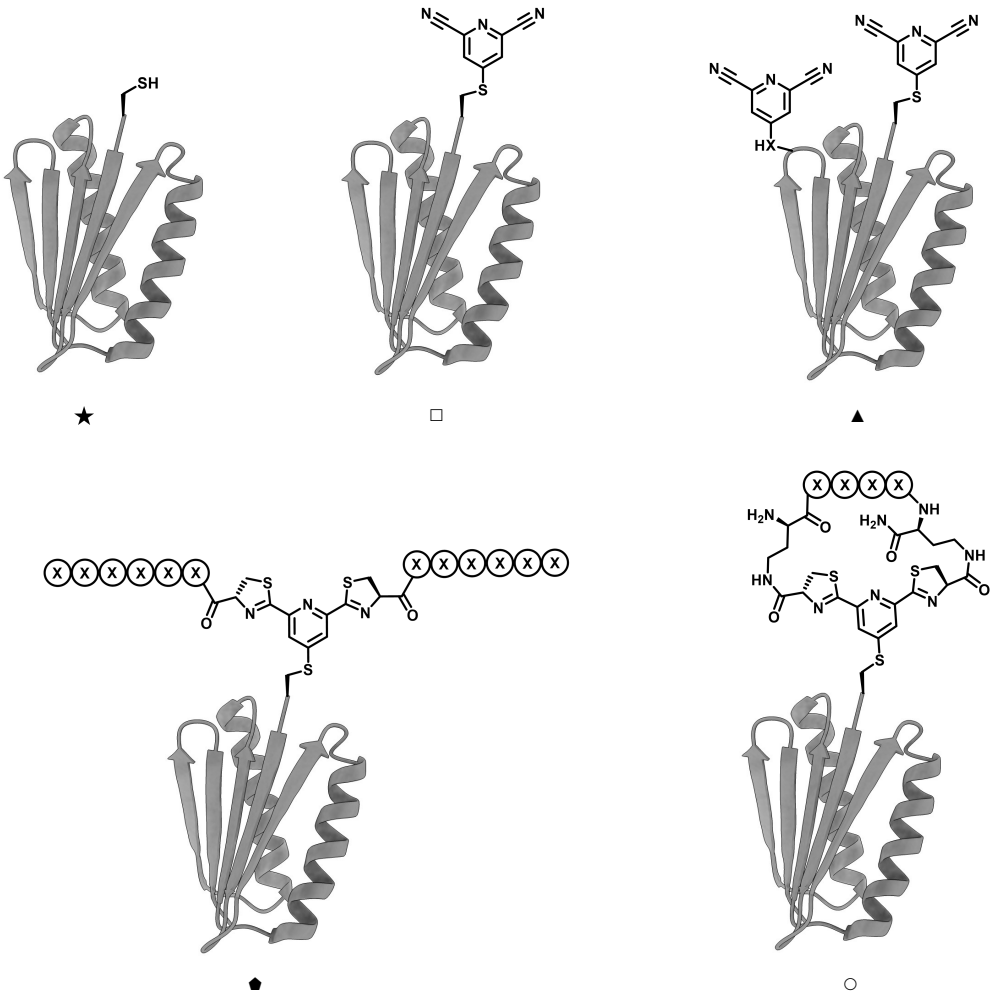


Figure S8. Legend for the following intact protein mass spectra, showing cartoon representations of single cysteine-containing protein (★), DCP-tagged protein (□), multi-DCP-tagged protein (▲; with second FDCP likely attaching to lysine or the N-terminus, X = N), branched peptide-protein conjugates (⬟) and cyclic peptide-protein conjugates (○). The protein X-ray crystal structure (PDB: 1QYS)[21] shown here is intended to represent any of the proteins used in this study.

Table S5. Conditions of reactions between 1C-proteins and FDCP.

Condition	Buffer	pH	Additive	FDCP	Temp.	Time
i	50 mM Tris-HCl, 2 mM TCEP	7.5	-	15 eq.	r.t.	24 h
ii	50 mM Tris-HCl, 2 mM TCEP	7.5	300 mM NaCl	15 eq.	r.t.	24 h
iii	50 mM Tris-HCl, 2 mM TCEP	7.5	-	30 eq.	r.t.	24 h
iv	50 mM Tris-HCl, 2 mM TCEP	7.5	-	15 eq.	37 °C	24 h
v	50 mM Tris-HCl, 2 mM TCEP	7.5	-	15 eq.	4 °C	72 h
vi	50 mM HEPES-KOH, 2 mM TCEP	8.0	-	15 eq.	r.t.	24 h
vii	50 mM MES, 2 mM TCEP	6.0	-	15 eq.	r.t.	24 h
viii	50 mM Tris-HCl, 2 mM TCEP	7.5	-	15 eq.	r.t.	3 h
ix	50 mM Tris-HCl, 2 mM TCEP	7.5	300 mM NaCl	15 eq.	r.t.	3 h
x	50 mM Tris-HCl, 2 mM TCEP	7.5	-	30 eq.	r.t.	3 h
xi	50 mM Tris-HCl, 2 mM TCEP	7.5	-	30 eq.	37 °C	24 h
xii	50 mM Tris-HCl, 2 mM TCEP	7.5	-	15 eq.	4 °C	24 h
xiii	50 mM Tris-HCl, 2 mM TCEP	7.5	-	15 eq.	4 °C	48 h
xiv	50 mM HEPES-KOH, 2 mM TCEP	8.0	-	15 eq.	r.t.	3 h
xv	50 mM HEPES-KOH, 2 mM TCEP	8.0	-	30 eq.	r.t.	3 h
xvi	50 mM HEPES-KOH, 2 mM TCEP	8.0	-	30 eq.	r.t.	24 h
xvii	50 mM Tris-HCl, 2 mM TCEP	7.5	300 mM NaCl	30 eq.	r.t.	3 h
xviii	50 mM Tris-HCl, 2 mM TCEP	7.5	300 mM NaCl	30 eq.	r.t.	24 h
xix	50 mM HEPES-KOH, 2 mM TCEP	8.0	300 mM NaCl	15 eq.	r.t.	3 h
xx	50 mM HEPES-KOH, 2 mM TCEP	8.0	300 mM NaCl	15 eq.	r.t.	24 h

