Supplementary Information

Linkage-specific ubiquitin chain formation depends on a lysine hydrocarbon ruler

Joanna Liwocha, David T. Krist, Gerbrand J. van der Heden van Noort, Fynn M. Hansen, Vinh H. Truong, Ozge Karayel, Nicholas Purser, Daniel Houston, Nicole Burton, Mark Bostock, Michael Sattler, Matthias Mann, Joseph S.Harrison, Gary Kleiger, Huib Ovaa, Brenda A. Schulman

Table of Contents

Item	Page
Supplementary Tables	
Supplementary Table 1. UB-conjugating enzymes (E2s) and their associated E3s employed in this study.	2
Supplementary Table 2. Final chase reaction acceptor UB concentrations performed in the absence of an E3 ligase.	3
Supplementary Table 3. Final chase reaction concentrations of acceptor UBs and the RING domain from the E3 RNF4 or the HECT E3 ligases NEDD4 or Rsp5p.	4
Supplementary Table 4. Chase reaction final concentrations for acceptor UBs and multi-subunit E3s including NEDD8-CUL1/RBX1 (NEDD8-CRL1), NEDD8-CUL4/RBX1 (NEDD8-CRL4) or APC/C with its coactivator CDH1.	5
Supplementary Table 5. Target peptide list.	6
Supplementary Notes	
Supplementary Note 1. UB mutant synthesis and analysis.	7
Supplementary Note 2. 1D and 2D proton NMR.	18
Supplementary Note 3. Molecular Modeling Methods.	19

UB-conjugating enzyme (E2)	E3 ligase	Lys specificity	Substrate receptor
UBE2N/UBE2V1	RNF4	K63	-
UBE2G1	NEDD8-CUL4/RBX1(NEDD8-CRL4)	K48	CRBN/DDB1+pomelidomide
UBE2R2	NEDD8-CUL1/RBX1(NEDD8-CRL1)	K48	SKP1/FBW7
UBE2D3	RNF4	Mixed	-
UBE2D2 (used to load NEDD4)	NEDD4 (HECT E3 ligase)	K63	-
UBE2D2 (used to load Rsp5p)	Rsp5p (HECT E3 ligase)	K63	-
UBE2S_IsoT	APC	K11	

Supplementary Table 1 | UB-conjugating enzymes (E2s) and their associated E3s employed in this study. Substrate receptor identities are shown for CRLs. The lysine specificities of each enzyme complex are listed.

E2 enzyme	UB [μM]
UBE2N/UBE2V1	300
UBE2R2	750
UBE2G1	550
UBE2S_IsoT	50

Supplementary Table 2 | Final chase reaction acceptor UB concentrations performed in the absence of an E3 ligase. These reactions all had an approximate final concentration of 0.5μM E2~UB that had been generated in the pulse step.

E2 enzyme	UB [μM]	RNF4 [μM]
UBE2N/UBE2V1	30	0.5
UBE2D3	100	1.0

Assay type	UB [μM]	HECT E3 [μM]	UB-substrate [μM]
UBE2D2/NEDD4	30	1.0	-
UBE2D2/Rsp5p	10	0.3	-
UBE2D2/Rsp5p	-	0.3	0.5

Supplementary Table 3 | Final chase reaction concentrations of acceptor UBs and the RING domain from the E3 RNF4 or the HECT E3 ligases NEDD4 or Rsp5p. These reactions had an approximate final concentration of 0.5 μM E2~UB that had been generated in the pulse step, with the exception of the Rsp5p reactions with UB-Sna4p, that had an approximate final concentration of 0.3 μM.

E2 enzyme	UB [μM]	UB-substrate [μM]	NEDD8-CRL1 [μM]	SKP1/FBW7 [μM]	NEDD8-CRL4 [μM]	CRBN [μM]	APC/C +CDH1 [μM]
UBE2R2	40	-	0.4	-	-	-	
UBE2R2	-	0.4	0.2	0.2	-	-	
UBE2G1	100	-	-	-	1.0	-	
UBE2G1	-	2.0	-	-	1.0	1.0	
UBE2S	20	-	-	-	-	-	0.1

Supplementary Table 4 | Chase reaction final concentrations for acceptor UBs and multi-subunit E3s including NEDD8-CUL1/RBX1 (NEDD8-CRL1), NEDD8-CUL4/RBX1 (NEDD8-CRL4) or APC/C with its coactivator CDH1.

Peptide	Sequence	Mass [m/z]	Charge State [z]
NleQIFVK	NleQIFVK	374.24181	2
K06_GG (light)	NleQIFVK^{GG}TLTGK	454.60989	3
K11_GG (light)	TLTGK^{GG}TITLEVEPSDTIENVK	801.42688	3
K11_GG (heavy)	TLTGK^{GG}TI***T***LEVEPSDTIENVK	803.09703	3
K27_GG (light)	TITLEVEPSDTIENVK^{GG}AK	701.03895	3
K27_GG (heavy)	TITLEVEPSDTIEN***V***K^{GG}AK	703.04356	3
K29_GG (light)	AK^{GG}IQDK	408.73233	2
K29_GG (heavy)	AK^{GG}***I***QDK	412.24092	2
K33_GG (light)	IQDK^{GG}EGIPPDQQR	546.61291	3
K33_GG (heavy)	IQDK^{GG}EGIP***P***DQQR	548.61751	3
K48_GG (light)	LIFAGK^{GG}QLEDGR	487.60005	3
K48_GG (heavy)	LIFAGK^{GG}QLEDG***R***	490.93614	3
K63_GG (light)	TLSDYNIQK^{GG}ESTLHLVLR	561.80503	4
K63_GG (heavy)	TLSDYNIQK^{GG}ESTLHLV***L***R	563.55932	4

Supplementary Table 5 | Target peptide list. Superscripted "GG" and bold/italic characters mark Gly-Gly modified lysines and stable isotope labeled amino acids, respectively.

Supplementary Note 1

UB mutant synthesis:

Solid Phase Peptide Synthesis (SPPS)

SPPS was performed on a Syro II MultiSyntech Automated Peptide synthesizer using standard 9-fluorenylmethoxycarbonyl (Fmoc) based solid phase peptide chemistry at 25 µmol scale, using fourfold excess of amino acids relative to pre-loaded Fmoc amino acid trityl resin (0.2 mmol/g, Rapp Polymere GmbH). Fmoc-Dap(Boc)-OH (C1), Fmoc-Dab(Boc)-OH (C2) and Fmoc-hLys(Boc)-OH (C5) were purchased from Iris Biotech GmbH and Fmoc-Orn(Boc)-OH (C3) was purchased from ChemImpex Int'l Inc. The Ub (mutant) peptide sequences were synthesized on resin following the procedures as described before[1]. After automated synthesis the peptides were treated with TFA/TIS/H_2O/Phenol for 2.5 hours followed by precipitation from Et_2O/Pentane and subsequently purified using RP-HPLC. Appropriate fractions as judged by LC-MS were pooled, lyophilized and stored at -20 ^{O}C.

RP-HPLC purifications

Waters preparative RP-HPLC system, equipped with a Waters C18-Xbridge 5 µm OBD (30 x 150 mm) column at a flowrate of 37.5 mL/min using 3 mobile phases: A: MQ, B: CH_3CN and C: 1% TFA in MQ. Gradient: 20 -> 45% B, 5% C.

LC-MS measurements*:*

Waters Acquity UPLC using a Waters Acquity BEH C18-column (2.1x50 mm, 1.7 µm) coupled to a LCT^{TM} ESI-Mass Spectrometer. Samples were run using 2 mobile phases: A = 1% CH_3CN, 0.1% formic acid in water and B = 1% water and 0.1% formic acid in CH_3CN. Flow rate= 0.6 mL/min, runtime= 6 min, column T= 60°C. Gradient: 0 - 95% B. Data processing was performed using Waters MassLynx Mass Spectrometry Software 4.1 (deconvolution with MaxEnt1 function).

