Supporting Information

Tertiary amide bond formation by an engineered asparaginyl ligase

Simon J. de Veer,†* Yan Zhou,† Thomas Durek, David J. Craik,* Fabian B. H. Rehm*

Institute for Molecular Bioscience, Australian Research Council Centre of Excellence for Innovations in Peptide and Protein Science, The University of Queensland, Brisbane, QLD 4072, Australia

† Authors contributed equally to this work

Email: s.deveer@imb.uq.edu.au; d.craik@imb.uq.edu.au; fbhrehm@gmail.com

1. Experimental Section

Reagents

All reagents were obtained from commercial sources and used as supplied. Solvents for peptide synthesis were from ChemSupply (N,N-dimethylformamide [DMF], piperidine) or Merck (dichloromethane [DCM], N,N-diisopropylethylamine [DIPEA]). Fmoc-protected amino acids were from CSBio, except for the following non-canonical amino acids: Fmoc-sarcosine (Fmoc-*N*-methylglycine, Sigma Aldrich), Fmoc-OtBu-(2*S*,4*R*)-4-hydroxyproline, Nα-Fmoc-Nε-biotinyl-L-lysine (Chem-Impex), Fmoc-L-pipecolic acid, Fmoc-L-azetidine-2-carboxylic acid, Fmoc-(2*S*,4*R*)-4-azidoproline (Combi-Blocks), Fmoc-(2*S*,4*S*)-4-azidoproline (Acrotein Chembio), Fmoc-D-proline, Fmoc-D-leucine (Mimotopes). Coupling reagents were from CSBio (O-(1H-6-Chlorobenzotriazole-1-yl)-1,1,3,3-tetramethyluronium hexafluorophosphate, HCTU) or Mimotopes (1-[Bis(dimethylamino)methylene]-1H-1,2,3-triazolo[4,5-b]pyridinium 3-oxid hexafluorophosphate, HATU). Trifluoroacetic acid (TFA) was from ChemSupply and scavengers (thioanisole, triisopropylsilane) were from Sigma Aldrich. Acetonitrile was from Fisher Scientific. Dibenzocyclooctyne-AF488 (DBCO-AF488) was from Lumiprobe.

Solid-phase peptide synthesis

Peptides were assembled on 2-chlorotrityl chloride resin (CSBio, 0.125 mmol scale, resin loading: 0.7 mmol g^{-1}) using a Symphony multiplex automated synthesiser (Protein Technologies). The dry resin was briefly swelled in DCM, then derivatised with the first Fmoc-amino acid (4 eq.) in DCM:DMF (2:3) containing 8 eq. DIPEA (3 h for Gly, 6 h for Leu or His). Unreacted sites were capped using methanol (10% v/v in DMF, 15 min), then the resin was washed 6 times with DMF. Synthesis proceeded via the following protocols, unless stated otherwise:

Deprotection: The N-terminal Fmoc group was removed using 30% (v/v) piperidine in DMF (2 × 3 min), then the resin was washed 6 times with DMF.

Coupling: Solutions of Fmoc-amino acid, HCTU, and DIPEA (all prepared in DMF) were added to the resin and mixed for 10 min (twice). Each repeat was performed using 4 eq. Fmoc-Xaa-OH, 4 eq. HCTU and 8 eq. DIPEA. After the second repeat, the resin was washed 6 times with DMF.

After removing the Fmoc group of the final amino acid, the resin was either:

(i) Washed 3 times with DMF, 2 times with DCM, then dried under N_2.

or

(ii) Washed 4 times with DMF, then the N-terminal amine was acetylated by adding acetic anhydride (20 eq.) and DIPEA (20 eq.) in DMF (30 min). The resin was subsequently washed 3 times with DMF, 2 times with DCM, and dried under N_2.

Cleavage from the resin and side chain deprotection

Peptides were cleaved from the resin and deprotected using 92% TFA, 4% thioanisole, 2% triisopropylsilane, 2% H_2O (3 h). After removing the resin by filtration, the crude product was precipitated in eight volumes of ice-cold diethyl ether (ChemSupply). The recovered product was dissolved in a mixture of acetonitrile and H_2O then lyophilised.

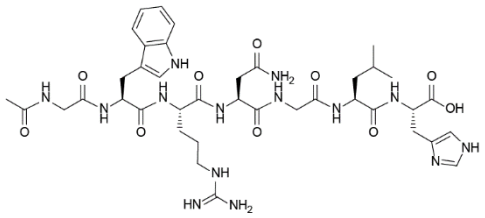
Peptide purification and validation

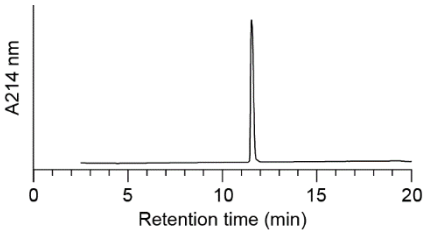
Crude peptides were purified by reverse-phase high performance liquid chromatography (RP-HPLC) using a Shimadzu Prominence system equipped with a Phenomenex Gemini C18 column (5 µm, 110 Å, 250 × 21.2 mm). Briefly, lyophilised peptides were dissolved in a mixture of acetonitrile and H_2O, passed through a 0.45 µm filter, and loaded onto the column. HPLC was performed using a gradient of 5-50% solvent B (90% acetonitrile, 0.1% TFA) in solvent A (0.1% TFA) over 60 min. Fractions containing the product mass were identified by electrospray ionisation-mass spectrometry (Shimadzu Prominence), then pooled and lyophilised. Peptide purity was verified by analytical RP-HPLC (A214 nm) using a Shimadzu Nexera system (Phenomenex Jupiter C18 column, 5 µm, 300 Å, 150 × 2 mm). The table below lists the peptides synthesised for this study, their calculated mass, and observed mass from matrix-assisted laser desorption ionisation-time of flight (MALDI-TOF) mass spectrometry (SCIEX TOF/TOF 5800 MALDI mass spectrometer, α-cyano-4-hydroxycinnamic acid matrix).

Sequence	Calculated $[M+H]^+$ (Da)	Observed $[M+H]^+$ (Da)
Ac-GWRNGLH	881.4	881.5
GLRL	458.3	458.3
NMe-GLRL	472.3	472.3
PLRL	498.3	498.4
Hyp-LRL	514.3	514.3
Pip-LRL	512.3	512.4
Aze-LRL	484.3	484.4
(2*S*,4*R*)Azp-LRL	539.3	539.4
(2*S*,4*S*)Azp-LRL	539.3	539.4
[D-Pro]LRL	498.3	498.4
P[D-Leu]RL	498.3	498.3
Ac-GWRN[NMe-Gly]LH	895.4	895.6
Ac-GWRNPLH	921.5	921.6
Ac-GWRN[Hyp]LH	937.5	937.6
Ac-GWRN[Aze]LH	907.5	907.5
Ac-GWRN[(2*S*,4*R*)Azp]LH	962.5	962.6
Ac-GWRN[(2*S*,4*S*)Azp]LH	962.5	962.6
PLGK(biotin)RG	853.5	853.6
GLGK(biotin)RG	813.5	813.6
Hyp-LGK(biotin)RG	869.5	869.6
(2*S*,4*R*)Azp-LGK(biotin)RG	894.5	894.6
TAMRA-GRNGLH	1065.5	1065.6

Ac-GWRNGLH

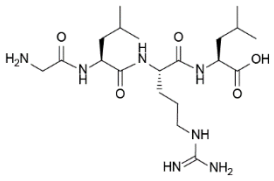
Peptide assembled using the standard protocols detailed above.

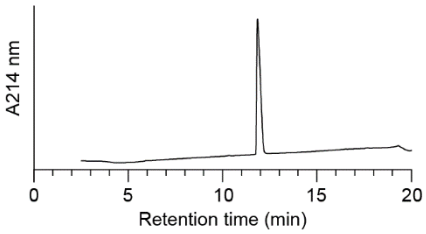




GLRL

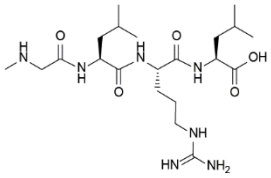
Peptide assembled using the standard protocols detailed above.