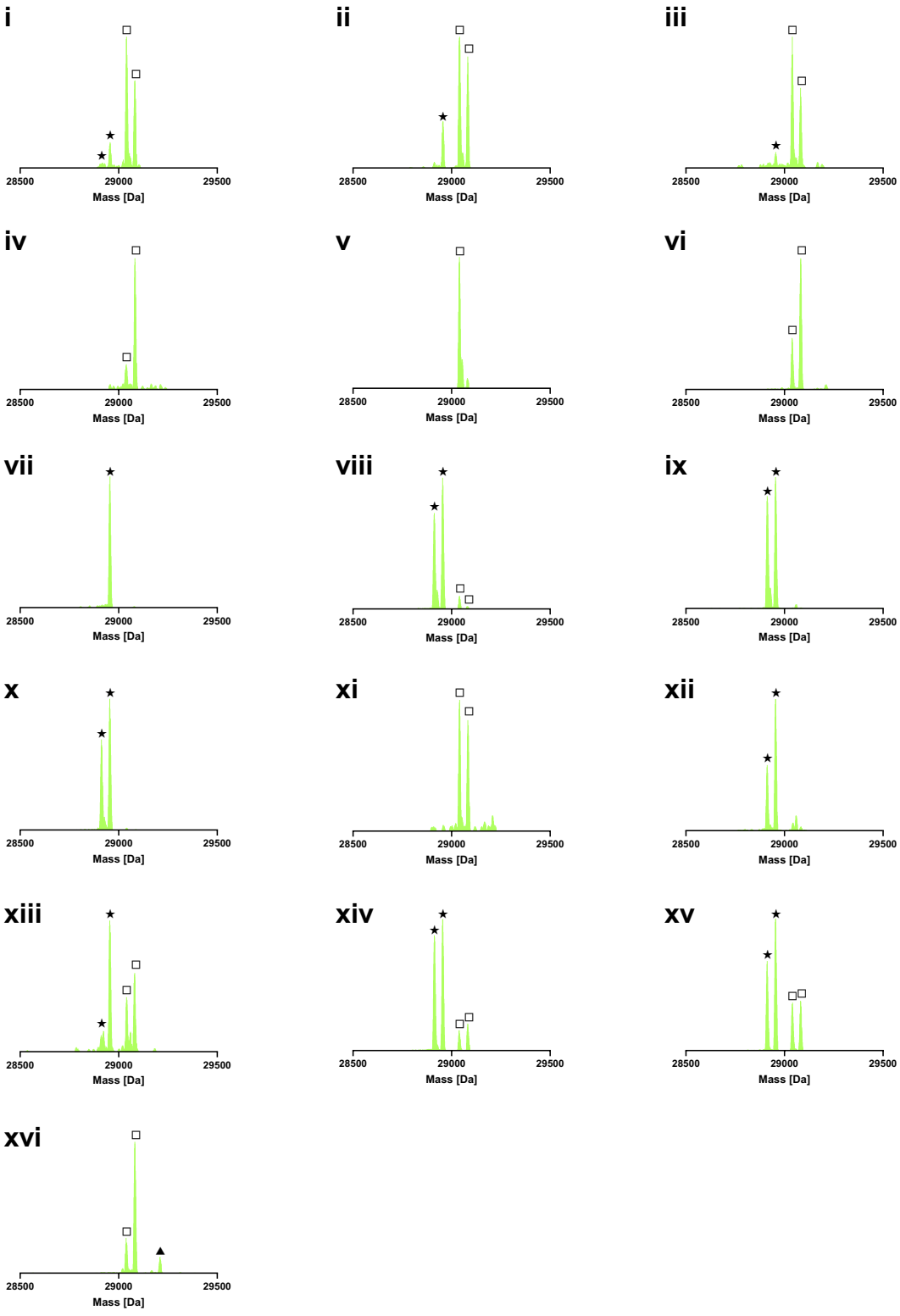


Figure S9. MS of GFP-DCP syntheses, whereby roman numeral indicates the reaction condition.

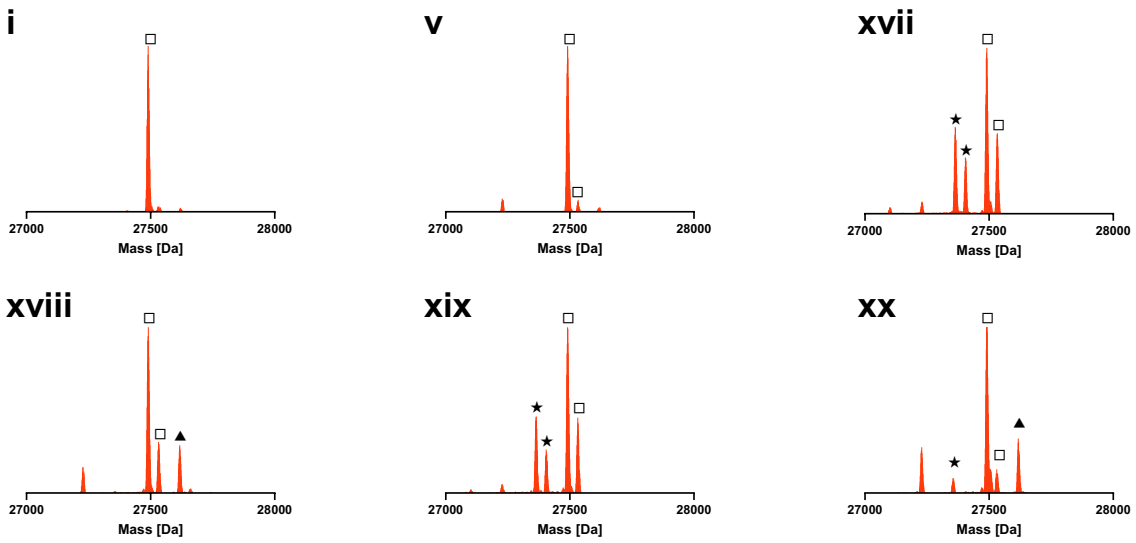


Figure S10. MS of ZiP-DCP syntheses, whereby roman numeral indicates the reaction condition.

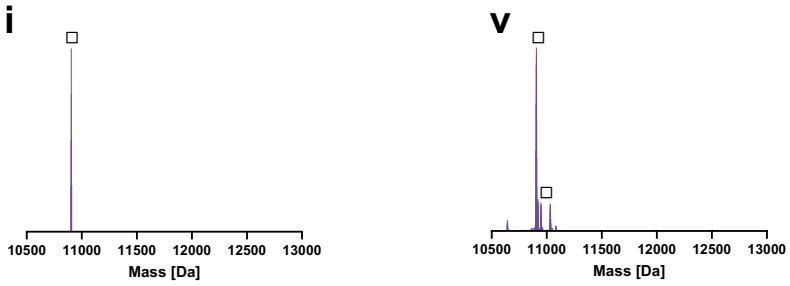


Figure S11. MS of Ubq-DCP syntheses, whereby roman numeral indicates the reaction condition.

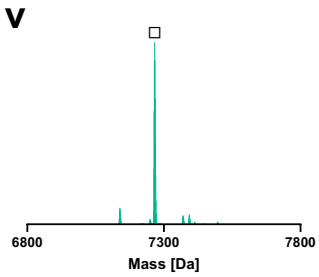


Figure S12. MS of GB1-DCP synthesis, whereby roman numeral indicates the reaction condition.

