Supporting Information

Enzymatic C-to-C Protein Ligation

F. B. H. Rehm*, T. J. Tyler, S. J. de Veer, D. J. Craik*, T. Durek*
Supporting Information

Contents:

1. Experimental section page S2
2. Supporting tables page S10
3. Supporting figures page S13
4. Supporting references page S22
1. Experimental section

Peptide synthesis

All reagents were obtained from commercial sources and used as purchased: dichloromethane (DCM), N,N-diisopropylethylamine (DIPEA) and methanol (MeOH) from MiliporeSigma; dimethylformamide (DMF), diethyl ether and acetonitrile from RCI Labscan; trifluoroacetic acid (TFA), piperidine and acetic anhydride from Chem-Supply; PyBOP from Mimotopes; HATU from ChemImpex Int; triisopropylsilane (TIS), biotin, ethylene diamine (Eda) and hydrazine monohydrate from Sigma. All Fmoc-protected amino acids were supplied from CSBio, with the exception of the following unnatural amino acids: Fmoc-β-Ala-OH, Tri-Boc-hydrazoneacetic acid, Fmoc-β-homolysine(Boc)-OH and Fmoc-(S)-3-amino-3-(2-nitrophenyl)propionic acid from ChemImpex Int; Fmoc-d-Val-OH, Fmoc-d-Leu-OH and Fmoc-d-Arg(Pbf)-OH from Mimotopes. Water was purified using a Milli-Q direct water purification system. Unless noted otherwise, all reactions were carried out at room temperature and purified on a Zorbax C18, 7 µm, 21.2 x 250 mm; Agilent column. Following standard Fmoc-solid-phase peptide synthesis (Fmoc-SPPS) protocols, all peptide sequences were synthesized on 2-chlorotriyl chloride resin (ChemImpex Int) at a 125 µmol scale.

General resin loading and capping protocol:

The resin was first swollen in dry DCM for 1 h and then washed with dry DMF. A solution of Fmoc-Xaa-OH (4 eq.) and DIPEA (8 eq.) in DMF was then added to the resin (1 eq., 0.1 M) and agitated. After 12 h, the resin was washed with DMF, DCM and DMF. Unreacted resin sites were then capped by adding a freshly prepared solution of DCM/MeOH/DIPEA (17:2:1, v/v/v) to the resin and agitating for 30 min. Following capping, the resin was then washed with DMF, DCM and DMF.

Synthesis of C-terminal Eda peptides:

The resin was first swollen in dry DCM for 1 h and then washed with dry DMF. The resin was then treated with a freshly prepared solution of 10% Eda (v/v) and DIPEA (2 eq. relative to resin loading) in DMF and agitated overnight. After 12 h, the resin was then washed with DMF, DCM and DMF. Peptide assembly was then performed in accordance with the standard Fmoc-SPPS protocols below.

Synthesis of C-terminal hydrazide peptides:

The resin was first swollen in dry DCM for 1 h and then washed with dry DMF. The resin was then treated with a freshly prepared solution of 5% (v/v) hydrazine monohydrate in DMF (2 x 45 min). The resin was washed with DMF, DCM and DMF. Unreacted resin sites were then capped by adding a freshly prepared solution of 10% (v/v) MeOH in DMF (1 x 10 min). Peptide assembly was then performed in accordance with the standard Fmoc-SPPS protocols below.

Standard peptide assembly (Fmoc-SPPS):

The following general protocol of peptide elongation was followed for canonical and non-canonical amino acids using iterative Fmoc-SPPS, unless stated otherwise:

1. Deprotection

   The resin was treated with 20% piperidine in DMF (2 x 5 min) and washed with DMF, DCM and DMF.

2. General amino acid coupling

   A pre-activated solution of protected amino acid (4 eq.), PyBOP (4 eq.), and DIPEA (8 eq.) in DMF was added to the resin (1 eq., 0.1 M). After 1 h, the resin was washed with DMF, DCM and DMF.
N-terminal amine acetylation:
Following the final amino acid coupling and subsequent Fmoc deprotection, the resin was treated with a solution of DMF, DIPEA and acetic anhydride (17:2:1 v/v/v, 3 mL) (2 x 3 min) and then washed with DMF, DCM and DMF.

