Electronic Supplementary Material (ESI) for Organic & Biomolecular Chemistry.

Electronic supplementary information for

Rapid sodium periodate cleavage of an unnatural amino acid enables unmasking of a highly reactive α-oxo aldehyde for protein bioconjugation

Robin L. Brabham[a†], Tessa Keenan[a†], Annika Husken[a], Jacob Bilsborrow[a], Ryan McBerney[b], Vajinder Kumar[b], W. Bruce Turnbull[b] and Martin A. Fascionea[*].

[a] Department of Chemistry, University of York, Heslington, YO10 5DD, York, United Kingdom

[b] School of Chemistry and Asbury Centre for Structural Molecular Biology, University of Leeds, Leeds LS2 9JT, United Kingdom

[†] These authors contributed equally

[*] *Corresponding author:* martin.fascione@york.ac.uk

1. Materials, Reagents and Instrumentation

1.1 General procedures and materials

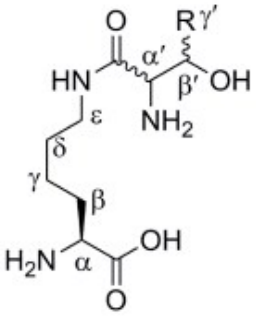
All solvents were dried prior to use according to standard methods, with the exception of solvents used for flash chromatography purposes, where GPR-grade solvents were used. All commercially-available reagents were used as received. Analytical grade reagents were supplied by Sigma-Aldrich, Fisher Scientific, VWR International, and TCI. **5** was purchased from Sirius Fine Chemicals. All solution-phase reactions were carried out under a dry nitrogen atmosphere using oven-dried glassware unless otherwise stated. All concentrations were performed *in vacuo* unless otherwise stated. Thin layer chromatography was carried out on Merck silica gel 60 F_{254} pre-coated aluminium foil sheets and these were visualised using UV light (254 nm) or charred following immersion in ninhydrin stain (1.5 g in 100 mL *n*-butanol and 3 mL acetic acid).

1.2 Mass spectrometry and NMR

Small-molecule high resolution mass spectrometry (HRMS) data were obtained at room temperature on a Bruker Daltonics microTOF instrument. Analytical high performance liquid chromatography (HPLC) of peptides was performed on a Shimadzu Prominence HPLC equipped with a Shimadzu photodiode array using an Accucore C18 2.6 µm column, 2.1 × 150 mm**.** All samples were analysed by using gradients of HPLC-grade H_2O and MeCN, spiked with 0.1% (v/v) FA. Infra-red (IR) spectra were recorded on a PerkinElmer UATR 2 spectrometer. Optical rotations were recorded at 20 °C using a Bellingham & Stanley ADP450 series polarimeter and values reported in units of 10^{-1} deg cm^2 g^{-1}.

Protein electrospray ionisation (ESI) mass spectra were obtained on a Bruker Solarix XR 9.4 T FTICR mass spectrometer. Samples were desalted and analysed at a final concentration of 0.3-10 µM in 50:50:1 (v/v) H_2O:MeCN:FA. Mass spectra were analysed and deconvoluted using Bruker DataAnalysis 4.4.

^{1}H and ^{13}C NMR spectra were recorded at 500 MHz and 126 MHz respectively on a Bruker AV500 NMR spectrometer using an internal deuterium lock at room temperature. Signals were assigned using additional DEPT-135, COSY, HSQC and HMBC experiments. Chemical shifts are reported in ppm according to the following references: DMSO-d6: δH 2.50; δC 39.52, centre of septet. The following abbreviations were used to describe signal multiplicities or appearances: s, singlet; d, doublet; t, triplet; dt, doublet of triplets; q, quartet; qd, quartet of doublets; quint, quintet; dq, doublet of quintets; m, multiplet. Acyclic amino acid carbon atoms are labelled with Greek letters (in subscript) in accordance with IUPAC conventions, with hydrogen and nitrogen atoms labelled with the Greek letter of the carbon atom to which they are attached. Prime symbols (') are used to label atoms of the second residue in a dipeptide.



2. Protein expression and characterisation

2.1 Plasmids

The vectors pEVOL, encoding the wild type *M. mazei* pylT and pylRS genes[1], and pBAD encoding the sfGFP(N150TAG) gene (Addgene plasmid #85483)[2] have been previously described.

2.2 tRNA synthetase NCAA incorporation assay

The pBAD vector (Amp^R) encoding the Ser-EGFP(Y39TAG) gene, together with the pEVOL vector (Cm^R) encoding the wild type *M. mazei* pylT and pylRS genes, were co-transformed into electrocompetent *E. coli* Top10 cells and selected on LB agar plates containing ampicillin (100 μg/ml) and chloramphenicol (35 μg/ml) at 37 ˚C for 16 h. Starter cultures were prepared by picking single colonies into Terrific Broth medium (TB) and grown at 37 ˚C for 16 h with shaking (220 rpm). TB (100 mL) containing ampicillin (100 μg/ml) and chloramphenicol (35 μg/ml) in a 250 mL baffled conical flask was inoculated with 1.0 mL of starter culture. The preparative cultures were grown at 37 °C with shaking (220 rpm). Typically within 3 h, the cultures reached an OD_{600} of 0.2-0.3. At this point, 10 mL from each preparative culture was transferred into a 50 mL Falcon tube, to which 400 μL of test NCAA (**1**, **2**, **3** or **4**) or positive control (**5**) was added (stock concentration 50 mM in 0.1 M NaOH, final concentration 2 mM). For negative controls, 400 μL of a solution of 0.1 M NaOH was added. The cultures were grown at 37 °C with shaking (220 rpm) for 30 min, at which point 10 μL of L-arabinose (stock 20% (w/v), final concentration 0.02% (w/v)) was added to induce gene expression and the cultures were left to grow for 16-18 h at 37 °C with shaking (220 rpm). The cells from 1.5 mL of each culture were isolated by centrifugation (7000 × *g*, 10 min). After discarding the supernatant, each pellet was resuspended in 75 μL of Bugbuster Protein Extraction Reagent (Novagen) with 0.75 u Benzonase nuclease and a protease inhibitor cocktail, and incubated at r.t. for 1 h with gentle rotation. The lysate was clarified by centrifugation (17 000 × *g*, 30 min) and the supernatant was visualised by fluorescence or analysed by SDS-PAGE with coomassie staining.

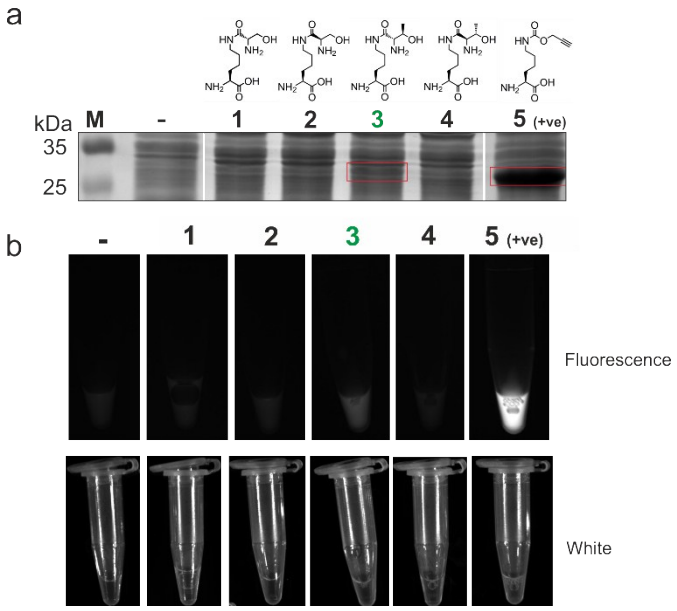


Figure S1: NCAAs **1, 2, 3**, and **4** screened for uptake in amber stop codon suppression by the *M. mazei* Pyl tRNA-RS pair into EGFP. SDS-PAGE (a) and fluorescent imaging (b) of cell lysates, with negative (-) and positive (**5**) controls. The circled bands in lanes with compounds **3** and **5** are attributed to full-length EGFP.

2.3 Protein production and purification of sfGFP(N150ThrK) 6

The pBAD vector (Amp^R) encoding the sfGFP(N150TAG) gene, together with the pEVOL vector (Cm^R) encoding the wild type *M. mazei* pylT and pylRS genes, were co-transformed into electrocompetent *E. coli* Top10 cells and selected on LB agar plates containing ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml) at 37 ˚C for 16 h. Starter cultures were prepared by picking single colonies into TB and grown at 37 ˚C for 16 h with shaking (220 rpm). 10 mL of starter culture was inoculated into 1 L of fresh TB containing ampicillin (100 µg/ml) and chloramphenicol (35 µg/ml), and grown at 37 °C with shaking (220 rpm). The cells typically grew to an OD_{600} of 0.2-0.3 within 3 h, at which point NCAA (stock solution 80 mM in 0.1 M NaOH (aq.)) was added to a final concentration of 1.5 mM. The cultures were allowed to grown at 37 °C with shaking (220 rpm) until the OD_{600} reached 0.4-0.6, at which point protein expression was induced by the addition of L-arabinose (stock solution 20% (w/w)) at a final concentration of 0.02% (w/w). The cultures were then grown at 37 °C with shaking (220 rpm) for 16-18 h. The cultures were harvested by centrifugation (6000 × g, 4 °C, 20 min). Pellets were resuspended in binding buffer (4 × sodium phosphate buffer (PB), 150 mM NaCl, 10 mM imidazole, pH 8) containing a protease inhibitor cocktail, and lysed by sonication on ice (30 s on/30 s off for 6 - 10 min). The lysate was clarified by centrifugation (20 000 × g, 4 °C, 20 min) and loaded onto a Ni HiTrap Chelating HP column (5 ml, GE Healthcare) pre-equilibrated in binding buffer. The column was washed with 10 column volumes (CV) of binding buffer and the recombinant protein was eluted on a gradient of 0 - 100% elution buffer (4 × PB, 150 mM NaCl, 500 mM imidazole, pH 8) over 7.5 CV. Fractions containing full-length protein (as determined by SDS-PAGE) were pooled, dialysed into dialysis buffer (1 × PB, 15 mM NaCl, pH 7.4), concentrated (Vivaspin Centrifugal Concentrator, 10 000 MWCO) to a final concentration of 330 µM (as determined by UV-visible spectroscopy, $\varepsilon_{280} = 2.0 \times 10^4$ $dm^3\,mol^{-1}\,cm^{-1}$) and stored at -80 °C. A typical yield was 18 mg/L.

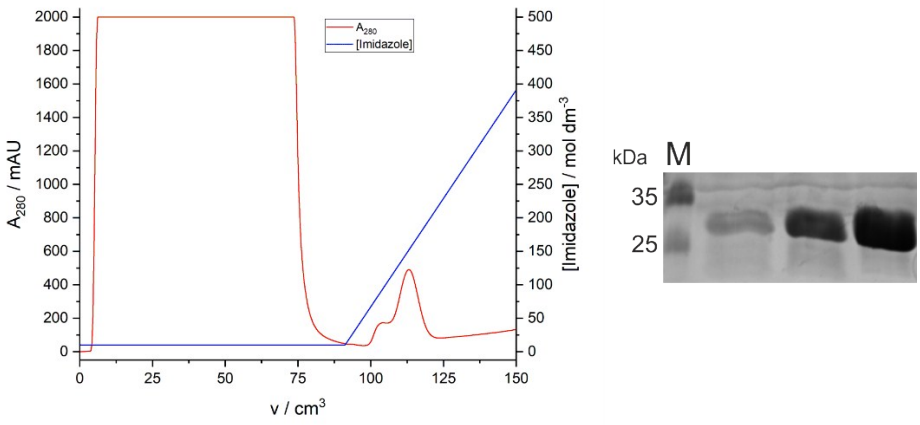


Figure S2: (left) 280 nm chromatogram of the purification of **6** by nickel affinity chromatography, eluting with an imidazole gradient; (right) SDS-PAGE of fractions containing **6.** Expected molecular mass: 28.6 kDa mass. Full SDS-PAGE gel shown in **Figure S8**.

2.4 Characterisation of sfGFP(N150ThrK) 6 by mass spectrometry

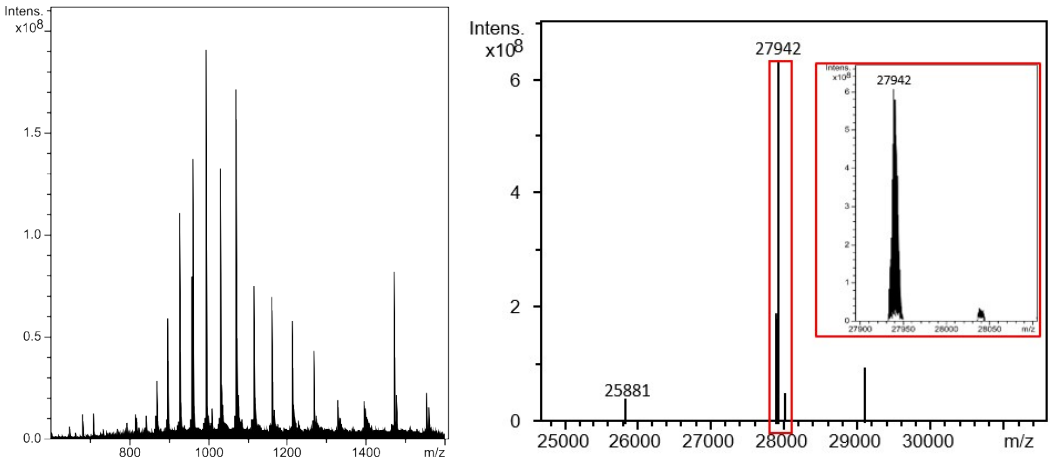


Figure S3: (left) ESI-FTICRMS of **6** raw data; (right) deconvoluted spectrum of **6**, calc. 27943 Da; found 27942 Da.

2.5 Sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE)

SDS-PAGE was performed using handcast 12% polyacrylamide gels. Samples were reduced by boiling for 10 min in reducing sample buffer (2% SDS, 2 mM 2-mercaptoethanol, 4% glycerol, 40 mM Tris-HCl pH 6.8, 0.01% bromophenol blue). Molecular weight markers used were PageRuler Plus Prestained Protein Ladder (Thermo Scientific). Each gel was run at 200 volts for ~45 min. For Coomassie staining, gels were washed with fixing solution (40% (v/v) MeOH, 10% AcOH), stained with 0.1% (w/v) Coomassie Brilliant Blue R-250 (50% (v/v) MeOH, 10% AcOH), and then washed with destain solution (50% (v/v) MeOH, 10% AcOH). Images of the gels were captured using a Syngene

G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera, with GeneSys software (Version 1.5.7.0).

2.6 Fluorescent imaging

For fluorescent imaging, SDS-PAGE gels were washed with fixing solution (40% (v/v) MeOH, 10% (v/v) AcOH). Visualisation of fluorescence was carried out on the Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera in line with GeneSys software (Version 1.5.7.0), at an excitation wavelength of 302 nm.

2.7 Western blot analysis of biotinylated proteins

For western blot analysis, biotinylated proteins and non-biotinylated negative controls (2.5 µg) were separated by SDS-PAGE (12% acrylamide gels) and transferred onto a nitrocellulose membrane (0.45 µm, Amersham Protran Sandwich, GE Healthcare) using an electroblot apparatus (Bio-Rad, Hercules, CA) at 100 V, 350 mA for 1 h in cooled transfer buffer (25 mM Tris–HCl pH 8.3, 192 mM glycine, 0.1% SDS, 20% (v/v) methanol). The membrane was blocked in PBS + 5% (w/v) milk for 16 h at 4 °C, with agitation. The membrane was then incubated in alkaline phosphatase anti-biotin (goat, Vector Labs, CA) (1:1000 in PBS) for 1 h at r.t. and then washed 3 x 5 min in PBS + 0.01% (v/v) Tween-20. Detection of biotinylated proteins was performed using the BCIP/NBT Alkaline Phosphatase Substrate Kit (Vector Labs, CA) until immunoreactive proteins on the membrane were visible (*ca*. 20 min). The reaction was stopped by washing the membrane in distilled water. The membranes were imaged using a Syngene G:BOX Chemi XRQ equipped with a Synoptics 4.0 MP camera, with GeneSys software (Version 1.5.7.0).

2.8 Determination of protein concentration

Protein concentration was estimated using UV-visible absorption spectroscopy and the predicted molar absorption coefficient at 280 nm (ExPASy ProtParam).