HRMS-measurements:

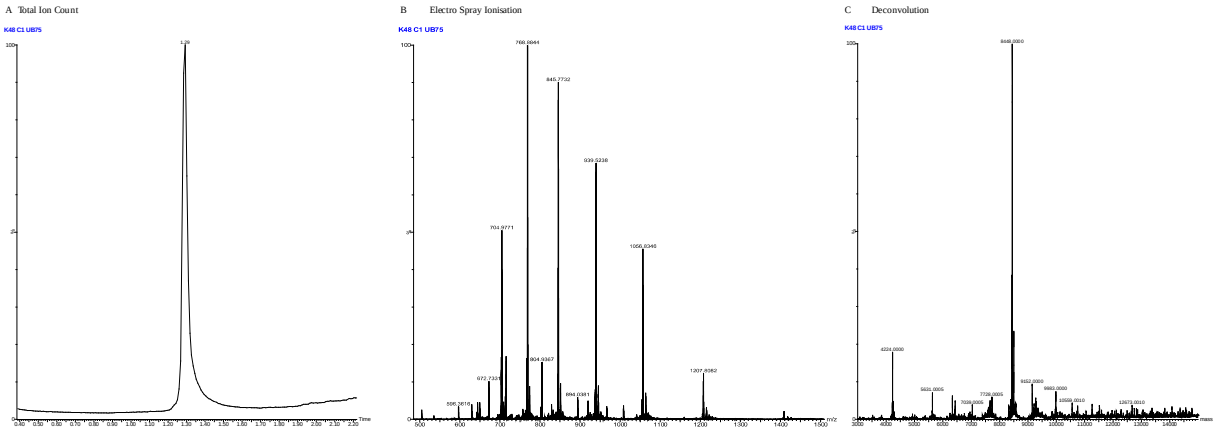
High resolution mass spectra of Ub mutants were recorded on a Waters XEVO-G2 XS Q-TOF) mass spectrometer equipped with an electrospray ion source in positive mode (source voltage 3.0 kV, desolvation gas flow 900 L/hr, temperature 250 °C) with resolution R = 22000 (mass range m/z = 50-2000) and 200 pg/uL Leu-Enk (m/z = 556.2771) as a "lock mass".

UB mutant analysis:

K48 C1 Ub_{75}

LC-MS: Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8447.70, found 8448.00.

HRMS: $[C_{374}H_{622}N_{104}O_{117} + 6H]^{6+}$: found 1408.7876, calc. 1408.7750, $[C_{374}H_{622}N_{104}O_{117} + 7H]^{7+}$: found 1207.6732, calc. 1207.6654, $[C_{374}H_{622}N_{104}O_{117} + 8H]^{8+}$: found 1056.8346, calc. 1056.8333, $[C_{374}H_{622}N_{104}O_{117} + 9H]^{9+}$: found 939.5238, calc. 939.5193, $[C_{374}H_{622}N_{104}O_{117} + 10H]^{10+}$: found 845.6780, calc. 845.6682, $[C_{374}H_{622}N_{104}O_{117} + 11H]^{11+}$: found 768.8844, calc. 768.8809, $[C_{374}H_{622}N_{104}O_{117} + 12H]^{12+}$: found 704.8956, calc. 704.8914.

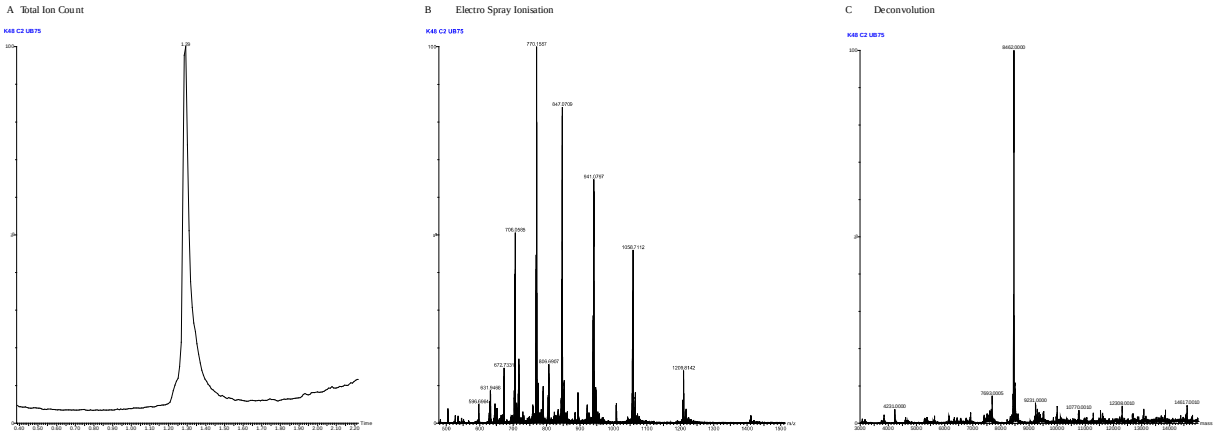


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K48 C2 Ub_{75}

LC-MS: Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8461.73, found 8462.00.

HRMS: $[C_{375}H_{624}N_{104}O_{117} + 6H]^{6+}$: found 1411.1154, calc. 1411.1110, $[C_{375}H_{624}N_{104}O_{117} + 7H]^{7+}$: found 1209.6720, calc. 1209.6676, $[C_{375}H_{624}N_{104}O_{117} + 8H]^{8+}$: found 1058.5913, calc. 1058.5852, $[C_{375}H_{624}N_{104}O_{117} + 9H]^{9+}$: found 941.0797, calc. 941.0766, $[C_{375}H_{624}N_{104}O_{117} + 10H]^{10+}$: found 847.0709, calc. 847.0697, $[C_{375}H_{624}N_{104}O_{117} + 11H]^{11+}$: found 770.1557, calc. 770.1550, $[C_{375}H_{624}N_{104}O_{117} + 12H]^{12+}$: found 706.0585, calc. 706.0594.

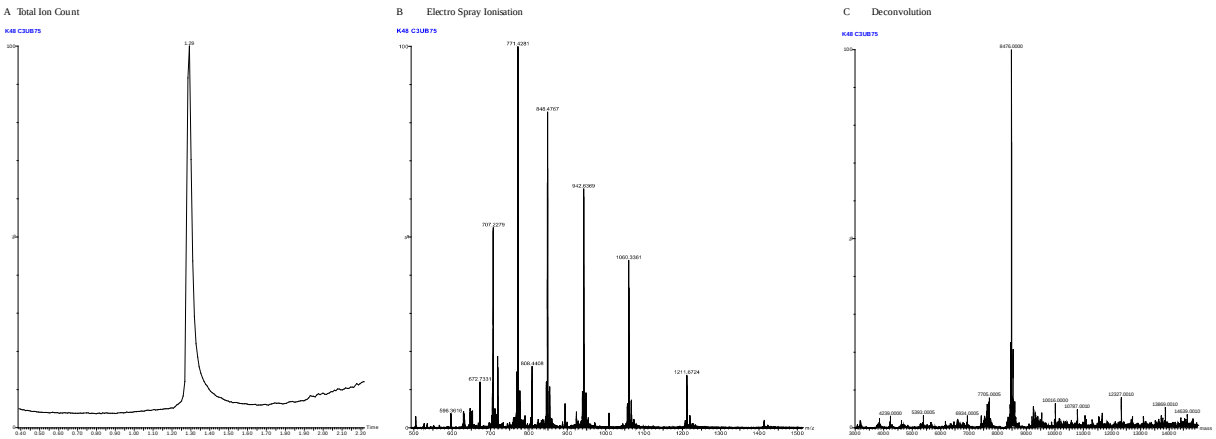


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K48 C3 Ub_{75}

LC-MS: Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8475.76, found 8476.00.