Cleavage:
A mixture of TFA/TIS/water (95:2.5:2.5 v/v/v, 3 mL) was added to the resin. After 2 h, the resin was washed with TFA (3 mL).

Work-up:
The combined TFA cleavage solutions were concentrated under a stream of nitrogen gas. The remaining residue was treated with pre-chilled diethyl ether to precipitate the crude peptide, which was subsequently dissolved in water and acetonitrile containing 0.1% TFA, filtered and purified by RP-HPLC with a gradient of 0-50% in 45 min. Fractions containing target product mass at the highest purity (>95%) were collected and lyophilized. Peptide purity was assessed via analytical RP-HPLC (Phenomenex Jupiter 5 μm C18 300 Å, 35 min 0-60% acetonitrile gradient). The list of peptides synthesized, with calculated mass and observed mass, are summarized in the table below.

| Peptide sequences | Calculated [M+H]+ (Da) | Observed [M+H]+ (Da) |
|-------------------|-------------------------|-----------------------|
| Ac-RWRGWRNGLH*    | 1379.7                  | 1379.8                |
| GLRL*             | 458.3                   | 458.3                 |
| (α G)-LRL         | 473.3                   | 473.3                 |
| (β A)-LRL         | 472.3                   | 472.3                 |
| gIrl              | 458.3                   | 458.3                 |
| (β A)(βHo L)(βHo K)(βHo L) | 486.4                   | 486.4                 |
| Ac-GLRL-Hydrazide | 514.3                   | 514.4                 |
| Ac-GLRL-Eda       | 542.4                   | 542.4                 |
| Ac-GLRI-Eda       | 542.4                   | 542.4                 |
| Ac-GLRV-Eda       | 528.4                   | 528.3                 |
| Ac-GLRv-Eda       | 528.4                   | 528.4                 |
| Ac-GLRG-Eda       | 486.3                   | 486.3                 |
| GLHRL-Eda         | 637.4                   | 637.4                 |
| GLHG-Anp-RL-Eda   | 886.5                   | 886.4                 |

h G = hydrazino Gly; αX = beta-amino acid; βHox = beta-homo-amino acid
*Also used by us in previous work[1]
The peptide was assembled following the standard protocols above.

**Ac-RWRGWRNGLH**

The peptide was assembled following the standard protocols above.

**GLRL**

The peptide was assembled following the standard protocols above.

**(h G)-LRL**

The peptide was assembled until the leucine residue following the protocol described above. Following Fmoc deprotection, a pre-activated solution of Tri-Boc-hydrazinoacetic acid (4 eq.), HATU (4 eq.), and DIPEA (8 eq.) in DMF was added to the resin (1 eq., 0.1M) and agitated overnight. After 12 h, the resin was washed with DMF, DCM and DMF, and the crude peptide was cleaved and purified in accordance with the protocols described above.

**(β A)-LRL**

The peptide was assembled until the leucine residue following the protocol described above. Following Fmoc deprotection, a pre-activated solution of Fmoc-β-Ala-OH (4 eq.), HATU (4 eq.), and DIPEA (8 eq.) in DMF was added to the resin (1 eq., 0.1M) and left to agitate. After 1 h, the resin was washed with DMF, DCM and DMF, and the crude peptide was cleaved and purified in accordance with the protocols described above.
The peptide was assembled following the standard protocols above.

\((\beta A)(\beta Ho L)(\beta Ho K)(\beta Ho L)\)

The peptide was assembled following the standard protocols above, except the more reactive coupling reagent HATU was used in the place of PyBOP.

**Ac-GLRL-Hydrazide**

The peptide was assembled following the standard protocols above.

**Ac-GLRL-Eda**

The peptide was assembled following the standard protocols above.

**Ac-GLRL-Eda**

The peptide was assembled following the standard protocols above.
The peptide was assembled following the standard protocols above.

The peptide was assembled following the standard protocols above.