2.9 Trypsin digest

All trypsin digests were performed by the University of York Bioscience Technology Facility. A solution of protein (50 µg/mL in 100 mM ammonium bicarbonate) was reduced with 5 mM DTT for 30 min at 50 °C and alkylated with 15 mM iodoacetamide for 30 min, in the dark at r.t. After the addition of trypsin (0.2 µg sequencing grade, Promega), samples were incubated overnight at 37 °C. Protease activity was stopped with the addition of aqueous trifluoroacetic acid (1% (v/v)). The resulting mixture of peptides was loaded onto a Pepmap (50 cm × 2 µm, 100 Å) and a Thermo C18 EasyNano nanocapillary column (15 cm × 75 µm) and eluted over a gradient of aqueous 3-35% (v/v) MeCN over 35 min into the Thermo Orbitrap Fusion hybrid mass spectrometer. MS1 and MS2 spectra were acquired in the Orbitrap mass analyser with Easy-IC internal calibration. Data-dependent acquisition was performed in top speed mode using a fixed 1 s cycle, selecting the most intense precursors with charge states 2-5. HCD was used for peptide fragmentation with 32% activation energy. Resulting spectral data were searched against the expected protein sequences using the PEAKS-DB search program. Search criteria specified: enzyme, trypsin; peptide tolerance, 3

ppm; MS/MS tolerance, 0.01 Da. Carbamidomethylation (C) was set as a fixed modification. Peptide matches were filtered to achieve a global false discovery method of < 5 %.

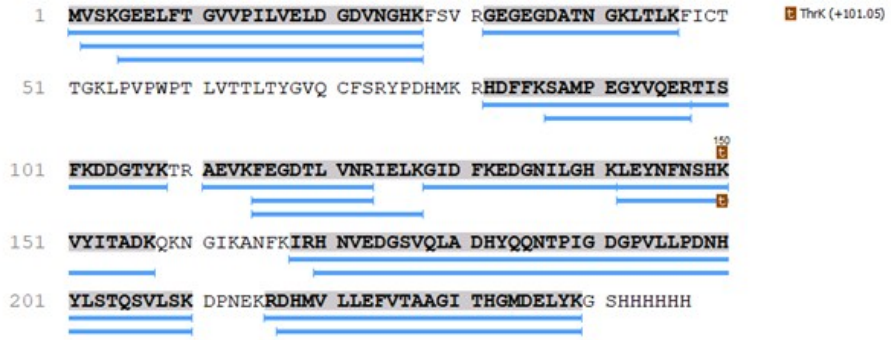
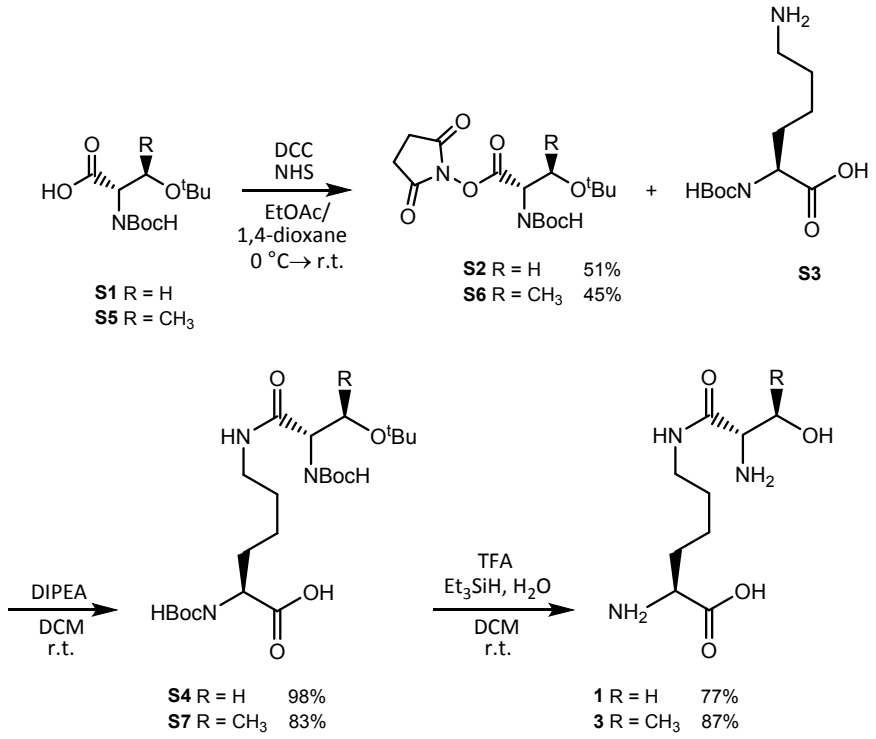


Figure S4: Trypsin digest of **6**, displaying found peptides in blue to map the sequence of the protein. The NCAA ThrK is confirmed in the correct location, highlighted in brown.

3. Synthesis

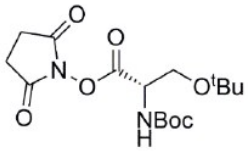
3.1 Solution phase synthesis

3.1.1 L-configured dipeptides



Scheme S1: Synthesis of the L-configured dipeptides

***N*-hydroxysuccinimidyl (2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoate S2**

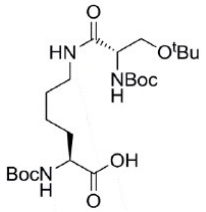


N-hydroxysuccinimide (0.73 g, 6.3 mmol, 1.05 eq.) was added to a solution of (2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoic acid **S1** (1.57 g, 6.0 mmol, 1.0 eq.) in 1:1 (v/v) EtOAc:1,4-dioxane (18 mL) at 0 °C. Dicyclohexylcarbodiimide (1.30 g, 6.3 mmol, 1.05 eq.) was added in one portion and the reaction left to warm to r.t. and stirred for 3 h. The reaction mixture was filtered through a Celite pad and concentrated. The crude product was redissolved in EtOAc (20 mL) and washed sequentially with 5% (w/w) NaHCO3 (10 mL), H2O (10 mL) and brine (10 mL). The organic layer was dried over MgSO4, filtered and concentrated to afford crude *N*-hydroxysuccinimidyl (2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoate **S2** as a white powder of sufficient purity for further manipulation (1.10 g, 51%).

δ_H (500 MHz, DMSO-d6): 7.32 (major rotamer) and 6.92 (minor) (d, 1H, 3*J* 8.3 Hz, NH), 4.45-4.52 (major) and 4.32-4.37 (minor) (m, 1H, Hα), 3.62-3.68 (m, 2H, Hβ), 2.80 (s, 4H, Su CH2), 1.39 (major)

and 1.37 (minor) (s, 9H, Boc CH3), 1.14 (major) and 1.08 (minor) (s, 9H, ether CH3); δC (126 MHz, DMSO-d6): 170.2 (minor) and 169.8 (major) (Su CO, mixture of rotamers), 167.0 (major) and 166.3 (minor) (CO2Su), 155.1 (major) and 154.7 (minor) (CO2NH), 79.4 (minor) and 78.8 (major) (Boc C(CH3)3), 73.2 (ether C(CH3)3), 61.0 (major) and 60.7 (minor) (Cβ) 54.3 (minor) and 53.0 (major) (Cα), 28.1 (major) and 27.6 (minor) (Boc CH3), 27.1 (minor) and 135, 27.0 (major) (ether CH3), 25.4 (Su CH2); IR (ATR): 3410, 2982, 2935, 1820, 1783, 1735, 1714, 1475, 1444, 1427, 1364, 1196, 1170, 1102, 1083, 1064, 1046, 1036, 995, 912, 867, 835, 810, 774, 749, 647, 576, 553; HRMS: found [M+Na]+ 381.1633; $C_{16}H_{26}N_2O_7Na$ requires 381.1632 (Δ = -1.0 ppm). Characterisation data in agreement with previous literature reports[3].

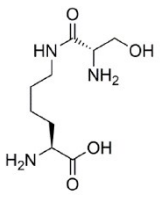
(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid S4



DIPEA (1.75 mL, 10 mmol, 4.0 eq.) and a solution of *N*-hydroxysuccinimidyl (2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoate **S2** (0.90 g, 2.5 mmol, 1.0 eq.) in DCM (8 mL) were added sequentially at r.t. to a stirred suspension of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-aminohexanoic acid **S3** (0.62 g, 2.5 mmol, 1.0 eq.) in DCM (12 mL). The reaction mixture was stirred at r.t. for 3 h, filtered and concentrated. The residue was redissolved in DCM (30 mL) and washed with sat. citric acid (10 mL) and brine (10 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford crude (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid **S4** as an off-white powder of sufficient purity for further manipulation (1.20 g, 98%).

δ_H (500 MHz, DMSO-d_6) 7.80 (t, 1H, ^3J 5.0 Hz, HN_ε), 6.98 (d, 1H, ^3J 8.0 Hz, HN_α), 6.85 (d, 1H, ^3J 8.3 Hz, $HN_{\alpha'}$), 3.95 (dt, 1H, ^3J 8.4 Hz, $^3J_{\alpha'\text{-}\beta'}$ 5.6 Hz, $H_{\alpha'}$), 3.80 (dt, 1H, ^3J 7.9 Hz, $^3J_{\alpha\text{-}\beta}$ 7.0 Hz, H_α), 3.38-3.43 (m, 2H, $H_{\beta'}$), 2.94-3.12 (m, 2H, H_ε), 1.20-1.65 (m, 24H, Boc CH_3, Boc CH_3, H_β, H_γ, H_δ), 1.09 (s, 9H, ether CH_3); δ_C (126 MHz, DMSO-d_6) 174.3 (CO_2H), 169.9 (CONH), 155.6, 155.0 (Boc CO_2N), 78.1, 77.9 (Boc $C(CH_3)_3$), 72.6 (ether $C(CH_3)_3$), 62.0 ($C_{\beta'}$), 54.9 ($C_{\alpha'}$), 53.4 (C_α), 38.2 (C_ε), 30.3 (C_β), 28.6 (C_γ), 28.2, 28.1 (Boc CH_3), 27.2 (ether CH_3), 22.9 (C_δ); IR (ATR): 3321, 2976, 2933, 1705, 1661, 1501, 1392, 1365, 1247, 1161, 1088, 1019, 860, 779, 735; HRMS: found $[M+Na]_+$ 512.2959; $C_{23}H_{43}N_3O_8Na$ requires 512.2942 (Δ = -3.5 ppm).

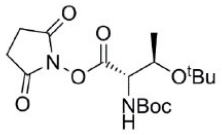
(2*S*)-2-amino-6-((2*S*)-2-amino-3-hydroxypropanamido)hexanoic acid 1



Deprotection cocktail 95:2.5:2.5 (v/v) TFA:TIS:H_2O (4 mL) was added dropwise to a stirred solution of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid **S4** (1.01 g, 2.06 mmol) in DCM (8 mL) at 0 °C and left to reach rt. Once the reaction was determined complete by TLC, usually within 3 h, the reaction mixture was triturated with ice-cold Et_2O (30 mL. The suspension was centrifuged at 4 000 × *g* for 5 min, supernatant decanted, and the peptide pellet resuspended in ice-cold Et_2O (20 mL), with this process repeated 5 times in total, after which the pellet was left to air dry. The resulting off-white powder was redissolved in 10% (v/v) aqueous acetic acid and lyophilised to afford (2*S*)-2-amino-6-((2*S*)-2-amino-3-hydroxypropanamido)-hexanoic acid **1** as an off-white foam (560 mg, diacetate salt, 77%); $[\alpha]_D$ +21.1 (c 1.0, H_2O).

δ_H (500 MHz, DMSO-d_6): 8.41 (t, 1H, ^3J 5.5 Hz, HN_ε), 3.55-3.63 (m, 4H, H_α, $H_{\alpha'}$, $H_{\beta'}$), 3.08-3.12 (m, 2H, H_ε), 1.69-1.79 (m, 2H, H_β), 1.28-1.47 (m, 4H, H_γ, H_δ); δ_C (126 MHz, DMSO-d_6): 171.1 (CO_2H), 166.6 (CONH), 60.3 ($C_{\beta'}$), 54.4 ($C_{\alpha'}$), 52.1 (C_α), 38.5 (C_ε), 29.7 (C_β), 28.3 (C_γ), 21.7 (C_δ); IR (ATR): 3316, 2944, 1661, 1572, 1521, 1432, 1273, 1251, 1184, 1127, 1037, 836, 799, 721, 600, 542; HRMS: found $[M+H]^+$ 234.1451; $C_9H_{20}N_3O_4$ requires 234.1448 (Δ = -0.6 ppm).

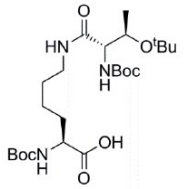
_N_-hydroxysuccinimidyl (2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonyl-amino)-butanoate S6



N-hydroxysuccinimide (0.91 g, 7.9 mmol, 1.05 eq.) was added to a solution of (2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanoic acid **S5** (2.06 g, 7.5 mmol, 1.0 eq.) in 1:1 (v/v) EtOAc:1,4-dioxane (22 mL) at 0 °C. Dicyclohexylcarbodiimide (1.62 g, 7.9 mmol, 1.05 eq.) was added in one portion and the reaction left to warm to r.t. and stirred for 3 h. The reaction mixture was filtered through a Celite pad and concentrated. The crude product was redissolved in EtOAc (20 mL) and washed sequentially with 5% (w/w) $NaHCO_3$ (10 mL), H_2O (10 mL) and brine (10 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford crude *N*-hydroxysuccinimidyl (2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanoate **S6** as a white powder of sufficient purity for further manipulation (1.25 g, 45%).

δ_H (500 MHz, DMSO-d_6): 6.97 (d, 1H, ^3J 9.3 Hz, $N_\alpha H$), 4.42 (dd, 1H, ^3J 9.3 Hz, $^3J_{\alpha\text{-}\beta}$ 4.1 Hz, H_α), 4.02-4.07 (m, 1H, H_β), 2.80 (s, 4H, Su H), 1.40 (s, 9H, Boc CH_3), 1.19 (d, 3H, $^3J_{\beta\text{-}\gamma}$ 6.2 Hz, H_γ), 1.14 (s, 9H, ether CH_3); δ_C (126 MHz, DMSO-d_6): 169.9 (Su CO), 166.8 (CO_2H), 155.2 (Boc CO_2NH), 78.9 (Boc $C(CH_3)_3$), 73.9 (ether $C(CH_3)_3$), 67.2 (C_β), 57.9 (C_α), 28.2 (Boc CH_3), 28.1 (ether CH_3), 25.4 (Su CH_2), 19.6 (C_γ); IR (ATR): 3351, 2978, 2935, 2119, 1819, 1781, 1732, 1711, 1512, 1456, 1368, 1241, 1205, 1151, 1120, 1090, 1059, 957, 914, 868, 825, 792, 644, 605, 546; HRMS: found $[M+Na]^+$ 395.1793; $C_{17}H_{28}N_2O_7Na$ requires 395.1789 (Δ = -1.0 ppm).

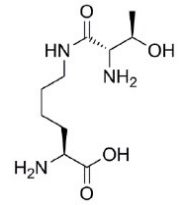
(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)-hexanoic acid S7



DIPEA (2.1 mL, 12 mmol, 4.0 eq.) and a solution of *N*-hydroxysuccinimidyl (2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanoate **S6** (1.12 g, 3.0 mmol, 1.0 eq.) in DCM (9 mL) were added sequentially at r.t. to a stirred suspension of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-aminohexanoic acid **S3** (738 mg g, 3.0 mmol, 1.0 eq.) in DCM (15 mL). The reaction mixture was stirred at r.t. for 3 h, filtered and concentrated. The residue was redissolved in DCM (30 mL) and washed with sat. citric acid (10 mL) and brine (10 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford crude (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)-hexanoic acid **S7** as a white fluffy powder of sufficient purity for further manipulation (1.25 g, 83%).