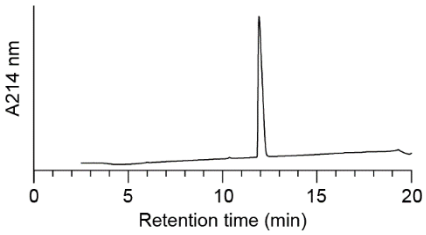




NMe-GLRL

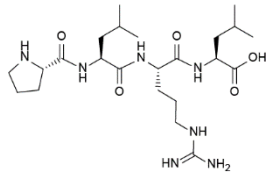
Peptide assembled using the standard protocols detailed above and Fmoc-*N*-methylglycine.

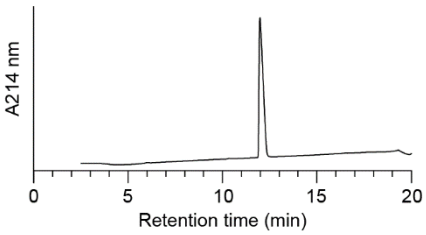




PLRL

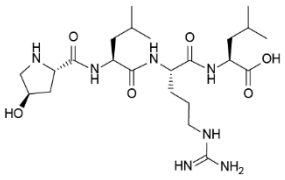
Peptide assembled using the standard protocols detailed above.

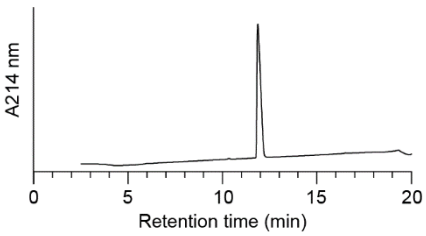




Hyp-LRL

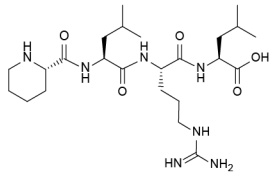
Peptide assembled using the standard protocols detailed above and Fmoc-OtBu-(2*S*,4*R*)-4-hydroxyproline.

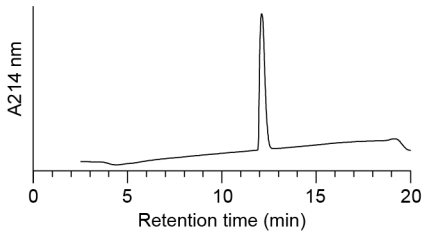




Pip-LRL

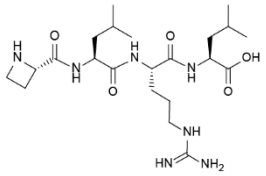
Peptide assembled using the standard protocols detailed above and Fmoc-L-pipecolic acid.

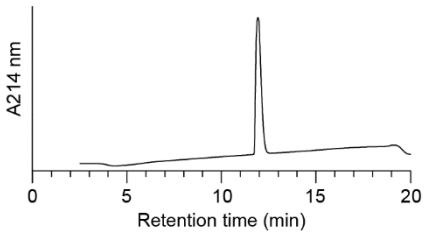




Aze-LRL

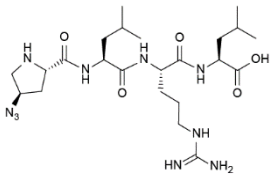
Peptide assembled using the standard protocols detailed above and Fmoc-L-azetidine-2-carboxylic acid.

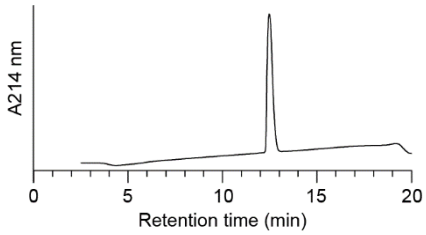




(2*S*,4*R*)Azp-LRL

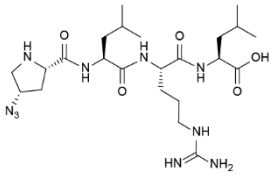
Peptide assembled using the standard protocols detailed above and Fmoc-(2*S*,4*R*)-4-azidoproline.

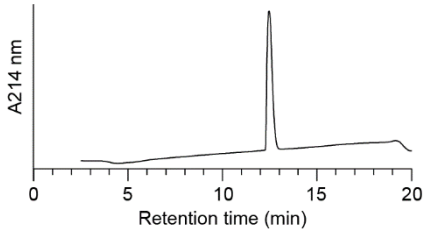




(2*S*,4*S*)Azp-LRL

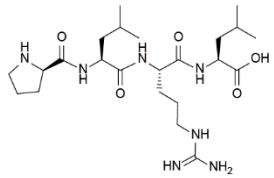
Peptide assembled using the standard protocols detailed above and Fmoc-(2*S*,4*S*)-4-azidoproline.

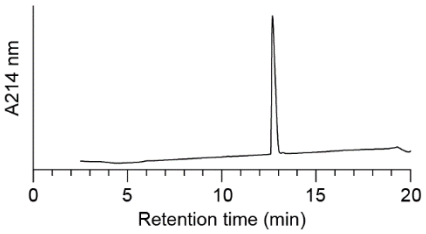




[D-Pro]LRL

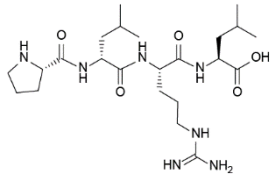
Peptide assembled using the standard protocols detailed above and Fmoc-D-proline.

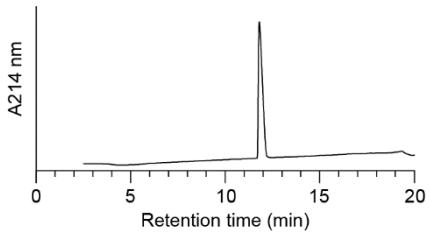




P[D-Leu]RL

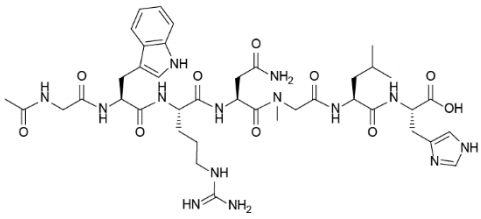
Peptide assembled using the standard protocols detailed above and Fmoc-D-leucine.

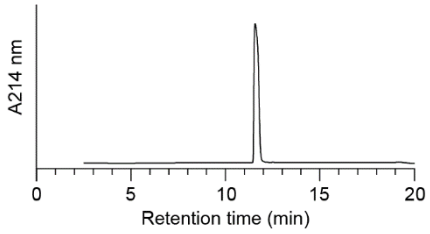




Ac-GWRN[NMe-Gly]LH

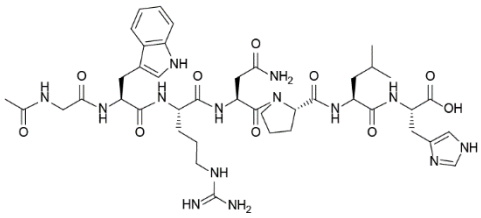
Peptide assembled using the standard protocols detailed above and Fmoc-*N*-methylglycine.

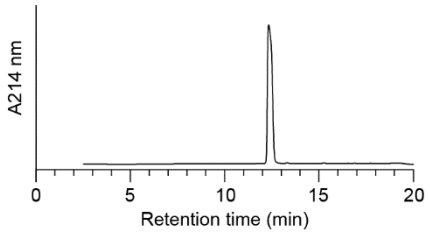




Ac-GWRNPLH

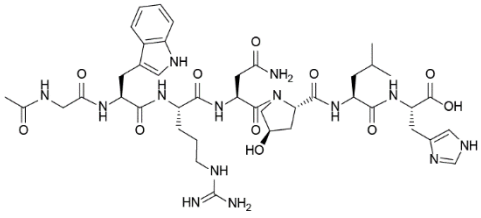
Peptide assembled using the standard protocols detailed above.