The peptide was assembled following the standard protocols above.

The peptide was assembled following the standard protocols above.
GLHG-Anp-RL-Eda

The peptide was assembled until the arginine residue following the protocol described above. Following Fmoc deprotection, a pre-activated solution of Fmoc-((S)-3-amino-3-(2-nitrophenyl)propionic acid (Anp) (4 eq.), HATU (4 eq.), and DIPEA (8 eq.) in DMF was added to the resin (1 eq., 0.1M) and agitated overnight under protection from the light. After 12 h, the resin was washed with DMF, DCM and DMF, and the crude peptide was cleaved and purified in accordance with the protocols described above, whilst minimising exposure to ambient light.
**Recombinant protein production:**

Nanobody substrates and [C247A]OaAEP1 were produced in *E. coli* SHuffle whereas all other substrates were produced in *E. coli* BL21, essentially as previously described.[1-2] Nanobody expression in the cytoplasm of *E. coli* SHuffle has previously been demonstrated.[3] Plasmids encoding the substrates (pET14b, listed below Table S2) or enzyme (pHUE) were transformed into the appropriate strain and expressed by induction with 0.3 mM IPTG after growth at 37°C (BL21) or 30°C (SHuffle), with 200 rpm shaking, in Luria Bertani broth containing 100 µg/mL ampicillin once OD$_{600}$ of 0.6-0.8 was reached. Following induction, cells were either grown overnight at 18°C (enzyme and non-intein-containing substrates) or at the initial temperature for 8 h (intein-containing substrate precursors). Cells were collected by centrifugation, suspended in lysis buffer (50 mM Na$_2$PO$_4$, 300 mM NaCl, 10 mM imidazole, pH 8 for His-tagged proteins; 100 mM Tris-HCl, 150 mM NaCl, pH 8 for Streptagged proteins; 20 mM HEPES, 50 mM NaOAc, 100 mM NaCl, pH 6.5 for intein-containing chitin binding domain-carrying proteins), lysed at 32 kpsi with a Constant Systems cell disruptor, and clarified via centrifugation at 40,000 x g for 30 min. Soluble protein was then bound to NiNTA agarose beads (Qiagen), StrepTactin Superflow resin, or chitin resin (NEB) and washed (50 mM Na$_2$PO$_4$, 300 mM NaCl, 20 mM imidazole, pH 8 for His-tagged proteins; 100 mM Tris-HCl, 150 mM NaCl, pH 8 for Streptagged proteins; 20 mM HEPES, 50 mM NaOAc, 100 mM NaCl, pH 6.5 for intein-containing chitin binding domain-carrying proteins) before elution (50 mM Na$_2$PO$_4$, 300 mM NaCl, 250 mM imidazole, pH 8 for His-tagged proteins; 100 mM Tris-HCl, 150 mM NaCl, 2.5 mM desthiobiotin, pH 8 for Streptagged proteins; 20 mM HEPES, 50 mM NaOAc, 100 mM NaCl, 100 mM MESNa, pH 6.5 with on-column incubation at 25°C for 18 h for intein-containing chitin binding domain-carrying proteins). Eluates were further purified via size exclusion chromatography (SEC) on a HiLoad 16/600 Superdex 75 pg column equilibrated with 20 mM HEPES, 150 mM NaCl, pH 7 (substrates) or 20 mM Tris-HCl buffer containing 100 mM NaCl and 10% glycerol (pH 8) (enzyme). TEV protease (NEB) cleavage of eGFP bearing a C-terminal ENLYFQ-H6 sequence was conducted according to the manufacturer’s instructions prior to rebinding to NiNTA and SEC of the unbound, cleaved fraction. Aminolysis of protein-MES thioesters generated by intein splicing was conducted by buffer exchanging the chitin column eluates into aminolysis buffer (20 mM HEPES, 50 mM NaOAc, 100 mM NaCl, 20 mM MESNa, 500 mM Eda, pH 7.5) and incubating at 25°C for at least 6-8 h, as assessed by ESI-MS, prior to SEC as above. Auto-activation of the SEC-purified enzyme was carried out via addition of acetic acid (~1:500 v/v acid:buffer, to pH 4.0-4.5) and incubation at 25°C for 18 h. NanoDrop A$_{280}$ readings, using calculated protein extinction coefficients and molecular weights, were used to determine protein substrate and enzyme concentrations. Where necessary, higher concentrations were achieved using Amicon spin columns. All proteins were stored at -80°C until use.