δ_H (500 MHz, DMSO-d_6) 7.71 (t, 1H, ^3J 5.6 Hz, HN_ε), 6.95 (d, 1H, ^3J 8.0 Hz, HN_α), 5.95 (t, 1H, ^3J 9.1 Hz, $HN_{\alpha'}$), 3.79-3.90 (m, 3H, H_α, $H_{\alpha'}$, $H_{\beta'}$), 2.93-3.12 (m, 2H, H_ε), 1.23-1.65 (m, 24H, Boc CH_3, Boc CH_3, H_β, H_γ, H_δ), 1.08 (s, 9H, ether CH_3), 1.00 (d, 3H, $^3J_{\beta'\text{-}\gamma'}$ 6.2 Hz, $H_{\gamma'}$); δ_C (126 MHz, DMSO-d_6) 174.2 (CO_2H), 169.8 (CONH), 155.5, 155.1 (Boc CO_2N), 78.3, 77.8 (Boc $C(CH_3)_3$), 73.3 (ether $C(CH_3)_3$), 67.4 ($C_{\beta'}$), 59.4 ($C_{\alpha'}$), 53.4 (C_α), 38.4 (C_ε), 30.4 (C_β), 28.5 (C_γ), 28.2, 28.1 (Boc CH_3), 28.0 (ether CH_3), 23.1 (C_δ), 19.9 ($C_{\gamma'}$); IR (ATR): 3436, 2980, 1722, 1706, 1493, 1453, 1424, 1391, 1364, 1312, 1248, 1216, 1160, 1091, 1068, 1038, 946, 864, 836, 793, 775, 754, 683; HRMS: found $[M+Na]^+$ 526.3089; $C_{24}H_{45}N_3O_8Na$ requires 526.3099 (Δ = 3.2 ppm).

(2*S*)-2-amino-6-((2*S*,3*R*)-2-amino-3-hydroxybutanamido)-hexanoic acid 3

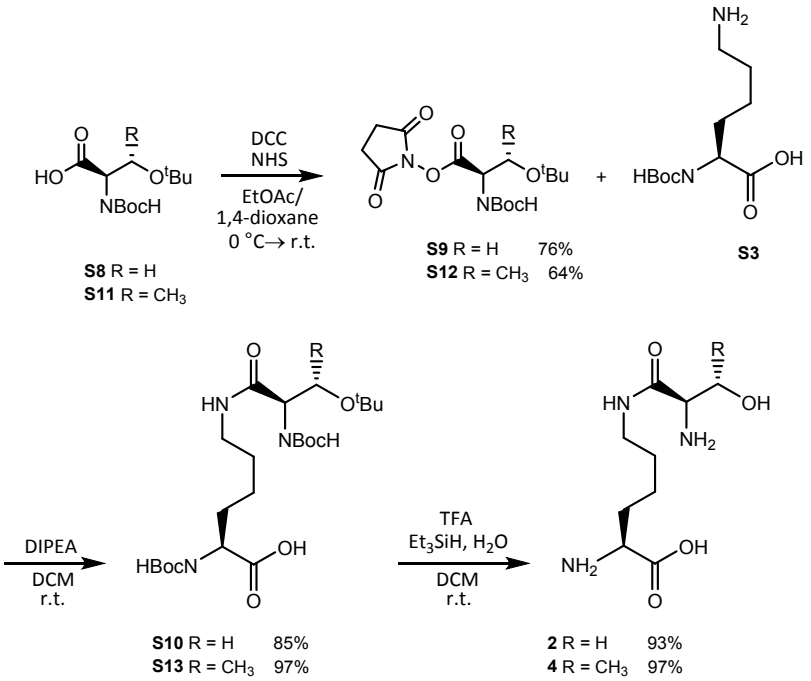


Deprotection cocktail 95:2.5:2.5 (v/v) TFA:TIS:H_2O (4 mL) was added dropwise to a stirred solution of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)-hexanoic acid **S7** (1.11 g, 2.2 mmol) in DCM (6 mL) at 0 °C and left to reach rt. Once the reaction was determined complete by TLC, *ca.* 3 h, the reaction mixture was split into two portions and each portion triturated with ice-cold Et_2O (30 mL). The suspension was centrifuged at 4 000 × *g* for 5 min, supernatant decanted, and the peptide pellet resuspended in ice-cold Et_2O (20 mL), with this process repeated 5 times in total, after which the pellet was left to air dry. The resulting off-white solid was redissolved in 10% (v/v) aqueous acetic acid and lyophilised to afford (2*S*)-2-amino-6-((2*S*,3*R*)-2-amino-3-hydroxybutanamido)-hexanoic acid **3** as a fluffy white foam (706 mg, diacetate salt, 87%).

$[\alpha]_D$ +9.5 (c 1.1, H_2O); δ_H (500 MHz, DMSO-d_6): 8.53 (t, 1H, ^3J 5.6 Hz, HN_ε), 3.85 (app. quint., 1H, $^3J_{\alpha'\text{-}\beta'}$, $_{\beta'\text{-}\gamma'}$ 6.5 Hz, $H_{\beta'}$), 3.79 (t, 1H, $^3J_{\alpha\text{-}\beta}$ 6.3 Hz, H_α), 3.48 (d, 1H, $^3J_{\alpha'\text{-}\beta'}$ 6.8 Hz, $H_{\alpha'}$), 3.14 (app. dq, 1H, ^2J 12.7 Hz,

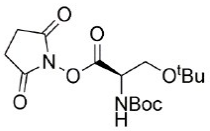
^3J 6.6 Hz, H_ε), 3.05 (app. dq, 1H, ^2J 12.9 Hz, ^3J 6.5 Hz, H'_ε), 1.70-1.82 (m, 2H, H_β), 1.24-1.47 (m, 4H, H_γ, H_δ) 1.12 (d, 3H, $^3J_{\beta'\text{-}\gamma'}$ 6.4 Hz, $H_{\gamma'}$); δ_C (126 MHz, DMSO-d_6): 171.1 (CO_2H), 166.8 (CONH), 65.8 ($C_{\beta'}$), 58.5 ($C_{\alpha'}$), 52.1 (C_α), 38.5 (C_ε), 29.7 (C_β), 28.3 (C_γ), 21.8 (C_δ), 20.0 ($C_{\gamma'}$); IR (ATR): 2944, 1660, 1512, 1432, 1250, 1182, 1131, 918, 838, 798, 721, 599, 517; HRMS: found $[M+H]_+$ 248.1604; $C_{10}H_{22}N_3O_4$ requires 248.1605 (Δ = 1.0 ppm).

3.1.2 S-configured dipeptides



Scheme S2: Synthesis of the S-configured dipeptides

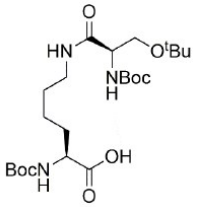
N-hydroxysuccinimidyl (2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoate S9



N-hydroxysuccinimide (0.44 g, 3.9 mmol, 1.1 eq.) was added to a solution of (2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoic acid **S8** (0.91 g, 3.5 mmol, 1.0 eq.) in 1:1 (v/v) EtOAc:1,4-dioxane (10 mL) at 0 °C. Dicyclohexylcarbodiimide (0.79 g, 3.9 mmol, 1.1 eq.) was added in one portion and the reaction left to warm to r.t. and stirred for 3 h. The reaction mixture was filtered through a Celite pad and concentrated. The crude product was redissolved in EtOAc (20 mL) and washed sequentially with 5% (w/w) $NaHCO_3$ (10 mL), H_2O (10 mL) and brine (10 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford crude *N*-hydroxysuccinimidyl (2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoate **S9** as a white powder of sufficient purity for further manipulation (0.96 g, 76%).

δ_H (500 MHz, DMSO-d_6): 7.32 (major rotamer) and 6.92 (minor) (d, 1H, ^3J 8.3 Hz, NH), 4.48-4.52 (major) and 4.33-4.37 (minor) (m, 1H, H_α), 3.62-3.68 (m, 2H, H_β), 2.80 (s, 4H, Su CH_2), 1.40 (major) and 1.37 (minor) (s, 9H, Boc CH_3), 1.14 (major) and 1.08 (minor) (s, 9H, ether CH_3); δ_C (126 MHz, DMSO-d_6): 169.8 (Su CO), 167.0 (major) and 166.3 (minor) (CO_2Su, mixture of rotamers), 155.1 (major) and 154.7 (minor) (CO_2NH), 79.4 (minor) and 78.8 (major) (Boc $C(CH_3)_3$), 73.2 (ether $C(CH_3)_3$), 61.0 (major) and 60.7 (minor) (C_β) 54.3 (minor) and 53.0 (major) (C_α), 28.1 (major) and 27.6 (minor) (Boc CH_3), 27.0 (ether CH_3), 25.4 (Su CH_2); IR (ATR): 3410, 2980, 2934, 1819, 1783, 1735, 1713, 1475, 1445, 1426, 1365, 1195, 1165, 1102, 1082, 1065, 1046, 1035, 995, 912, 866, 810, 775, 749, 647, 577, 553; HRMS: found $[M+Na]^+$ 381.1638; $C_{16}H_{26}N_2O_7Na$ requires 381.1632 (Δ = -1.3 ppm).

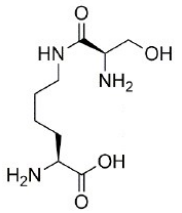
(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid S10



DIPEA (1.75 mL, 10 mmol, 4.0 eq.) and a solution of *N*-hydroxysuccinimidyl (2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoate **S9** (0.90 g, 2.5 mmol, 1.0 eq.) in DCM (8 mL) were added sequentially at r.t. to a stirred suspension of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-aminohexanoic acid **S3** (0.62 g, 2.5 mmol, 1.0 eq.) in DCM (12 mL). The reaction mixture was stirred at r.t. for 3 h, filtered and concentrated. The residue was redissolved in DCM (30 mL) and washed with sat. citric acid (10 mL) and brine (10 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford crude (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid **S10** as a white foam of sufficient purity for further manipulation (1.04 g, 85%).

δ_H (500 MHz, DMSO-d_6) 7.77 (t, 1H, ^3J 5.7 Hz, HN_ε), 6.97 (d, 1H, ^3J 8.0 Hz, HN_α), 6.42 (d, 1H, ^3J 8.2 Hz, $HN_{\alpha'}$), 3.78-3.97 (m, 2H, $H_{\alpha'}$, H_α), 3.38-3.44 (m, 2H, $H_{\beta'}$), 2.96-3.09 (m, 2H, H_ε), 1.20-1.65 (m, 24H, Boc CH_3, Boc CH_3, H_β, H_γ, H_δ), 1.09 (s, 9H, ether CH_3); δ_C (126 MHz, DMSO-d_6) 174.3 (CO_2H), 169.8 (CONH), 155.6, 155.0 (Boc CO_2N), 78.1, 77.9 (Boc $C(CH_3)_3$), 72.6 (ether $C(CH_3)_3$), 62.0 ($C_{\beta'}$), 54.9 ($C_{\alpha'}$), 53.4 (C_α), 38.2 (C_ε), 30.3 (C_β), 28.6 (C_γ), 28.2, 28.1 (Boc CH_3), 27.2 (ether CH_3), 22.9 (C_δ); IR (ATR): 3331, 2976, 1705, 1501, 1392, 1365, 1247, 1161, 1088, 1019, 860, 779, 735; HRMS: found $[M+Na]^+$ 512.2941; $C_{23}H_{43}N_3O_8Na$ requires 512.2942 (Δ = 0.3 ppm).

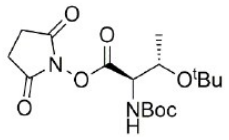
(2*S*)-2-amino-6-((2*R*)-2-amino-3-hydroxypropanamido)-hexanoic acid 2



Deprotection cocktail 95:2.5:2.5 (v/v) TFA:TIS:H_2O (4 mL) was added dropwise to a stirred solution of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid **S10** (1.01 g, 2.06 mmol) in DCM (6 mL) at 0 °C and left to reach rt. Once the reaction was determined complete by TLC, usually within 3 h, the reaction mixture was triturated with ice-cold Et_2O (30 mL). The suspension was centrifuged at 4 000 × *g* for 5 min, supernatant decanted, and the peptide pellet resuspended in ice-cold Et_2O (20 mL), with this process repeated 5 times in total, after which the pellet was left to air dry. The resulting off-white powder was redissolved in 10% (v/v) aqueous acetic acid and lyophilised to afford (2*S*)-2-amino-6-((2*R*)-2-amino-3-hydroxypropanamido)-hexanoic acid **2** as a fluffy off-white foam (657 mg, diacetate salt, 93%).

$[\alpha]_D$ +31.8 (c 1.0, H_2O); δ_H (500 MHz, DMSO-d_6): 8.45 (t, 1H, ^3J 5.6 Hz, HN_ε), 3.66-3.82 (m, 4H, H_α, $H_{\alpha'}$, $H_{\beta'}$), 3.07-3.11 (m, 2H, H_ε), 1.71-1.82 (m, 2H, H_β), 1.28-1.46 (m, 4H, H_γ, H_δ); δ_C (126 MHz, DMSO-d_6): 171.2 (CO_2H), 166.8 (CONH), 60.5 ($C_{\beta'}$), 54.6 ($C_{\alpha'}$), 52.2 (C_α), 38.6 (C_ε), 29.8, (C_β), 28.4 (C_γ), 21.8 (C_δ); IR (ATR): 3317, 2946, 1661, 1581, 1519, 1428, 1278, 1255, 1184, 1129, 1039, 840, 802, 720, 600, 518; HRMS: found $[M+H]^+$ 234.1448; $C_9H_{20}N_3O_4$ requires 234.1448 (Δ = 0.3 ppm).

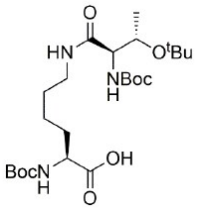
_N_-hydroxysuccinimidyl (2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonyl-amino)-butanoate S12



N-hydroxysuccinimide (257 mg, 2.2 mmol, 1.1 eq.) was added to a solution of (2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanoic acid **S11** (561 mg, 2.0 mmol, 1.0 eq.) in 1:1 (v/v) EtOAc:1,4-dioxane (6 mL) at 0 °C. Dicyclohexylcarbodiimide (460 mg, 2.2 mmol, 1.1 eq.) was added in one portion and the reaction left to warm to r.t. and stirred for 3 h. The reaction mixture was filtered through a Celite pad and concentrated. The crude product was redissolved in EtOAc (10 mL) and washed sequentially with 5% (w/w) $NaHCO_3$ (10 mL), H_2O (5 mL) and brine (5 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford crude *N*-hydroxysuccinimidyl (2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanoate **S12** as a white foam of sufficient purity for further manipulation (483 mg, 64%).

δ_H (500 MHz, DMSO-d_6): 6.97 (d, 1H, ^3J 9.3 Hz, $N_\alpha H$), 4.42 (dd, 1H, ^3J 9.3 Hz, $^3J_{\alpha\text{-}\beta}$ 4.1 Hz, H_α), 4.02-4.07 (m, 1H, H_β), 2.80 (s, 4H, Su H), 1.40 (s, 9H, Boc CH_3), 1.19 (d, 3H, $^3J_{\beta\text{-}\gamma}$ 6.2 Hz, H_γ), 1.14 (s, 9H, ether CH_3); δ_C (126 MHz, DMSO-d_6): 169.9 (Su CO), 166.8 (CO_2H), 155.2 (Boc CO_2NH), 78.9 (Boc $C(CH_3)_3$), 73.9 (ether $C(CH_3)_3$), 67.2 (C_β), 57.9 (C_α), 28.2 (Boc CH_3), 28.1 (ether CH_3), 25.4 (Su CH_2), 19.6 (C_γ); IR (ATR): 3350, 2978, 2933, 1819, 1781, 1731, 1710, 1512, 1456, 1367, 1242, 1205, 1152, 1120, 1090, 1059, 957, 914, 868, 825, 791, 644, 605, 546; HRMS: found $[M+Na]^+$ 395.1792; $C_{17}H_{28}N_2O_7Na$ requires 395.1789 (Δ = -0.5 ppm).

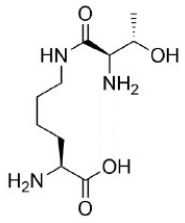
(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)-hexanoic acid S13



DIPEA (0.6 mL, 5.2 mmol, 4.0 eq.) and a solution of *N*-hydroxysuccinimidyl (2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanoate **S12** (483 mg, 1.3 mmol, 1.0 eq.) in DCM (6 mL) were added sequentially at r.t. to a stirred suspension of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-aminohexanoic acid **S3** (320 mg g, 1.3 mmol, 1.0 eq.) in DCM (9 mL). The reaction mixture was stirred at r.t. for 3 h, filtered and concentrated. The residue was redissolved in DCM (15 mL) and washed with sat. citric acid (5 mL) and brine (5 mL). The organic layer was dried over $MgSO_4$, filtered and concentrated to afford crude (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)hexanoic acid **S13** as a white foam of sufficient purity for further manipulation (635 mg, 97%).