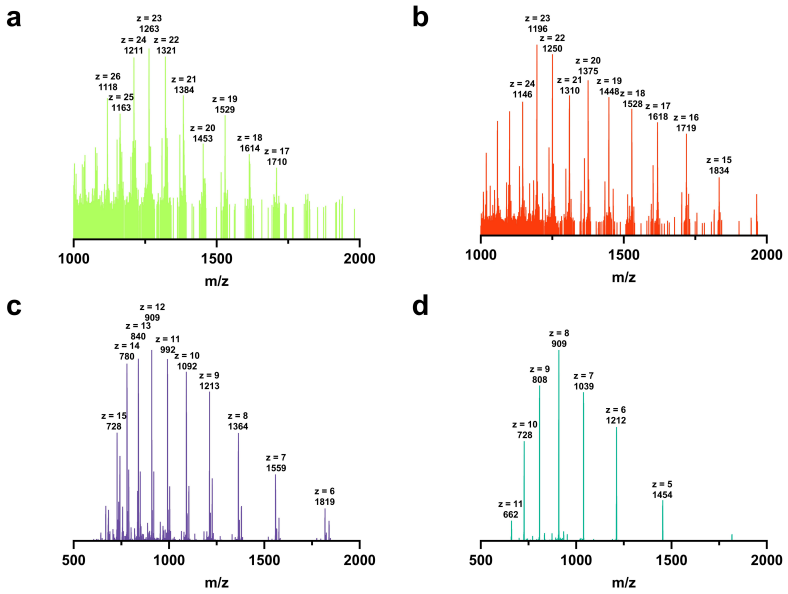


Figure S13. Original (non-deconvoluted) intact protein mass spectra of syntheses of (a) GFP-DCP, (b) ZiP-DCP, (c) Ubq-DCP (all condition i), and (d) GB1-DCP (condition v).

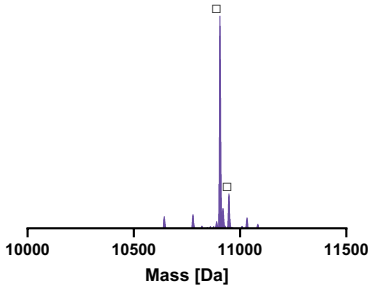


Figure S14. Deconvoluted MS of mixture of Ubq-1C and FDCP at lower protein concentration than typically tested (60 µM) under reaction condition (i), showing ~90% desired product mass.

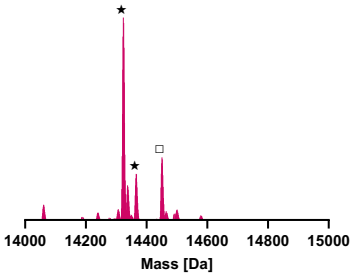


Figure S15. Deconvoluted MS of mixture of SC3 and FDCP under reaction condition (i), showing ~80% remaining unreacted starting material.

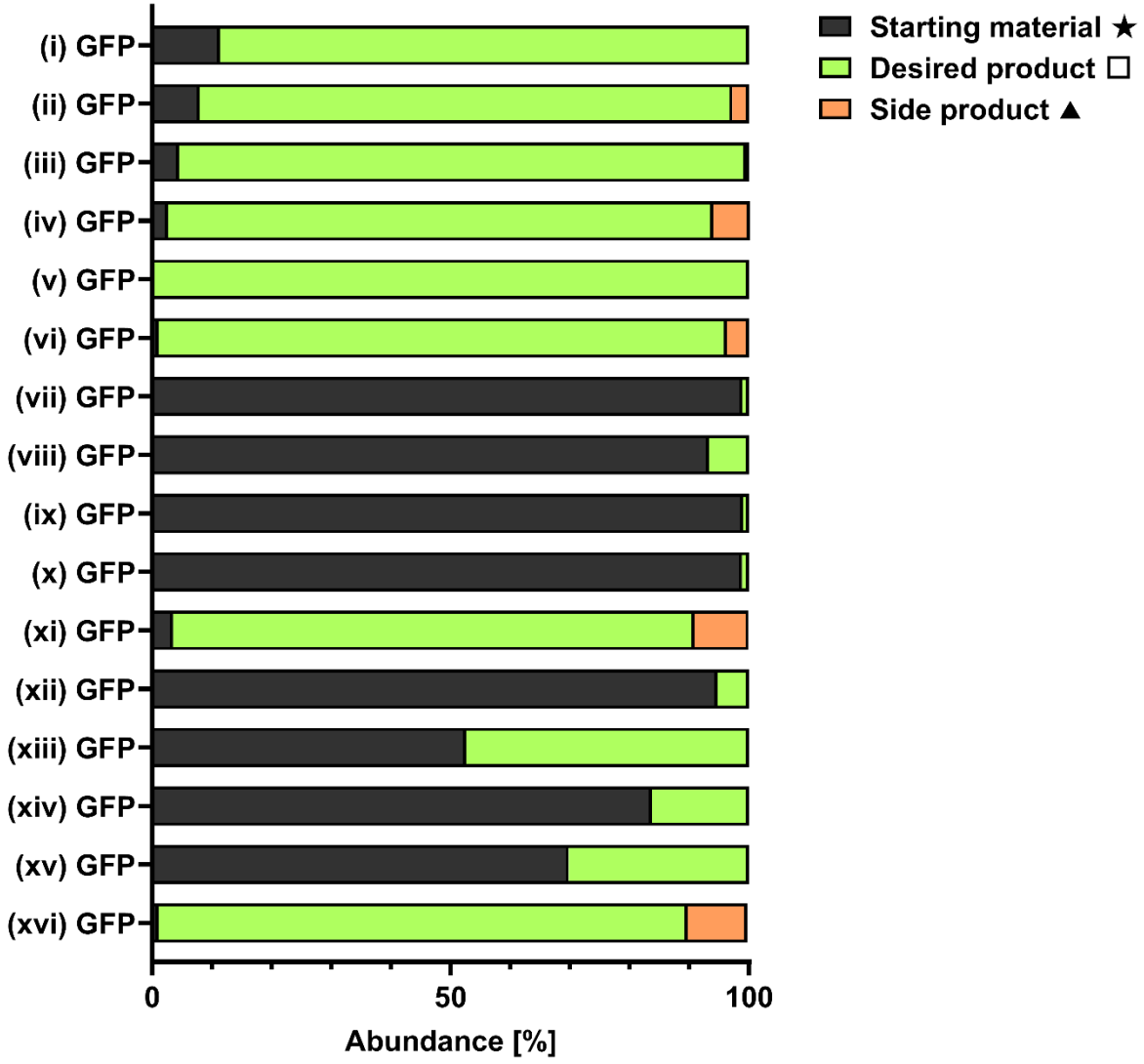


Figure S16. Percent abundance of protein starting material (black), desired product (green) and side product (orange) after exposure to FDCP under various conditions.

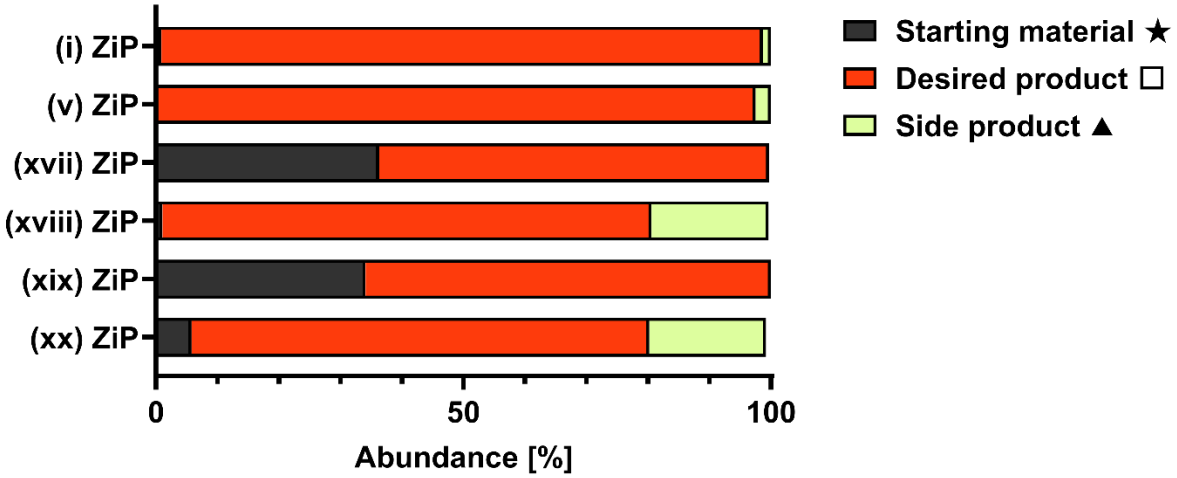


Figure S17. Percent abundance of protein starting material (black), desired product (red) and side product (yellow) after exposure to FDCP under various conditions.