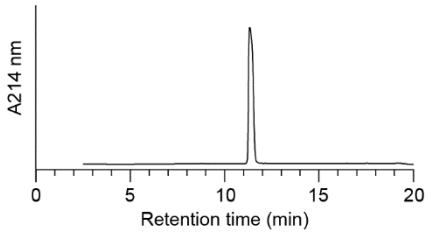




Ac-GWRN[Hyp]LH

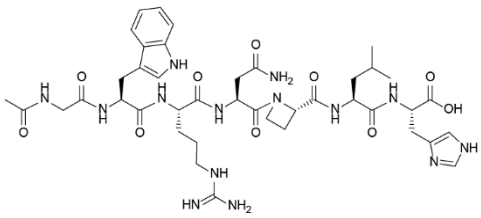
Peptide assembled using the standard protocols detailed above and Fmoc-OtBu-(2*S*,4*R*)-4-hydroxyproline.

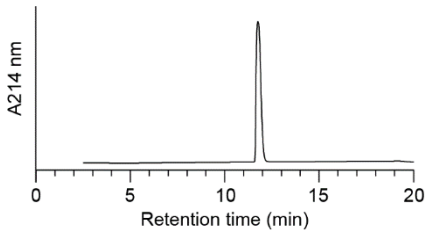




Ac-GWRN[Aze]LH

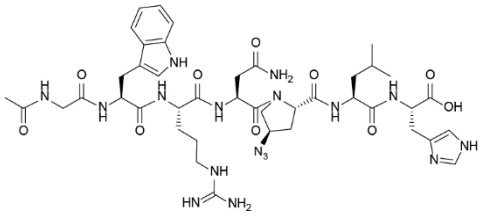
Peptide assembled using the standard protocols detailed above and Fmoc-L-azetidine-2-carboxylic acid.

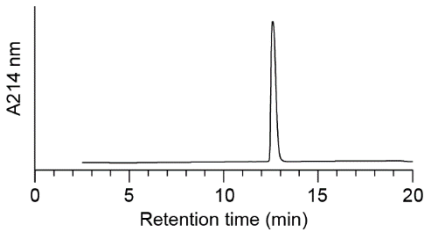




Ac-GWRN[(2*S*,4*R*)Azp]LH

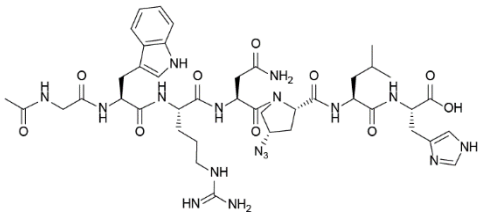
Peptide assembled using the standard protocols detailed above and Fmoc-(2*S*,4*R*)-4-azidoproline.

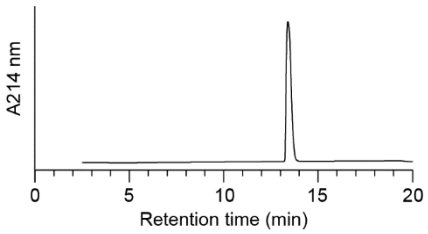




Ac-GWRN[(2*S*,4*S*)Azp]LH

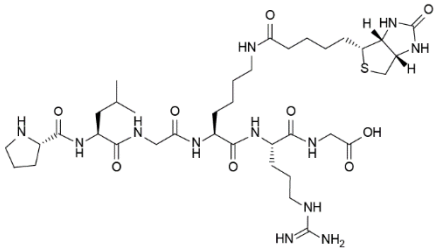
Peptide assembled using the standard protocols detailed above and Fmoc-(2*S*,4*S*)-4-azidoproline.

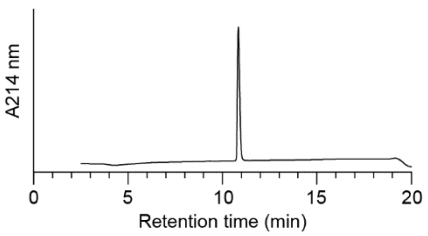




PLGK(biotin)RG

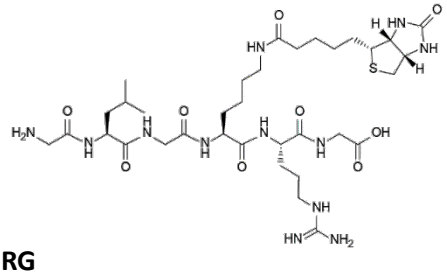
Peptide assembled using the standard protocols detailed above and Nα-Fmoc-Nε-biotinyl-L-lysine.

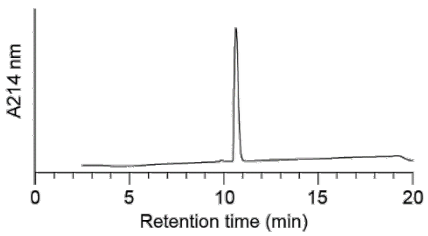




GLGK(biotin)RG

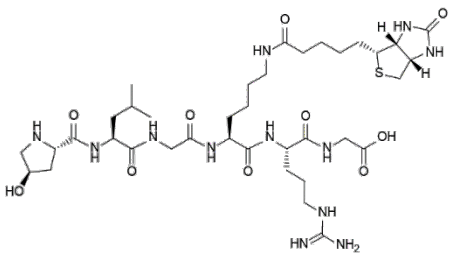
Peptide assembled using the standard protocols detailed above and Nα-Fmoc-Nε-biotinyl-L-lysine.

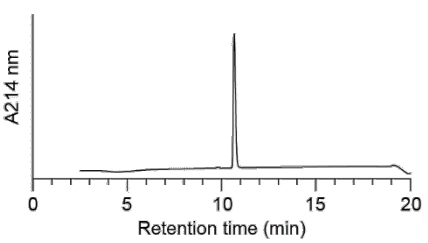




Hyp-LGK(biotin)RG

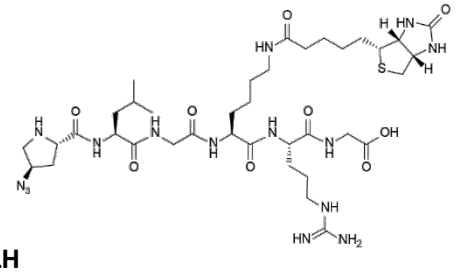
Peptide assembled using the standard protocols detailed above and Nα-Fmoc-Nε-biotinyl-L-lysine as well as Fmoc-OtBu-(2*S*,4*R*)-4-hydroxyproline.

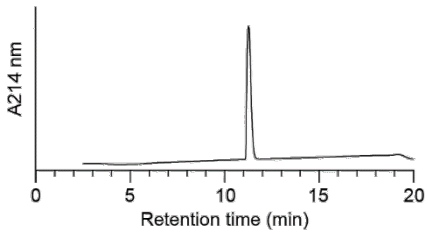




(2*S*,4*R*)Azp-LGK(biotin)RG

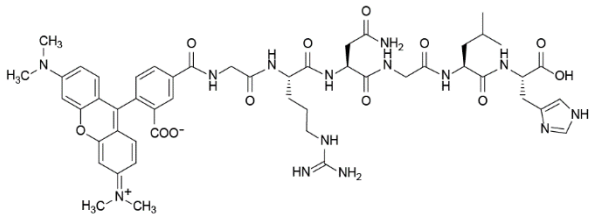
Peptide assembled using the standard protocols detailed above and Nα-Fmoc-Nε-biotinyl-L-lysine as well as Fmoc-(2*S*,4*R*)-4-azidoproline.

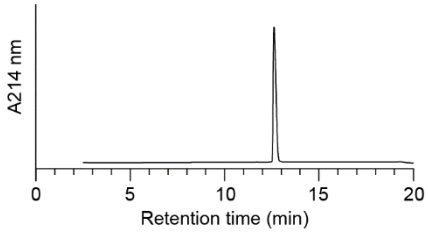




TAMRA-GRNGLH

Peptide assembled using the standard protocols detailed above, except that after deprotecting the final amino acid (Gly), an additional coupling reaction was performed using 3 eq. 5(6)-carboxytetramethylrhodamine (TAMRA, ChemPep), 2.5 eq. HATU, 6 eq. DIPEA in DMF (5 h), as previously described (Rehm et al. *J. Am. Chem. Soc.* **2019**, *141*, 17388–17393).





