Supporting information

Semisynthesis of an Active Enzyme by Quantitative Click Ligation

Christian Kofoed, Stephan Riesenberg, Jaroslava Šmolíková, Morten Meldal, and Sanne Schoffelen[†]*

Center for Evolutionary Chemical Biology, Department of Chemistry, University of Copenhagen, Universitetsparken 5, 2100 Copenhagen, Denmark

Table of contents

Supporting figures

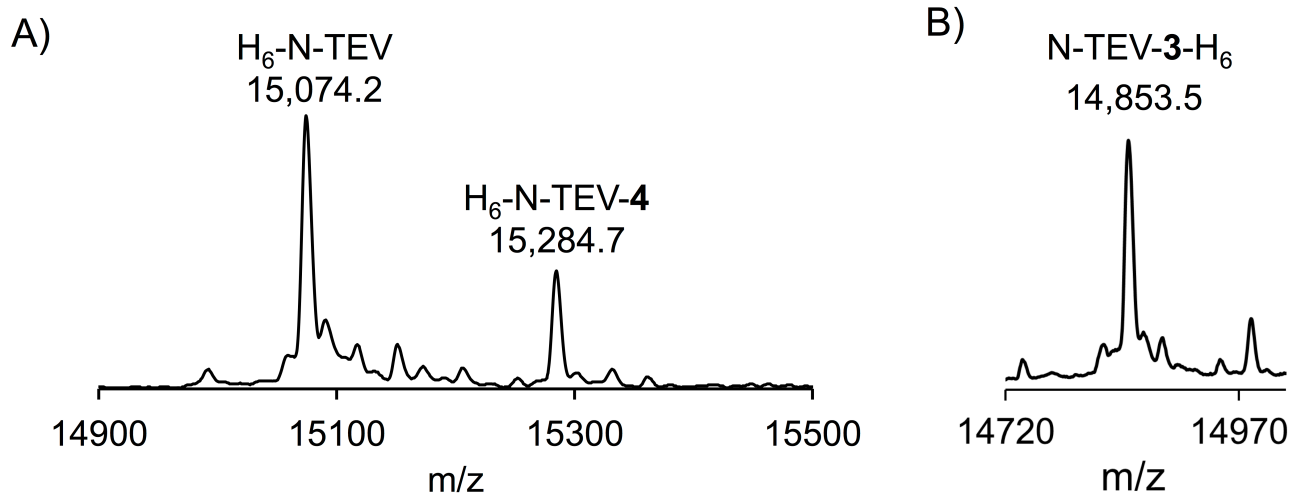


Figure S1. A) Deconvoluted mass spectrum of N-TEV with an N-terminal His-tag. The DNA encoding for this variant of N-TEV contained an amber (TAG) stop codon directly followed by an ochre (TAA) stop codon. ncAA **4** was incorporated via the amber suppression method. The purified sample was a mixture of the desired product (H_6-N-TEV-**4**) and pre-terminated protein (H_6-N-TEV) lacking **4**. Based on the relative abundance of both peaks, the amount of pre-terminated protein is ~70%. This undesired phenomenon of non-functionalized N-TEV being present in the purified sample was overcome by installing the H_6-tag at the C-terminus of N-TEV. B) Deconvoluted mass spectrum of N-TEV-**3**-H_6. This is the same spectrum as shown in Fig. 2B.