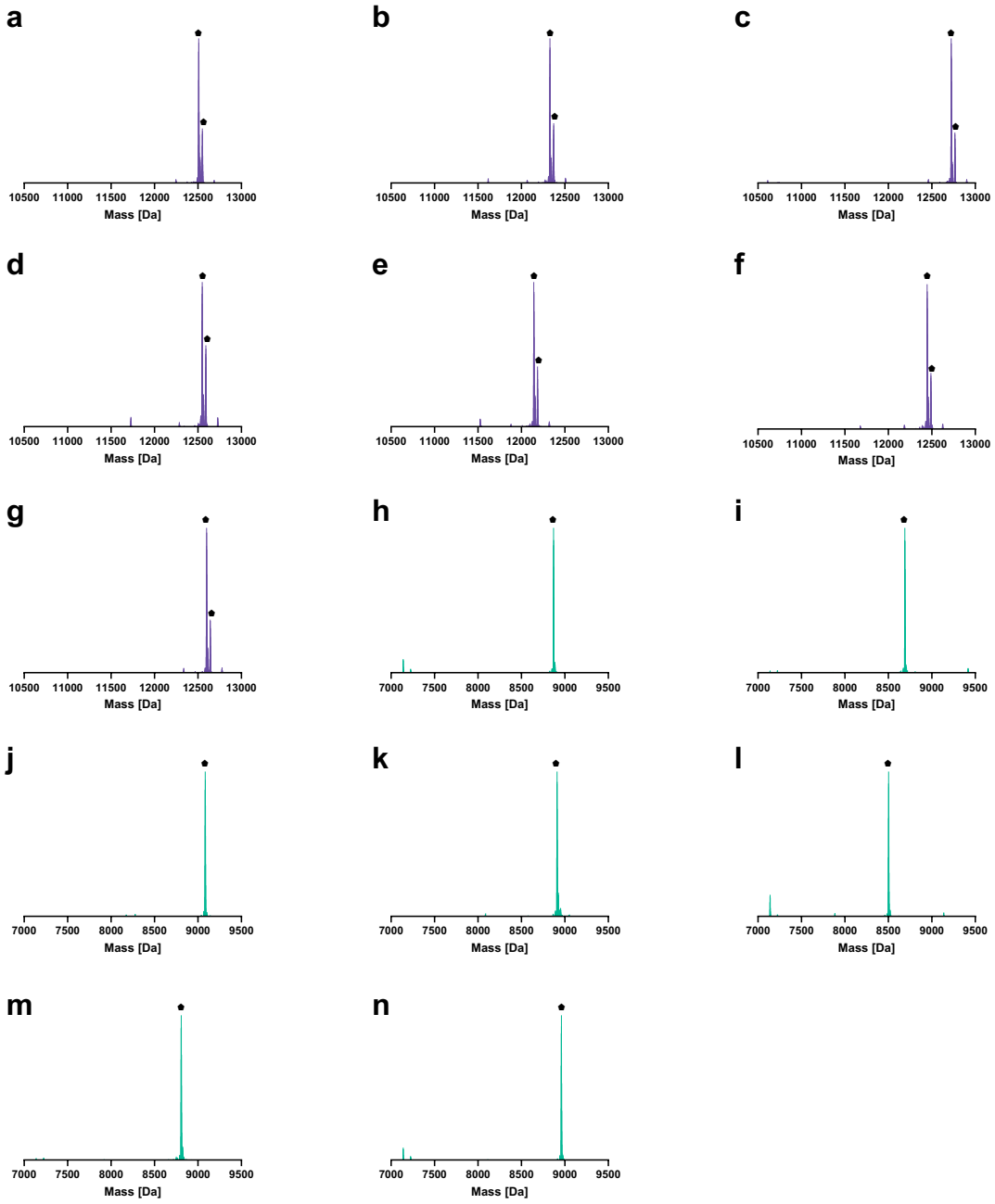


Figure S18. Deconvoluted intact protein mass spectra of branched peptide–protein conjugates. (a) Ubq–P1 (12509.1 Da), (b) Ubq–P2 (12329.0 Da), (c) Ubq–P3 (12723.4 Da), (d) Ubq–P4 (12549.1 Da), (e) Ubq–P5 (12143.0 Da), (f) Ubq–P6 (12447.0 Da), (g) Ubq–P7 (12599.2 Da), (h) GB1–P1 (8870.0 Da), (i) GB1–P2 (8689.9 Da), (j) GB1–P3 (9084.3 Da), (k) GB1–P4 (8909.9 Da), (l) GB1–P5 (8503.8 Da), (m) GB1–P6 (8807.9 Da), (n) GB1–P7 (8960.1 Da).

Cross-linking of free cysteines on GB1–P1 peptide–protein conjugate

GB1-P1 (60 µM) was incubated with 30 eq. of 1,3-bis(bromomethyl)benzene (BBMB; Tokyo Chemical Industry, Japan), in aqueous buffer (50 mM Tris-HCl, 2 mM TCEP, pH 7.5) for 24 h at room temperature before analysis using deconvoluted intact protein mass spectrometry. Excess reagent was removed by buffer exchange using centrifugal filters (Millipore, USA) with a 3 kDa molecular weight cut-off.