HRMS: $[C_{376}H_{626}N_{104}O_{117} + 6H]^{6+}$: found 1413.4528, calc. 1413.4469, $[C_{376}H_{626}N_{104}O_{117} + 7H]^{7+}$: found 1211.6724, calc. 1211.6699, $[C_{376}H_{626}N_{104}O_{117} + 8H]^{8+}$: found 1060.3361, calc. 1060.3372, $[C_{376}H_{626}N_{104}O_{117} + 9H]^{9+}$: found 942.6369, calc. 942.6339, $[C_{376}H_{626}N_{104}O_{117} + 10H]^{10+}$: found 848.4767, calc. 848.4713, $[C_{376}H_{626}N_{104}O_{117} + 11H]^{11+}$: found 771.4281, calc. 771.4291, $[C_{376}H_{626}N_{104}O_{117} + 12H]^{12+}$: found 707.2279, calc. 707.2274.



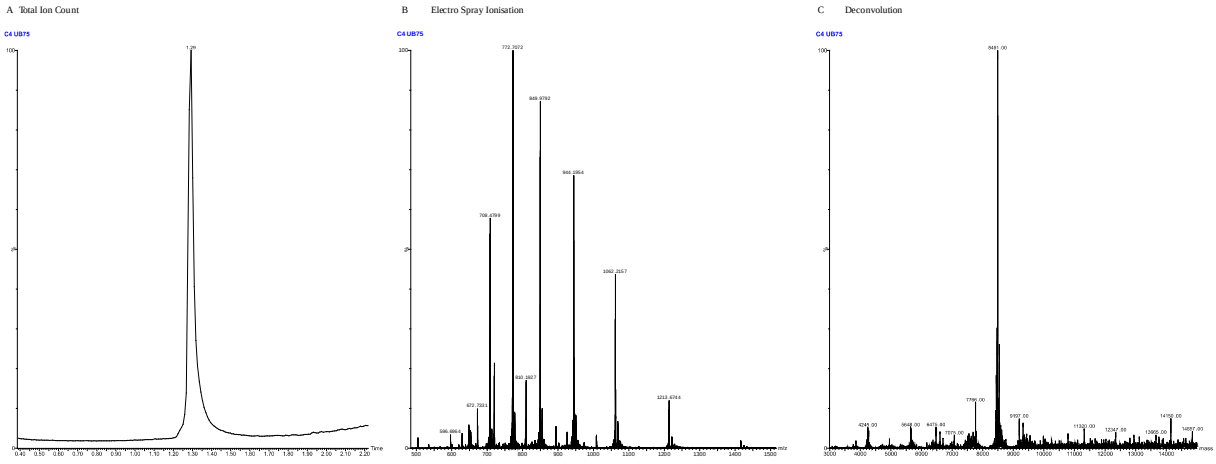
High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K48 C4 Ub_{75} (native lysine)

LC-MS**:** Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8489.62, found 8491.00.

HRMS: $[C_{377}H_{628}N_{104}O_{117} + 6H]^{6+}$: found 1415.7921, calc. 1415.7828, $[C_{377}H_{628}N_{104}O_{117} + 7H]^{7+}$: found 1213.6744, calc. 1213.6721, $[C_{377}H_{628}N_{104}O_{117} + 8H]^{8+}$: found 1062.0957, calc. 1062.0891, $[C_{377}H_{628}N_{104}O_{117} + 9H]^{9+}$: found 944.1954, calc. 944.1912, $[C_{377}H_{628}N_{104}O_{117} +$

10H]$^{10+}$: found 849.8779, calc. 849.8729, [$C_{377}H_{628}N_{104}O_{117}$ + 11H]$^{11+}$: found 772.7072, calc. 772.7033, [$C_{377}H_{628}N_{104}O_{117}$ + 12H]$^{12+}$: found 708.3982, calc.708.3953.

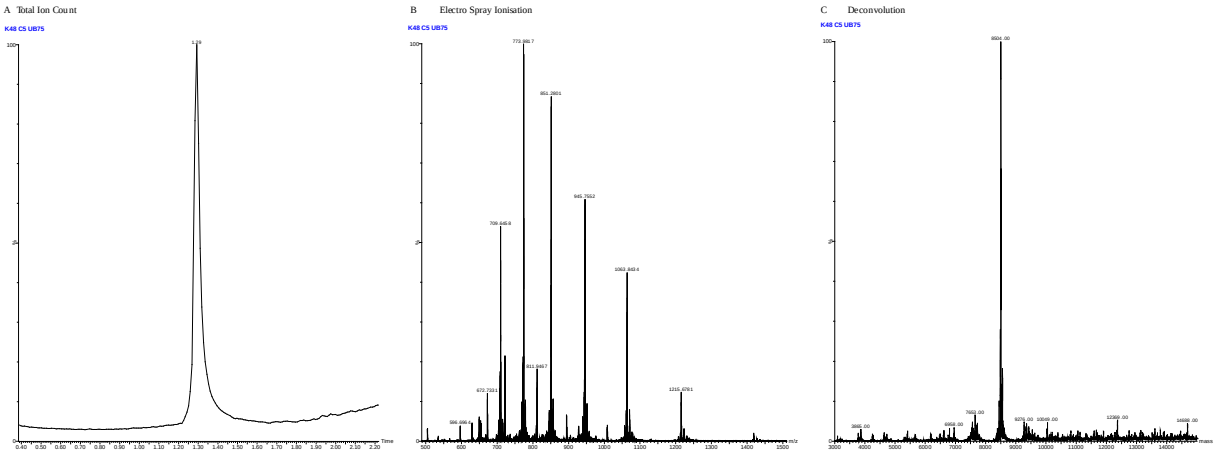


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K48 C5 Ub$_{75}$

LC-MS: Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8503.65, found 8504.00

HRMS: [$C_{378}H_{630}N_{104}O_{117}$ + 6H]$^{6+}$: found 1418.1257, calc. 418.1188, [$C_{378}H_{630}N_{104}O_{117}$ + 7H]$^{7+}$: found 1215.6781, calc. 1215.643, [$C_{378}H_{630}N_{104}O_{117}$ + 8H]$^{8+}$: found 1063.8434, calc. 1063.811, [$C_{378}H_{630}N_{104}O_{117}$ + 9H]$^{9+}$: found 945.7552, calc. 945.7485, [$C_{378}H_{630}N_{104}O_{117}$ + 10H]$^{10+}$: found 851.2801, calc. 851.2744, [$C_{378}H_{630}N_{104}O_{117}$ + 11H]$^{11+}$: found 773.9817, calc. 773.9774, [$C_{378}H_{630}N_{104}O_{117}$ + 12H]$^{12+}$: found 709.5640, calc. 709.5633.

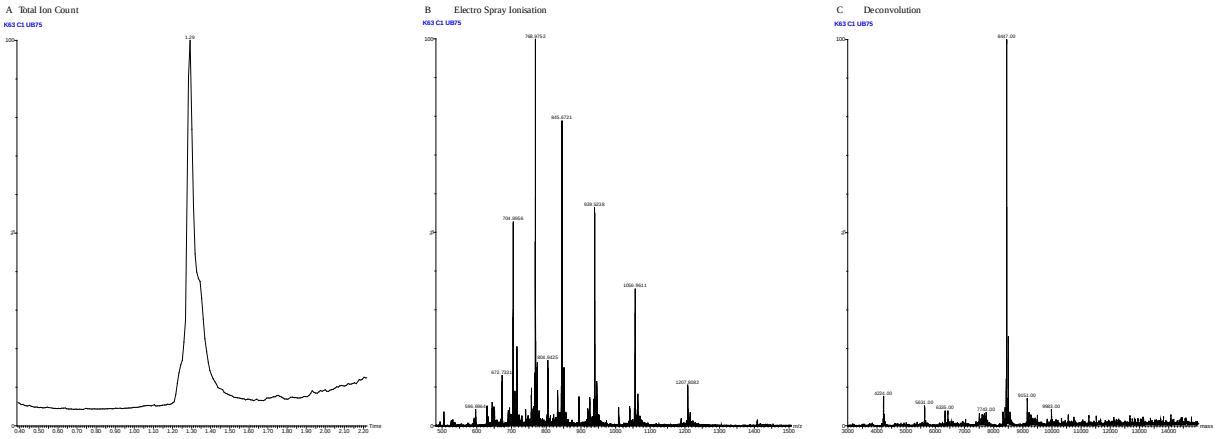


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K63 C1 Ub$_{75}$

LC-MS: Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8447.70, found 8447.00.