Recombinant protein production

Proteins were produced recombinantly in *E. coli* BL21 (all substrate proteins except VHH_{MHCII}) or SHuffle ([C247A]*Oa*AEP1 and VHH_{MHCII}), essentially as previously reported (Rehm et al. *Angew. Chem. Int. Ed.* **2021**, *60*, 4004–4008; Yap et al. *Nat. Protoc.* **2021**, 16, 1740–1760). Plasmids (pET14b for substrates or pHUE for the enzyme) were transformed into the appropriate strain and expression was induced via IPTG addition to 0.3–0.4 mM after cultures (30°C for SHuffle, 37°C for BL21) reached an OD_{600} of 0.6–0.8. Subsequently, His-tagged or Strep-tagged proteins were purified via NiNTA (Qiagen) or Strep-Tactin resin (IBA Lifesciences), following the manufacturer's guidelines. Where needed, proteins were further purified via size exclusion chromatography (HiLoad 16/600 Superdex 75 pg). For His-tagged proteins, we cleaved off the His-tag using Factor Xa protease (NEB) at 25°C overnight. Subsequently, uncleaved proteins were removed via addition of NiNTA (Qiagen) for 1 h followed by centrifugation at 800 × g. Protein substrates were buffer exchanged (PD10 or NAP5 columns) into the reaction buffer, or a buffer of equivalent pH. [C247A]*Oa*AEP1 was auto-activated via addition of acetic acid (~1:500 v/v acid:buffer, to pH 4.0–4.5) and incubation at 37°C for 2–4 h (monitored by SDS-PAGE). Where necessary, proteins were concentrated using Amicon spin columns. Protein concentrations were determined via NanoDrop A_{280} readings using calculated protein extinction coefficients and molecular weights. All proteins were stored at −80°C until use.

Analysis of peptide ligation reactions

Reactions for ligation of various tetrapeptide nucleophiles (200 µM) to model NGLH-containing substrates (Ac-GWRNXLH, 100 µM) were run using 100 nM [C247A]*Oa*AEP1 with or without $NiCl_2$ (300 µM) in 100 mM HEPES pH 7.5 at 25°C until they were quenched via TFA addition to 1% (v/v). The quenched solutions were analysed by RP-HPLC using a Shimadzu Nexera system (Phenomenex Jupiter C18 column, 5 µm, 300 Å, 150 × 2 mm, 1–50% acetonitrile gradient) and MALDI-TOF MS (SCIEX TOF/TOF 5800 MALDI mass spectrometer, α-cyano-4-hydroxycinnamic acid matrix).

Mass spectrometry analysis of protein labelling reactions

Protein labelling reactions (generally quenched in 2% TFA) were loaded on a Zorbax 300SBC18 column (Agilent) and eluted over a 15 min 1–50%, or similar, acetonitrile gradient. The column outflow was directed to a QSTAR Elite (SCIEX) or 5600 TripleTOF (SCIEX) electrospray ionisation mass spectrometer. Reconstructed spectra were generated in the Analyst software (SCIEX). Percent conversion to product was calculated based on relative peak areas determined by the same Analyst software.

N-terminal protein labelling

N-terminal reactions contained 50 µM substrate protein, 500 µM TAMRA-GRNGLH, 2 mM $NiCl_2$ and 200 nM [C247A]*Oa*AEP1 in 100 mM HEPES pH 7 at 25°C.

C-terminal protein labelling

C-terminal labelling reactions contained 50 µM substrate protein, 200 µM peptide (XLGK(biotin)RG, where X is proline, (2*S*,4*R*)-4-hydroxyproline or (2*S*,4*R*)-4-azidoproline), 250 µM $NiCl_2$ and 500 nM [C247A]*Oa*AEP1 in 100 mM HEPES pH 7 at 25°C. For (2*S*,4*R*)Azp-LGK(biotin)RG, ligation of the biotinylated peptide (4 h) was followed by strain-promoted azide-alkyne cycloaddition using 250 µM DBCO-AF488 (4 h) to generate a nanobody that was dual labelled directly at the ligation junction.

Protein dual labelling at the N- and C-termini

Initially the CTC-445.2d protein was C-terminally labelled in a reaction containing 50 µM protein, 500 µM PLGK(biotin)RG peptide, 250 µM $NiCl_2$ and 500 nM [C247A]*Oa*AEP1 in 100 mM HEPES pH 7 for 4 h at 25°C. Next, the reaction was buffer exchanged into the same reaction buffer (NAP5 column) and then N-terminally deprotected via TEV protease (NEB) cleavage and concentrated back to the original volume via an Amicon spin column. Next, the protein was N-terminally labelled in a reaction containing 50 µM protein, 200 µM TAMRA-GRNGLH, 1 mM $NiCl_2$ and 100 nM [C247A]*Oa*AEP1 in 100 mM HEPES, pH 7, for 10 min to yield the dual labelled product.

2. Supporting Tables

Table S1. Monoisotopic masses for model peptide substrates (Ac-GWRNXLH where X = Gly, NMe-Gly, Pro, Hyp, Aze or Azp) and conjugate products

		Calculated $[M+H]^+$ (Da)	Observed $[M+H]^+$ (Da)
X = Gly	Starting material	881.4	881.5
	Hydrolysed substrate	574.3	574.4
	Ac-GWRNGLRL	1013.6	1013.6
	Ac-GWRN[NMe-Gly]LRL	1027.6	1027.6
	Ac-GWRNPLRL	1053.6	1053.6
	Ac-GWRN[Hyp]LRL	1069.6	1069.6
	Ac-GWRN[Pip]LRL	1067.6	n.d.
	Ac-GWRN[Aze]LRL	1039.6	1039.7
	Ac-GWRN[Azp]LRL	1094.6	1094.7
X = NMe-Gly	Starting material	895.4	895.6
	Hydrolysed substrate	574.3	n.d.
	Ac-GWRNGLRL	1013.6	1013.7
X = Pro	Starting material	921.5	921.6
	Hydrolysed substrate	574.3	n.d.
	Ac-GWRNGLRL	1013.6	1013.7
X = Hyp	Starting material	937.5	937.6
	Hydrolysed substrate	574.3	n.d.
	Ac-GWRNGLRL	1013.6	1013.7
X = Aze	Starting material	907.5	907.5
	Hydrolysed substrate	574.3	n.d.
	Ac-GWRNGLRL	1013.6	1013.7
X = Azp	Starting material	962.5	962.6
	Hydrolysed substrate	574.3	n.d.
	Ac-GWRNGLRL	1013.6	1013.7

n.d. = not detected

Table S2. Average masses for model protein substrates and conjugate products

	Calculated mass (Da)	Observed mass (Da)
N-terminal labelling		
PL-SUMO	12,343.9	12,343
TAMRA-GRNPL-SUMO	13,083.7	13,082
PL-sfGFP	27,950.5	27,950
TAMRA-GRNPL-sfGFP	28,690.3	28,690
C-terminal labelling		
VHH_{MHCII}-NGLH	15,820.4	15,820
VHH_{MHCII}-NPLH	15,860.5	15,860
VHH_{MHCII}-NPLGK(biotin)RG	15,308.0	15,308
VHH_{MHCII}-NGLGK(biotin)RG	15,268.0	15,268
VHH_{MHCII}-N[Hyp]LGK(biotin)RG	15,324.0	15,323
VHH_{MHCII}-N[Azp]LGK(biotin)RG	15,349.0	15,349
C-terminal labelling + SPAAC		
VHH_{MHCII}-N[Azp-AF488]LGK(biotin)RG	16,183.9	16,183
Dual labelling		
PENLYFQGL-CTC-445.2d-NGLHGIEGR	38,215.5	38,215
PENLYFQGL-CTC-445.2d-NPLGK(biotin)RG	38,230.6	38,230
GL-CTC-445.2d-NPLGK(biotin)RG	37,338.7	37,339
TAMRA-GRNGL-CTC-445.2d-PLGK(biotin)RG	38,078.5	38,078