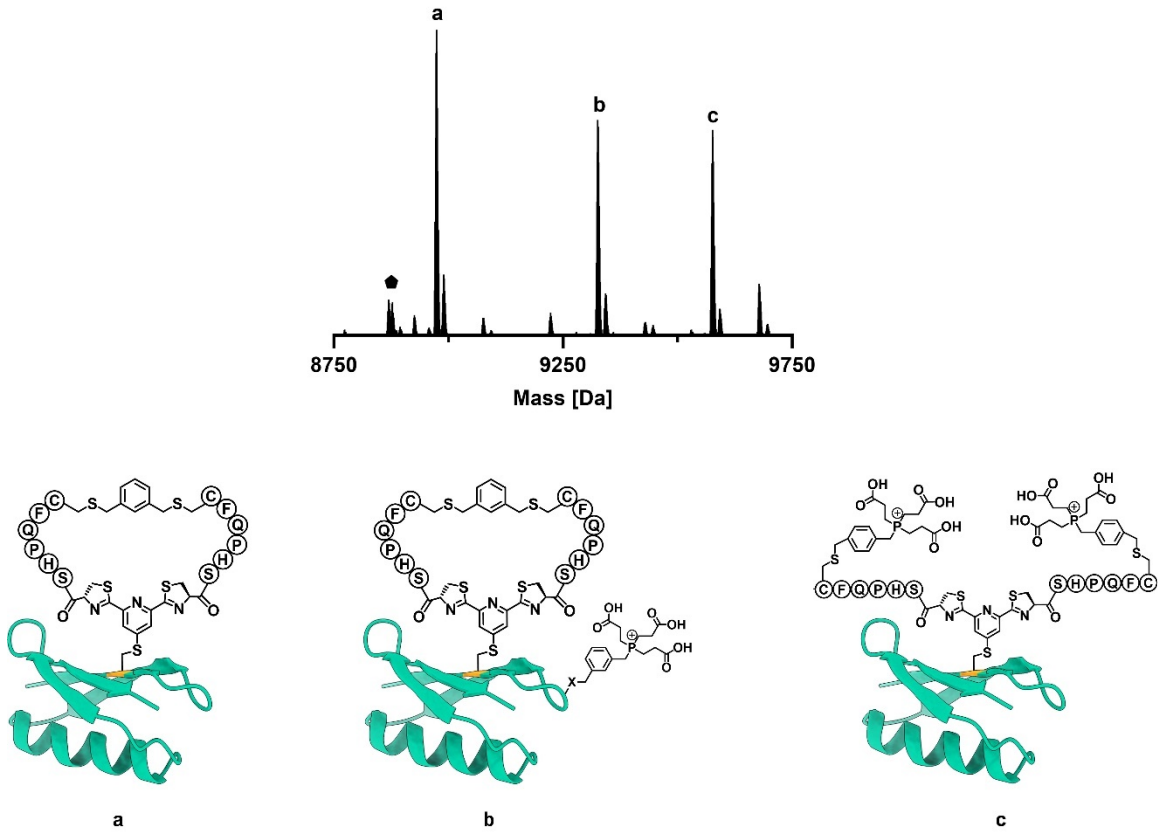


Figure S19. Deconvoluted intact protein mass spectrum of the reaction of GB1–P1 and BBMB. A side reaction between BBMB and the reducing agent TCEP generates a BBMB-TCEP adduct that leads to undesired conjugation byproducts. a = GB1–P1-BBMB cyclised product (8974.0 Da). b = GB1–P1-BBMB cyclised product with additional BBMB-TCEP adduct likely attached to lysine or the N-terminus (X = NH; 9326.1 Da). c = GB1–P1 double reacted product with the BBMB-TCEP adduct (9576.2 Da).

Zika virus protease activity assay

This protocol was adapted and modified from previous research.[1] Experiments were monitored using an Infinite 200 PRO M Plex plate reader (Tecan, Switzerland), with enzymatic reactions performed at a volume of 100 µL in black polypropylene 96-well plates (Greiner Bio-One, Austria) in triplicate. *In situ* generated ZiP-DCP (using condition i) was diluted to 10 nM and exposed to 5 µM protease substrate (Bz-Nle-Lys-Lys-Arg-AMC; Mimotopes, Australia). Acting as control experiments (ZiP-DCP −), the components of the FDCP tagging reaction (condition i) without ZiP-1C (at matching concentration) were also exposed to 5 µM substrate. The mixtures were observed for 20 min at 460 nm, using 360 nm as the excitation wavelength. Fluorescence increase rates were determined from the slope of relative fluorescence units over time in triplicate. The data were analysed and visualised in GraphPad Prism 10 (Dotmatics, USA).

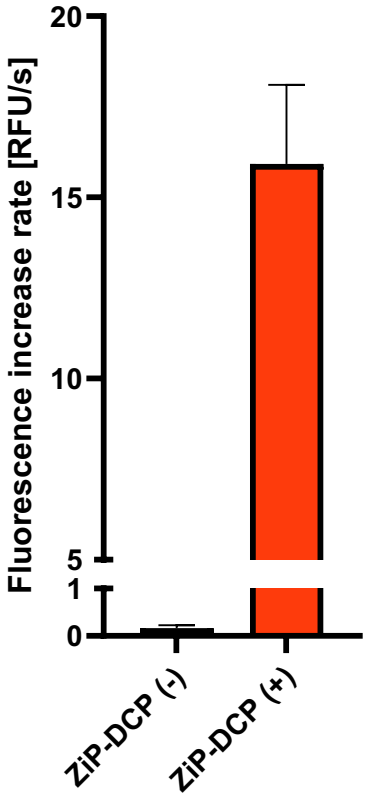


Figure S20. Enzymatic activity of *in situ* generated ZiP-DCP (red, ZiP-DCP +) compared to control reaction (condition i) without protein (black, ZiP-DCP −).

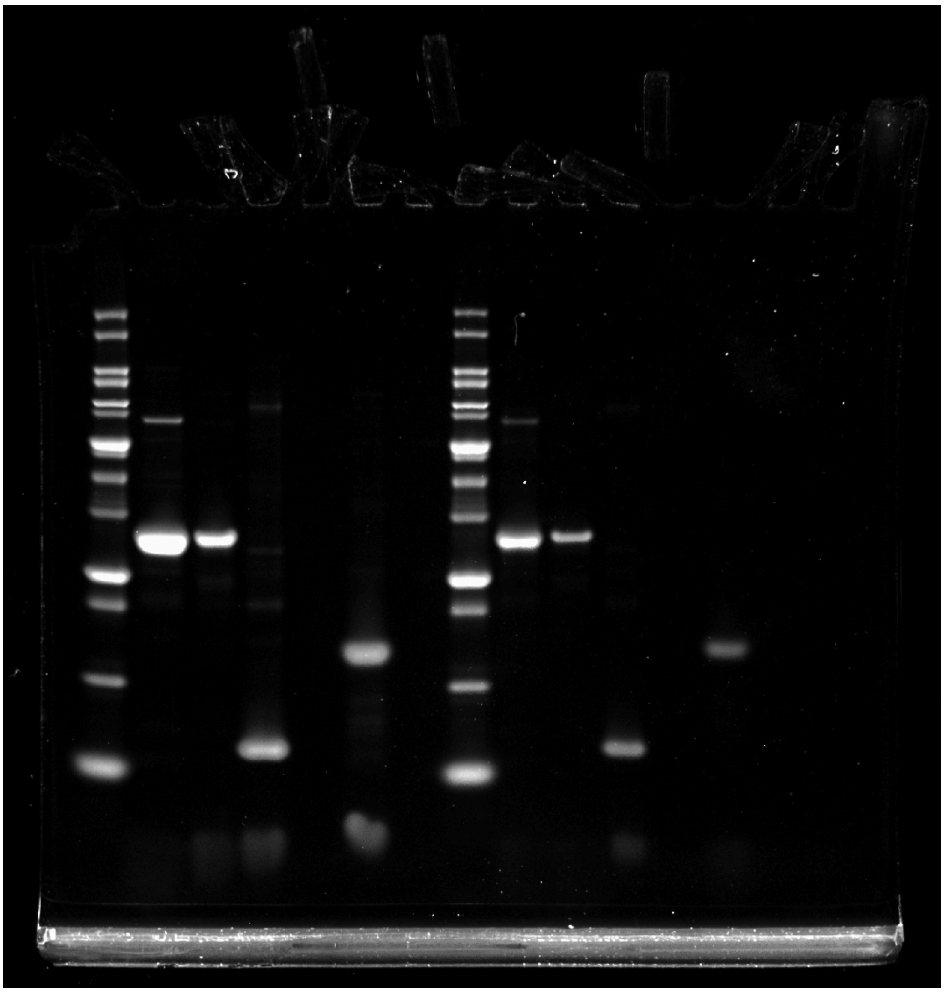


Figure S21. Complete Coomassie Blue-stained SDS-PAGE gel, from which an excerpt is shown (lanes 8–13 from the left) in Figure S1 (left).