δ_H (500 MHz, DMSO-d_6) 7.71 (t, 1H, ^3J 5.6 Hz, HN_ε), 6.98 (d, 1H, ^3J 8.0 Hz, HN_α), 5.95 (t, 1H, ^3J 9.2 Hz, $HN_{\alpha'}$), 3.70-3.90 (m, 3H, H_α, $H_{\alpha'}$, $H_{\beta'}$), 2.97-3.08 (m, 2H, H_ε), 1.21-1.73 (m, 24H, Boc CH_3, Boc CH_3, H_β, H_γ, H_δ), 1.08 (s, 9H, ether CH_3), 1.00 (d, 3H, $^3J_{\beta'\text{-}\gamma'}$ 6.2 Hz, $H_{\gamma'}$); δ_C (126 MHz, DMSO-d_6) 174.2 (CO_2H), 169.9 (CONH), 155.6, 155.1 (Boc CO_2N), 78.3, 77.9 (Boc $C(CH_3)_3$), 73.3 (ether $C(CH_3)_3$), 67.3 ($C_{\beta'}$), 59.4 ($C_{\alpha'}$), 53.4 (C_α), 38.5 (C_ε), 30.4 (C_β), 28.5 (C_γ), 28.2, 28.1 (Boc CH_3), 28.0 (ether CH_3), 23.1 (C_δ), 20.0 ($C_{\gamma'}$); IR (ATR): 3320, 2979, 1708, 1663, 1496, 1392, 1366, 1249, 1161, 1065, 1022, 959, 857, 779; HRMS: found $[M+Na]_+$ 526.3108; $C_{24}H_{45}N_3O_8Na$ requires 526.3099 (Δ = -0.3 ppm).

(2*S*)-2-amino-6-((2*S*,3*R*)-2-amino-3-hydroxybutanamido)-hexanoic acid 4



Deprotection cocktail 95:2.5:2.5 (v/v) TFA:TIS:H_2O (2 mL) was added dropwise to a stirred solution of (2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)-hexanoic acid **S13** (588 mg, 1.2 mmol) in DCM (3 mL) at 0 °C and left to reach rt. Once the reaction was determined complete by TLC, *ca.* 3 h, the reaction mixture was triturated with ice-cold Et_2O (30 mL). The suspension was centrifuged at 4 000 × *g* for 5 min, supernatant decanted, and the peptide pellet resuspended in ice-cold Et_2O (20 mL), with this process repeated 5 times in total, after which the pellet was left to air dry. The resulting off-white solid was redissolved in 10% (v/v) aqueous acetic acid and lyophilised to afford (2*S*)-2-amino-6-((2*R*,3*S*)-2-amino-3-hydroxybutanamido)hexanoic acid **4** as a fluffy off-white foam (539 mg, di(trifluoroacetate) salt, 97%).

$[\alpha]_D$ -52.2 (c 1.0, H_2O); δ_H (500 MHz, DMSO-d_6): 8.50 (t, 1H, ^3J 5.6 Hz, HN_ε), 3.85 (app. quint., 1H, $^3J_{\alpha'\text{-}\beta', \beta'\text{-}\gamma'}$ 6.4 Hz, $H_{\beta'}$), 3.77 (t, 1H, $^3J_{\alpha\text{-}\beta}$ 6.3 Hz, H_α), 3.46 (d, 1H, $^3J_{\alpha'\text{-}\beta'}$ 6.8 Hz, $H_{\alpha'}$), 3.15 (app. dq, 1H, ^2J 12.6 Hz, ^3J 6.5 Hz, H_ε), 3.05 (app. dq, 1H, ^2J 13.0 Hz, ^3J 6.5 Hz, H'_ε), 1.70-1.82 (m, 2H, H_β), 1.24-1.47 (m, 4H, H_γ, H_δ) 1.12 (d, 3H, $^3J_{\beta'\text{-}\gamma'}$ 6.3 Hz, $H_{\gamma'}$); δ_C (126 MHz, DMSO-d_6): 171.0 (CO_2H), 166.7 (CONH), 65.8 ($C_{\beta'}$), 58.5 ($C_{\alpha'}$), 52.1 (C_α), 38.4 (C_ε), 29.7 (C_β), 28.2 (C_γ), 21.8 (C_δ), 20.0 ($C_{\gamma'}$); IR (ATR): 2980, 1662, 1581, 1519, 1429, 1250, 1184, 1130, 1038, 932, 838, 799, 721, 599, 517; HRMS: found $[M+H]^+$ 248.1606; $C_{10}H_{22}N_3O_4$ requires 248.1605 (Δ = 0.0 ppm).

4. Solid phase synthesis

4.1 General methods

Preloaded resin preparation

The preloaded 2-chlorotrityl resin was weighed out into a 2 mL SPPS cartridge fitted with a PTFE stopcock, swollen in DCM for 30 min and then filtered.

Fmoc deprotection

A solution of 20% piperidine in DMF was added to the resin and gently agitated by rotation for 2 minutes. The resin was filtered off and repeated four more times, followed by washes with DMF (5 × 2 min with rotation).

Amino acid coupling

DIPEA (11 eq.) was added to a solution of amino acid (5 eq.) and HCTU (5 eq.) dissolved in the minimum volume of DMF and the solution added to the resin. The reaction mixture was gently agitated by rotation for 1 h, and the resin filtered off and washed with DMF (3 × 2 min with rotation).

Cleavage Cocktails

Deprotection and resin cleavage: 95:2.5:2.5 (v/v) TFA:H_2O:triisopropylsilane. Cleavage only: 4:1 (v/v) DCM: 1,1,1,3,3,3-hexafluoroisopropanol.

Cleavage and Isolation

The resin was washed with DCM (3 × 2 min with rotation) and MeOH (3 × 2 min with rotation). The resin was dried on a vacuum manifold and further dried on a high vacuum line overnight. A solution of cleavage cocktail was added to the resin and gently agitated by rotation for 60 min.

For peptides cleaved and deprotected: the reaction mixture was drained into ice-cold Et_2O and centrifuged at 5000 × *g* at 4 °C until pelleted (5-10 min). The supernatant was carefully decanted and subsequently resuspended, centrifuged and supernatant decanted three more times. The peptide pellet was dissolved in 10% (v/v) aq. AcOH and lyophilised.

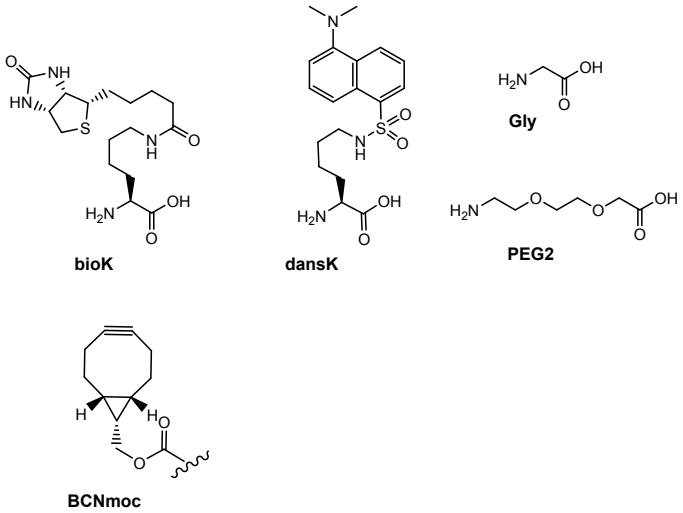
For proteins undergoing further couplings: the reaction mixture was drained and concentrated *in vacuo*.

Nomenclature

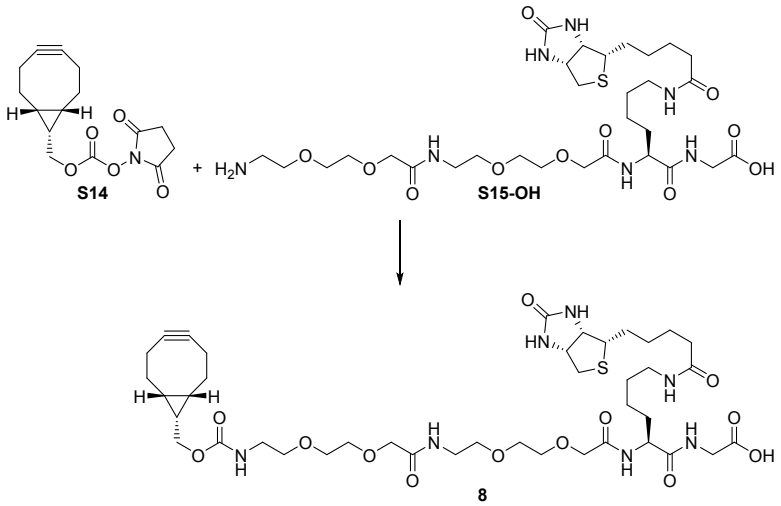
The names and structures of all standard and non-standard residues and *N*-terminal capping moieties featuring in the peptides synthesised are listed below.

Residues: bioK, N_ε-(+)-biotinyl L-lysine; dansK, N_ε-dansyl L-lysine; G, glycine; PEG2, 8-amino-2,6-dioxaoctanoic acid.

Caps: BCNmoc, *endo*-bicyclo[6.1.0]non-4-yn-9-ylmethoxycarbonyl.

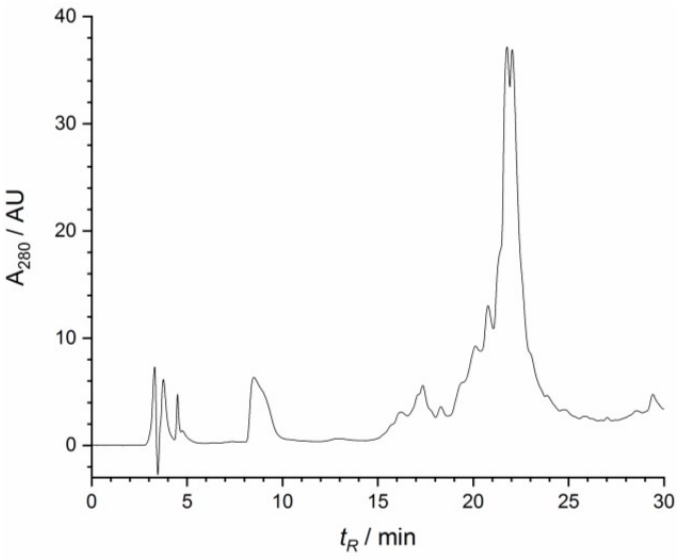


4.2 BCNmoc-PEG2-PEG2-bioK-G-OH 8

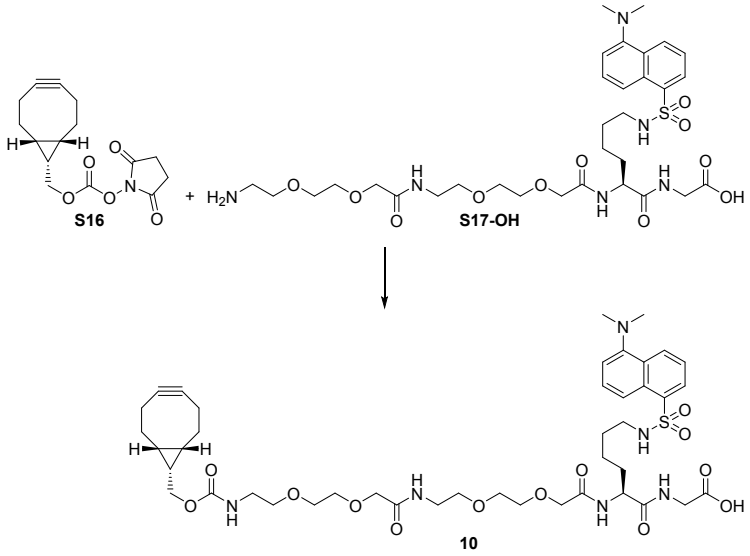


Precursor peptide **S15-OH** was synthesised according to the general SPPS protocol using preloaded H-Gly-2ClTrt resin (75 mg, loading 0.52 mmol g^{-1}, 0.039 mmol), Fmoc-bioK-OH and Fmoc-PEG2-OH and isolated following resin cleavage (no deprotection), purification and concentration. The residue was dissolved in DMF (1 mL) and BCNmoc-OSu **S14** (13 mg, 0.044 mmol, 1.1 eq.) and DIPEA (9 µL, 0.049 mmol, 1.2 eq.) added. The reaction was left stirring at r.t. for 16 h, then concentrated, redissolved in MeOH and purified by size-exclusion chromatography (Sephadex LH-20, equilibrated in MeOH, eluted with isocratic MeOH) to afford **8** as a fluffy white powder (24 mg, 68%).

HRMS: found $[M+Na]^+$ 918.4283; $C_{41}H_{65}N_7O_{13}S$ requires 918.4253 (Δ = -2.5 ppm); HPLC: t_R = 21.91 min.

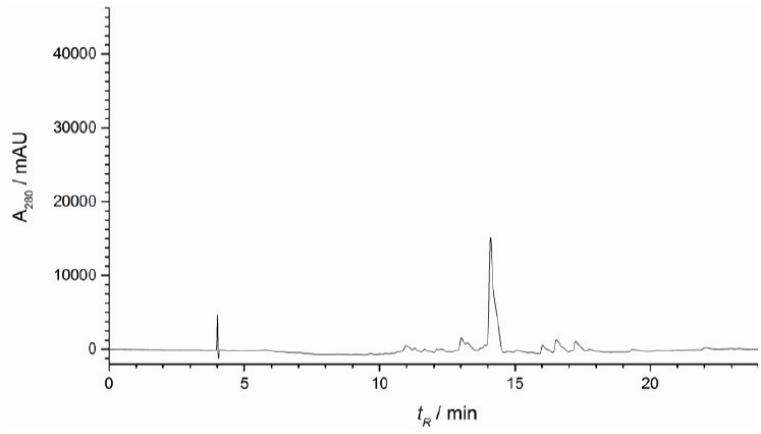


4.3 BCNmoc-PEG2-PEG2-dansK-G-OH 10



Precursor peptide **S17-OH** was synthesised according to the general SPPS protocol using preloaded H-Gly-2ClTrt resin (127 mg, loading 0.52 mmol g^{-1}, 0.066 mmol), Fmoc-bioK-OH and Fmoc-PEG2-OH and isolated following resin cleavage (no deprotection), purification and concentration. The residue was dissolved in DMF (2 mL) and BCNmoc-OSu **S16** (21 mg, 0.072 mmol, 1.1 eq.) and DIPEA (14 µL, 0.079 mmol, 1.2 eq.) added. The reaction was left stirring at r.t. for 16 h, then concentrated, redissolved in MeOH and purified by size-exclusion chromatography (Sephadex LH-20, equilibrated in MeOH, eluted with isocratic MeOH) to afford **10** as a fluffy yellow powder (33 mg, 55%).

HRMS: found $[M+Na]^+$ 925.4010; $C_{43}H_{62}N_6O_{13}SNa$ requires 925.3988 (Δ = -1.1 ppm); HPLC: t_R = 14.12 min.


40000
30000
20000
10000
0
A_{280} / mAU
0
5
10
15
20
t_R / min

5. Protein modification

5.1 Oxidation screening conditions

For a 100 µL scale reaction, solutions of L-methionine and $NaIO_4$ (stock solutions made in PBS, pH 7.4 or 20 mM PB, 150 mM NaCl, pH 7.4) were added to protein 1,2-aminoalcohol **6** (stock 100 µM in PBS, pH 7.4 or 20 mM PB, 150 mM NaCl, pH 7.4) in varying molar equivalents (**Table S1**). After mixing by gentle pipetting, the reactions were incubated at 0 °C in the dark for 4 - 10 min, after which the oxidised protein was purified by desalting using a PD Minitrap G-25 (GE Healthcare), gravity method, into H_2O for analysis.