HRMS: $[C_{374}H_{622}N_{104}O_{117} + 6H]^{6+}$: found 1408.7799, calc. 1408.7750, $[C_{374}H_{622}N_{104}O_{117} + 7H]^{7+}$: found 1207.6732, calc. 1207.6654, $[C_{374}H_{622}N_{104}O_{117} + 8H]^{8+}$: found 1056.8346, calc. 1056.8333, $[C_{374}H_{622}N_{104}O_{117} + 9H]^{9+}$: found 939.5238, calc. 939.5193, $[C_{374}H_{622}N_{104}O_{117} + 10H]^{10+}$: found 845.6721, calc. 845.6682, $[C_{374}H_{622}N_{104}O_{117} + 11H]^{11+}$: found 768.8844, calc. 768.8809, $[C_{374}H_{622}N_{104}O_{117} + 12H]^{12+}$: found 704.8956, calc. 704.8914.

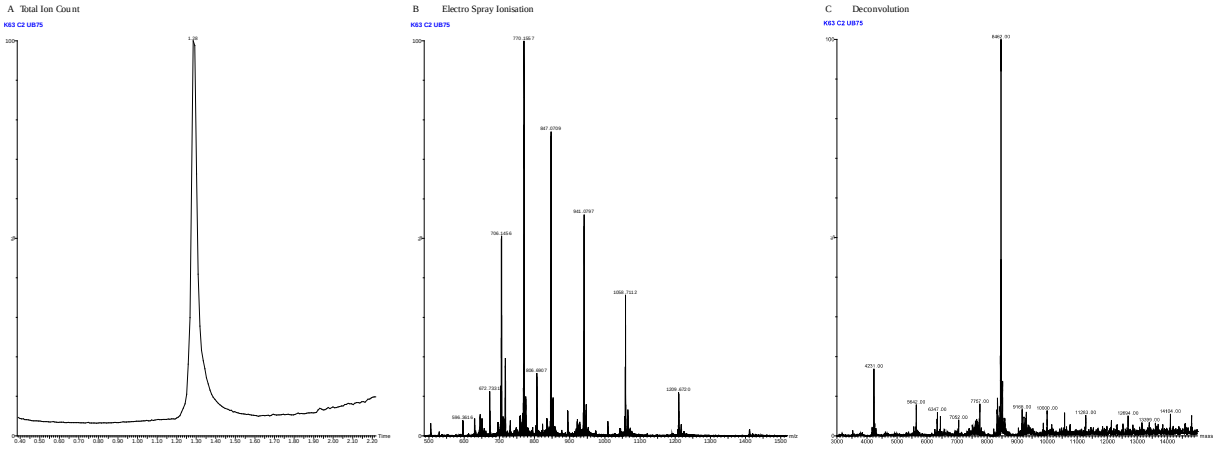


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K63 C2 Ub_{75}

LC-MS**:** Rt = 1.28 min., ESI MS+ (amu) deconvolution calcd: 8461.73, found 8462.00.

HRMS: $[C_{375}H_{624}N_{104}O_{117} + 6H]^{6+}$: found 1411.1154, calc. 1411.1110, $[C_{375}H_{624}N_{104}O_{117} + 7H]^{7+}$: found 1209.6720, calc. 1209.6676, $[C_{375}H_{624}N_{104}O_{117} + 8H]^{8+}$: found 1058.5913, calc. 1058.5852, $[C_{375}H_{624}N_{104}O_{117} + 9H]^{9+}$: found 941.0797, calc. 941.0766, $[C_{375}H_{624}N_{104}O_{117} + 10H]^{10+}$: found 847.0709, calc. 847.0697, $[C_{375}H_{624}N_{104}O_{117} + 11H]^{11+}$: found 770.1557, calc. 770.1550, $[C_{375}H_{624}N_{104}O_{117} + 12H]^{12+}$: found 706.0640, calc. 706.0594.

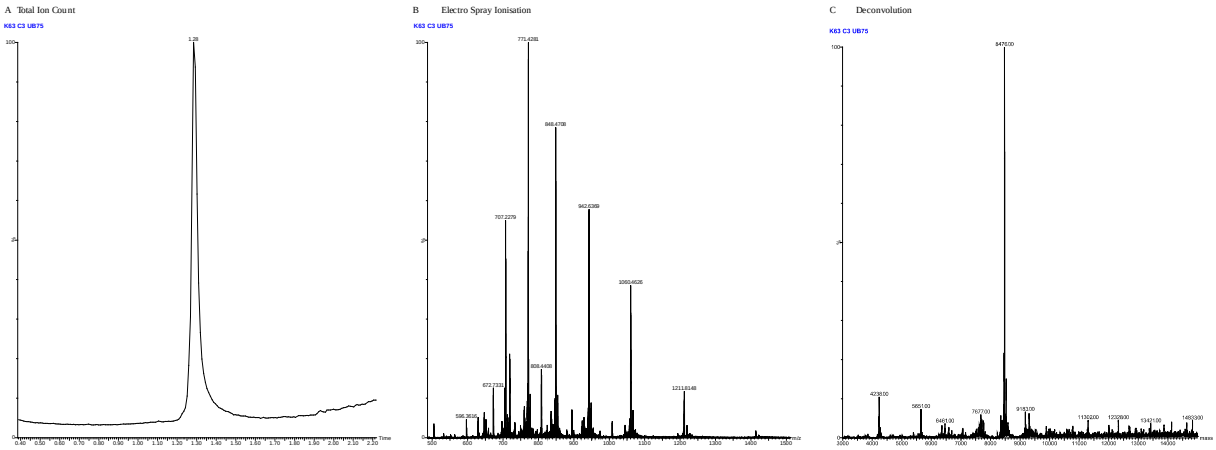


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K63 C3 Ub_{75}

LC-MS: Rt = 1.28 min., ESI MS+ (amu) deconvolution calcd: 8475.76, found 8475.00.

HRMS: $[C_{376}H_{626}N_{104}O_{117} + 6H]^{6+}$: found 1413.4528, calc. 1413.4469, $[C_{376}H_{626}N_{104}O_{117} + 7H]^{7+}$: found 1211.6794, calc. 1211.6699, $[C_{376}H_{626}N_{104}O_{117} + 8H]^{8+}$: found 1060.3428, calc. 1060.3372, $[C_{376}H_{626}N_{104}O_{117} + 9H]^{9+}$: found 942.6369, calc. 942.6339, $[C_{376}H_{626}N_{104}O_{117} + 10H]^{10+}$: found 848.4767, calc. 848.4713, $[C_{376}H_{626}N_{104}O_{117} + 11H]^{11+}$: found 771.4338, calc. 771.4291, $[C_{376}H_{626}N_{104}O_{117} + 12H]^{12+}$: found 707.2279, calc. 77.2274.

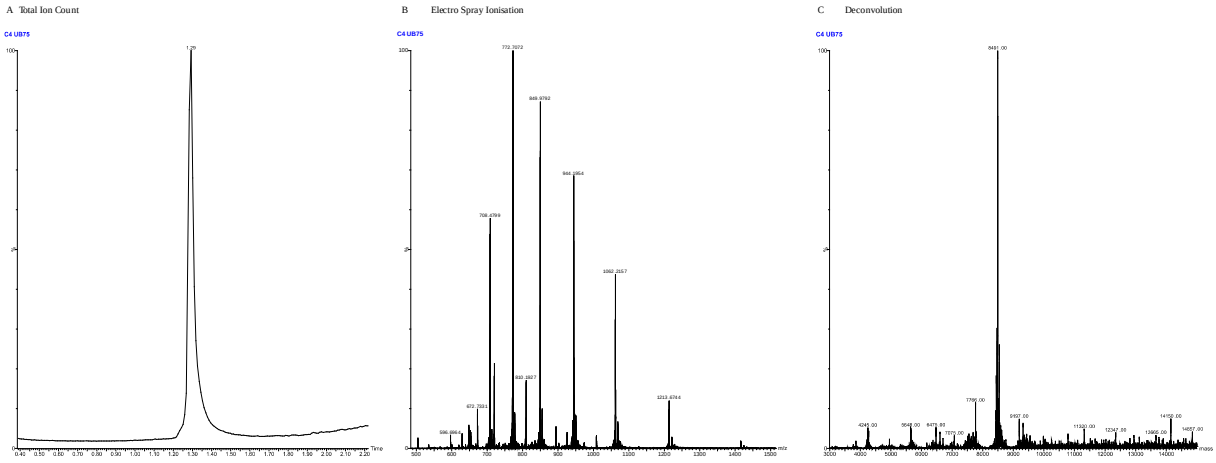


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K63 C4 Ub_{75} (native lysine)

LC-MS**:** Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8489.62, found 8491.00.