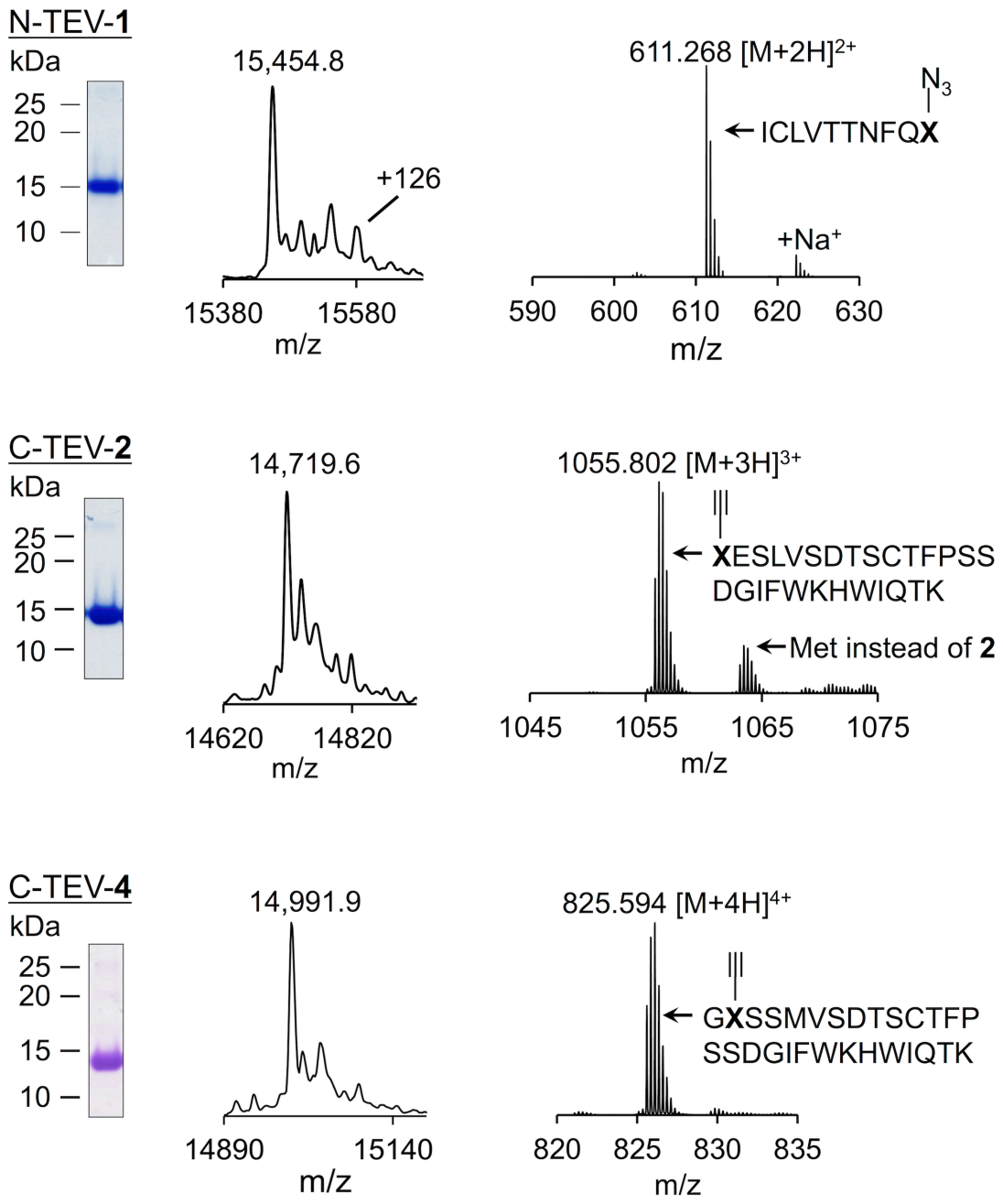


Figure S2. Characterization of purified fragments N-TEV-**1**, C-TEV-**2** and C-TEV-**4**. For each fragment, an SDS-PAGE gel with purified protein is shown, as well as a deconvoluted spectrum of the intact protein and a mass spectrum of the relevant tryptic fragment containing the respective ncAA (referred to as **X** in all cases). Calculated masses of intact proteins: N-TEV-**1** = 15,455.4 Da, C-TEV-**2** = 14,721.1 Da, C-TEV-**4** = 14,992.5 Da. Calculated masses of tryptic fragments (all modified with carbamidomethyl on Cys): ICLVTTNFQ-**1** = 1221.6045 Da, **2**-ESLVSDTSCTFPSSDGIFWKHIQTK = 3165.4986 Da, G-**4**-SSMVSDTSCTFPSSDGIFWKHIQTK = 3299.5136 Da. Incomplete removal of the N-terminal azidohomoalanine residue in N-TEV-**1** was observed in the mass spectrum of the intact N-TEV-**1** (+126 Da). Incorporation of azidohomoalanine (**1**) in N-TEV-**1** proceeded with 100% efficiency as concluded from the fact that no mass for the tryptic fragment containing methionine instead of **1** (calculated mass shift: +5 Da) was detected. Partial incorporation of methionine instead of homopropargyl glycine (**2**) in C-TEV-**2** had occurred as concluded from the presence of the peak corresponding to the tryptic fragment containing methionine instead of **2** (calculated mass shift: +22 Da). The difference in the efficiency of replacement of methionine with either **1** or **2** is in line with previous observations in our laboratory (data not shown). The additional peaks in the mass spectra of the intact proteins correspond to oxidation (+16) and multiple acetonitrile adducts (+42, +64 and +83, which are common ESI+ adducts). See Figure 2B for the characterization of N-TEV-**3**.

TEV[**1**-Tz-**2**]

Fragment	Sequence	m/z calc.	m/z detected
[17-27]	GSGESLFKGPR	1134.5902	1134.5917
[28-63]	DYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNK	3969.8964	3969.8887
[86-104]	NTTTLQQHLIDGRDLIIIR	2220.2459	2220.2392
[158-175]	HWIQTKDGQCGSPLVSTR	2070.0185	2070.0131
[164-175]	DGQCGSPLVSTR	1276.5950	1276.5942
[164-200]	DGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPK	3987.8930	3987.8896
[176-200]	DGFIVGIHSASNFTNTNNYFTSVPK	2730.3158	2730.3128
[176-219]	DGFIVGIHSASNFTNTNNYFTSVPKNFAELLTNQEAQQWVSGWR	4988.4027	4988.4038
[201-219]	NFAELLTNQEAQQWVSGWR	2277.1047	2277.1018
[220-231]	LNADSVLWGGHK	1296.6695	1296.6667
[220-245]	LNADSVLWGGHKVFAVKPEEPFQPVK	2892.5407	2892.5281
[232-257]	VFAVKPEEPFQPVKEATQLLNELVPR	2978.6350	2978.6275

```
         10          20          30          40          50          60          70
ASSHHHHHHS SGLVPRGSGE SLFKGPRDYN PISSTICHLT NESDGHTTSL YGIGFGPFII TNKHLFRRNN

         80          90         100         110         120         130         140
GTLLVQSLHG VFKVKNTTTL QQHLIDGRDL IIIRLPKDFP PFPQKLKFRE PQREERICLVTTNFQX‖XESL

        150         160         170         180         190         200         210
VSDTSCTFPS SDGIFWKHWI QTKDGQCGSP LVSTRDGFIV GIHSASNFTN TNNYFTSVPK NFAELLTNQE

        220         230         240         250         260      268
AQQWVSGWRL NADSVLWGGH KVFAVKPEEP FQPVKEATQL LNELVPRGSG SSHHHHHH
```

TEV[**3**-Tz-**4**]

Fragment	Sequence	m/z calc.	m/z detected
[10-45]	DYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNK	3969.8964	3969.8935
[66-86]	VKNTTTLQQHLIDGRDMIIIR	2465.3657	2465.3658
[68-86]	NTTTLQQHLIDGRDMIIIR	2238.2023	2238.2021
[151-168]	HWIQTKDGQCGSPLVSTR	2070.0185	2070.0088
[157-168]	DGQCGSPLVSTR	1276.5950	1276.5944
[169-193]	DGFIVGIHSASNFTNTNNYFTSVPK	2730.3158	2730.3136
[169-212]	DGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGWR	5048.4061	5048.4120
[194-212]	NFMELLTNQEAQQWVSGWR	2337.1081	2337.1093
[213-250]	LNADSVLWGGHKVFMVKPEEPFQPVKEATQLMNELVPR	4334.2464	4334.2509
[213-261]	LNADSVLWGGHKVFMVKPEEPFQPVKEATQLMNELVPRGSGSSHHHHHH	5531.7389	5531.7257
[225-250]	VFMVKPEEPFQPVKEATQLMNELVPR	3056.5948	3056.5924
[225-261]	VFMVKPEEPFQPVKEATQLMNELVPRGSGSSHHHHHH	4254.0872	4254.0837

```
         10          20          30          40          50          60          70
 GESLFKGPRD YNPISSTICH LTNESDGHTT SLYGIGFGPF IITNKHLFRR NNGTLLVQSL HGVFKVKNTT

         80          90         100         110         120         130         140
 TLQQHLIDGR DMIIIRMPKD FPPFPQKLKF REPQREERIC LVTTNXLVPR GSHHHHHH‖GXSSMVSDTSCT

        150         160         170         180         190         200         210
 FPSSDGIFWK HWIQTKDGQC GSPLVSTRDG FIVGIHSASN FTNTNNYFTS VPKNFMELLT NQEAQQWVSG

        220         230         240         250         260
WRLNADSVLW GGHKVFMVKP EEPFQPVKEA TQLMNELVPR GSGSSHHHHH H
```

Figure S3. Overview of tryptic fragments detected by MS analysis of TEV[**1**-Tz-**2**] and TEV[**3**-Tz-**4**]. Detected fragments are highlighted in gray in the protein sequences provided below the tables. The symbol '||' represents the end of the N-terminal (N-TEV-**1** or N-TEV-**3**) and start of the C-terminal fragment (C-TEV-**2** or C-TEV-**4**).

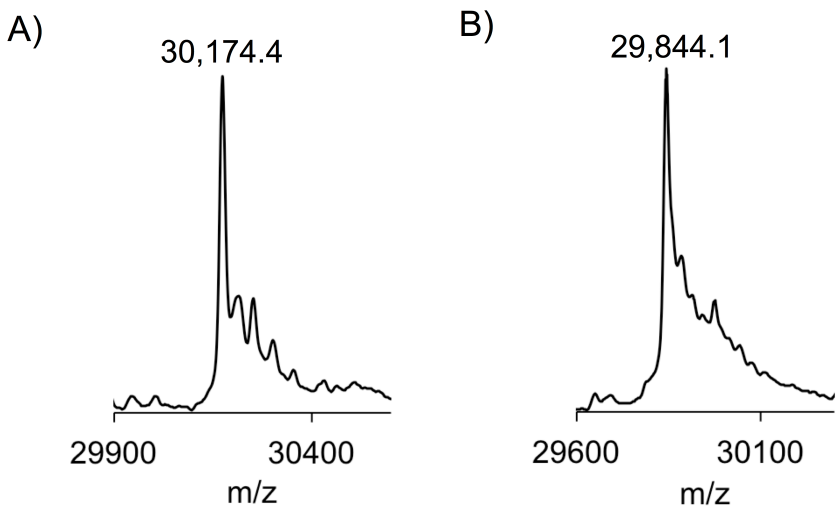


Figure S4. Deconvoluted mass spectra of TEV[**1**–Tz–**2**] and TEV[**3**–Tz–**4**]. A) TEV[**1**–Tz–**2**], calculated mass: 30,176.5 Da. B) TEV[**3**–Tz–**4**], calculated mass: 29,845.5 Da.