Table S1: Screening conditions for the oxidation of **6**

Buffer system	$NaIO_4$ (eq.)	L-Met (eq.)	t/min
PBS	3	6	4
PBS	3	6	10
PBS	5	10	4
PBS	5	10	8
PBS	10	20	4
20 mM PB + 150 mM NaCl pH 7.4	5	10	4

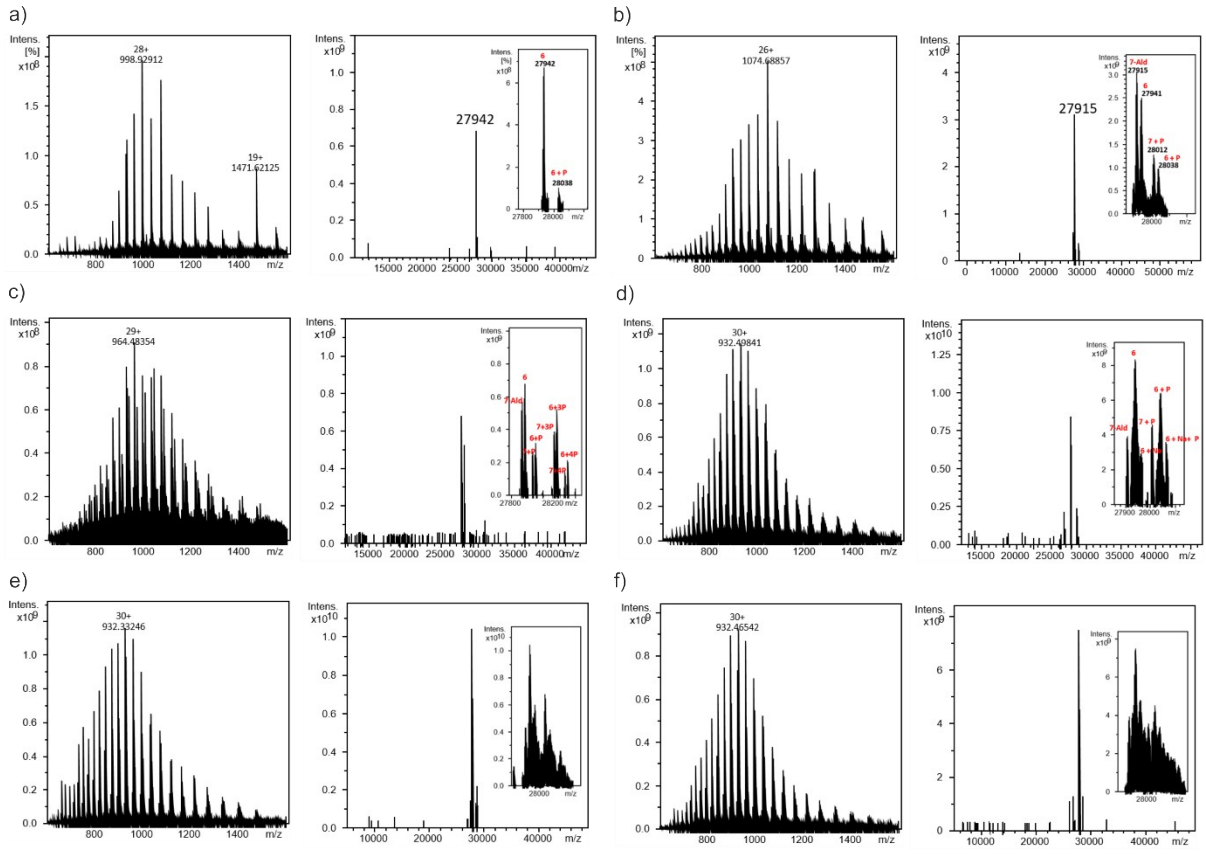
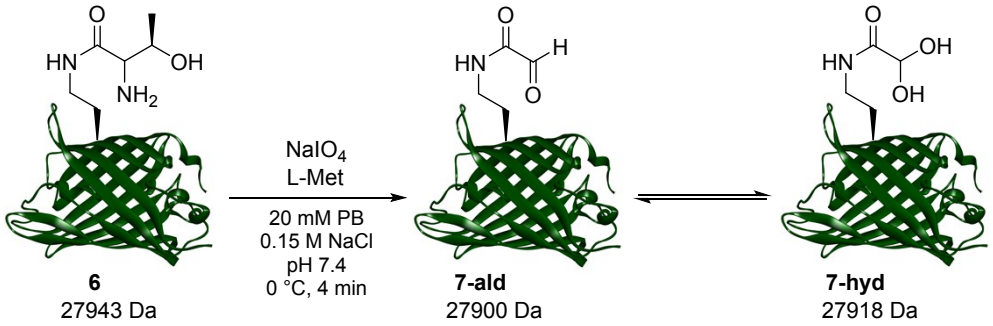


Figure S5: Raw ESI-FTICR-MS (left) and deconvoluted spectra (right) collected in the screening experiments to determine optimal conditions for the oxidation of **6** (27943 Da) to **7** (Ald – 27900 Da, Hyd – 27918 Da). a) non-oxidised **6** (control); b) PBS, 3 eq. $NaIO_4$, 6 eq. L-Met 4 min; c) PBS, 3 eq. $NaIO_4$, 6 eq. L-Met 10 min; d) PBS, 5 eq. $NaIO_4$, 10 eq. L-Met 4 min; e) PBS, 5 eq. $NaIO_4$, 10 eq. L-Met 8 min; f) PBS, 10 eq. $NaIO_4$, 20 eq. L-Met 4 min

5.2 Optimised procedure for oxidation



For a typical 100 µL scale reaction, solutions of L-methionine (stock 102 mM in 20 mM sodium phosphate buffer, 150 mM NaCl, pH 7.4, 1 µL, final concentration 1 mM, 10 eq.) and $NaIO_4$ (stock 51 mM in 20 mM phosphate buffer, 150 mM NaCl, pH 7.4, 1 µL, final concentration 0.5 mM, 5 eq.) were added to protein 1,2-aminoalcohol **6** (stock 100 µM in 20 mM phosphate buffer, 150 mM NaCl, pH 7.4, 100 µL, final concentration 100 µM, 1 eq.). After mixing by gentle pipetting, the reaction mixture was incubated at 0 °C in the dark for 4 min, after which the oxidised protein was purified by desalting using a PD Minitrap G-25 (GE Healthcare), gravity method, into H_2O for analysis and further manipulation.

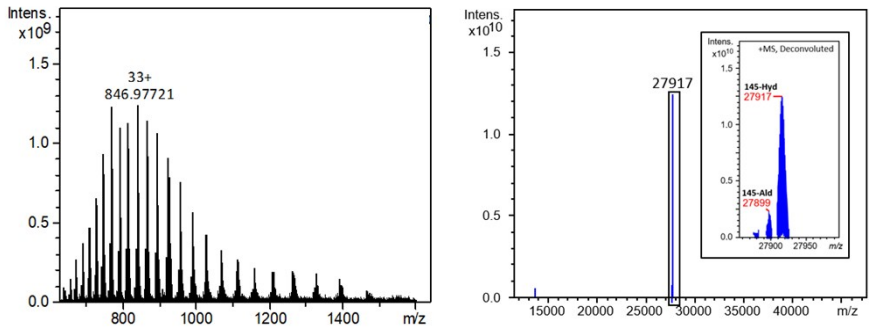
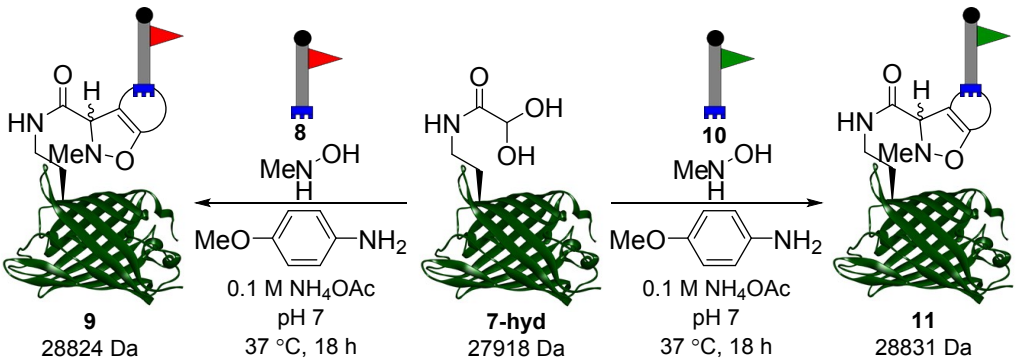


Figure S6: Raw ESI-FTICR-MS (left) and deconvoluted spectrum (right) of **7**, observed predominantly as the hydrated aldehyde (calc. 27916 Da, found 27917 Da).

5.3 SPANC ligation



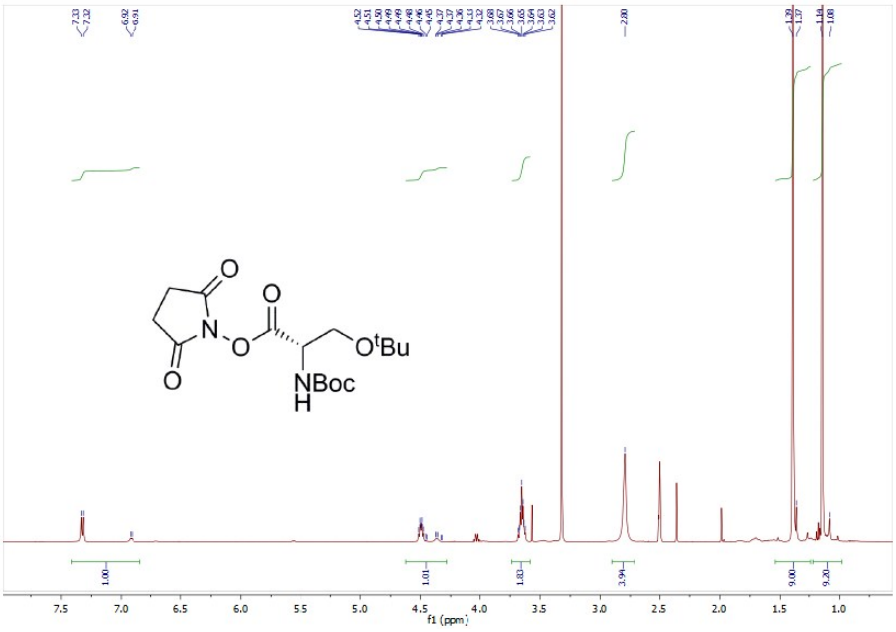
For a typical 100 µL scale reaction, solutions of *p*-anisidine (stock 1 M in 50% aq. MeCN, 10 µL, final concentration 100 mM), *N*-methylhydroxylamine hydrochloride (stock 75 mM in 0.1 M NH_4OAc, pH

6.8, 2 µL, final concentration 1.5 mM, 50 eq.) and BCN probe **8** or **10** (stock 3 mM in 50% aq. MeCN, 20 µL, final concentration 600 µM, 20 eq.) were added to protein aldehyde **7** (stock 100 µM in H_2O, 30 µL, final concentration 30 µM, 1 eq.) and the solution made up to 100 µL with buffer (38 µL, 0.1 M NH_4OAc, pH 6.8). After mixing by gentle pipetting, the reaction mixture was incubated at 37 °C for 16 h, after which the modified protein was purified by desalting using a PD Minitrap G-25 (GE Healthcare), gravity method, into H_2O for analysis and further manipulation.

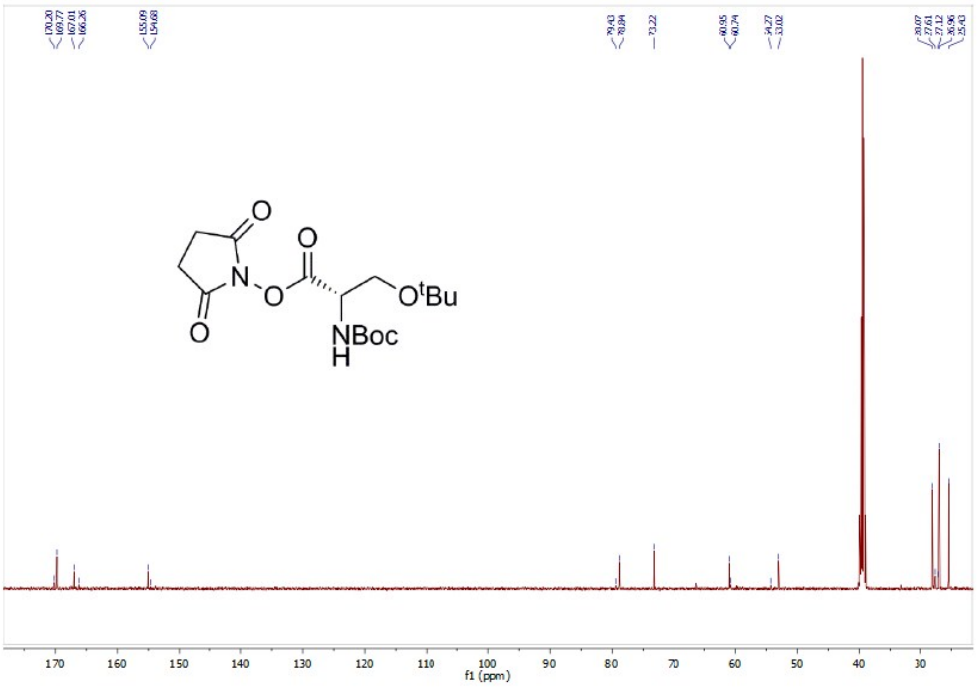
6. NMR Spectra

_N_-hydroxysuccinimidyl (2_S_)-3-(_tert_-butoxy)-2-(_tert_-butoxycarbonyl-amino)-propanoate S2

^{1}H (500 MHz, DMSO-d$_6$):

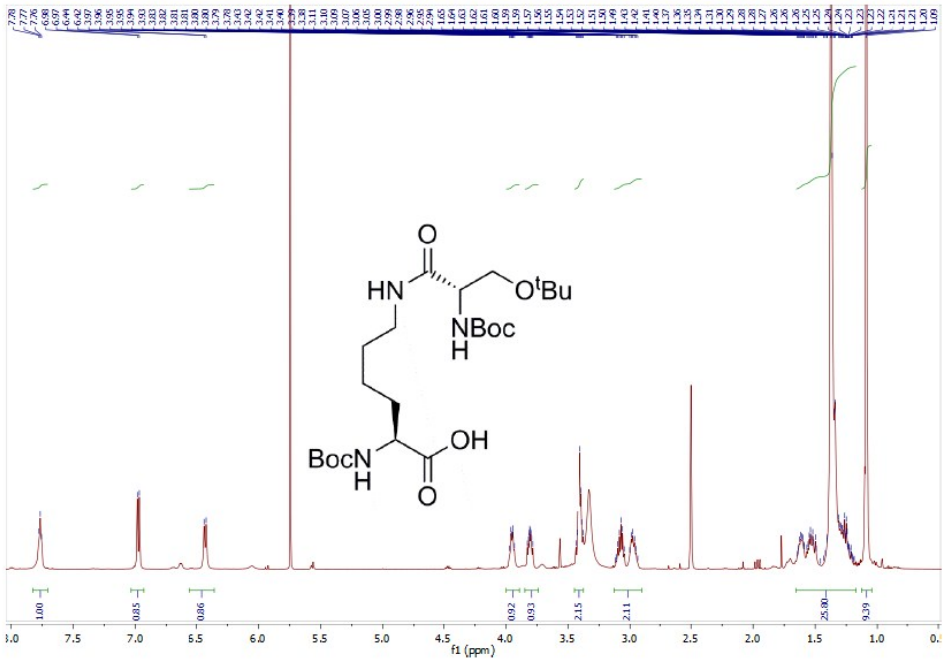


^{13}C (126 MHz, DMSO-d$_6$):

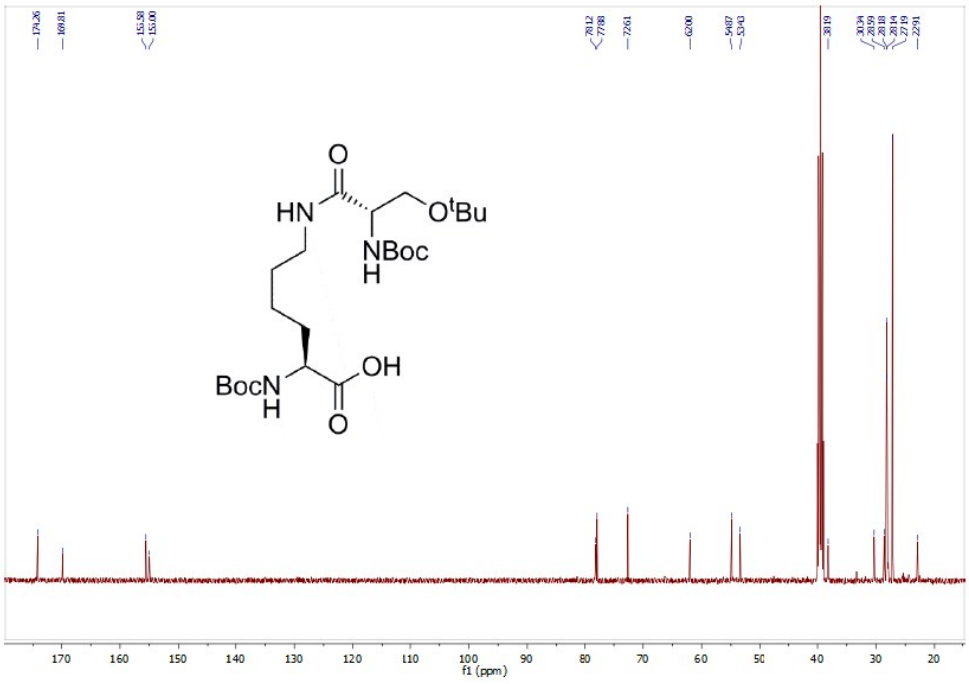


(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid S4

^{1}H (500 MHz, DMSO-d_6):