Sequences of protein substrates produced in this work (note: calculated masses take disulfide or chromophore formation into consideration; final substrates are underlined)

VHH_{MHCII}-NGLH

MPQVQLQESGGGLVQAGDSLRLSCAASGRTFSRGVMGWFRRAPGKEREFVAIFSGSSWSGRSTYYSDSVKGRFTI
SRDNAKNTVYLQMNGLKPEDTAVYYCAAGYPEAYSAYGRESTYDYWGQGTQVTVSSGGNGLHWSHPQFEK

VHH_{MHCII}-NPLH

MPQVQLQESGGGLVQAGDSLRLSCAASGRTFSRGVMGWFRRAPGKEREFVAIFSGSSWSGRSTYYSDSVKGRFTI
SRDNAKNTVYLQMNGLKPEDTAVYYCAAGYPEAYSAYGRESTYDYWGQGTQVTVSSGGNPLHWSHPQFEK

PL-SUMO

MPLASMSDSEVNQEAKPEVKPEVKPETHINLKVSDGSSEIFFKIKKTTPLRRLMEAFAKRQGKEMDSLRFLYDGIRIQ
ADQTPEDLDMEDNDIIEAHREQIGGGSGGIEGRHHHHHH

PL-sfGFP

MPLSRKGEELFTGVVPILVELDGDVNGHKFSVRGEGEGDATNGKLTLKFICTTGKLPVPWPTLVTTLTYGVQCFARYP
DHMKQHDFFKSAMPEGYVQERTISFKDDGTYKTRAEVKFEGDTLVNRIELKGIDFKEDGNILGHKLEYNFNSHNVYI
TADKQKNGIKANFKIRHNVEDGSVQLADHYQQNTPIGDGPVLLPDNHYLSTQSVLSKDPNEKRDHMVLLEFVTAA
GITHGMAELYKGSGALGSGIEGRGHHHHHH

ENLYFQ-GL-CTC-445.2d-NGLH

MPENLYFQGLSVEIDLGKGDFREIRASEDAREAAEALAEAARAMKEALEILREIAEKLRDSSRASEAAKRIAKAIRKAA
DAIAEAAKIAARAAKDGDAARNAENAARKAKEFAEEQAKLADMYAELAKNGDKSSVLEQLKTFADKAFHEMEDLF
YQAALAVFEAAEAAAGGGGSGGSGSGGSGGGSPGSVEIDLGKGDFREIRASEDAREAAEALAEAARAMKEALEILR
EIAEKLRDSSRASEAAKRIAKAIRKAADAIAEAAKIAARAAKDGDAARNAENAARKAKEFAEEQAKLADMYAELAKN
GDKSSVLEQLKTFADKAFHEMEDLFYQAALAVFEAAEAAAGGSGWGSGSNGLHGIEGRHHHHHH

3. Supporting Figures

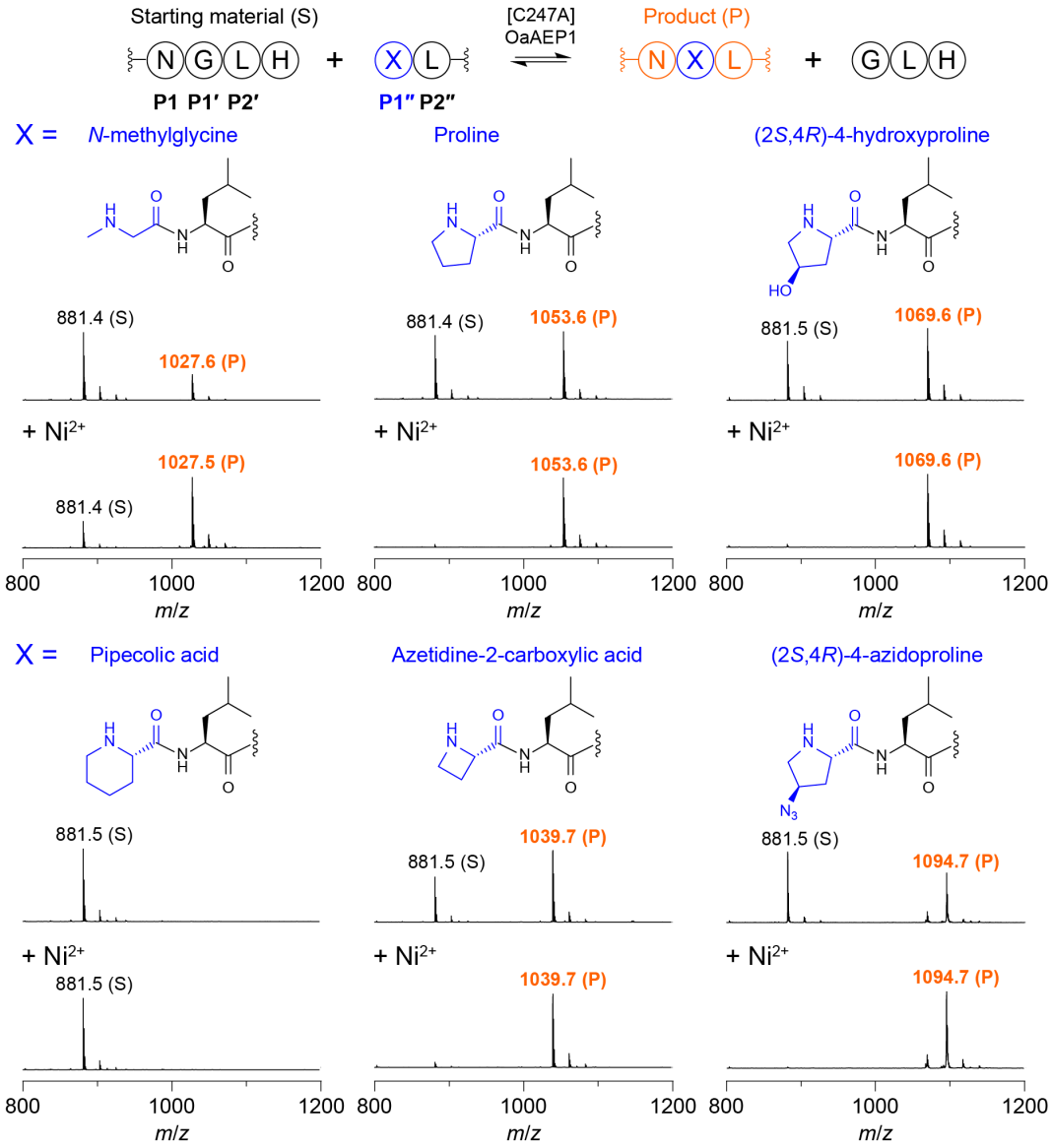


Figure S1. Enzymatic tertiary amide bond formation with model peptides. Reactions were run using 100 nM [C247A]*Oa*AEP1, an NGLH-containing substrate (Ac-GWRNGLH, 100 µM) and tetrapeptides bearing an N-terminal secondary amine (XLRL where X = *N*-methylglycine, proline, (2*S*,4*R*)-4-hydroxyproline, pipecolic acid, azetidine-2-carboxylic acid or (2*S*,4*R*)-4-azidoproline, 200 µM, 2 equiv.) in 100 mM HEPES pH 7.5 (1 h at 25°C). Shown are MALDI-TOF MS spectra. Peaks for starting material (S, black) and product (P, orange) are indicated, as well as the observed masses. Analytical reverse-phase HPLC traces are shown in Figure 1. Reactions labelled + Ni^{2+} (lower panels) were run with 300 µM $NiCl_2$ included.