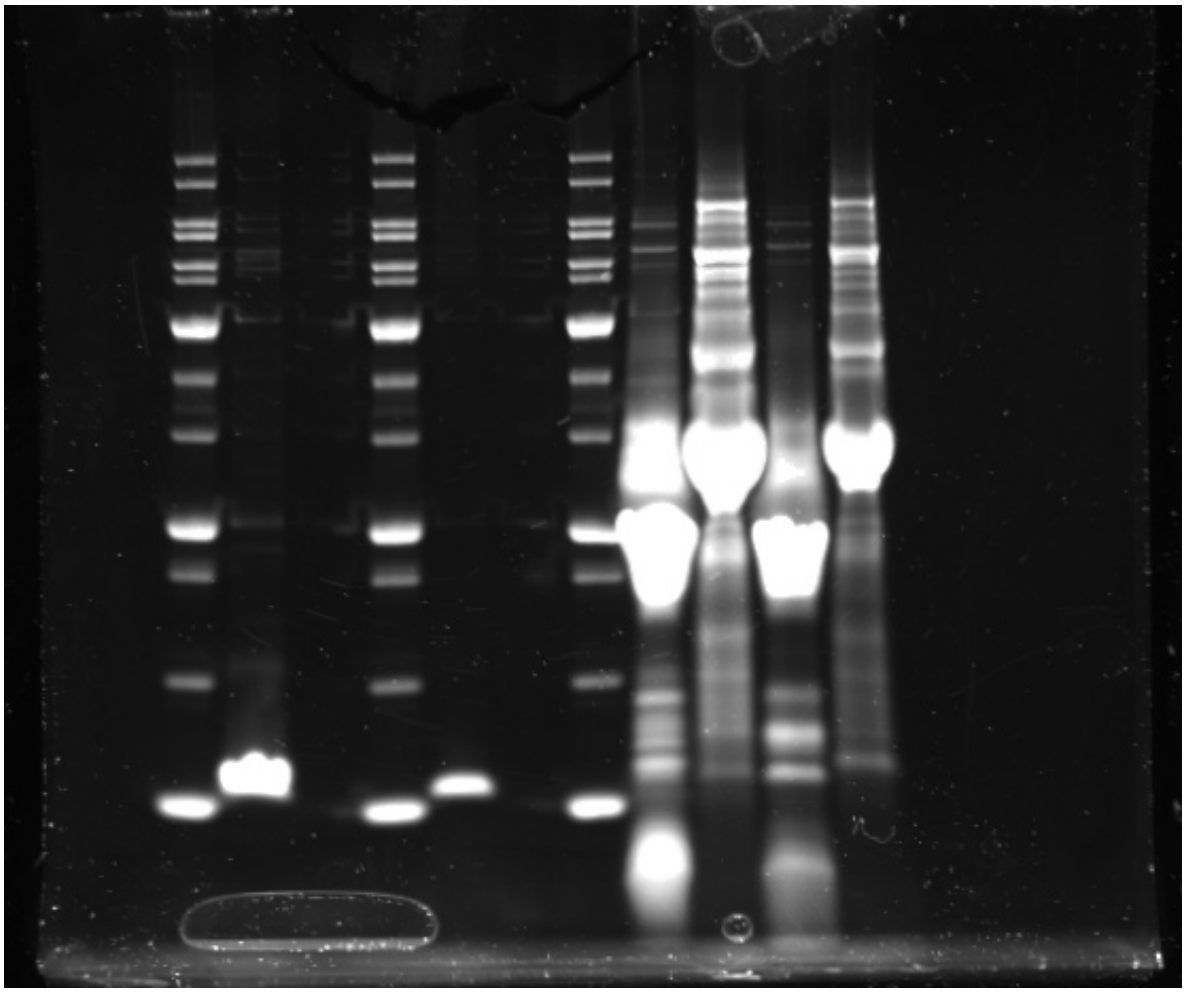


Figure S22. Complete Coomassie Blue-stained SDS-PAGE gel, from which an excerpt is shown (lanes 4 & 5 from the left) in Figure S1 (right).

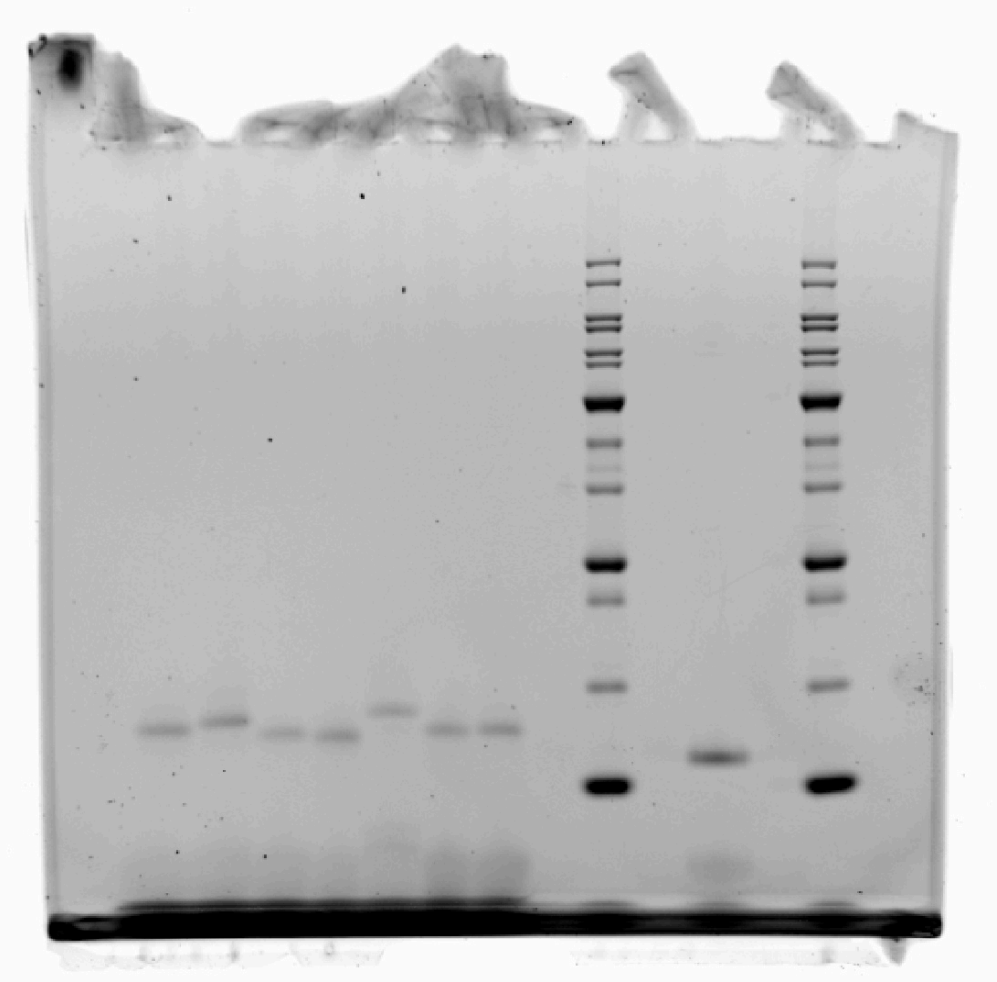


Figure S23. Complete Coomassie Blue-stained SDS-PAGE gel, from which an excerpt is shown (lanes 1–11 from the left) in Figure 4 of the main text.