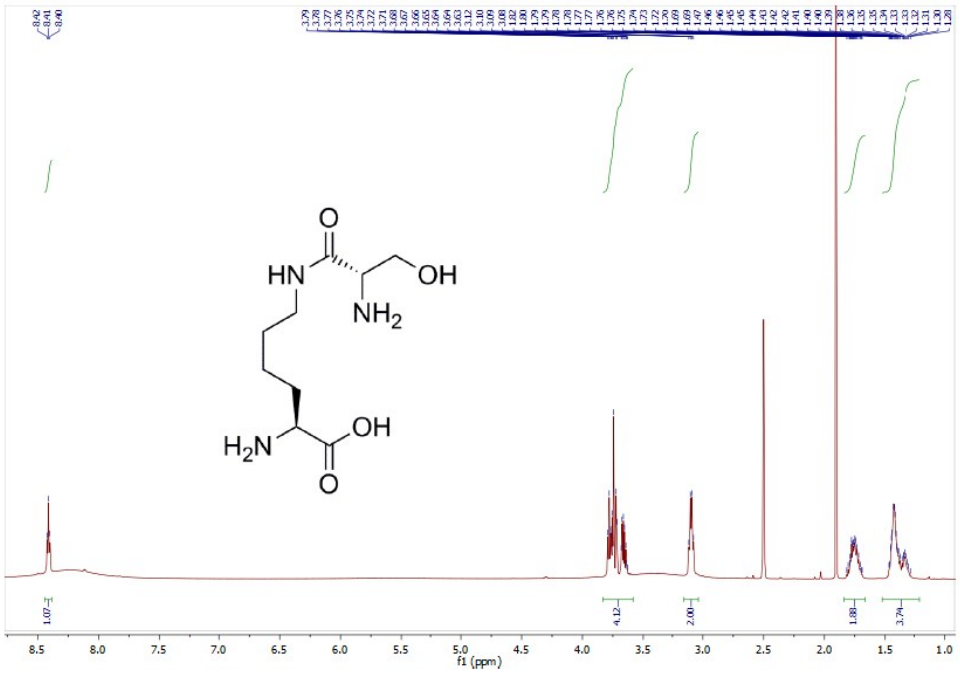


^{13}C (126 MHz, DMSO-d_6):

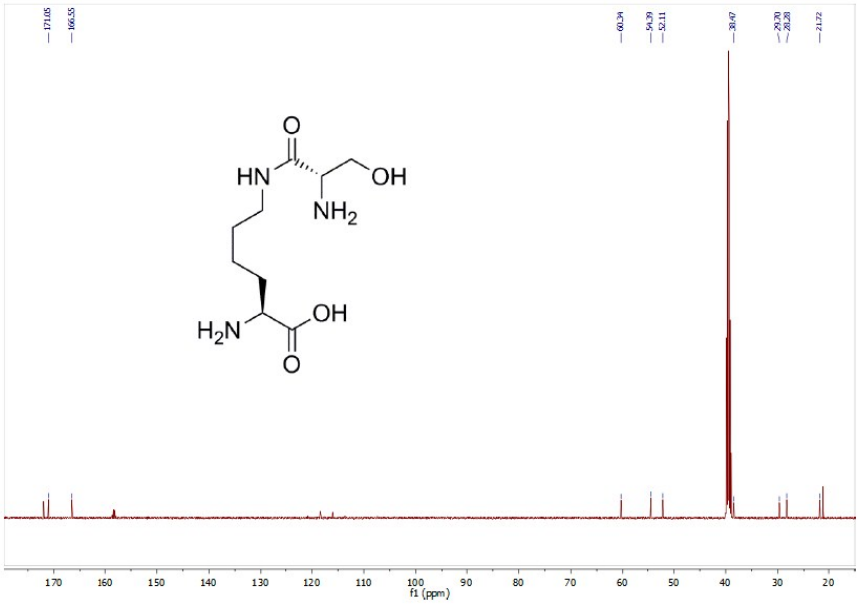


(2*S*)-2-amino-6-((2*S*)-2-amino-3-hydroxypropanamido)-hexanoic acid 1

^{1}H (500 MHz, DMSO-d_6):

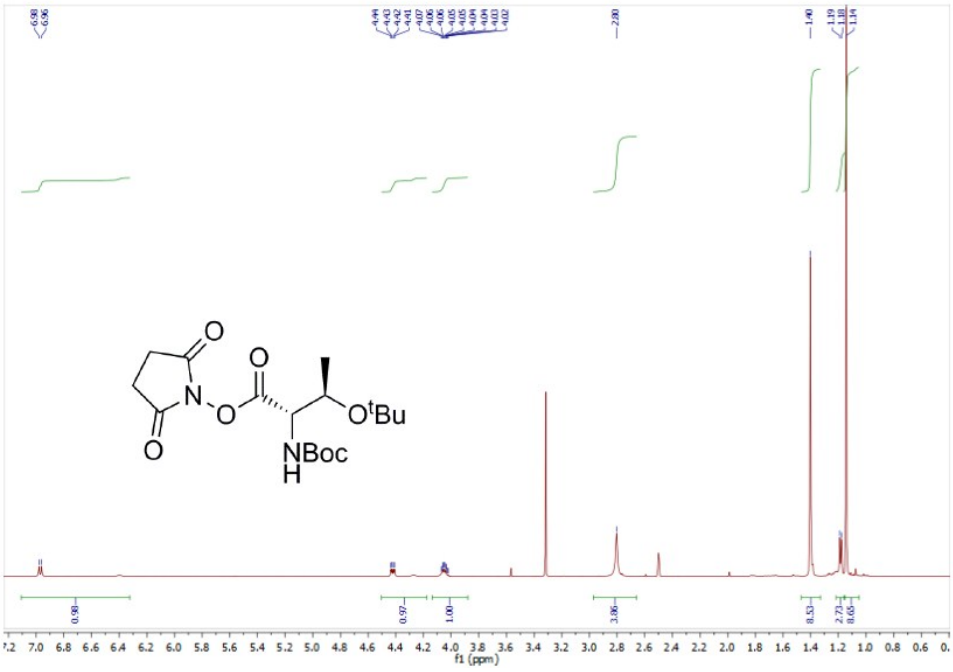


^{13}C (126 MHz, DMSO-d_6):

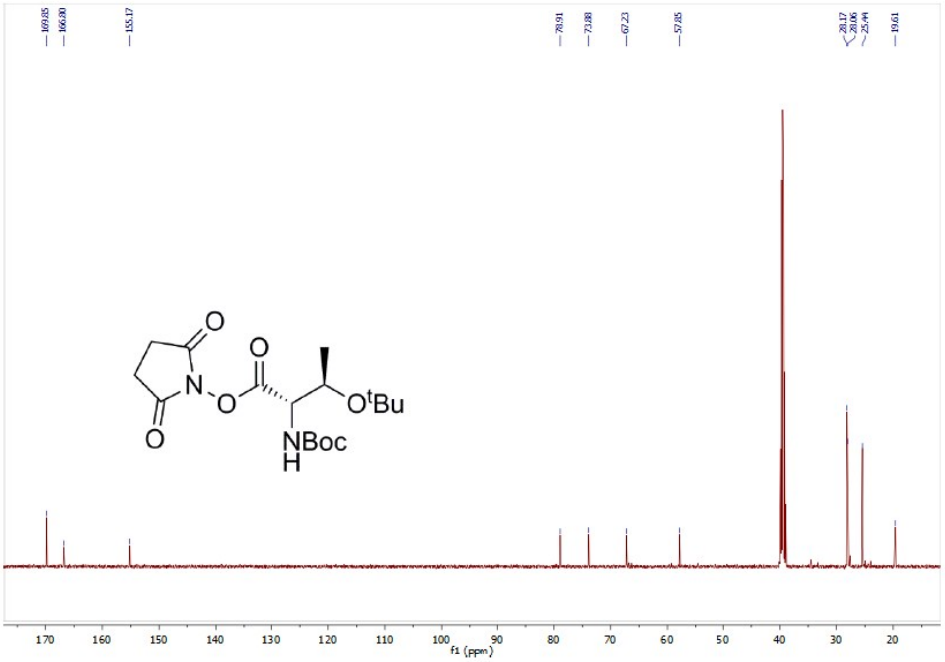


***N*-hydroxysuccinimidyl (2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonyl-amino)-butanoate S6**

^{1}H (500 MHz, DMSO-d_6):

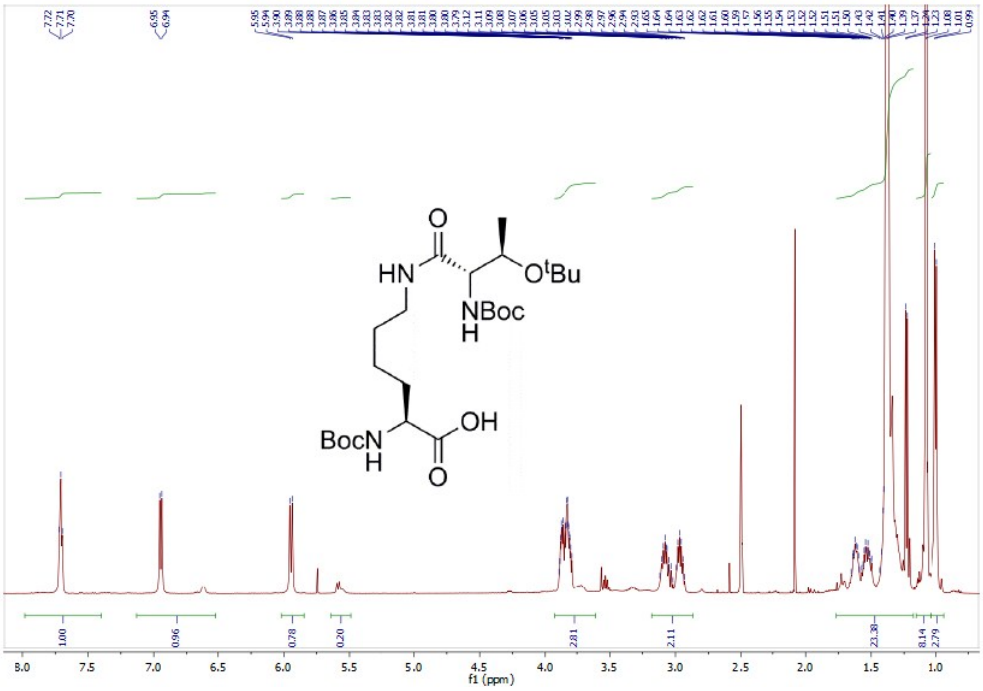


^{13}C (126 MHz, DMSO-d_6):

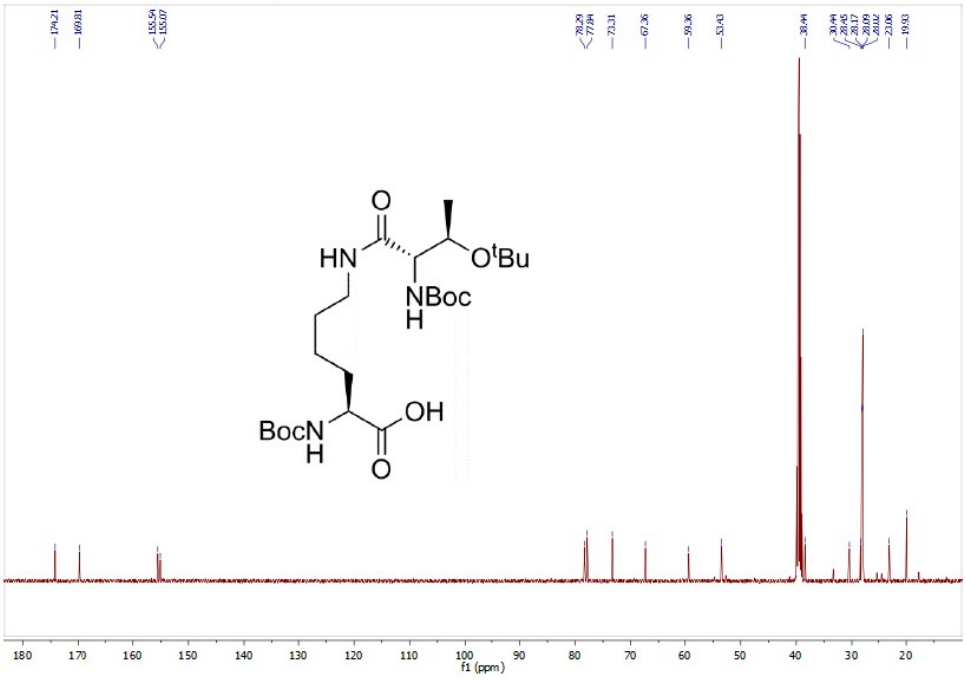


(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*S*,3*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)-hexanoic acid S7

^{1}H (500 MHz, DMSO-d$_6$):

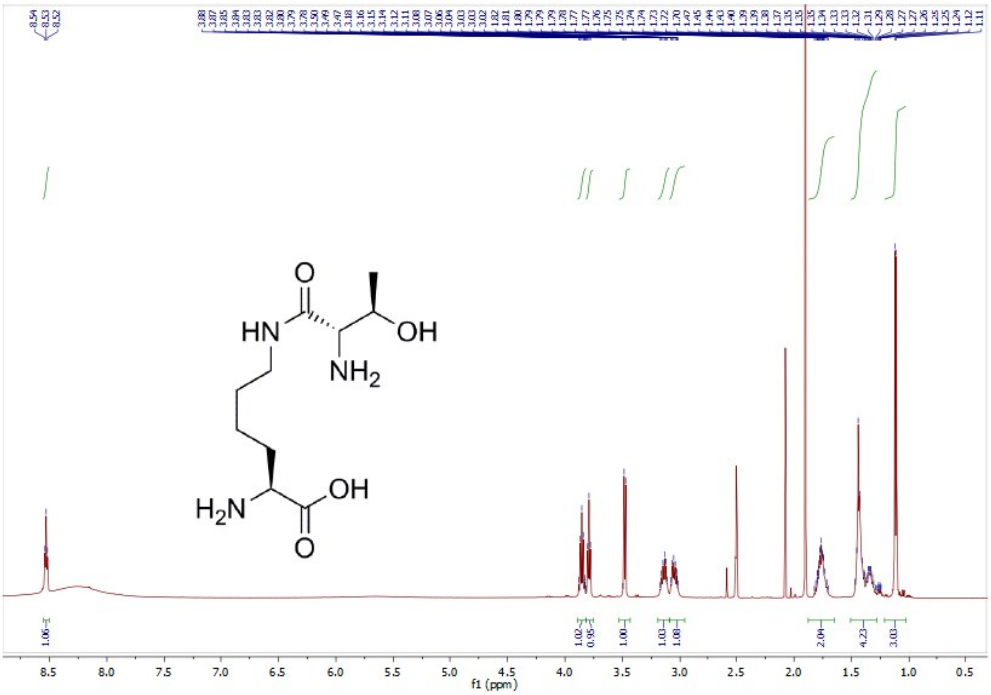


^{13}C (126 MHz, DMSO-d$_6$):