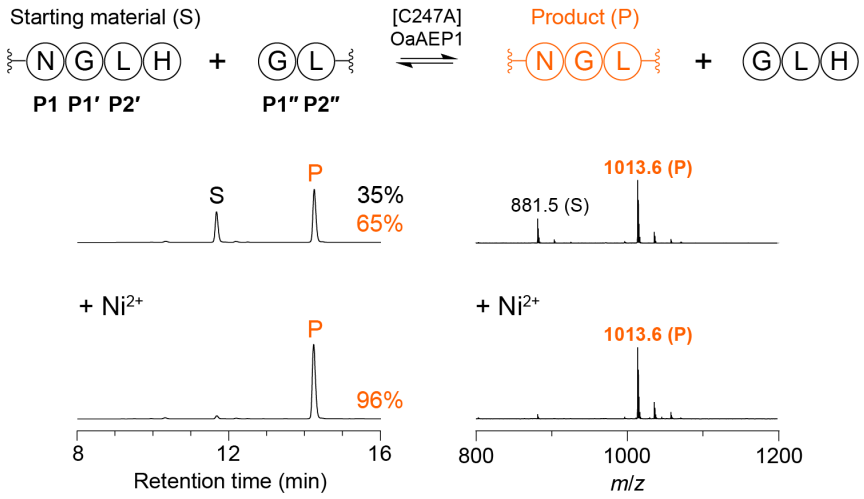


Figure S2. Conventional ligation reactions with an optimal peptide nucleophile. Reactions were run using 100 nM [C247A]*Oa*AEP1, an NGLH-containing substrate (Ac-GWRNGLH, 100 µM) and GLRL (200 µM, 2 equiv.) in 100 mM HEPES pH 7.5 (1 h at 25°C). Shown are (left panels) analytical reverse-phase HPLC traces (A280 nm), and (right panels) MALDI-TOF MS spectra. Peaks for starting material (S, black) and product (P, orange) are indicated, as well as the % conversion to product or observed mass. Reactions labelled + Ni^{2+} (lower panels) were run with 300 µM $NiCl_2$ included.

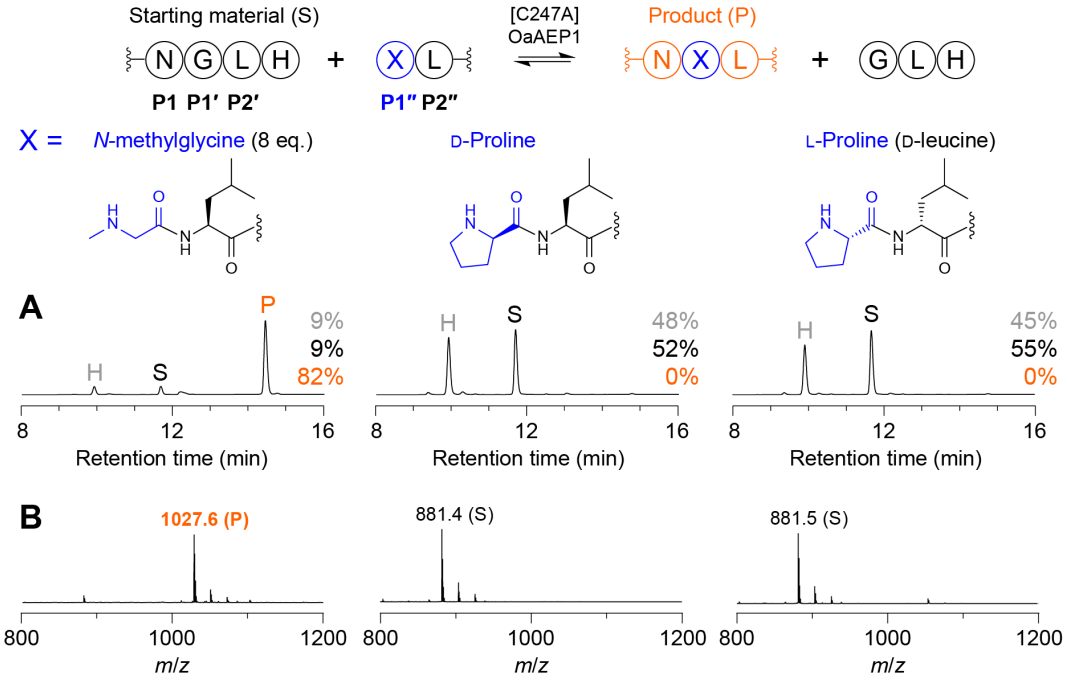


Figure S3. Additional ligation reactions with model peptides. Reactions were run using 100 nM [C247A]*Oa*AEP1, an NGLH-containing substrate (Ac-GWRNGLH, 100 µM), tetrapeptides (XLRL where X = *N*-methylglycine (800 µM, 8 equiv.), D-proline or L-proline (but D-leucine), 200 µM, 2 equiv.) and 300 µM $NiCl_2$ in 100 mM HEPES pH 7.5 (1 h at 25°C). Shown are (A) analytical reverse-phase HPLC traces (A280 nm) and (B) MALDI-TOF MS spectra. Peaks for starting material (S, black), product (P, orange) and hydrolysed substrate (H, grey) are indicated, as well as the % conversion to product or observed mass.

Figure S4. Time course for ligation of GLRL or PLRL to a model NGLH substrate. Reactions were run as described in Figure S1 in the presence (+) or absence (-) of 300 μM $NiCl_2$ (as indicated). At the indicated time points (left panels), reactions were quenched by addition of TFA (1%) prior to analysis. Shown are representative analytical reverse-phase HPLC traces (A280 nm) from three experiments. Peaks for starting material (S, black) and product (P, orange) are indicated.

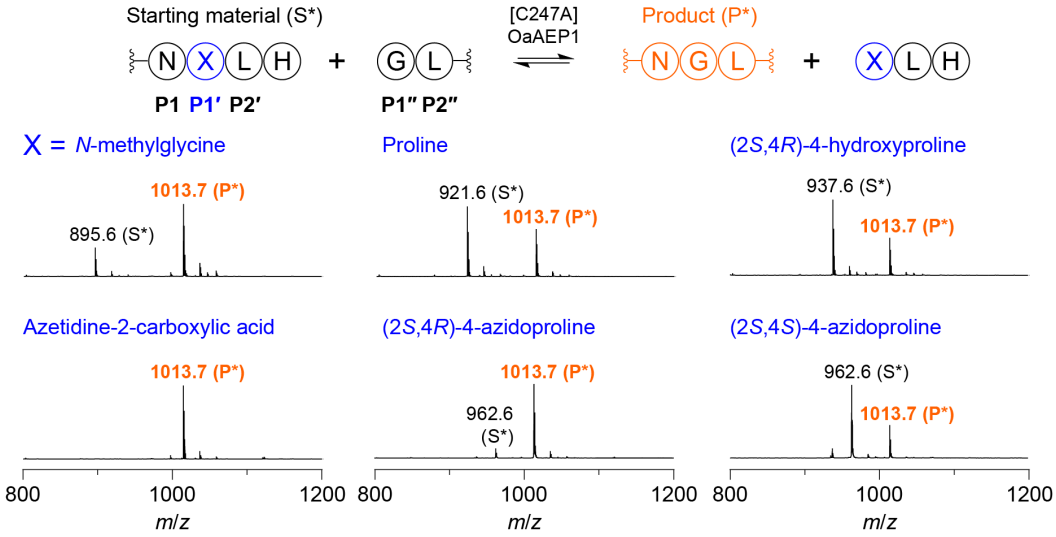


Figure S5. Recognition of product mimetics bearing a tertiary amide bond. Reactions were run using 100 nM [C247A]OaAEP1, NXLH-containing substrates (Ac-GWRNXLH where X = *N*-methylglycine, proline, (2*S*,4*R*)-4-hydroxyproline, azetidine-2-carboxylic acid, (2*S*,4*R*)-4-azidoproline or (2*S*,4*S*)-4-azidoproline, 100 µM), GLRL (200 µM, 2 equiv.) and $NiCl_2$ (300 µM) in 100 mM HEPES pH 7.5 (18 h at 25°C). Shown are MALDI-TOF MS spectra. Peaks for starting material (S*, black) and product (P*, orange) are indicated, as well as the observed masses.