HRMS: $[C_{377}H_{628}N_{104}O_{117} + 6H]^{6+}$: found 1415.7921, calc. 1415.7828, $[C_{377}H_{628}N_{104}O_{117} + 7H]^{7+}$: found 1213.6744, calc. 1213.6721, $[C_{377}H_{628}N_{104}O_{117} + 8H]^{8+}$: found 1062.0957, calc. 1062.0891, $[C_{377}H_{628}N_{104}O_{117} + 9H]^{9+}$: found 944.1954, calc. 944.1912, $[C_{377}H_{628}N_{104}O_{117} + 10H]^{10+}$: found 849.8779, calc. 849.8729, $[C_{377}H_{628}N_{104}O_{117} + 11H]^{11+}$: found 772.7072, calc. 772.7033, $[C_{377}H_{628}N_{104}O_{117} + 12H]^{12+}$: found 708.3982, calc.708.3953.

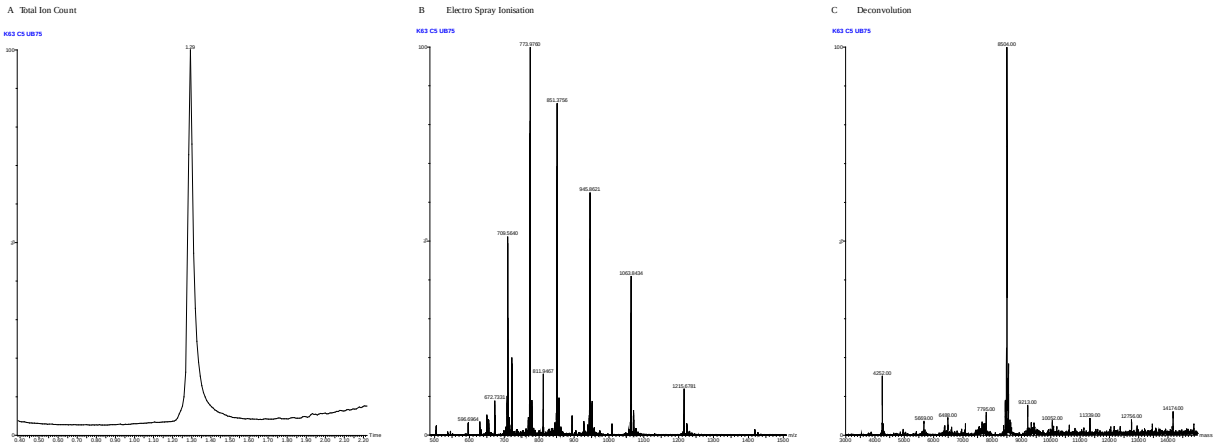


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K63 C5 Ub_{75}

LC-MS: Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8503.81, found 8504.00

HRMS: $[C_{378}H_{630}N_{104}O_{117} + 6H]^{6+}$: found 1418.1257, calc. 418.1188, $[C_{378}H_{630}N_{104}O_{117} + 7H]^{7+}$: found 1215.6781, calc. 1215.6743, $[C_{378}H_{630}N_{104}O_{117} + 8H]^{8+}$: found 1063.8434, calc. 1063.8411, $[C_{378}H_{630}N_{104}O_{117} + 9H]^{9+}$: found 945.7489, calc. 945.7485, $[C_{378}H_{630}N_{104}O_{117} + 10H]^{10+}$: found 851.2801, calc. 851.2744, $[C_{378}H_{630}N_{104}O_{117} + 11H]^{11+}$: found 773.9760, calc. 773.9774, $[C_{378}H_{630}N_{104}O_{117} + 12H]^{12+}$: found 709.5640, calc. 709.5633.

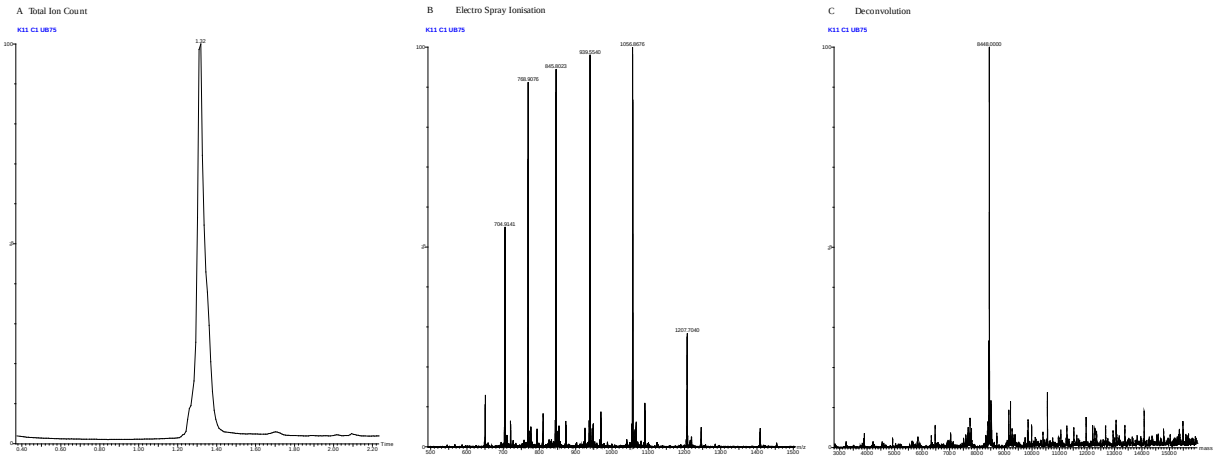


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K11 C1 Ub_{75}

LC-MS: Rt = 1.32 min., ESI MS+ (amu) deconvolution calcd: 8447.54, found 8448.00.

HRMS: $[C_{374}H_{622}N_{104}O_{117} + 6H]^{6+}$: found 1408.8263, calc. 1408.7750, $[C_{374}H_{622}N_{104}O_{117} + 7H]^{7+}$: found 1207.7040, calc. 1207.6654, $[C_{374}H_{622}N_{104}O_{117} + 8H]^{8+}$: found 1056.8676, calc. 1056.8333, $[C_{374}H_{622}N_{104}O_{117} + 9H]^{9+}$: found 939.5540, calc. 939.5193, $[C_{374}H_{622}N_{104}O_{117} + 10H]^{10+}$: found 845.7011, calc. 845.6682, $[C_{374}H_{622}N_{104}O_{117} + 11H]^{11+}$: found 768.9076, calc. 768.8809, $[C_{374}H_{622}N_{104}O_{117} + 12H]^{12+}$: found 704.9141, calc. 704.8914.