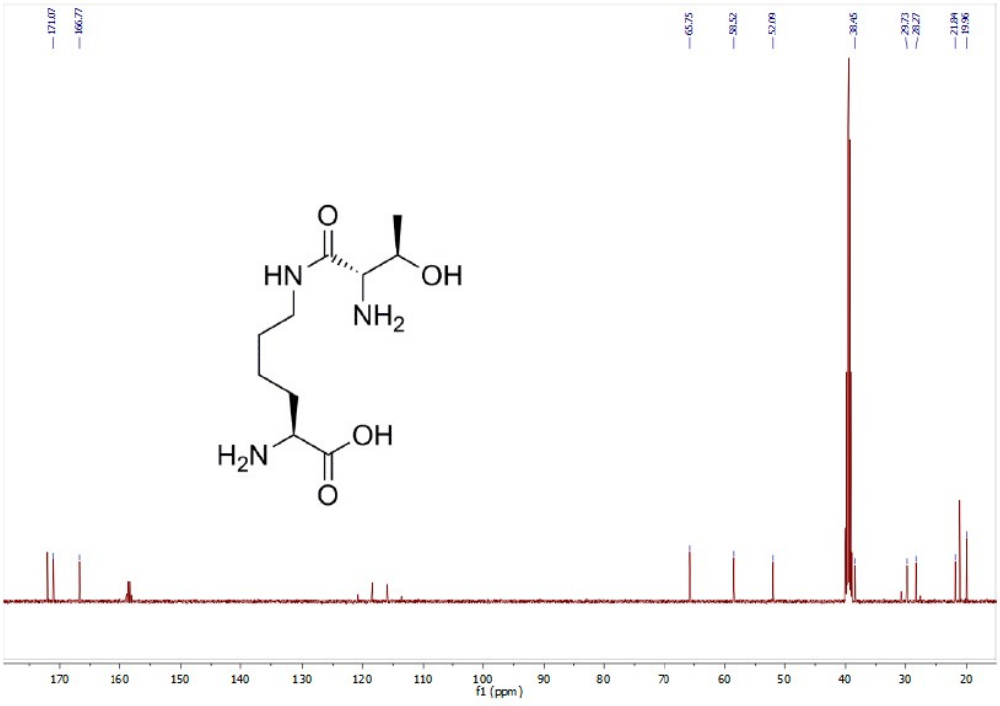


(2*S*)-2-amino-6-((2*S*,3*R*)-2-amino-3-hydroxybutanamido)-hexanoic acid 3

^{1}H (500 MHz, DMSO-d_6):

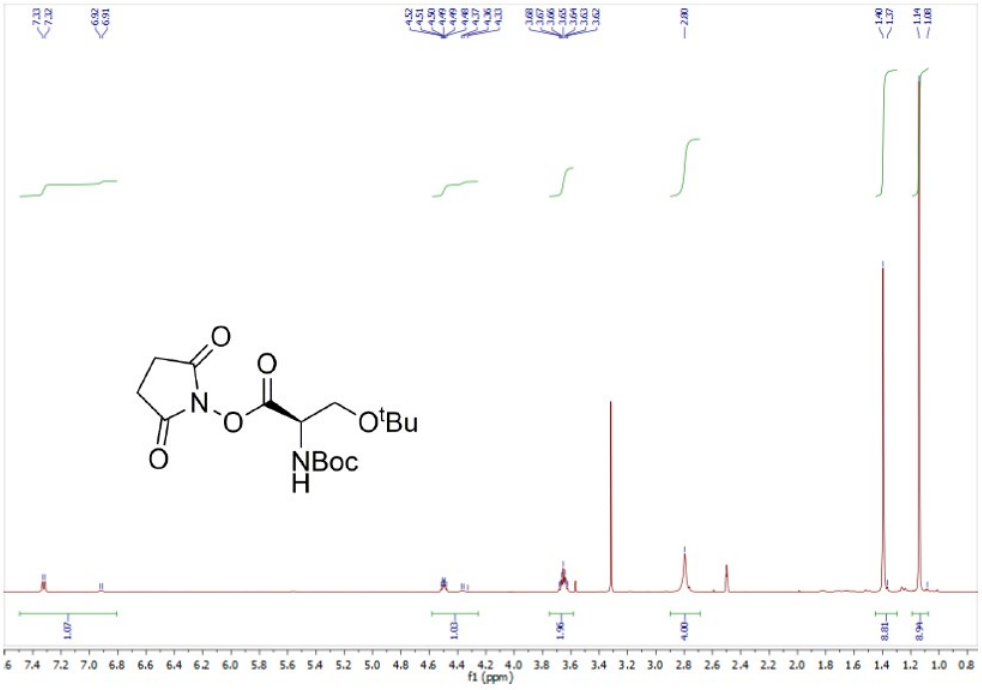


^{13}C (126 MHz, DMSO-d_6):

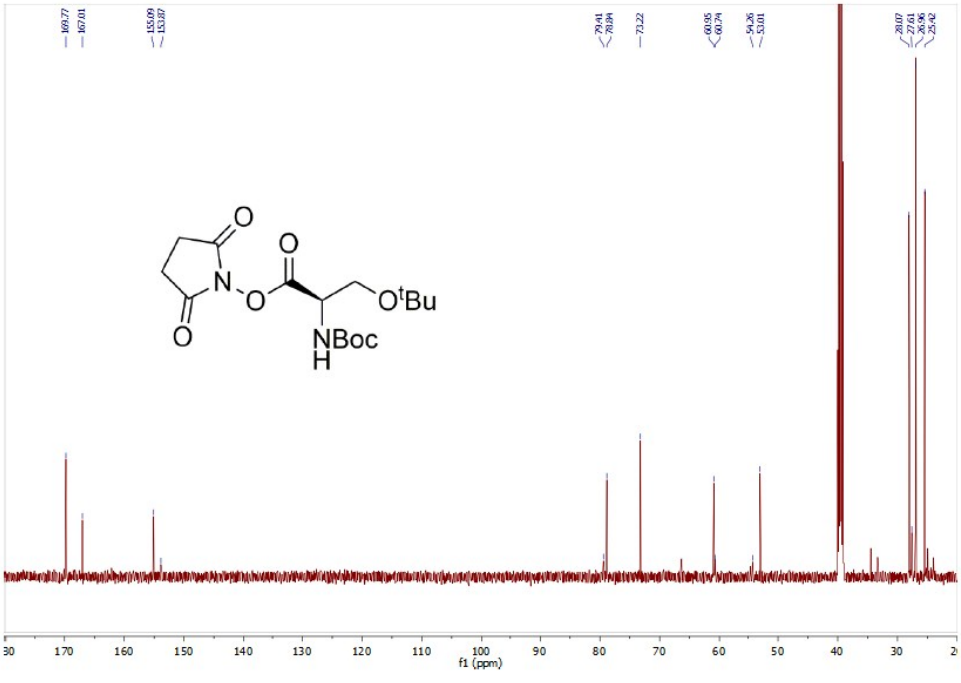


***N*-hydroxysuccinimidyl (2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanoate S9**

^{1}H (500 MHz, DMSO-d_6):

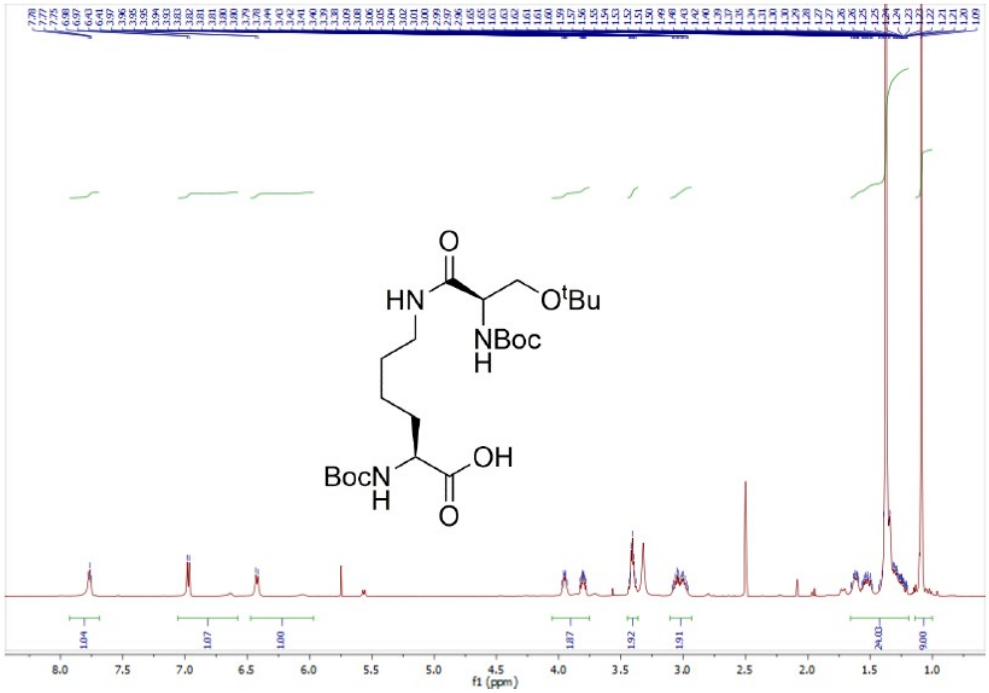


^{13}C (126 MHz, DMSO-d_6):

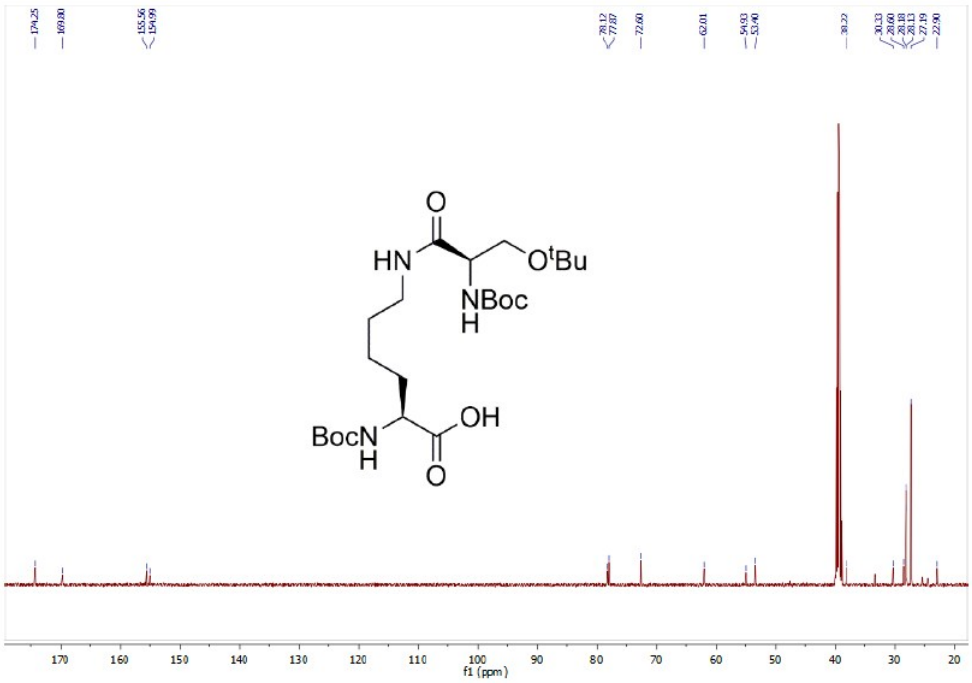


(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-propanamido)-hexanoic acid S10

^{1}H (500 MHz, DMSO-d$_6$):

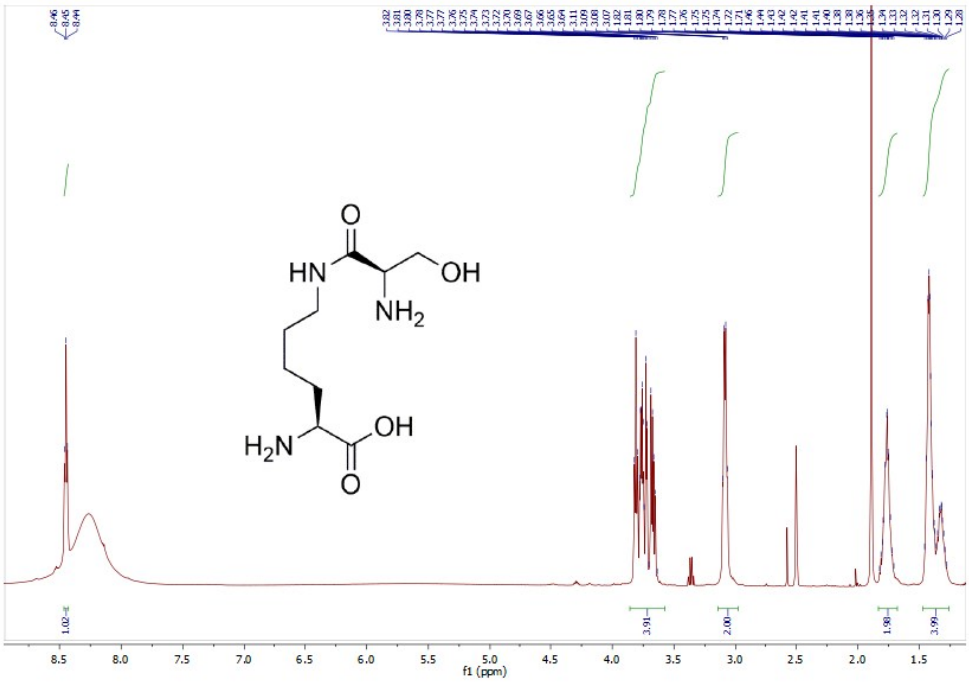


^{13}C (126 MHz, DMSO-d$_6$):

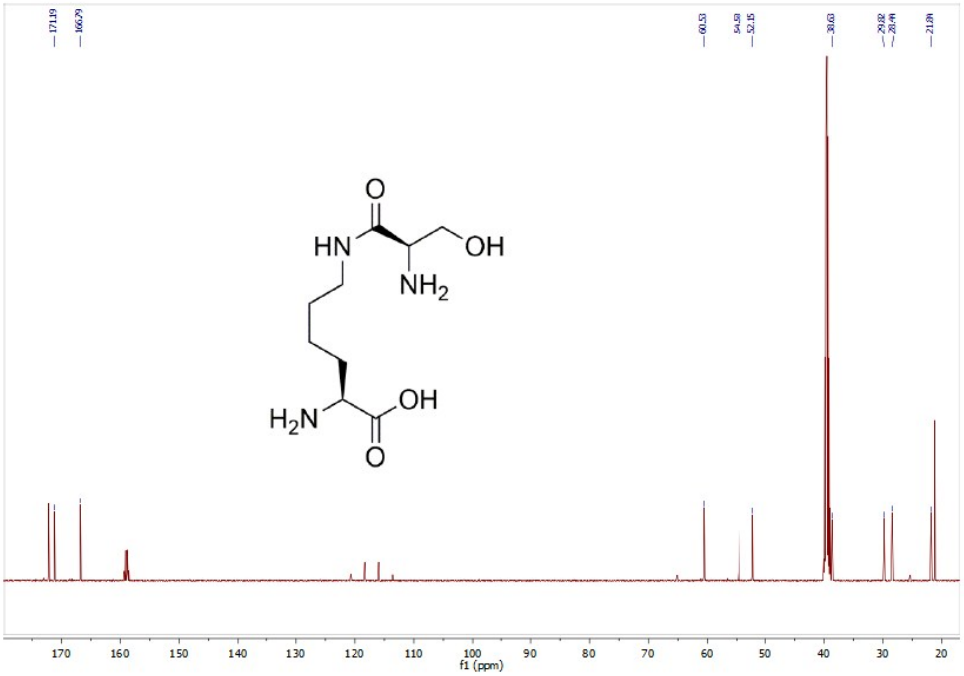


(2*S*)-2-amino-6-((2*R*)-2-amino-3-hydroxypropanamido)-hexanoic acid 2

^{1}H (500 MHz, DMSO-d_6):

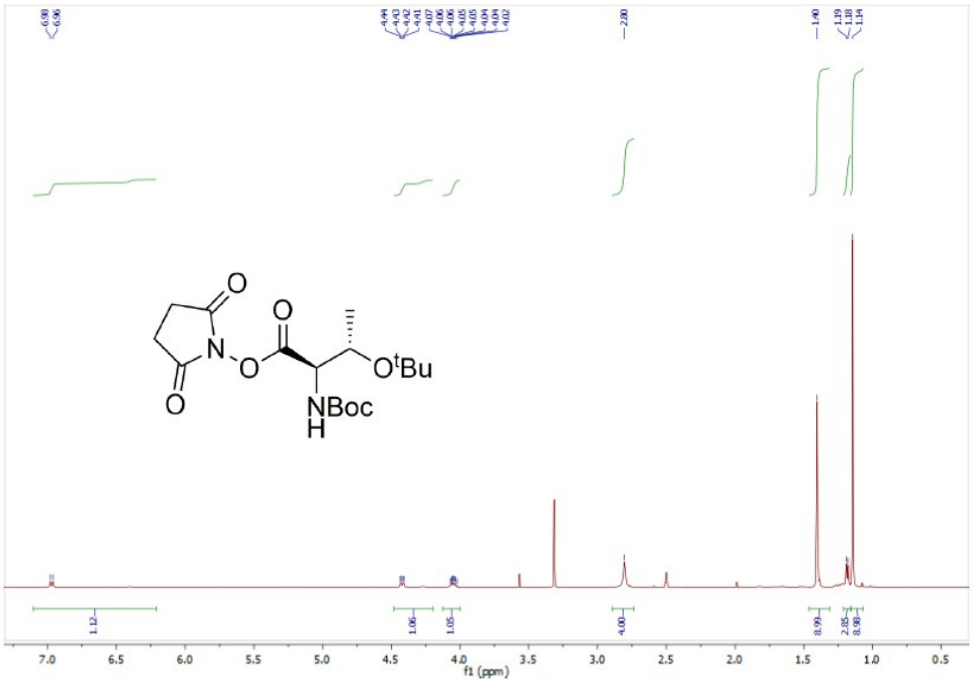


^{13}C (126 MHz, DMSO-d_6):