**Analysis of peptide ligation reactions:**

The various ligation reactions of substrate/substrate mimetics (2 eq. for native substrate, 5 eq. for substrate mimetics, 10 eq. for the bifunctional linker peptide) to the model NGLH-containing substrate (Ac-RWGRWRNGGLH, 100 µM) as catalyzed by 100 nM [C247A]OaAEP1 with/without NiSO$_4$ (2 mM, or 5 mM for the bifunctional linker peptide; all NiSO$_4$ solutions were pre-adjusted to the reaction pH) were run in 100 mM Tris-HCl, pH 7.5, at 25°C until they were quenched via TFA addition to 1% (v/v). Before addition of enzyme, the bifunctional linker peptide reactions containing NiSO$_4$ were incubated at 25°C for at least 15 min. The quenched solutions were then analyzed by analytical RP-HPLC (Phenomenex Jupiter 5 µm C18 300 Å, 15 min 20-50% acetonitrile gradient) and MALDI-TOF MS (SCIEX 5800 MALDI–TOF–MS, α-Cyano-4-hydroxycinnamic acid matrix).

S8
**C-terminal protein labeling:**

Proteins bearing a C-terminal **NGLH** sequence (50 µM) were incubated with 500 µM substrate mimetic peptide and 400 nM [C247A]OaAEP1 in 100 mM HEPES pH 7, 25°C before being quenched via TFA addition to 1% (v/v). These reactions were then loaded onto a Zorbax 300SB18 column (Agilent) and eluted over a 15 min 1-50% acetonitrile gradient using a Shimadzu Nexera X2 LC system. The liquid chromatography outflow was connected to a 5600 Triple TOF mass spectrometer (SCIEX) (Turbo V ion source) or a QSTAR Elite (SCIEX). Reconstructed spectra were generated in the Analyst software (SCIEX) (other proteins were analyzed in the same way).

**Linker-based protein-protein C-to-C ligations:**

Proteins bearing a C-terminal **NGLH** sequence (50 µM) were incubated with 500 µM bifunctional linker peptide and 5 mM NiSO$_4$ in 100 mM HEPES pH 7, at 25°C for 15 min, before addition of 800 nM [C247A]OaAEP1 and further incubation for 2 h or until the reaction had reached completion as assessed by ESI-MS (as described in the above section). Subsequently, excess linker peptide and NiSO$_4$ were removed from the reaction via buffer exchange into phosphate buffered saline (NAP5 column), and Ni$^{2+}$ was removed from the peptide-metal complex by 5-fold diluting with reaction buffer that had been adjusted to pH 4.5 and incubating the solution on ice for 30 min. Another buffer exchange step into PBS was then conducted and the resulting protein-linker conjugates were concentrated (Amicon) as needed for protein-linker-protein ligation. For protein-linker-protein ligations, the protein-linker substrate was supplied at 10-20 µM, another **NGLH**-bearing protein at 20-40 µM, and [C247A]OaAEP1 at 100 nM and reactions were run in 100 mM HEPES buffer, pH 7, at 25°C. Reaction progress was analyzed on 4-12% SDS-PAGE gradient gels stained with InstantBlue. To cleave photocleavable protein-linker-protein conjugates (Anp-containing bifunctional linker), reactions were diluted 10-fold with PBS and transferred into clear Agilent glass vials before being irradiated by UV for 10 min on a Bio-Rad GelDoc.