Supporting references

1 S. Ullrich, B. Panda, U. Somathilake, D. J. Lawes and C. Nitsche, *Chem. Commun.*, 2025, **61**, 933–936.

2 R. Morewood and C. Nitsche, *Chem. Sci.*, 2021, **12**, 669–674.

3 G. R. Fulmer, A. J. M. Miller, N. H. Sherden, H. E. Gottlieb, A. Nudelman, B. M. Stoltz, J. E. Bercaw and K. I. Goldberg, *Organometallics*, 2010, **29**, 2176–2179.

4 M. Loos, C. Gerber, F. Corona, J. Hollender and H. Singer, *Anal. Chem.*, 2015, **87**, 5738–5744.

5 S. Ullrich, U. Somathilake, M. Shang and C. Nitsche, *Commun. Chem.*, 2024, **7**, 143.

6 E. Gasteiger, C. Hoogland, A. Gattiker, S. Duvaud, M. R. Wilkins, R. D. Appel and A. Bairoch, in *The Proteomics Protocols Handbook*, 2005, DOI: 10.1385/1-59259-890-0:571, ch. 52, pp. 571–607.

7 S. Nagasundarapandian, L. Merkel, N. Budisa, R. Govindan, N. Ayyadurai, S. Sriram, H. Yun and S. G. Lee, *ChemBioChem*, 2010, **11**, 2521–2524.

8 J. Y. Choi, T. H. Jang and H. H. Park, *FEBS Lett.*, 2016, **591**, 442–447.

9 W. W. Phoo, Y. Li, Z. Zhang, et al., *Nat. Commun.*, 2016, **7**, 13410.

10 W. Becker, L. A. Adams, B. Graham, G. E. Wagner, K. Zangger, G. Otting and C. Nitsche, *J. Biomol. NMR*, 2018, **70**, 211–218.

11 S. Vijay-Kumar, C. E. Bugg and W. J. Cook, *J. Mol. Biol.*, 1987, **194**, 531–544.

12 M. C. Mahawaththa, H. W. Orton, I. Adekoya, T. Huber, G. Otting and C. Nitsche, *Chem. Commun.*, 2022, **58**, 701–704.

13 A. M. Gronenborn, D. R. Filpula, N. Z. Essig, A. Achari, M. Whitlow, P. T. Wingfield and G. M. Clore, *Science*, 1991, **253**, 657–661.

14 V. Vitali, K. Ackermann, G. Hagelueken and B. E. Bode, *Appl. Magn. Reson.*, 2023, **55**, 187–205.

15 B. Zakeri and M. Howarth, *J. Am. Chem. Soc.*, 2010, **132**, 4526–4527.

16 A. H. Keeble, P. Turkki, S. Stokes, I. N. A. Khairil Anuar, R. Rahikainen, V. P. Hytönen and M. Howarth, *Proc. Natl. Acad. Sci. USA*, 2019, **116**, 26523–26533.

17 Association of Biomolecular Resource Facilities, A database of protein post translational modifications, https://abrf.org/resources/delta-mass/, (accessed November 28, 2024).

18 S. Ullrich, J. George, A. E. Coram, R. Morewood and C. Nitsche, *Angew. Chem. Int. Ed.*, 2022, **61**, e202208400.

19 M. Mirdita, K. Schütze, Y. Moriwaki, L. Heo, S. Ovchinnikov and M. Steinegger, *Nat. Meth.*, 2022, **19**, 679–682.

20 E. C. Meng, T. D. Goddard, E. F. Pettersen, G. S. Couch, Z. J. Pearson, J. H. Morris and T. E. Ferrin, *Prot. Sci.*, 2023, **32**, e4792.