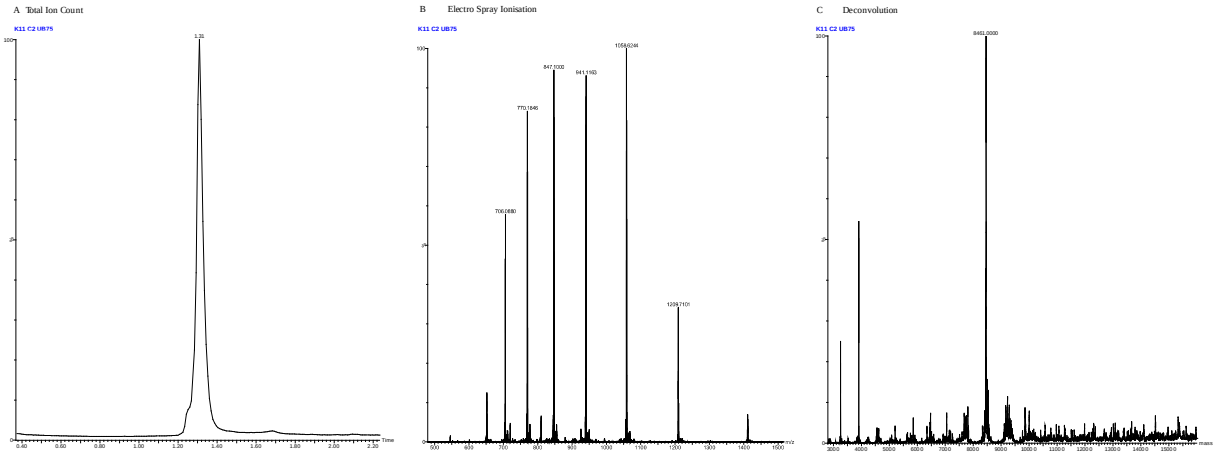


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K11 C2 Ub_{75}

LC-MS: Rt = 1.31 min., ESI MS+ (amu) deconvolution calcd: 8461.57, found 8461.00.

HRMS: $[C_{375}H_{624}N_{104}O_{117} + 6H]^{6+}$: found 1411.1619, calc. 1411.1110, $[C_{375}H_{624}N_{104}O_{117} + 7H]^{7+}$: found 1209.7101, calc. 1209.6676, $[C_{375}H_{624}N_{104}O_{117} + 8H]^{8+}$: found 1058.6244, calc. 1058.5852 $[C_{375}H_{624}N_{104}O_{117} + 9H]^{9+}$: found 941.1163, calc. 941.0766, $[C_{375}H_{624}N_{104}O_{117} + 10H]^{10+}$: found 847.1000, calc. 847.0697, $[C_{375}H_{624}N_{104}O_{117} + 11H]^{11+}$: found 770.1846, calc. 770.1550, $[C_{375}H_{624}N_{104}O_{117} + 12H]^{12+}$: found 706.0880, calc. 706.0594.



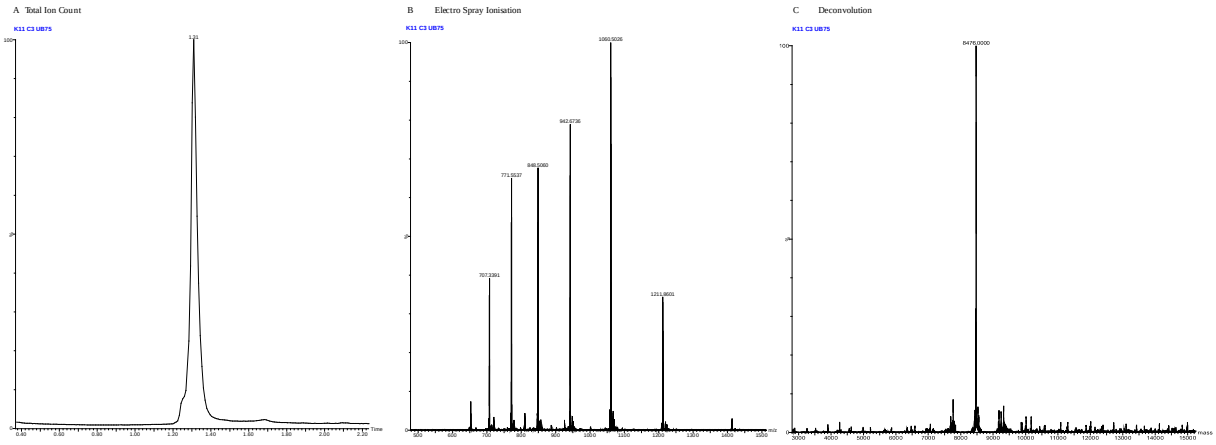
High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K11 C3 Ub_{75}

LC-MS: Rt = 1.31 min., ESI MS+ (amu) deconvolution calcd: 8475.60, found 8476.00.

HRMS: $[C_{376}H_{626}N_{104}O_{117} + 6H]^{6+}$: found 1413.4995, calc. 1413.4469, $[C_{376}H_{626}N_{104}O_{117} + 7H]^{7+}$: found 1211.7177, calc. 1211.6699, $[C_{376}H_{626}N_{104}O_{117} + 8H]^{8+}$: found 1060.3826, calc. 1060.3372, $[C_{376}H_{626}N_{104}O_{117} + 9H]^{9+}$: found 942.6736, calc. 942.6339, $[C_{376}H_{626}N_{104}O_{117} +$

10H]$^{10+}$: found 848.5060, calc. 848.4713, [$C_{376}H_{626}N_{104}O_{117}$ + 11H]$^{11+}$: found 771.4628, calc. 771.4291, [$C_{376}H_{626}N_{104}O_{117}$ + 12H]$^{12+}$: found 707.2520, calc. 707.2274.

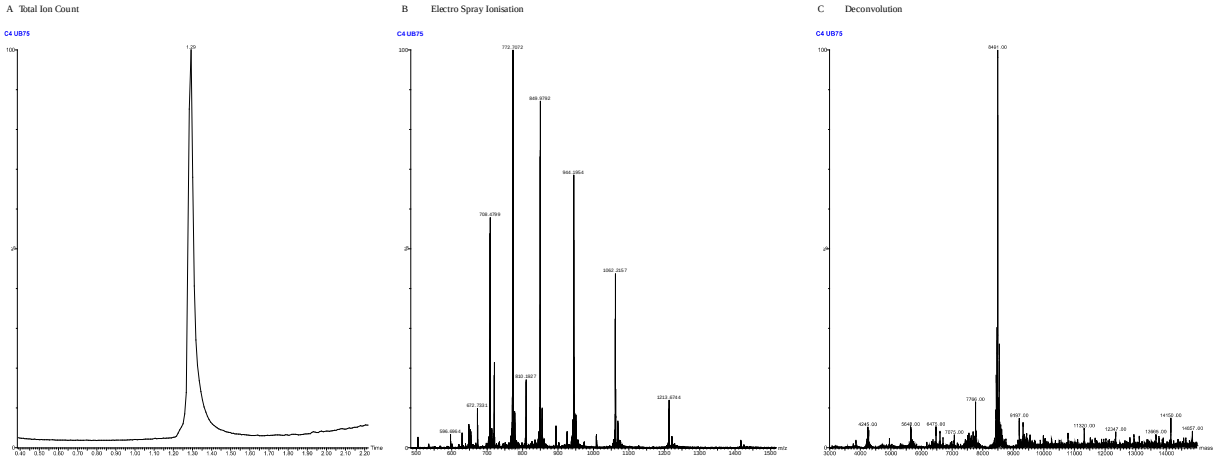


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K11 C4 Ub$_{75}$ (native lysine)

LC-MS**:** Rt = 1.29 min., ESI MS+ (amu) deconvolution calcd: 8489.62, found 8491.00.

HRMS: [$C_{377}H_{628}N_{104}O_{117}$ + 6H]$^{6+}$: found 1415.7921, calc. 1415.7828, [$C_{377}H_{628}N_{104}O_{117}$ + 7H]$^{7+}$: found 1213.6744, calc. 1213.6721, [$C_{377}H_{628}N_{104}O_{117}$ + 8H]$^{8+}$: found 1062.0957, calc. 1062.0891, [$C_{377}H_{628}N_{104}O_{117}$ + 9H]$^{9+}$: found 944.1954, calc. 944.1912, [$C_{377}H_{628}N_{104}O_{117}$ + 10H]$^{10+}$: found 849.8779, calc. 849.8729, [$C_{377}H_{628}N_{104}O_{117}$ + 11H]$^{11+}$: found 772.7072, calc. 772.7033, [$C_{377}H_{628}N_{104}O_{117}$ + 12H]$^{12+}$: found 708.3982, calc.708.3953.