**Direct C-to-C protein-protein ligations:**

Reactions containing protein-L-Eda (30-40 µM) and protein-**NGLH** (60 µM) substrates were incubated with 400 nM [C247A]OaAEP1 in 100 mM HEPES buffer, pH 7, with or without NiSO$_4$ (2 mM) for 1-2 h at 25°C before analysis by SDS-PAGE as described in the above section. For the GLH-SUMO-L-Eda reaction supplemented with NiSO$_4$, the reaction was incubated for 15 min prior to enzyme addition.

**Validation of bifunctional linker orientation:**

Reactions containing 100 µM Ac-RWRGWR**NGLH**, 1 mM GLHRL-Eda, and 200 nM [C247A]OaAEP1 in 100 mM pH 7.5 Tris-HCl with/without 5 mM NiSO$_4$ were run for 1 h at 25°C and then 2-fold diluted with ammonium bicarbonate (100 mM, pH 8). Sequencing grade trypsin (Promega) was then added at 2.5 ng/µL. Following 16 h digestion at 37°C, reactions were quenched via addition of 1% (v/v) formic acid and run on a Zorbax 300SB-C18 column (Agilent; 75 min 1-80% acetonitrile gradient) interfaced with a 5600 Triple TOF mass spectrometer (SCIEX) equipped with a Turbo V ion source (SCIEX). MS/MS data were acquired in Information Dependent Acquisition experiments (m/z range of 80-1800 Da; 250 ms accumulation time). The ‘Rolling collision energy’ option was enabled, allowing for dynamic variation in collision energy based on precursor ion m/z and z. Extracted ion chromatograms were generated in the Analyst software (SCIEX).
2. Supporting tables

**Table S1.** Observed and calculated monoisotopic masses of the model NGLH-containing peptide substrate (Ac-RWGRW/RNGLH) and the conjugate products.

| Peptide conjugate sequences | Calculated [M+H]^+ (Da) | Observed [M+H]^+ (Da) |
|----------------------------|-------------------------|-----------------------|
| Unmodified substrate       | 1379.7                  | 1379.8                |
| Hydrolysis                 | 1072.5                  | 1072.6                |
| GLRL                       | 1511.8                  | 1511.9                |
| (h,G)-LRL                  | 1526.8                  | 1527.0                |
| (β A)-LRL                  | 1525.9                  | 1526.0                |
| gIrl                       | 1511.8                  | 1511.9                |
| (β A)(βho L)(βho K)(βho L) | 1539.9                  | 1540.0                |
| Ac-GLRL-Hydrazide          | 1567.9                  | 1568.0                |
| Ac-GLRL-Eda                | 1595.9                  | 1596.1                |
| Ac-GLRI-Eda                | 1595.9                  | 1596.1                |
| Ac-GLRV-Eda                | 1581.9                  | 1582.1                |
| Ac-GLRv-Eda                | 1581.9                  | 1582.1                |
| Ac-GLRG-Eda                | 1539.8                  | 1540.0                |
| GLHRL-Eda (peptide-linker) | 1691.0                  | 1690.9                |
| GLHRL-Eda (peptide-linker-peptide) | 2745.5 | 2745.3 |

\(^{h}G = \) hydrazino Gly; \(^{β}X = \) beta-amino acid; \(^{βho}X = \) beta-homo-amino acid
Table S2. Calculated and observed average masses of the protein substrates and conjugates.

| Protein               | Calculated* average mass (Da) | Observed mass (Da) |
|-----------------------|-------------------------------|--------------------|
| eGFP-NGLH            | 28,248                        | 28,250             |
| VHHfer-NGLH          | 13,660                        | 13,661             |
| eGFP-N-grl           | 27,586                        | 27,589             |
| VHHfer-N-grl         | 12,753                        | 12,754             |
| eGFP-N-(βA)(βHo L)(βHo K)(βHo L) | 27,614                  | 27,617             |
| VHHfer-N-(βA)(βHo L)(βHo K)(βHo L) | 12,781                  | 12,782             |
| eGFP-N-Eda-LRL-Ac    | 27,670                        | 27,671             |
| VHHfer-N-Eda-LRL-Ac  | 12,837                        | 12,838             |
| eGFP-N-Eda-LRHLG-(N-term.) | 27,765               | 27,767             |
| eGFP-N-Eda-LR-Anp-GHLG-(N-term.) | 28,014               | 28,016             |
| VHHfer-N-Eda-LR-Anp-GHLG-(N-term.) | 13,181               | 13,183             |
| SUMO-L-Eda           | 11,946                        | 11,945             |
| VHHfer-L-Eda         | 12,585                        | 12,584             |
| VHHMHCII-NGLH        | 15,820                        | 15,819             |
| GLH-SUMO-L-Eda       | 12,157                        | 12,157             |