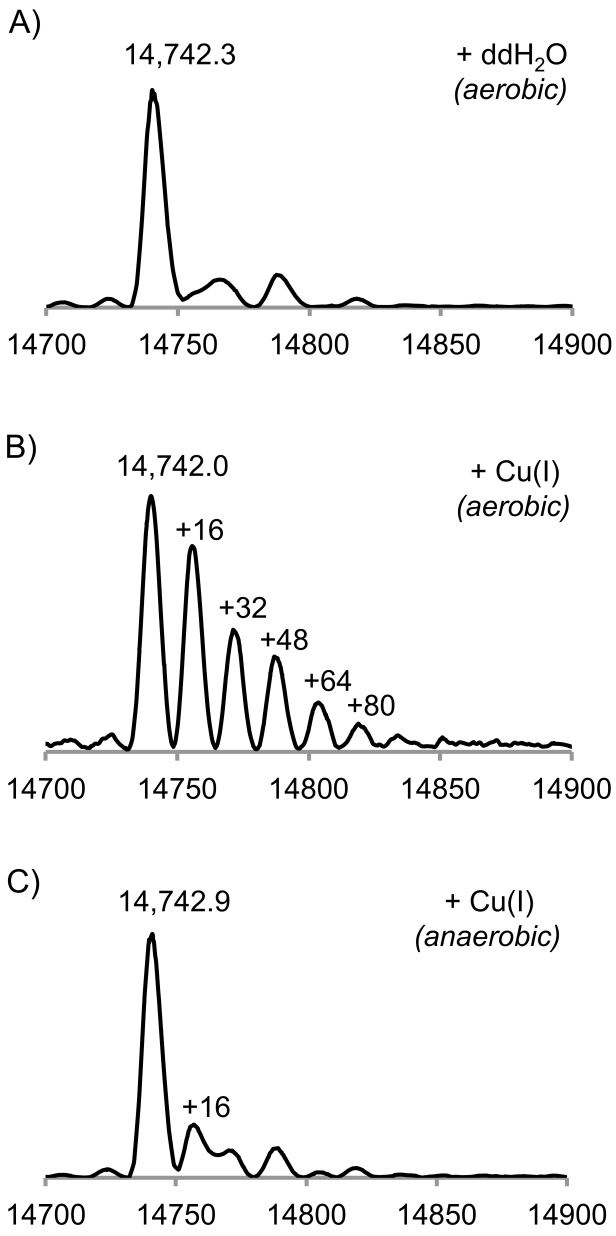
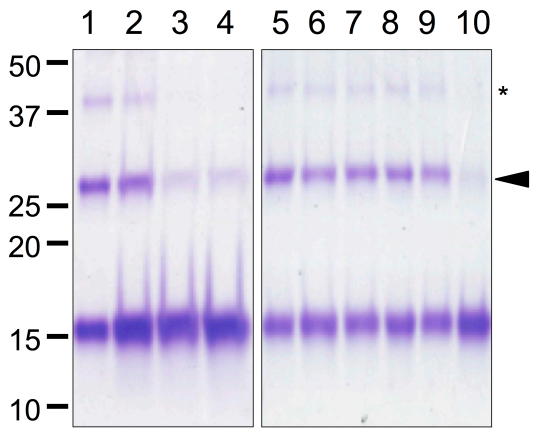


Figure S5. Effect of the employed CuAAC reaction conditions on protein oxidation. C-TEV was mixed with either water (A) or an aqueous solution of 250 μM CuSO4 and 3.75 mM sodium ascorbate (B and C). The samples were left at room temperature for 1 hour before being analyzed by ESI-TOF MS. A) Negative control showing no significant signs of oxidation. B) The sample was incubated under aerobic conditions in a closed tube resulting in multiple levels of oxidation. C) Thorough deaeration in standard test tubes was performed prior to mixing of the sample, and anaerobic conditions were maintained during the entirety of the incubation period, significantly suppressing the oxidation.



#	urea (M)	NaCl (mM)	$CuSO_4$ (μM)	cosolvent (40% v/v)	product (%)
1	2	150	250	-	31
2	4	150	250	-	23
3	6	150	250	-	7
4	7	150	250	-	7
5	4	0	500	-	37
6	4	300	250	-	27
7	4	150	250	-	32
8	4	0	250	-	24
9	4	0	250	MeOH	26
10	4	0	250	MeCN	6

Figure S6. Conjugation of N-TEV-**1** to C-TEV-**2** using various buffer compositions. The arrow indicates the product. The amount of product formed was estimated by densitometric analysis using ImageJ software. The third protein band (*) between 37 and 50 kDa is hypothesized to correspond to a conjugate consisting of two C-TEV-**2** chains attached to one N-TEV-**1** molecule. Due to incomplete processing of N-TEV-**1** by bacterial methionine aminopeptidase (MetAP), 10% of N-TEV-**1** contained two azide moieties, as observed by mass spectrometry (see Figure S2). The MetAP processing step, which constitutes the hydrolytic cleavage of the N-terminal methionine residue, corresponding to the start codon ATG, takes place in *E. coli* when the second residue is small in size.[1] The process is known to occur even if the N-terminal methionine is replaced by **1** as long as Ala is present at the second position.[2] Nevertheless, in our case, cleavage of the N-terminal azidohomoalanine was found to be incomplete.

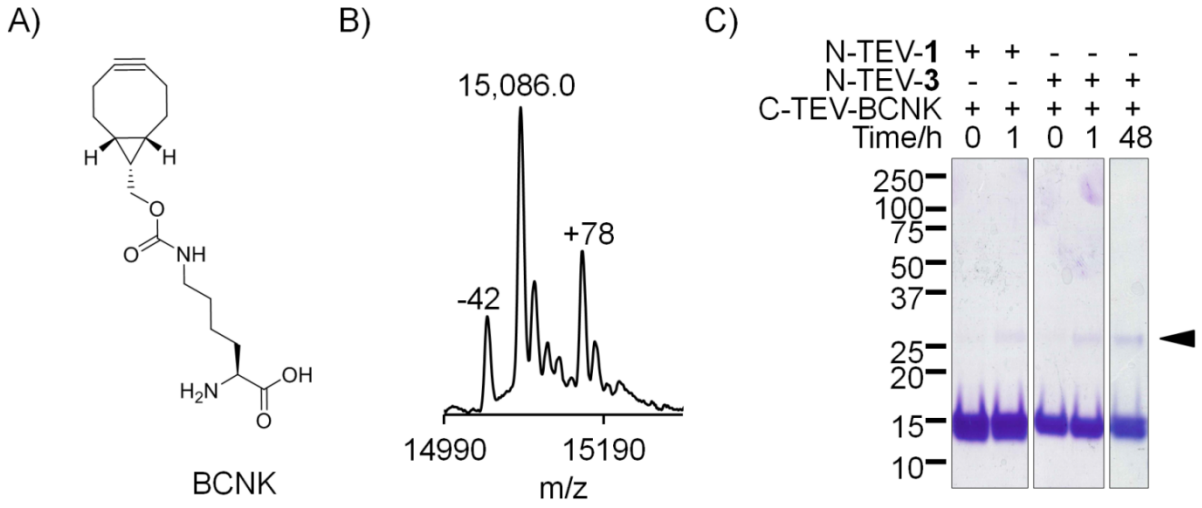


Figure S7. Conjugation of N-TEV-**1** or N-TEV-**3** to C-TEV-BCNK via the strain-promoted azide–alkyne cycloaddition (SPAAC). A) Structure of BCNK. B) Deconvoluted mass spectrum of purified C-TEV-BCNK (calculated M_r: 15,087.7 Da). The mass differences of the most predominant adduct peaks relative to the main peak are indicated. The nature of the -42 adduct is unknown. However, it was also detected for another protein in which BCNK was incorporated (data not shown). The +78 peak corresponds to a 2-mercaptoethanol adduct. C) SDS-PAGE gel of the reaction between C-TEV-BCNK and N-TEV-**1** or N-TEV-**3**. Reactions were performed at room temperature in 20 mM NaH_2PO_4 (pH 8.0), 150 mM NaCl and 4 M urea for the indicated amount of time. The arrow indicates the product.