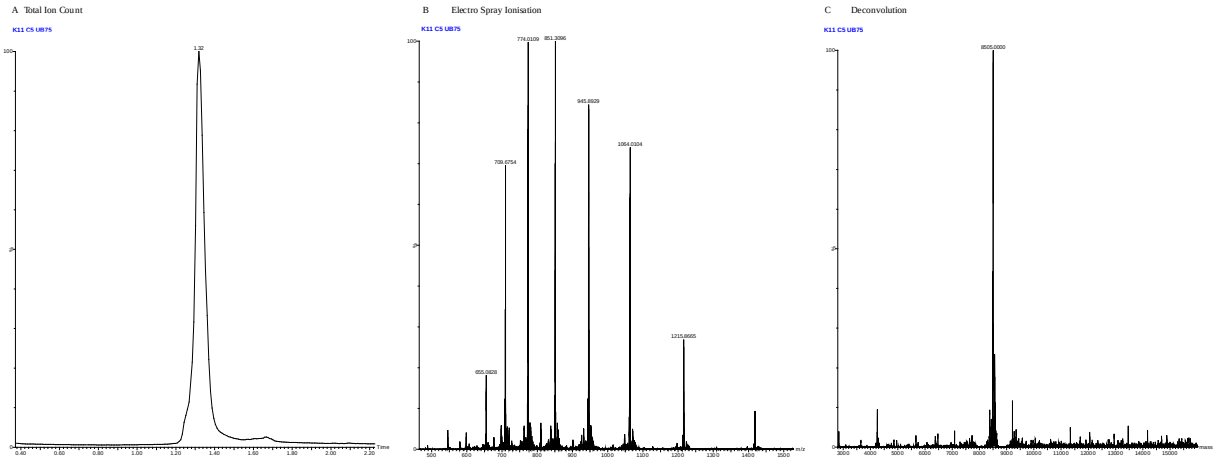
High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K11 C5 Ub$_{75}$

LC-MS: Rt = 1.32 min., ESI MS+ (amu) deconvolution calcd: 8503.65, found 8505.00

HRMS: [$C_{378}H_{630}N_{104}O_{117}$ + 6H]$^{6+}$: found 1418.1729, calc. 1418.1188, [$C_{378}H_{630}N_{104}O_{117}$ + 7H]$^{7+}$: found 1215.7238, calc. 1215.6743, [$C_{378}H_{630}N_{104}O_{117}$ + 8H]$^{8+}$: found 1063.8834, calc.

1063.8411, $[C_{378}H_{630}N_{104}O_{117} + 9H]^{9+}$: found 945.7485, calc. 945.7859, $[C_{378}H_{630}N_{104}O_{117} + 10H]^{10+}$: found 851.3096, calc. 851.2744, $[C_{378}H_{630}N_{104}O_{117} + 11H]^{11+}$: found 774.0109, calc. 773.9774, $[C_{378}H_{630}N_{104}O_{117} + 12H]^{12+}$: found 709.5937, calc. 709.5633.

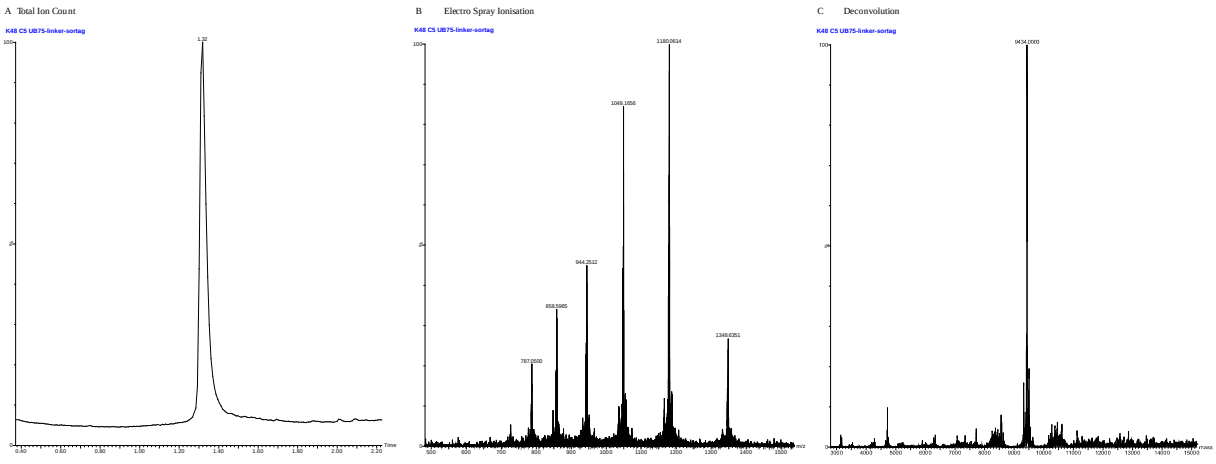


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K48 C5 Ub_{75}-SGSGS-LPETGG

LC-MS: Rt = 1.32 min., ESI MS+ (amu) deconvolution calcd: 9433.57, found 9434.00

HRMS: $[C_{415}H_{689}N_{115}O_{134} + 7H]^{7+}$: found 1348.6351, calc. 1348.5903, $[C_{415}H_{689}N_{115}O_{134} + 8H]^{8+}$: found 1180.1879, calc. 1180.1426, $[C_{415}H_{689}N_{115}O_{134} + 9H]^{9+}$: found 1049.1656, calc. 1049.1276, $[C_{415}H_{689}N_{115}O_{134} + 10H]^{10+}$: found 944.3455, calc. 944.3156, $[C_{415}H_{689}N_{115}O_{134} + 11H]^{11+}$: found 851.2801, calc. 851.2734, $[C_{415}H_{689}N_{115}O_{134} + 12H]^{12+}$: found 585.5965, calc. 858.5604, $[C_{415}H_{689}N_{115}O_{134} + 13H]^{13+}$: found 787.1303, calc. 787.0977.

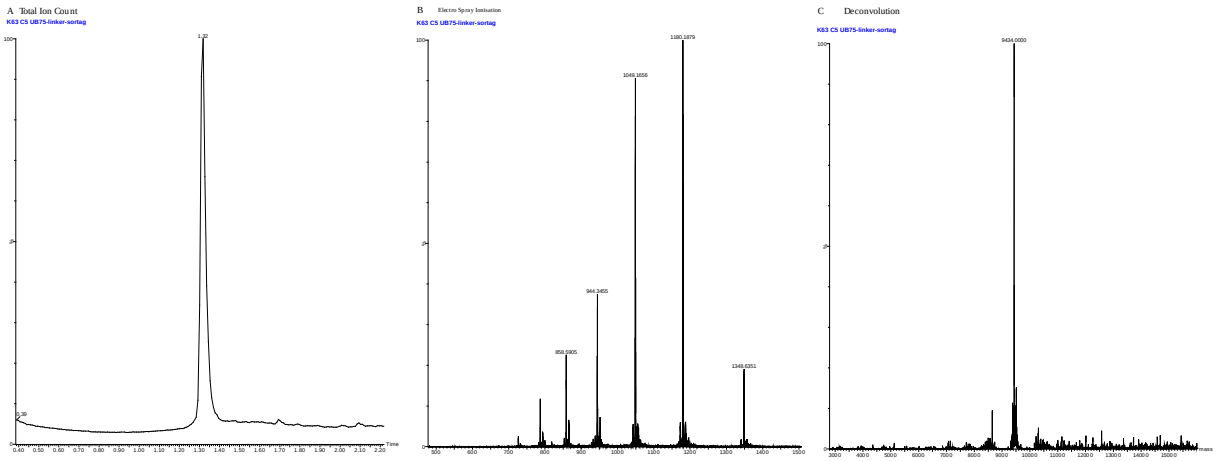


High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

K63 C5 Ub_{75}-SGSGS-LPETGG

LC-MS: Rt = 1.32 min., ESI MS+ (amu) deconvolution calcd: 9433.57, found 9434.00

HRMS: $[C_{415}H_{689}N_{115}O_{134} + 7H]^{7+}$: found 1348.6351, calc. 1348.5903, $[C_{415}H_{689}N_{115}O_{134} + 8H]^{8+}$: found 1180.1879, calc. 1180.1426, $[C_{415}H_{689}N_{115}O_{134} + 9H]^{9+}$: found 1049.1656, calc. 1049.1276, $[C_{415}H_{689}N_{115}O_{134} + 10H]^{10+}$: found 944.3455, calc. 944.3156, $[C_{415}H_{689}N_{115}O_{134} + 11H]^{11+}$: found 851.2801, calc. 851.2734, $[C_{415}H_{689}N_{115}O_{134} + 12H]^{12+}$: found 585.5905, calc. 858.5604, $[C_{415}H_{689}N_{115}O_{134} + 13H]^{13+}$: found 787.1246, calc. 787.0977.