21 B. Kuhlman, G. Dantas, G. C. Ireton, G. Varani, B. L. Stoddard and D. Baker, *Science*, 2003, **302**, 1364–1368.

Appendix

```
# Direct Conjugate Profiler
# Version 0.3.7
# 6-Dec-2024
# @s_ullr

import tkinter as tk
from tkinter import filedialog, ttk
import pandas as pd
import os
import sys

# Function to open file dialog and select Excel file
def select_file():
    filepath = filedialog.askopenfilename(filetypes=[("Excel files", "*.xlsx")])
    if filepath:
        excel_label.config(text=filepath)  # Display just the file path
        analyse_button.config(state=tk.NORMAL)  # Enable the "Analyse Data" button
        select_button.config(state=tk.DISABLED)  # Disable the "Select Excel File" button
        restart_button.config(state=tk.NORMAL)  # Enable the "Restart Program" button
    return filepath

# Function to clean and extract numeric values from columns 0 and 1 (A and B)
def clean_and_extract_numeric(df):
    df = df.iloc[:, [0, 1]]  # Focus only on columns 0 and 1 (A and B)
    df = df.apply(pd.to_numeric, errors='coerce')  # Convert invalid values to NaN
    df_clean = df.dropna()  # Drop any rows where either m/z or intensity is NaN
    return df_clean

# Function to dynamically generate dropdowns for each worksheet and pre-select based on the
worksheet name (case-insensitive)
def setup_worksheet_selection(filepath):
    xl = pd.ExcelFile(filepath)
    sheets = xl.sheet_names

    filename = filepath.split("/")[-1].lower()
    global filename_default_protein
    filename_default_protein = None
    if "ub" in filename:
        filename_default_protein = "Ub"
    elif "gfp" in filename:
        filename_default_protein = "esGFP"
    elif "gb1" in filename:
        filename_default_protein = "GB1"
    elif "zipro" in filename:
        filename_default_protein = "gZiPro"

    for i, sheet in enumerate(sheets):
        sheet_lower = sheet.lower()
        if "ub" in sheet_lower:
            default_protein = "Ub"
        elif "gfp" in sheet_lower:
            default_protein = "esGFP"
        elif "gb1" in sheet_lower:
            default_protein = "GB1"
        elif "zipro" in sheet_lower:
            default_protein = "gZiPro"
        else:
              default_protein = filename_default_protein if filename_default_protein else
"Select Protein for " + sheet

        protein_var = tk.StringVar(value=default_protein)
        protein_vars[sheet] = protein_var

        protein_label = tk.Label(scrollable_frame, text=f"Protein for {sheet}:")
        protein_label.pack()
```

```
            protein_menu  =  tk.OptionMenu(scrollable_frame,  protein_var,  *protein_options,
command=lambda val, sheet=sheet: handle_protein_selection(val, sheet))
        protein_menu.pack()
        protein_menus[sheet] = protein_menu

        handle_protein_selection(default_protein, sheet)

         mz_label = tk.Label(scrollable_frame, text=f"Enter m/z for {sheet} if 'Custom' is
selected:")
        mz_label.pack()
        mz_labels[sheet] = mz_label

        mz_entry = tk.Entry(scrollable_frame)
        mz_entries[sheet] = mz_entry
        mz_entry.pack()

        mz_label.pack_forget()
        mz_entry.pack_forget()

# Function to change colour of the dropdown button and show/hide m/z input based on selection
def handle_protein_selection(selected_protein, sheet):
    if selected_protein == "esGFP":
        protein_menus[sheet].config(bg="#AFFF5F")  # Green (GFP)
    elif selected_protein == "gZiPro":
        protein_menus[sheet].config(bg="#FE3B0C")  # Red (ZiPro)
    elif selected_protein == "Ub":
        protein_menus[sheet].config(bg="#6A4DA3")  # Purple (Ub)
    elif selected_protein == "GB1":
        protein_menus[sheet].config(bg="#00B894")  # Teal (GB1)
    elif selected_protein == "Custom":
        protein_menus[sheet].config(bg="#FFA726")
        mz_labels[sheet].pack()
        mz_entries[sheet].pack()
    elif selected_protein == "No Analysis":
        protein_menus[sheet].config(bg="lightgrey")
        mz_labels[sheet].pack_forget()
        mz_entries[sheet].pack_forget()

# Function to perform analysis for each worksheet and combine results into a single output
file
def analyse_worksheets():
    filepath = excel_label.cget("text")
    if filepath == "No file selected":
        output_label.config(text="Please select an Excel file first.")
        return

    xl = pd.ExcelFile(filepath)
    all_results = []

    for sheet in xl.sheet_names:
        df = xl.parse(sheet)
        df_clean = clean_and_extract_numeric(df)

        selected_protein = protein_vars[sheet].get()
        if selected_protein == 'No Analysis':
            continue
        if selected_protein == 'Custom':
            mz_value = float(mz_entries[sheet].get())
        else:
            mz_value = predefined_mz[selected_protein]

        tolerance = 3.0

        # Function to find highest intensity within a given tolerance
        def find_highest_intensity(mz_value, df_clean, tolerance=3.0):
            filtered_df = df_clean[(df_clean.iloc[:, 0] >= mz_value - tolerance) &
```

```
                                (df_clean.iloc[:, 0] <= mz_value + tolerance)]
            if filtered_df.empty:
                return None, None  # Return None if no valid data found
            max_idx = filtered_df.iloc[:, 1].idxmax()
            highest_intensity_row = filtered_df.loc[max_idx]
           return highest_intensity_row.iloc[0], highest_intensity_row.iloc[1]  # Return m/z
and intensity

        def get_combined_intensity(mz_value, df_clean, tolerance=3.0):
            intensity = 0
            peak_mz, orig_intensity = find_highest_intensity(mz_value, df_clean)
            if orig_intensity is not None:
                intensity += orig_intensity
           acylated_mz, acylated_intensity = find_highest_intensity(mz_value + 42, df_clean)
            if acylated_intensity is not None:
                intensity += acylated_intensity
            return intensity, peak_mz  # Also return the m/z of the highest intensity

        # Analysis for SM, DP, SPs
        sm_intensity, sm_mz = get_combined_intensity(mz_value, df_clean)
        dp_intensity, dp_mz = get_combined_intensity(mz_value + 127, df_clean)
        sp_intensities = []
        sp_mzs = []

        # Calculate side products intensities and m/z values
        for i in range(2, 6):
            sp_intensity, sp_mz = get_combined_intensity(mz_value + i * 127, df_clean)
            sp_intensities.append(sp_intensity)
            sp_mzs.append(sp_mz)

        # Sum the total intensity of side products
        sp_total_intensity = sum([i for i in sp_intensities if i is not None])

        # Calculate total intensity and percentages
        total_intensity = sm_intensity + dp_intensity + sp_total_intensity
        if total_intensity == 0:
                output_label.config(text=f"No  valid  intensities  found  for  sheet  {sheet}.
Skipping.")
            continue

        sm_percentage = (sm_intensity / total_intensity) * 100
        dp_percentage = (dp_intensity / total_intensity) * 100
        sp_percentage = (sp_total_intensity / total_intensity) * 100

        # Filter out None values from sp_mzs before finding the max.
        valid_sp_mzs = [mz for mz in sp_mzs if mz is not None]

        # Append results for the current sheet
        all_results.append({
            'Sheet': sheet,
            'Protein': selected_protein,
            'SM': round(sm_percentage, 1),
            'DP': round(dp_percentage, 1),
            'SP': round(sp_percentage, 1),
             'Highest Peak m/z': sm_mz if sm_intensity >= dp_intensity and sm_intensity >=
sp_total_intensity else (
                                        dp_mz  if  dp_intensity  >=  sp_total_intensity  else
(max(valid_sp_mzs) if valid_sp_mzs else None)),
            'File': filepath
        })

    # After looping through sheets, save results if any are present
    if all_results:
        combined_df = pd.DataFrame(all_results)
        output_file_path = filepath.replace('.xlsx', '_combined_percent-analysis.xlsx')
        combined_df.to_excel(output_file_path, index=False)
        output_label.config(text=f"Results saved to {output_file_path}")
```

```
        analyse_button.config(state=tk.DISABLED)
    else:
        output_label.config(text="No valid data to save.")

# Tkinter setup
root = tk.Tk()
root.title("Direct Conjugate Profiler")

# Canvas and scrollbar for scrollable content
canvas = tk.Canvas(root)
scrollbar = ttk.Scrollbar(root, orient="vertical", command=canvas.yview)
scrollable_frame = ttk.Frame(canvas)

scrollable_frame.bind(
    "<Configure>",
    lambda e: canvas.configure(scrollregion=canvas.bbox("all"))
)

canvas.create_window((0, 0), window=scrollable_frame, anchor="nw")
canvas.configure(yscrollcommand=scrollbar.set)

canvas.pack(side="left", fill="both", expand=True)
scrollbar.pack(side="right", fill="y")

protein_vars = {}
protein_menus = {}
mz_entries = {}
mz_labels = {}

# Frame for buttons at the bottom
button_frame = tk.Frame(root)
button_frame.pack(fill="x", pady=10)

# Label to show selected file
excel_label = tk.Label(scrollable_frame, text="No file selected")
excel_label.pack()

protein_options = ["esGFP", "gZiPro", "Ub", "GB1", "Custom", "No Analysis"]
predefined_mz = {"esGFP": 28911.4, "gZiPro": 27360.4, "Ub": 10777.5, "GB1": 7139.4}

# Buttons and output label in the button frame
output_label = tk.Label(button_frame, text="")
output_label.pack(side="left", padx=10)

select_button = tk.Button(button_frame, text="Select Excel File", command=lambda:
setup_worksheet_selection(select_file()))
select_button.pack(side="left", padx=10)

analyse_button = tk.Button(button_frame, text="Analyse Data", command=analyse_worksheets,
state=tk.DISABLED)
analyse_button.pack(side="left", padx=10)

restart_button = tk.Button(button_frame, text="Restart Program", command=lambda:
root.destroy())
restart_button.pack(side="left", padx=10)

root.mainloop()
```