Amino acid sequences of expressed proteins

Wildtype TEV protease = MG-H_6-TEV(1-237)
MGSSHHHHHHSSGLVPRGSGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRNNGTLLVQSL
HGVFKVKNTTTLQQHLIDGRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTTNFQTKSMSSMVSDTSCTFPSSDGIF
WKHWIQTKDGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEAQQWVSGWRLNADSVLWGG
HKVFMVKPEEPFQPVKEATQLMNE

TEV fragment N-TEV-1 = XA-H_6-TEV(1-117)ΔM-X
XASSHHHHHHSSGLVPRGSGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRNNGTLLVQSL
HGVFKVKNTTTLQQHLIDGRDLIIIRLPKDFPPFPQKLKFREPQREERICLVTTNFQ**X**

TEV fragment C-TEV-2 = X-TEV(122-237)ΔM-H_6
XESLVSDTSCTFPSSDGIFWKHWIQTKDGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFAELLTNQEAQQ
WVSGWRLNADSVLWGGHKVFAVKPEEPFQPVKEATQLLNELVPRGSGSSHHHHHH

TEV fragment N-TEV-3 = MG-TEV(1-115)-tag-H_6
MGESLFKGPRDYNPISSTICHLTNESDGHTTSLYGIGFGPFIITNKHLFRRNNGTLLVQSLHGVFKVKNTTTLQQHLID
GRDMIIIRMPKDFPPFPQKLKFREPQREERICLVTTN**X**LVPRGSHHHHHH

TEV fragment C-TEV-4 = MG-tag-TEV(122-237)-H_6
MG**X**SSMVSDTSCTFPSSDGIFWKHWIQTKDGQCGSPLVSTRDGFIVGIHSASNFTNTNNYFTSVPKNFMELLTNQEA
QQWVSGWRLNADSVLWGGHKVFMVKPEEPFQPVKEATQLMNELVPRGSGSSHHHHHH

Loop F116-M121 is underlined in the sequence of wildtype TEV protease. **X** represents the position of the ncAA. The histidine purification tag and thrombin recognition sequence are depicted in blue. The mutations M82L, M87L, M121L, M124L, M187A, M218A and M235L are underlined in N-TEV-**1** and C-TEV-**2**. The residues cleaved off by methionine aminopeptidase (MetAP) are depicted in red. In C-TEV-**2**, Ser at position 2 (position 122 in the wildtype TEV protease) was replaced with Glu such that the first residue (homopropargylglycine, **2**) was not cleaved off by MetAP.[2]

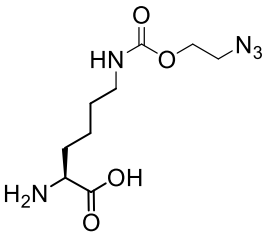
Experimental section

Materials and instrumentation

All commercially available reagents were used as received without further purification. Propargyl chloroformate, 2-chloroethyl chloroformate, sodium azide, *N*-ethylmorpholine (NEM), piperidine, *N,N*-diisopropylethylamine (DIPEA), triisopropylsilane (TIPS), trimethylamine (TEA), LB medium, ampicillin, chloramphenicol, arabinose, isopropyl β-D-1-thiogalactopyranoside (IPTG), lysozyme (from chicken egg white), sodium chloride, sodium phosphate, 2-mercaptoethanol, urea, guanidine hydrochloride (GdmCl), imidazole, iodoacetamide, trypsin (sequencing-grade, modified), Coomassie® Brilliant blue R-250, tris(3-hydroxypropyltriazolylmethyl)amine (THPTA), sodium ascorbate and $CuSO_4 \cdot H_2O$ were purchased from Sigma-Aldrich. $PEGA_{800}$ resin (loading: 0.30 mmol/g) was acquired from VersaMatrix A/S. 9-fluorenylmethyloxycarbonyl (Fmoc)-protected amino acids, N^{α}-Boc-L-lysine, 1-(2-mesitylenesulfonyl)-3-nitro-1H-1,2,4-triazole (MSNT), 4-hydroxymethyl benzoic acid (HMBA), 2-(1H-benzotriazole-1-yl)-1,1,3,3-tetramethyluronium tetrafluoroborate (TBTU), 1-methylimidazole (MeIm) and N^{γ}-azido-L-2,4-diaminobutyric acid hydrochloride (L-azidohomoalanine) were purchased from Bachem AG (Germany), Chem-Impex International, Inc. (USA) and Honeywell Fluka. Organic solvents and hydrochloride acid were purchased from VWR Chemicals. All 20 unprotected natural amino acids, trifluoroacetic acid (TFA) and thiamine were purchased from Carl Roth (Germany). BD Bacto™ tryptone and yeast extract were acquired from BD Biosciences, (*S*)-2-aminohex-5-ynoic acid (L-homopropargyl glycine) from Chiralix (The Netherlands), and (2*S*)-2-amino-6-((((1*R*,8*S*)-bicyclo[6.1.0]non-4-yn-9-ylmethoxy)carbonyl)amino)hexanoic acid (BCN-lysine) from Synaffix (The Netherlands).
The water used for all experiments was purified using an Ultra Clear Water System (Siemens) set at 0.055 µS/cm. LC-MS analysis was performed on a Dionex UltiMate 3000 (Thermo Scientific) equipped with an Acclaim™ RSLC 120 C18 column (2.2 µm, 120Å, 2.1 x 100 mm) coupled to a Bruker micrOTOF-QIII mass spectrometer. A linear gradient of CH_3CN in H_2O with 0.1% formic acid was used, running from 5% to 100% CH_3CN, 0.5 mL/min over 10 min. MALDI-TOF MS

spectra were recorded on a Bruker autoflex™ speed MALDI-TOF instrument. Analytical HPLC analysis was performed on an Agilent 1100 HPLC using a 4.6 x 100 mm XBrigde C18 column. A liner gradient of acetonitrile in water with 0.1% TFA was used, running from 0% to 90% acetonitrile, 1 mL/min over 10 min.

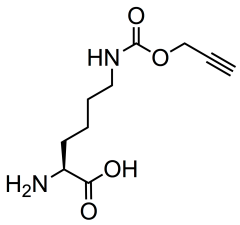
Synthesis of (*S*)-2-Amino-6-((2-azidoethoxy)carbonylamino)hexanoic acid (3)