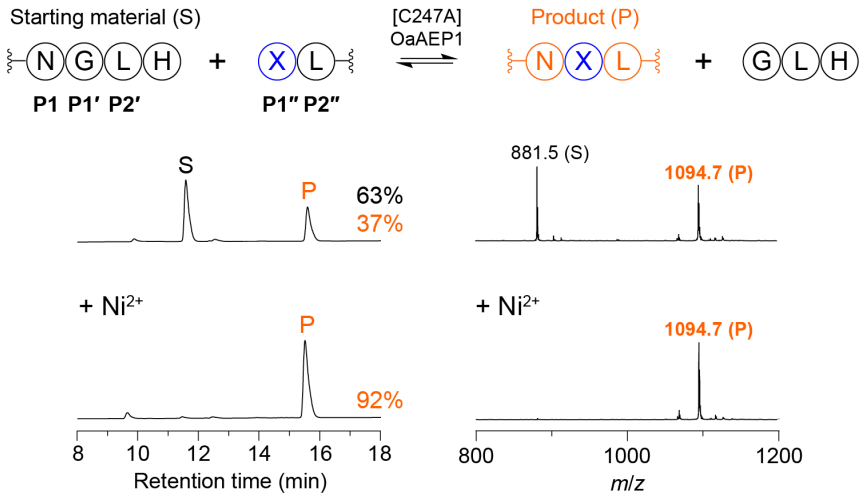


Figure S6. Ligation of (2*S*,4*S*)Azp-LRL to a model NGLH substrate. Reactions were run using 100 nM [C247A]*Oa*AEP1, an NGLH-containing substrate (Ac-GWRNGLH, 100 µM), tetrapeptides (XLRL where X = (2*S*,4*S*)-4-azidoproline, 200 µM, 2 equiv.) and 300 µM $NiCl_2$ in 100 mM HEPES pH 7.5 (1 h at 25°C). Shown are (left panels) analytical reverse-phase HPLC traces (A280 nm) and (right panels) MALDI-TOF MS spectra. Peaks for starting material (S, black) and product (P, orange) are indicated, as well as the % conversion to product or observed mass.

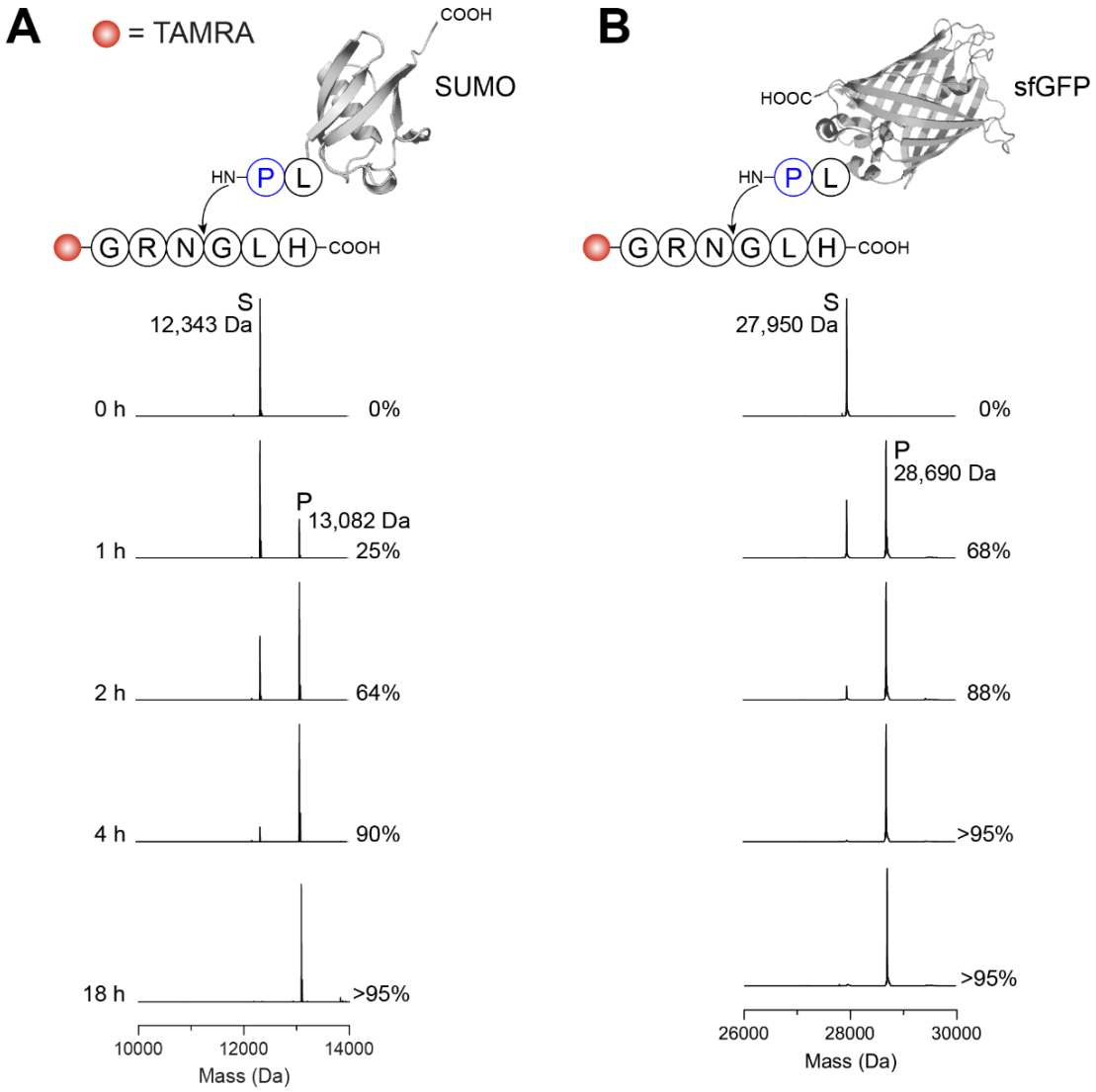


Figure S7. N-terminal protein labelling of (A) PL-SUMO or (B) PL-sfGFP using TAMRA-GRNGLH. Reactions contained 50 µM protein substrate, 500 µM TAMRA-GRNGLH, 2 mM $NiCl_2$ and 200 nM [C247A]*Oa*AEP1 in 100 mM HEPES pH 7 at 25°C and were run for 0, 1, 2, 4 or 18 h, then quenched via TFA addition (2%) prior to analysis by ESI-MS. Shown are reconstructed spectra for the indicated timepoints with the observed substrate (S) and product (P) masses indicated.