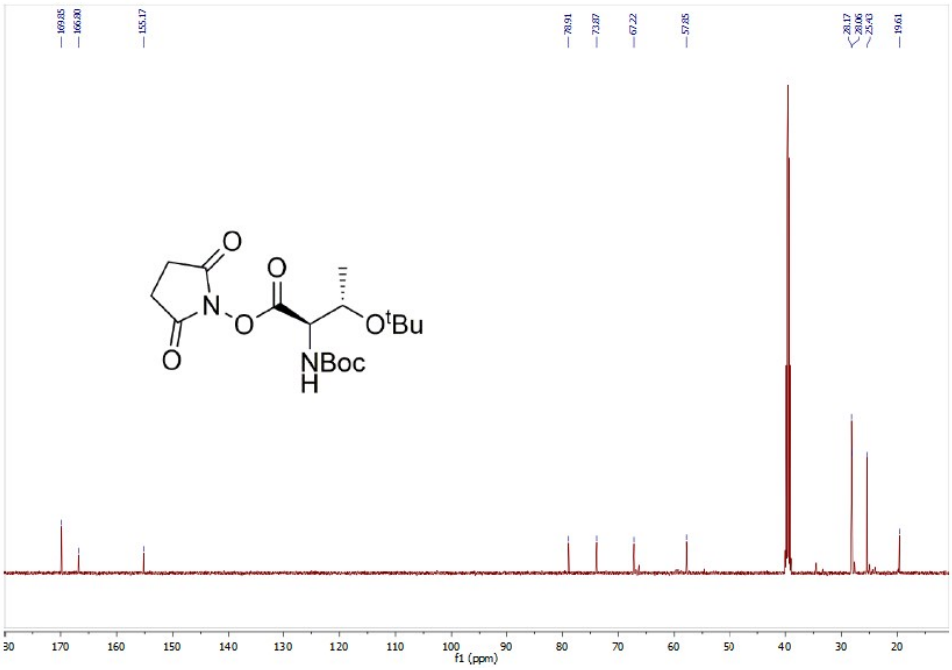


***N*-hydroxysuccinimidyl (2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonyl-amino)-butanoate S12**

^{1}H (500 MHz, DMSO-d$_6$):

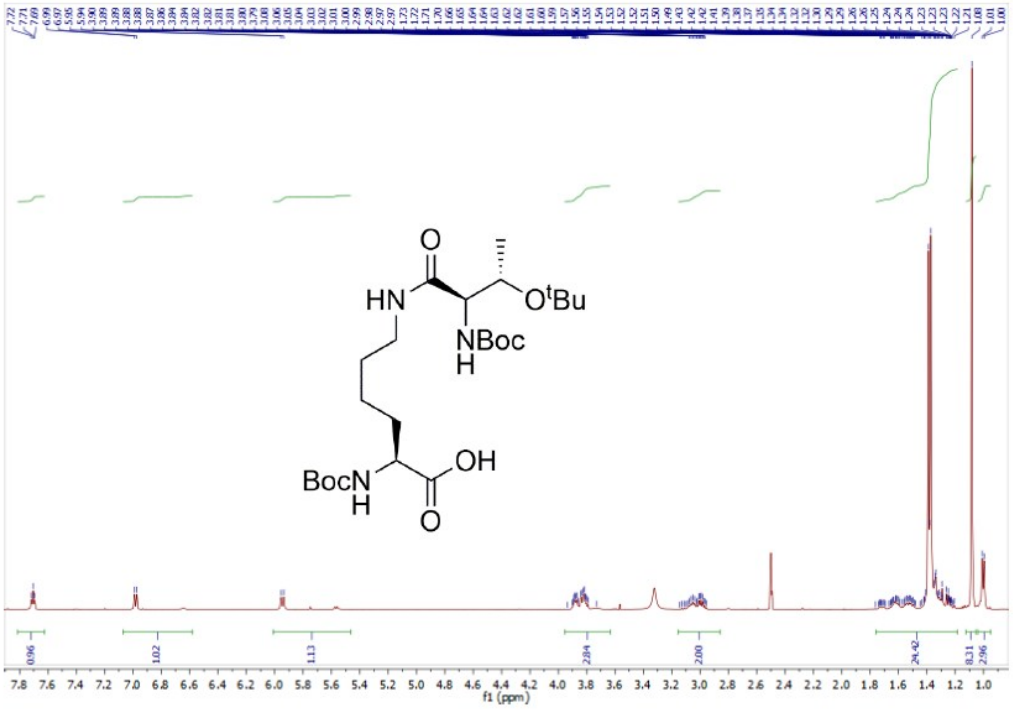


^{13}C (126 MHz, DMSO-d$_6$):

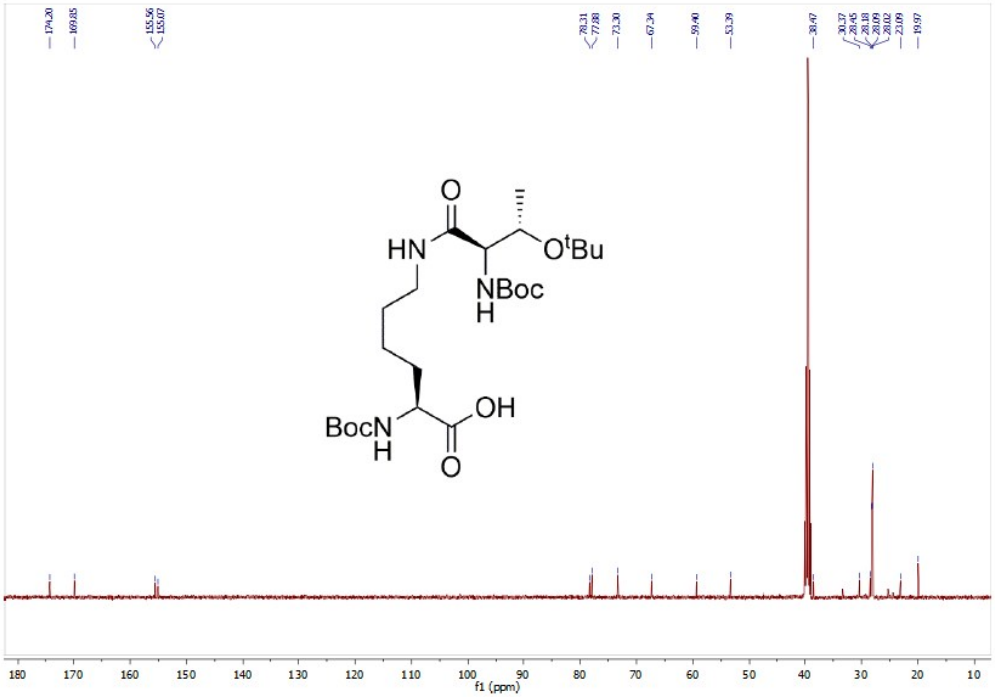


(2*S*)-2-(*tert*-butoxycarbonylamino)-6-((2*R*,3*S*)-3-(*tert*-butoxy)-2-(*tert*-butoxycarbonylamino)-butanamido)-hexanoic acid S13

^{1}H (500 MHz, DMSO-d$_6$):

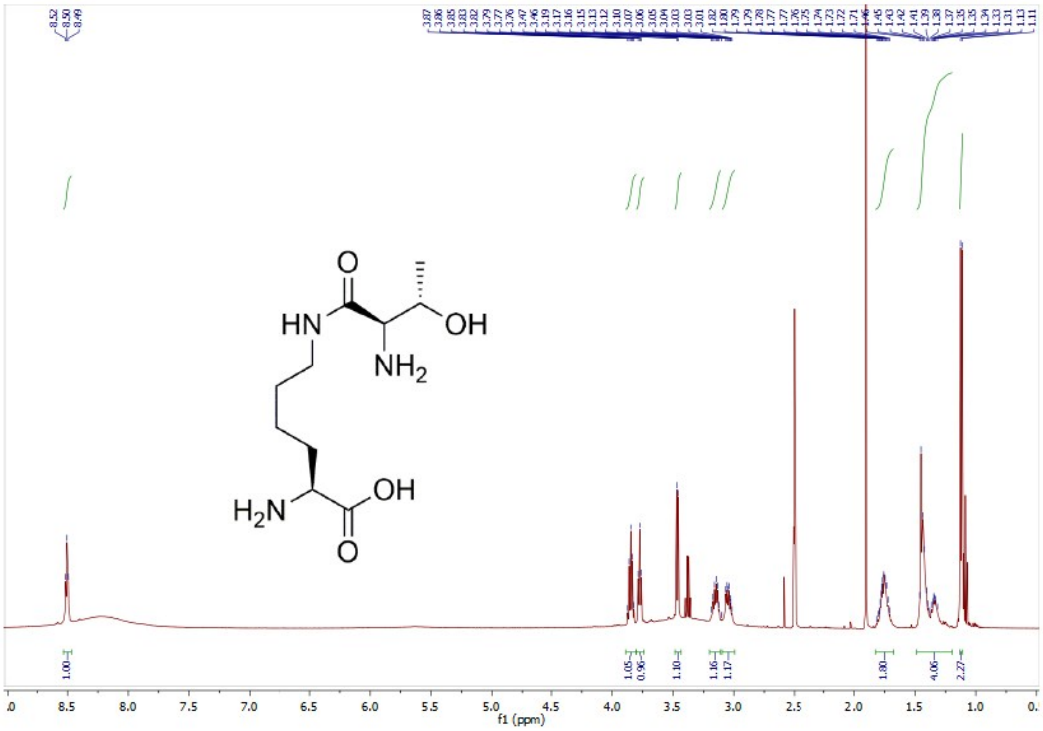


^{13}C (126 MHz, DMSO-d$_6$):

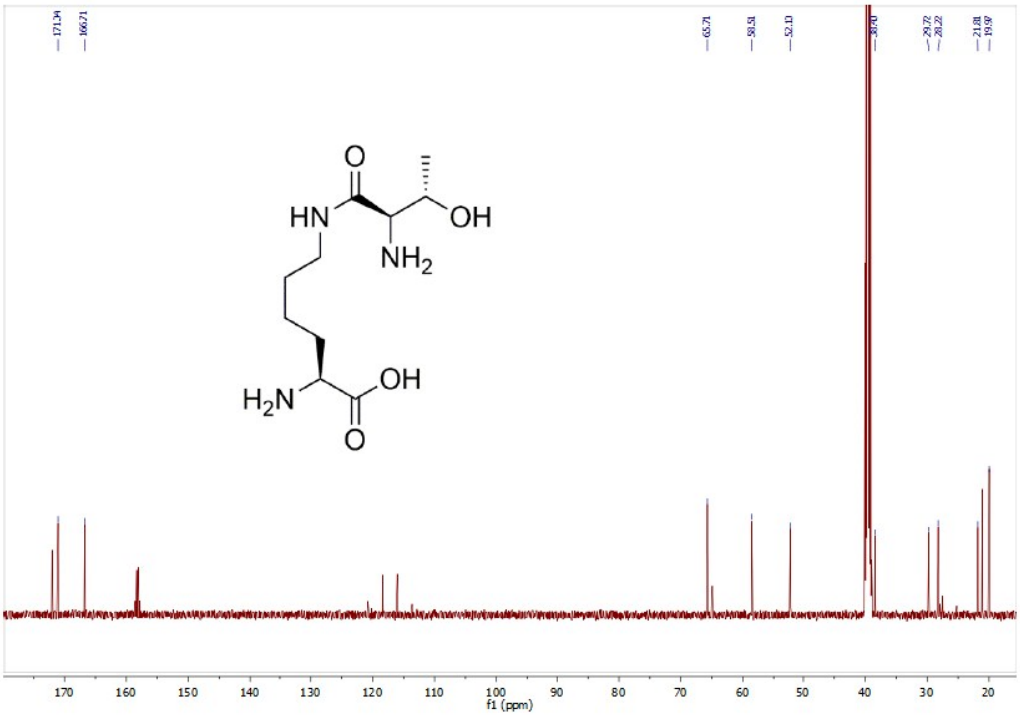


(2*S*)-2-amino-6-((2*S*,3*R*)-2-amino-3-hydroxybutanamido)-hexanoic acid 4

^{1}H (500 MHz, DMSO-d$_6$):



^{13}C (126 MHz, DMSO-d$_6$):



7. Uncropped SDS-PAGE gels and western blots

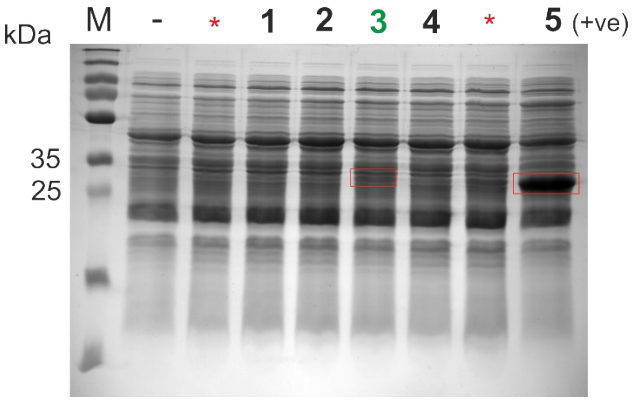


Figure S7: Full SDS-PAGE gel of the screening of lysine dipeptides **1-4** (reported in **Fig. 2**) for their incorporation into EGFP at position Y39 by the *M. mazei* Pyl tRNA-RS pair. Lanes indicated by a red * show data that has not been reported in the current manuscript.

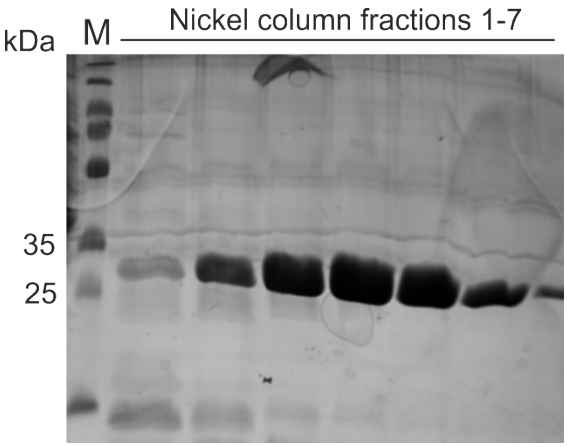


Figure S8: Full SDS-PAGE gel showing fractions containing **6**, after purification by nickel affinity chromatography.

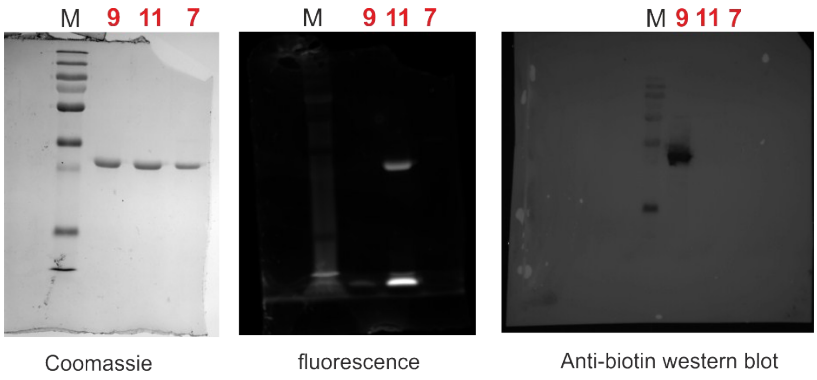


Figure S9: Full SDS-PAGE gels and western blot of the SPANC ligation of **7** with BCN biotin probe **8** or BCN dansyl probe **10** (reported in **Fig. 4**), to afford biotinylated protein **9** and dansylated protein **11** respectively.

References:

1. T. Plass, S. Milles, C. Koehler, C. Schultz and E. A. Lemke, *Angewandte Chemie International Edition*, 2011, **50**, 3878-3881.

2. R. L. Brabham, R. J. Spears, J. Walton, S. Tyagi, E. A. Lemke and M. A. Fascione, *Chemical communications*, 2018, **54**, 1501-1504.

3. C. M. Haney, M. T. Loch and W. S. Horne, *Chemical communications*, 2011, **47**, 10915-10917.