3

N^{α}-Boc-L-lysine (1.00 g, 4.06 mmol) was dissolved in deionized water (250 mL). DIPEA (2.0 mL, 11.48 mmol) was added and the solution cooled on an ice bath. 2-Chloroethyl chloroformate (650 µL, 6.18 mmol) was added dropwise under stirring and the reaction left for 15 min. EtOAc (50 mL) was added, and the combined phases acidified by dropwise addition of conc. HCl. The aqueous phase was extracted with EtOAc (3 × 50 mL) and the combined organic phases washed with acidified deionized water, then dried with Na_2SO_4 and further evaporated giving a clear oily residue (1.40 g, 3.98 mmol, 98% yield). The residue was dissolved in DMF (20 mL) and NaN_3 (780 mg, 12.00 mmol) added. The reaction was left overnight at 70 °C under stirring and protected from light sources. The residual solid was filtered off and the solvent evaporated. The resulting residue was dissolved in acidified deionized water (50 mL) and EtOAc (50 mL). The aqueous phase was extracted with EtOAc (3 × 50 mL) and the combined organic phases washed with acidified deionized water, then dried with Na_2SO_4 and further evaporated giving a yellow oily residue (1.16 g, 3.23 mmol, 81%). The residue was dissolved in DCM (10 mL) and left on ice. TFA (10 mL) was added dropwise and the reaction left under stirring for 1 hr. The solvents were evaporated. The resulting residue was first dissolved in MeOH (10 mL) and then precipitated in Et_2O. The precipitate was then dissolved in HCl (1 M, 50 mL) and freeze-dried providing the HCl salt of **3** as a yellow-white powder (0.79 g, 3.05 mmol, 94% yield). ^{1}H NMR (500 MHz, deuterium oxide), δ = 4.25 (t, *J* = 4.9 Hz, 2H), 4.06 (t, *J* = 6.3 Hz, 1H), 3.53 (t, *J* = 4.9 Hz, 2H), 3.17 (t, *J* = 6.8 Hz, 2H), 2.06 – 1.89 (m, 2H), 1.62 – 1.38 (m, 4H). ^{13}C NMR (126 MHz, deuterium oxide), δ = 172.36, 158.20, 63.80, 52.97, 49.99, 39.86, 29.42, 28.35, 21.40. ESI-TOF m/z: $[M+H]^+$ for $C_9H_{17}N_5O_4$: calculated = 260.1353; detected = 260.1350.

Synthesis of (*S*)-2-Amino-6-((prop-2-ynyloxy)carbonylamino)hexanoic acid (4)

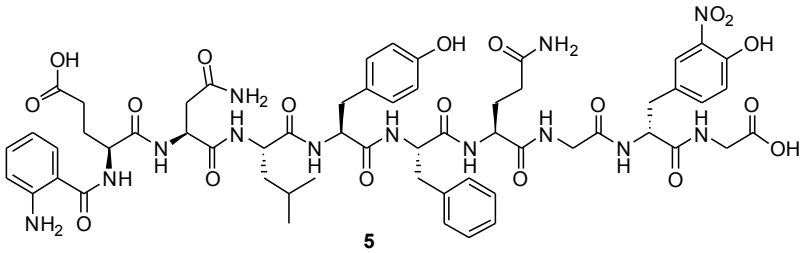


4

(*S*)-2-Amino-6-((prop-2-ynyloxy)carbonylamino)hexanoic acid (**4**) was synthesized according to a previously published procedure[3] with minor modifications. N^{α}-Boc-L-lysine (1.02 g, 4.14 mmol) was dissolved in deionized water (250 mL). DIPEA (2.0 mL, 11.48 mmol) was added and the solution was cooled on an ice bath. Propargyl chloroformate (813 µL, 8.33 mmol) was added dropwise under stirring and the reaction left for 30 min. EtOAc (50 mL) was added and the combined phases acidified by dropwise addition of conc. HCl. The aqueous phase was extracted with EtOAc (3 × 50 mL) and the combined organic phases were washed with acidified deionized water, then dried with Na_2SO_4 and further evaporated giving a clear oily residue (1.32 g, 4.02 mmol, 97% yield). The residue was dissolved in DCM (10 mL) and left on ice. TFA (10 mL) was added dropwise and the reaction left under stirring for 1 hr. The solvents were evaporated. The resulting residue was first dissolved in MeOH (10 mL) and then precipitated in Et_2O. The precipitate was then dissolved in HCl (1 M, 50 mL) and freeze-dried providing the HCl salt of **4** as a white powder (0.96 g, 3.63 mmol, 90% yield). ^{1}H NMR (500 MHz, deuterium oxide), δ = 4.68 (s, 2H), 3.74 (t, *J* = 6.1 Hz, 1H), 3.17 (t, *J* = 6.8 Hz, 2H), 2.97 – 2.86 (m, 1H), 1.97 – 1.81 (m, 2H), 1.63 – 1.50 (m, 2H), 1.50 – 1.35 (m, 2H). ^{13}C NMR (126 MHz, deuterium oxide), δ = 174.81, 157.65, 78.59, 75.55, 54.71, 52.64, 40.05, 30.07, 28.45, 21.57. ESI-TOF m/z: $[M+H]^+$ for $C_{10}H_{16}N_2O_4$: calculated = 229.1188; detected = 229.0981.

Synthesis of FRET-substrate Abz-ENLYFQGY(NO_2)G (5)

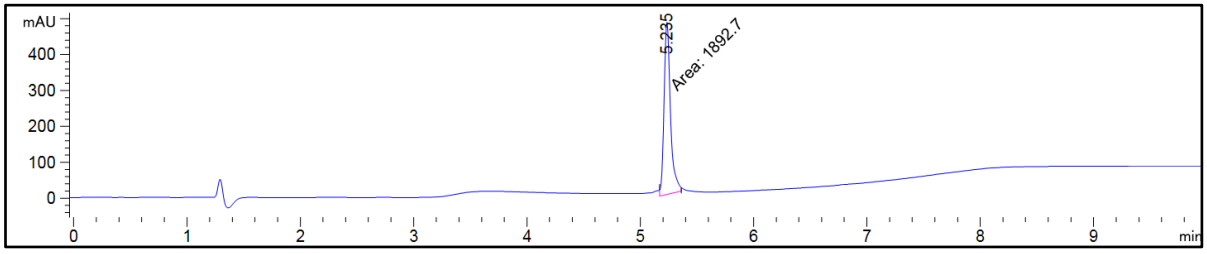
Substrate **5** was designed based on the TEV protease recognition sequence ENLYFQ-G.[4] In order to be able to use the substrate for analysis of TEV protease activity by fluorescence spectroscopy, the sequence was equipped with fluorescence resonance energy transfer (FRET) properties using 4-aminobenzoic acid (Abz) as a fluorophore and nitrotyrosine as a quencher.