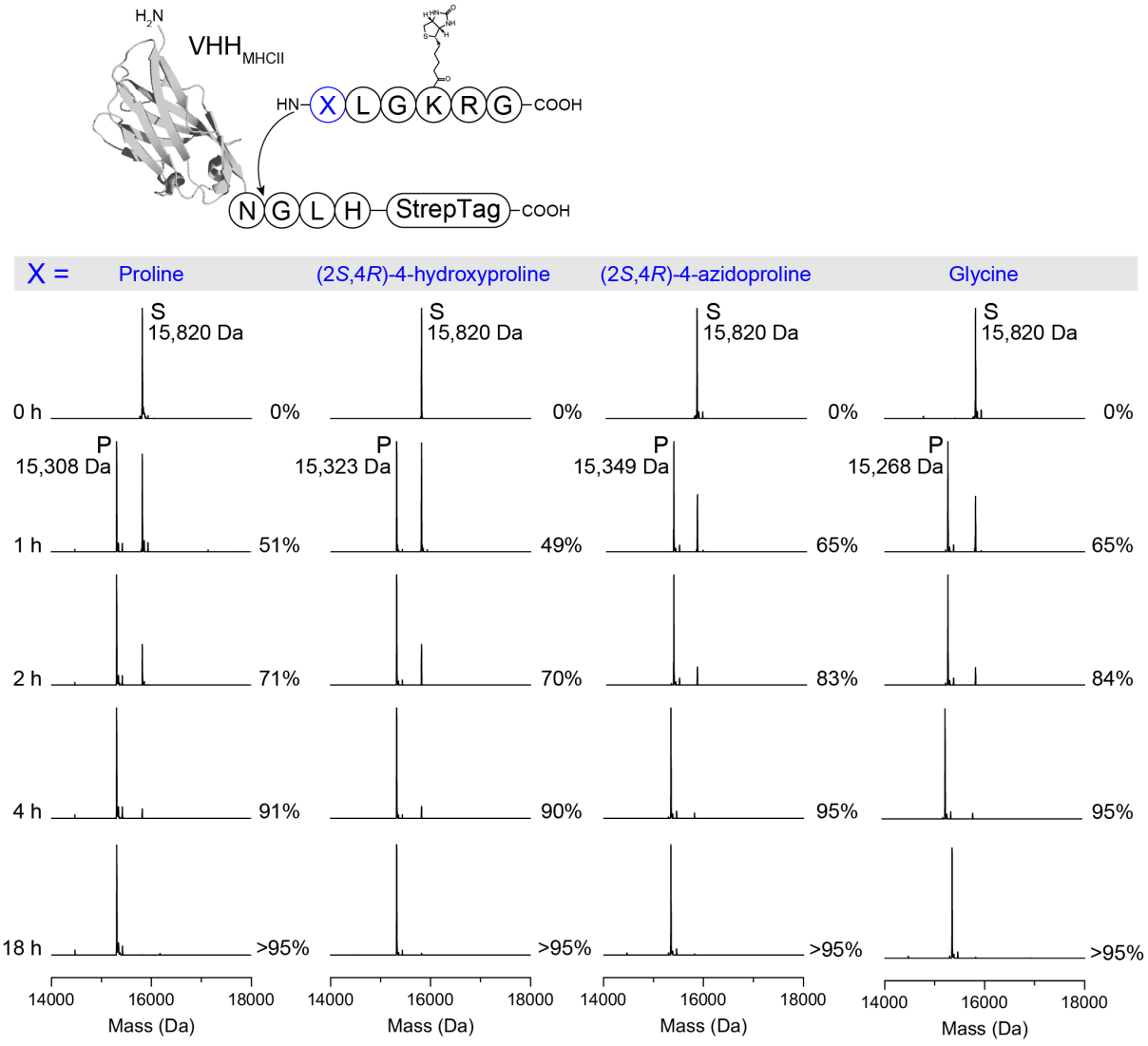


Figure S8. C-terminal protein labelling of VHH_{MHCII} using XLGK(biotin)RG where X = proline, (2*S*,4*R*)-4-hydroxyproline, (2*S*,4*R*)-4-azidoproline or glycine. Reactions contained 50 µM protein substrate, 200 µM XLGK(biotin)RG (X = Pro, Hyp, Azp or Gly as indicated), 250 µM $NiCl_2$ and 500 nM [C247A]*Oa*AEP1 in 100 mM HEPES pH 7 at 25°C and were run for 0, 1, 2, 4 or 18 h, then quenched via TFA addition (2%) prior to analysis by ESI-MS. Shown are reconstructed spectra for the indicated timepoints with the observed substrate (S) and product (P) masses indicated.

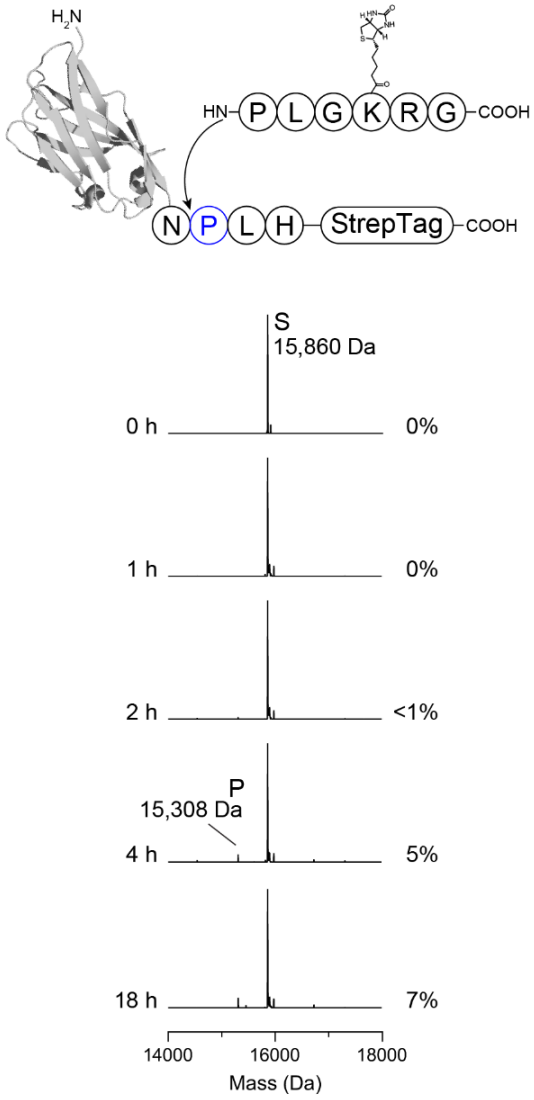


Figure S9. C-terminal protein labelling of VHH_{MHCII} using PLGK(biotin)RG after modifying the AEP recognition sequence to NPLH. Reactions contained 50 µM protein substrate, 200 µM PLGK(biotin)RG, 250 µM $NiCl_2$ and 500 nM [C247A]*Oa*AEP1 in 100 mM HEPES pH 7 at 25°C and were run for 0, 1, 2, 4 or 18 h, then quenched via TFA addition (2%) prior to analysis by ESI-MS. Shown are reconstructed spectra for the indicated timepoints (0 h, 1 h, 2 h, 4 h, and 18 h) with the observed substrate (S) and product (P) masses indicated.

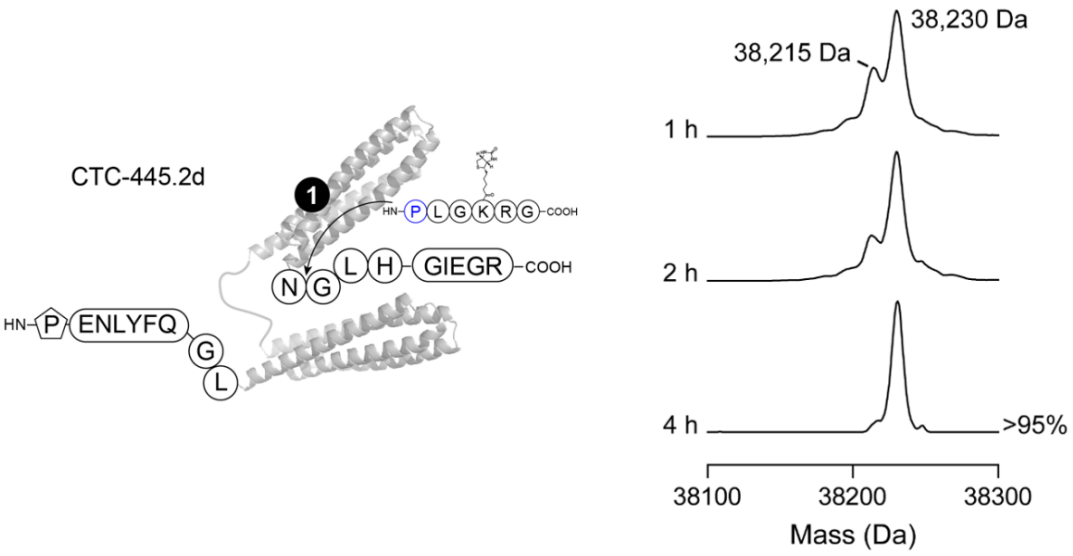


Figure S10. Higher resolution spectra for monitoring C-terminal labelling in the dual labelling reaction shown in Figure 5. The CTC-445.2d protein was C-terminally labelled with a PL-biotin peptide in a reaction comprising 50 µM protein, 500 µM PLGK(biotin)RG, 250 µM $NiCl_2$ and 500 nM [C247A]*Oa*AEP1 in 100 mM HEPES pH 7 at 25°C. At the indicated timepoints (1, 2 or 4 h), samples were quenched via TFA addition (2%) then analysed by ESI-MS. Shown are reconstructed spectra with the observed masses indicated. 38,230 Da is the C-terminally labelled product.

Unprocessed gel images (both images are of the same gel, it was first imaged for fluorescence and then stained with Instant Blue and imaged again)