β X = beta-amino acid; βHo X = beta-homo-amino acid

*Calculated masses assume removal of the N-terminal Met residue.

Reconstructed ESI-MS spectra of the unmodified NGL-containing proteins:

Full sequences of recombinant protein substrates produced in this study (final sequence after processing/cleavage underlined):

eGFP-NGLH:

MPVSKGEELFTGVVPILVELDGDVNGHKFSVSGEGEDATGKLTLKICTTGKLKPWPTLVTTLTYGVCQRSYRPDHMKQHDFFMAMQEGYVQERTIFKDDGNYKTRAERVKFGDTHVNRIELKGEDG
NILGHKLFEYNNSHNVYIMADKQNGKVNFTKTNIEDGSVQLNYQQNTPIGDPVLLFPDNYLS
TQSALSKEKDPRHDMLLEFVTAAITLGMDELKYSNFLHENLYFQGHHHHHH
VHH<sub>6e</sub>-NGLH:

MPQVLQESGGGLVQPGGLRLSCAASGFVFENSAWAMWRQAPGKERELEIAYGTIFIKLASVESVKGFTISRDNAKSTVYLQMNNKLKPEDTAYYCSKSGAYWQGTQVTGVTSSGGNGLHWSHPQFEK

SUMO-L-intein:

MPGSGSGMSDSeVQEAKEPKVPEKETHINLKVSDDSSIEFFKIKKTTPLRRLMEAFAKRQQKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIIGGLCITGDALVALPEGESVRIADIVPGarpNSDNAIDLKVLRHGNPVLADRFHSEHPVYTVRTVEGLRVTGTANHPPLLCLVDVAGVPITLWLKLIIDEIKPGDYAVIQRSAFSVDCAGFARGKPEAFPTTYTVGVPGLVRFLEAHRRDPDQAIADELTVDGFYYAKVASVTDAGQPVYSLRVTDADHAPFNGFVSHATGLTGLNSGLTTNPVGSAWQVTNTAYTAGQLYTVNGKTYKCLQPHTSLAGWFSNPALWQLQ

VHH<sub>6e</sub>-L-intein:

MPQVLQESGGGLVQPGGLRLSCAASGFVFENSAWAMWRQAPGKERELEIAYGTIFIKLASVESVKGFTISRDNAKSTVYLQMNNKLKPEDTAYYCSKSGAYWQGTQVTGVTSSGGNGLHWSHPQFEK

VHH<sub>MHcII</sub>-NGLH:

MPQVLQESGGGLVQAGDSRLSCAASGRFSRGWMGFRRAPKEREFVAIFSGSSWSGRSTYSDSVKGFRTISRDNAKNTVYLQMNGKLKPEDTAAYCAAGYPEASYGRESTDYWGQGTQVTGVTSSGGNGLHWSHPQFEK

GLH-SUMO-L-intein:

MGLHSSGSGMSDSeVQEAKEPKVPEKETHINLKVSDSSEIFFKIKKTTPLRRLMEAFAKRQQKEMDSLRFLYDGIRIQADQTPEDLDMEDNDIIEAHREQIIGGLCITGDALVALPEGESVRIADIVPGarpNSDNAIDLKVLRHGNPVLADRFHSEHPVYTVRTVEGLRVTGTANHPPLLCLVDVAGVPITLWLKLIIDEIKPGDYAVIQRSAFSVDCAGFARGKPEAFPTTYTVGVPGLVRFLEAHRRDPDQAIADELTVDGFYYAKVASVTDAGQPVYSLRVTDADHAPFNGFVSHATGLTGLNSGLTTNPVGSAWQVTNTAYTAGQLYTVNGKTYKCLQPHTSLAGWFSNPALWQLQ
Figure S1. RP-HPLC (280 nm) analysis of the crude reaction mixtures as described in Figure 1 B, with/without Ni^{2+}, as indicated. The chromatograms for reactions that included Ni^{2+} are the same as those shown in Figure 1 B. % conversion to product (*) is shown, as calculated from relative peak integrals.
Figure S2. MALDI-TOF MS traces of the reactions shown in Figure S1 and Figure 1 B. The observed monoisotopic product (*) masses are indicated. For the reaction with the hydrazino Gly analogue, a low intensity -17 Da product mass is observed and could indicate succinimide formation from Asn.
Figure S3. Synthesis scheme for the preparation of peptides bearing a C-terminal Leu-Eda motif. DIPEA is added at 2 eq. relative to resin loading.
Figure S4. RP-HPLC (280 nm) analysis of the crude reaction mixtures as described in Figure 2B, with/without Ni^{2+}, as indicated. The chromatograms for reactions that included Ni^{2+} are the same as those shown in Figure 2B. % conversion to product (*) is shown, as calculated from relative peak integrals. As the P2'' Gly product partially co-elutes with the peptide-NGLH substrate, the peaks were manually split when integrated and, thus, likely present an overestimation of conversion to product.
Figure S5. MALDI-TOF MS traces of the reactions shown in Figure S4 and Figure 2 B. The observed monoisotopic product (*) masses are indicated.
Figure S6. A) MALDI-TOF MS traces of the reactions shown in Figure 3 A. The observed monoisotopic product (*) masses are indicated. As 10 eq. of the bifunctional linker peptide (GLHRL-Eda) was used in the reactions, Ni\textsuperscript{2+} addition functions only to control conjugation orientation without substantially impacting the extent of conversion to product; B) MALDI-TOF MS traces of the ligation of the model substrate Ac-RWRGWRNGLH (400 µM) to the bifunctional linker peptide (100 µM) as catalyzed by 200 nM [C247A]OaAEP1 with/without (as indicated) NiSO\textsubscript{4} addition (10 mM) in 100 mM HEPES, pH 7.5, for 1 h. Spectra focused on the peptide-linker-peptide product (calculated [M+H]\textsuperscript{+} = 2745.5 Da) are shown on the right. Reactions containing Ni\textsuperscript{2+} were incubated for 15 min prior to enzyme addition; C) Reactions as in Figure 3 A were digested with trypsin and analyzed by ESI-MS. Shown are extracted ion chromatograms of the unique N-to-C and C-to-C product-derived peptides; D) MS/MS spectrum of the unique N-to-C product-derived peptide from (C).
Figure S7. Reconstructed ESI-MS spectra for the protein-linker conjugates generated in Figure 3.
**Figure S8.** SDS-PAGE analysis of eGFP-linker-eGFP C-to-C fusion protein formation over time when the linker attachment orientation was initially uncontrolled or controlled, without or with Ni$^{2+}$, respectively.
Figure S9. Reconstructed ESI-MS spectra for the protein-Leu-Eda substrates generated in Figure 4. A small amount of thioester hydrolysis (as indicated) rather than Eda aminolysis was observed, culminating in a -42 Da mass relative to product.
4. Supporting references

[1] F. B. H. Rehm, T. J. Tyler, K. Yap, T. Durek, D. J. Craik, *Angew Chem Int Ed Engl* **2021**, *60*, 4004-4008.

[2] K. Yap, J. Du, F. B. H. Rehm, S. R. Tang, Y. Zhou, J. Xie, C. K. Wang, S. J. de Veer, L. H. L. Lua, T. Durek, D. J. Craik, *Nat Protoc* **2021**, *16*, 1740-1760.

[3] B. Billen, C. Vincke, R. Hansen, N. Devoogdt, S. Muyldermans, P. Adriaensens, W. Guedens, *Protein Expr Purif* **2017**, *133*, 25-34.