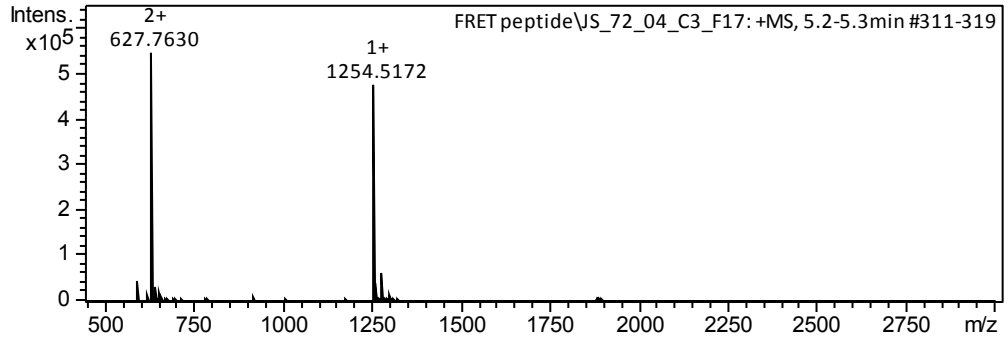
5

The peptide was synthesized on $PEGA_{800}$ resin (1 g dry resin, 0.4 mmol/g) according to standard Fmoc-based solid-phase peptide synthesis. The base-labile linker HMBA was coupled to the resin by preactivation of the linker (3 equiv) in DMF containing 4 equiv of NEM and 2.88 equiv of TBTU. After 5 min the solution was added to the resin. After 2 h the resin was washed 5× with DMF. The first amino acid, Fmoc-Gly-OH, was coupled to HMBA in a double MSNT coupling. 3 equiv of Fmoc-Gly-OH were dissolved in dry DCM and 2.25 equiv of MeIM were added. After dissolution 3 eq. of MSNT were added and after 30 s the mixture was added to the resin. The coupling was performed twice for 1 h each, and the resin was washed 5× with dry DCM after each reaction. The next amino acids were coupled by preactivation (3 min) of the respective Fmoc-amino acid (3 equiv) with TBTU (2.9 equiv) and NEM (4 equiv) in DMF. Every coupling was allowed a reaction time of 2 h and was tested for sufficiency by the Kaiser test.[5] After every coupling step, the resin was washed once with methanol and 5× with DMF. The amino acids were deprotected with 20% piperidine in DMF twice for 2 and 18 minutes, respectively, after which the resin was washed with 5× DMF. In order to prevent cyclization of the first amino acids, deprotection prior to coupling of the second glycine was performed 2× for only 2 min, respectively. Complete removal of piperdine was tested by adding 3-hydroxybenzo[d][1,2,3]triazin-4(3H)-one (DhbtOH) to the flow-through. The Fmoc protecting group of the last amino acid was deprotected with 20% piperidine in DMF twice for 2 and 18 minutes, respectively. *N*-Fmoc-2-aminobenzoic acid was used to introduce the Abz group. The resin was washed with DMF (5×) and DCM (3×) and dried with nitrogen. The side-chain protecting groups were removed using a mixture of TFA/water/phenol/TIPS (88/5/5/2). After deprotection, the resin was washed with 2% DIPEA in DMF for 20 minutes, 5× DMF and 3× DCM. The peptide was cleaved from the dried resin with 5% TEA and the resin was washed with 30–70% MeCN. The collected solution was lyophilized overnight (515 mg crude peptide, 85.7% yield). An aliquot of the yellow powder (48 mg) was then dissolved in 70% MeCN in water under slightly basic conditions and subjected to preparative high-pressure liquid chromatography (HPLC) purification starting at 5% of buffer B (100% MeCN, 0.1% TFA) increasing to 85% in 14 min with buffer A being ddH_2O with 0.1% TFA. Pure fractions were pooled, lyophilized (12 mg, yield: 25%) and analyzed by analytical HPLC and ESI-TOF MS. ESI-TOF m/z: $[M+H]^+$ = 1254.5172 (calcd. 1254.5062), $[M+2H]^{2+}$ = 627.7630 (calcd. 627.7567).

Analytical HPLC chromatogram of **5** (UV at 230 nm):



ESI-MS spectrum of **5**:



Molecular modeling

Protein structures were visualized and molecular modeling studies were performed in Yasara Dynamics, version 16.12.24. Structures of the triazole-containing linkages **1**–Tz–**2** and **3**–Tz–**4** were imported as SMILES strings. The geometry of each linkage was optimized such that the atoms to be linked to the split TEV protein were brought close to each other (5.8 Å). This optimization of geometry was achieved by pulling the relevant atoms towards each other followed by the addition of a force-field term, while running an interactive MD simulation in the annealing mode. Next, the structure of TEV protease (PDB 1LVM) was imported. The N-terminal His-tag was deleted from the structure. The structure of the triazole linkage was combined with the protein by deletion of the loop residues (T118–M121 for TEV[**1**–Tz–**2**] and F116–M121 for TEV[**3**–Tz–**4**]) followed by the formation of atom bonds between the newly formed protein termini and the triazole-containing linkage. For TEV[**3**–Tz–**4**], a Gly residue was added to the N-terminus of the C-terminal fragment, and the C-terminus of the N-terminal fragment was capped with an *N*-methyl group (NME). Then, a MD simulation was run on the split-clicked proteins and wildtype TEV protease using Yasara's simulation macro "md_run" employing the Amber14 force field. The simulations were run for 40 ns (TEV[**1**–Tz–**2**]) and 100 ns (wildtype and TEV[**3**–Tz–**4**], respectively.

Multiple sequence alignment

A BLAST search was performed using the BLOSUM62 matrix and the primary structure of TEV protease (Tobacco etch virus, NP_734212.1) as input. From this search, hits with an identity of 44% or higher were selected and aligned with the TEV protease sequence in Clustal Omega (version 1.2.4). The alignment included the following 26 viral proteases: TVMV (Tobacco vein mottling virus, AAF31455.1), PVA (Potato virus A, NP_734366.1), TMV (Turnip mosaic virus, AAA89116.1), OMV (Ornithogalum mosaic virus, YP_007001281.1), SMMV (Sunflower mild mosaic virus, APA32024.1), PVY (Panax virus Y, YP_003725726.1), SMV (Scallion mosaic virus, NP_734130.1), CMV (Celery mosaic virus, YP_006423980.1), HMV (Hippeastrum mosaic virus, YP_006390064.1), SPFMV (Sweet potato feathery mottle virus, NP_734316.1), AVY (Apium virus Y, YP_004123961.1), SPVC (Sweet potato virus C, YP_004046677.1), PPV (Plum pox virus, NP_734346.1), PMV (Pokeweed mosaic virus, YP_008719795.1), SPV2 (Sweet potato virus 2, YP_006438195.1), WOSV (Wild onion symptomless virus, YP_009259519.1), TLMV (Tamarillo leaf malformation virus, YP_009126868.1), SPVG (Sweet potato virus G, YP_006493341.1), NLSYV (Narcissus late season yellows virus, YP_009010965.1), KMV (Konjac mosaic virus, YP_529497.1), YMV (Yam mosaic virus, YP_022761.1), PRVW (Papaya ringspot virus W, AAG47346.1 polyprotein), SYSV (Shallot yellow stripe virus, YP_331421.1), BYMV (Bean yellow mosaic virus, BAA05979.1), PSBMV (pea seed-borne mosaic virus, P29152.1), JGMV (Johnson grass mosaic virus, ASZ83718.1). It was verified that all candidates contained the catalytic triad residues at a position corresponding to those in TEV protease.

Plasmid construction

Gene strands coding for full-length wildtype TEV protease (MG-H_6-TEV[1-237]), full-length methionine-deficient TEV protease (MG-H_6-TEV[1-237]ΔM) and the C-terminal fragments M-TEV[122-237]ΔM-H_6 and MG-tag-TEV[122-237]-H_6 (with 'tag' being the amber stop codon) were ordered codon-optimized for *E. coli* from Eurofins. The gene strands, flanked with *Nco*I and *Bam*HI restriction sites, were digested with *Nco*I and *Bam*HI (New England Biolab, NEB) and ligated in the corresponding sites of plasmid pET15b, treated with the same restriction enzymes. Ligation reactions were performed using T4 DNA ligase (NEB) at 16 °C overnight. Heat-shock competent *E. coli* DH5α cells were transformed with ligation reactions and plated on LB-agar supplemented with ampicillin. Plasmids were isolated from selected colonies using the

Wizard® Miniprep kit (Promega) following the manufacturer's protocol. All plasmids were sequenced using the T7 forward primer (Eurofins Genomics GmbH).

Plasmids containing the N-terminal fragments MG-H_6-TEV[1-117]ΔM-M and MG-TEV[1-115]-tag-H_6 were obtained by PCR amplification using Phusion Hot start II high fidelity polymerase (Thermo Fisher Scientific) and as template the plasmids pET15-MG-H_6-TEV[1-237]ΔM and pET15-MG-H_6-TEV[1-115], respectively. The sequences of the primers used in the PCR are shown in Table S1. *Nco*I and *Bam*HI restriction sites were included in the primers, allowing for ligation in the corresponding sites of plasmid pET15b. Transformation, plasmid isolation and sequencing steps were performed as described above.

The methionine-deficient N-terminal fragment initially contained a Gly residue as position 2. At a later stage, Gly2 was mutated into Ala, for more efficient removal of the residue at position 1 (azidohomoalanine in N-TEV-**1**),[2] providing the plasmid encoding MA-H_6-TEV[1-117]ΔM-M. It is this plasmid that was used for all work described in the present paper. The G2A mutation was introduced by site-directed mutagenesis, using pET15-MG-H_6-TEV[1-117]ΔM-M as templates and primers designed according to guidelines described by Liu *et al*.[6] The PCR reaction mixture was treated with *Dpn*I to digest non-mutated, parental DNA prior to transformation.

Table S1. Primer sequences.

Primer name	*Plasmid generated*	*Sequence* [a]
N-TEV-ΔM forward	pET15-MG-H_6-TEV[1-117]ΔM-M	5'-GATATACCATGGGCAGCAGC-3'
N-TEV-ΔM reverse	pET15-MG-H_6-TEV[1-117]ΔM-M	5'-ATATGGATCCTTACATCTGAAAATTGGTGGTAACCAG-3'
N-TEV-ΔM G2A forward	pET15-MA-H_6-TEV[1-117]ΔM-M	5'-CCATGGCCAGCAGCCATCATCATCACCATCATAG-3'
N-TEV-ΔM G2A reverse	pET15-MA-H_6-TEV[1-117]ΔM-M	5'-CTGCTGGCCATGGTATATCTCCTTCTTAAAGTTAAAC-3'
N-TEV-tag forward	pET15-MG-TEV[1-115]-tag-H_6	5'-ATATCCATGGGTGAAAGCCTGTTTAAAG-3'
N-TEV-tag reverse	pET15-MG-TEV[1-115]-tag-H_6	5'-TATGGATCCTTAATGATGGTGATGATGATGGCTACCACG CGGAACCAGCTAATTGGTGGTAACCAG-3'
T7 forward primer	-	5'-TAATACGACTCACTATAGGG-3'

[a] Restriction sites are underlined.

Residue-specific incorporation of ncAAs 1 and 2

Amino acids **1** and **2** were incorporated using auxotrophic bacterial strain *E. coli* B834(DE3) (Novagen/Merck Millipore). All solid and liquid broths were supplemented with ampicillin for selection. Heat-shock competent cells were transformed with pET15-MA-H_6-TEV[1-117]ΔM-M (N-TEV-**1**) or pET15-M-TEV[122-237]ΔM-H_6 (C-TEV-**2**) and plated on LB-agar. From a single colony, an overnight culture was grown in LB. This was used to prepare an expression culture in M9 minimal medium supplemented with all 20 canonical amino acids (40 mg/L each), glucose (0.4% w/v), thiamine (0.0005% w/v) and $MgSO_4$ (1 mM). The culture was grown at 37 °C to an optical density of 0.6–0.8, at which point IPTG was added to 1 mM, and the culture left for another 15 min. The cells were then pelleted and washed twice with 0.9% NaCl solution, before being suspended in fresh, defined medium lacking methionine, but supplemented with either **1** or **2**. At this stage, a volume corresponding to 10% of the initial culture volume was used.[7] After an additional 10 min at 37 °C, the temperature was shifted to room temperature and IPTG added to 1 mM. The culture was left overnight for expression before being harvested and frozen to -20 °C.

Incorporation of ncAAs 3 and 4 by amber suppression

Incorporation of **3** and **4** was accomplished using the bacterial strain *E. coli* BL21-AI (Invitrogen/Thermo Fischer Scientific). All solid and liquid broths were supplemented with ampicillin and chloramphenicol for selection. Heat-shock competent cells were co-transformed with pEVOL-pylRS and either pET15-MG-TEV[1-115]-tag-H_6 (N-TEV-**3**) or pET15-MG-tag-TEV[122-237]-H_6 (C-TEV-**4**). From a single colony, an overnight culture was grown in LB. This was used to prepare an expression culture in 2×YT. The culture was grown in 250 mL at 37 °C to an optical density of 0.6–0.8. The cells were then pelleted and suspended in fresh 50 mL 2×YT supplemented with either **3** or **4**.[8] The ncAA was dissolved in 0.1 M NaOH using sonication. The 5-fold concentrated culture was shaken at 37 °C for 10 min, after which IPTG and arabinose were added to 1 mM and 20 mM, respectively. The culture was then left for 8 h before being harvested and frozen to -20 °C.

The C-terminal fragment containing bicyclononyne-derivatized lysine (BCNK) instead of **4** was produced following the same protocol, except for the fact that cells were co-transformed with pEVOL-pylRS^{AF}, encoding an engineered pyrrolysine tRNA-synthetase.[9]

Protein purification

Cell paste was suspended in 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl and lysed by sonication for 5 min on ice using a cycle of 10 s at 60%, with 30 s pause. The lysate was centrifuged for 20 min at 20,000 g, 4 °C. The inclusion bodies were washed with 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, supplemented with 1% Triton-X100 and 20 mM 2-mercaptoethanol, followed by a wash with 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl alone. The washed inclusion bodies were then suspended in 6 M GdmCl, 50 mM NaH_2PO_4 (pH 8.0), 300 mM NaCl, supplemented with 20 mM imidazole and 20 mM 2-mercaptoethanol. After a centrifugation step for 20 min at 20,000 g, 4 °C the polypeptides were subjected to Ni^{2+}-NTA affinity chromatography (1-mL HisTrap™ HP column, GE Healthcare) and further purified by size-exclusion chromatography (Superdex Increase 75 10/300 GL column, GE Healthcare) on an ÄKTA Pure 25M system (GE Healthcare). The resulting fractions contained pure material in 8 M urea, 20 mM NaH_2PO_4 (pH 8.0), 150 mM NaCl which were stored at -80 °C until further use.

Gel electrophoresis

Protein samples were mixed with 4× Laemmli sample buffer (BioRad) supplemented with 2-mercaptoethanol, left at room temperature for 2–10 minutes and loaded on an any kD™ Mini-PROTEAN® TGX™ precast protein gel (Biorad). The Dual Color Precision Plus Protein™ Standard (BioRad) was used as reference. The gel was run at 200 V, stained in 0.1% (w/v) Coomassie Brilliant Blue R250 solution (10% (v/v) AcOH, 50% (v/v) MeOH in ddH_2O) and destained in destaining solution (10% (v/v) AcOH, 50% (v/v) MeOH in deionized water).

Protein concentration determination

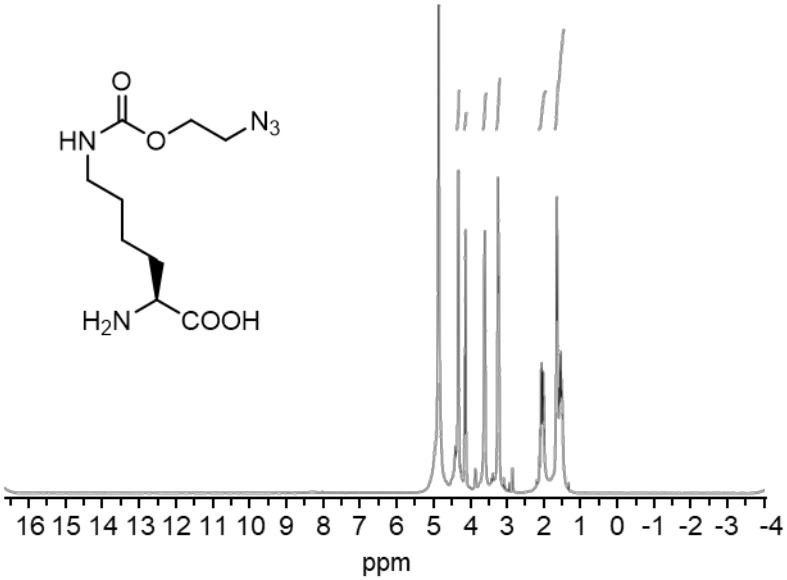
Protein concentrations were determined using the Pierce 660 nm protein assay reagent (Thermo Scientific). BSA was used as protein standard, and absorbance at 660 nm was measured on a NanoDrop™ 2000 spectrophotometer (Thermo Fisher Scientific).

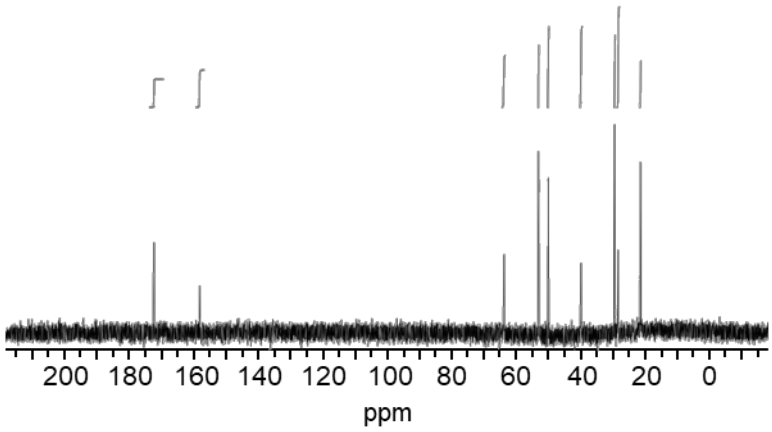
Tryptic digestion and mass spectrometric analysis

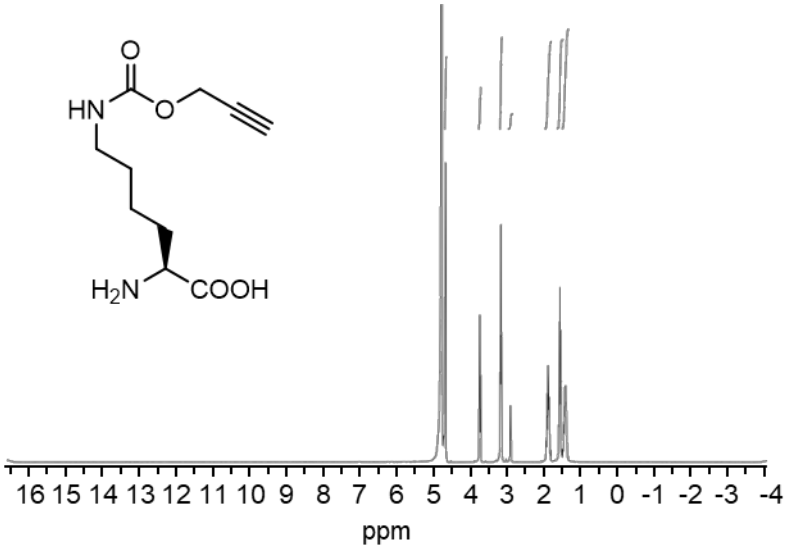
A sample of 10 µL was diluted 5-fold with ddH_2O. 1 µL 100 mM DTT was added, after which the solution was incubated for 30 min at 56 °C. The sample was then cooled to room temperature, before 1 µL 200 µM iodoacetamide was added. After 30 min incubation in the dark at room temperature, 1 µL trypsin was added and the sample left overnight at 37 °C to allow efficient digestion. For in-gel digestion, a band from an SDS-PAGE gel was cut out and destained before being treated first with 100 µL 5 mM DTT in 100 mM NH_4CO_3, for 30 min at room temperature, then 100 µL 15 mM iodoacetamide in 100 mM NH_4CO_3 for 30 min at room temperature in the dark, and finally overnight at 37 °C with 100 µL 10 µg/mL trypsin in 10 mM NH_4CO_3.

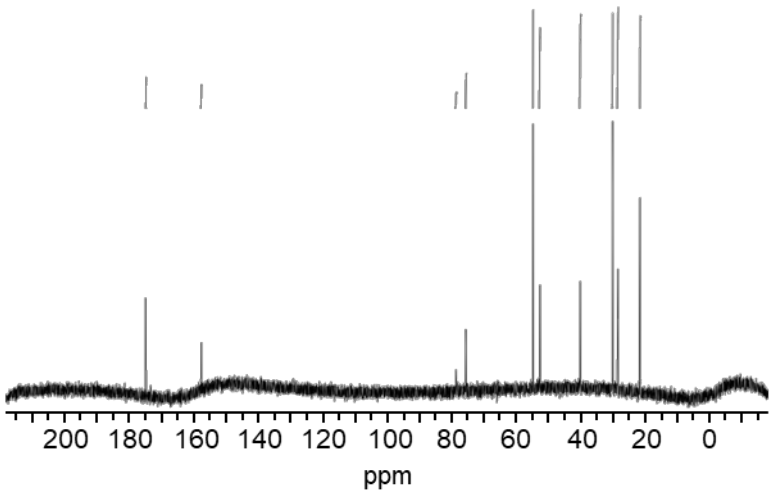
Digests and intact proteins were analyzed by LC-ESI-TOF and MALDI-TOF mass spectrometry. For MALDI-MS, the peptides and proteins were desalted with C18 ZipTip® pipette tips according to the manufacturer's protocol (Merck Millipore). Alpha-cyano-4-hydroxycinnamic acid was used as matrix.

1H- and 13C-NMR spectra of ncAAs 3 and 4









References

[1] Wiltschi, B., Merkel, L., Budisa, N. (2009) Fine tuning the N-terminal residue excision with methionine analogues. *ChemBioChem 10*, 217-220.

[2] Wang, A., Nairn, N. W., Johnson, R. S., Tirrell, D. A., Grabstein, K. (2008) Processing of N-terminal unnatural amino acids in recombinant human interferon-β in Escherichia coli. *ChemBioChem 9*, 324-330.

[3] Nguyen, D. P., Lusic, H., Neumann, H., Kapadnis, P. B., Deiters, A., Chin, J. W. (2009) Genetic encoding and labeling of aliphatic azides and alkynes in recombinant proteins via a pyrrolysyl-tRNA synthetase/tRNACUA pair and click chemistry. *J. Am. Chem. Soc. 131*, 8720-8721.

[4] Kapust, R. B., Tozser, J., Fox, J. D., Anderson, D. E., Cherry, S., Copeland, T. D., Waugh, D. S. (2001) Tobacco etch virus protease: mechanism of autolysis and rational design of stable mutants with wild-type catalytic proficiency. *Protein Eng. 14*, 993-1000.

[5] Kaiser, E., Colescot. R. L., Bossinge C. D., Cook, P. I. (1970) Color test for detection of free terminal amino groups in the solid-phase synthesis of peptides. *Anal. Biochem. 34*, 595-598.

[6] Liu, H. T., Naismith, J. H. (2008) An efficient one-step site-directed deletion, insertion, single and multiple-site plasmid mutagenesis protocol. *BMC Biotechnol. 8*.

[7] Schneider, D., Schneider, T., Rosner, D., Scheffner, M., Marx, A. (2013) Improving bioorthogonal protein ubiquitylation by click reaction. *Bioorg. Med. Chem. 21*, 3430-3435.
[8] Liu, J., Castaneda, C. A., Wilkins, B. J., Fushman, D., Cropp, T. A. (2010) Condensed E. coli cultures for highly efficient production of proteins containing unnatural amino acids. *Bioorg. Med. Chem. Lett. 20*, 5613-5616.
[9] Borrmann, A., Milles, S., Plass, T., Dommerholt, J., Verkade, J. M. M., Wiessler, M., Schultz, C., van Hest, J. C. M., van Delft, F. L., Lemke, E. A. (2012) Genetic encoding of a bicyclo[6.1.0]nonyne-charged amino acid enables fast cellular protein imaging by metal-free ligation. *ChemBioChem 13*, 2094-2099.