High-Resolution Mass Spectrometry data. A) Total Ion Count graph, B)Electro Spray Ionisation Spectrum, C.) Deconvoluted mass

Supplementary Note 2

1D and 2D proton NMR

NMR experiments were recorded on a Bruker Avance AVIII 600 MHz spectrometer (^{1}H frequency; 600 MHz) equipped with a 5 mm QCI cryoprobe at 298 K or 310 K. Samples of UB_{C4}, $UB_{C4\ Bio}$ and $^{K48}UB_{C5}$ were prepared in NMR buffer (50 mM sodium phosphate, pH 6.2, 50 mM NaCl) at concentrations of approximately 200 µM supplemented with 10% D_2O. 1D spectra were acquired with 64 scans, using the watergate W5 double echo sequence with gradients[2]. 2D [^{1}H,^{1}H] spectra were recorded using a 2D NOESY pulse sequence with water flip-back and watergate water-suppression[3,4], with a mixing time of 120 ms and 32 scans. NOESY spectra were acquired with a ^{1}H spectral width of 8417.5 Hz and 1024 x 256 complex points. NMR spectra were processed using the Azara package (W. Boucher, unpublished) and analysed using CcpNmr Analysis[5].

Supplementary Note 3

Molecular Modeling Methods:

Adapting C5 for Rosetta and Molecular Dynamics Packages

Neither the molecular modelling suite Rosetta[6] nor molecular dynamics software had existing support for the noncanonical C5 amino acid. To install C5 into ubiquitin we used Rosetta and protocol for inserting noncanonical amino acids into Rosetta[7]. An SDF file of 2,7-Diaminoheptanoic Acid with idealized atom coordinates was obtained from the chemical database PubChem, then converted to a params file using a the molfile2params script in Rosetta. We also made a deprotonated variant of 2,7-Diaminoheptanoic acid containing a neutral terminal amine using Spartan. These noncanonical amino acids were installed into the ubiquitin structure using the fixbb application in Rosetta.

To run MD simulations, all residues within a protein simulated must be described by a topology file, which includes information about the type, mass, and charge of every atom of each residue involved. The quantum chemistry software Spartan was used to calculate partial charges for C5. Full ab-initio calculations using the Hartree-Fock 6-31* G basis set, These calculations were done both for individual charged and neutral C5 residues. Entries for both types were inserted into topology files with their respective calculated charge values.

Building the UB~UBE2N/UBE2V1/UB Acceptor Complex

A thioester linkage between UBE2N and the donor UB molecule was created using the UBQ_E2_Thioester protocol of the Rosetta molecular modelling suite. Briefly, this protocol simultaneously samples the thioester-linked state of an E2 enzyme and also positions the approach of an incoming nucleophile[8]. The existing version of this code set the position of the amine in plane with the carbonyl bond, which does not mimic the tetrahedral geometry of the intermediate. Using 3-D tetrahedron geometry calculations, we determined the likely position of the acceptor with respect to the active site: the K63 amine nitrogen atom of the acceptor UB molecule and Ube2V were positioned at specific coordinates to form a tetrahedral conformation with the carbon atom of the donor UB G76 carbonyl carbon.

Using the modified protocol, 10000 models were developed that had a newly-created thioester bond and transition-state geometry. To select a model, filtering was done via a lowest-RMSD search comparing donor ubiquitin to the “closed state” found in the structure 4auq. We also used RMSD to the Ube2V position in the pdb 5ait to position the Ube2V/acceptor UB complex. We then combined the lowest RMSD Ube2N/donor Ub with the lowest RMSD Ube2V/acceptor UB. The Ube2V molecule of the revised complex structure was subjected to the Rosetta Relax

application while holding all other chains fixed to remove any side-chain clashes that may have resulted from merging structures.

This complex was used as a template to develop a C5 version of the model that maintained that amine acceptor position. To do so, a $^{K63}UB_{C5}$ version of the acceptor UB was formed using of the fixbb application of the Rosetta modelling suite. This model maintained identical atomic coordinates to the C4 acceptor UB model at all positions except for side-chain atoms at position 63. To determine the $^{K63}UB_{C5}$ side-chain torsion angle, side-chain torsion angle sets for C5 were retrieved from each frame of all free $^{K63}UB_{C5}$ simulations. We selected the best side-chain rotamer based on overlap in positioning of the side-chain nitrogen atom when comparing to the native acceptor UB structure's C4 residue. Additional manual reflection of the residue's side-chain was used to reduce clashing with other surrounding residues. This model was then used to replace the native acceptor UB structure to form the C5 complex model.

For the molecular dynamics simulations, additional values must be provided as parameter information for the thioester bond. Bond, angle, and dihedral entries were adapted from derived thioester bond values in Amaro *et al.* and implemented into parameter files[9]. For partial charges, values were adapted from thioester bond charges derived in Jones *et al.*[10]. These partial charge values were formatted into a patch residue topology entry for PSF generation.

Molecular Dynamics of UB and the UB~UBE2N/UBE2V1/UB complex

All MD was performed using the CHARMM36 force field and NAMD simulation software[11,12]. For free ubiquitin simulations, the structure was solvated in a rectilinear water box composed of TIP3P water molecules (r = 5.0 Å). PSF generation and solvation were done using the autopsf and solvate plug-ins of the VMD molecular visualization program. During simulation, the system was subjected to optimization stabilization, heating to 298 K, then an equilibration process of fifty nanoseconds using 2 femtosecond time steps. Runs were executed using Langevin dynamics, the Particle Mesh Ewald method, and constant pressure control.

Complex simulations were also executed with settings similar to the aforementioned. The structure was solvated in a rectilinear TIP3 water box (r = 12.5 Å). Complex simulations involved optimization, heating to 298 K, and an equilibration step of twenty-five nanoseconds using 2 femtosecond timesteps, with all other settings remaining the same as previously described for free ubiquitin runs.

Analysis of Molecular Dynamics Simulations

The Python library MDAnalysis was used to evaluate all MD simulations completed[13]. To determine structure preservation, a Python script was developed that utilizes the MDAnalysis module for RMSD alignment and calculations[14].A separate script was created to specifically calculate the RMSD for the gate loop region. To study residue flexibility in the free ubiquitin

runs, we analyzed the Chi torsion angles present at residue positions 11, 48 and 63. Using the dihedrals module from the MDAnalysis package, the torsion angles were retrieved at each frame of trajectory files associated with MD simulation runs.

Additional Python scripts were developed to assess the values based on the frequency and extent of angle change associated with each torsion angle. To study the degrees of freedom, we divided the conformational space into bins, similar to what has been done for side-chain dependent rotamer libraries[15], where phi and psi angles are divided into 10 degree bins (36 total bins) and the chi angles are divided into 120 degree bins (3 bins). We measured how frequently each torsion angle moved between bins during the simulation time of both the free and complex trajectories. We also used this framework to catalogue the unique rotameric states for each residue. To measure the C4 versus C5 methylene ruler we measured from the Cα atom to the side-chain nitrogen atom in all simulations. For the MD simulations of the ubiquitin complex, the distance between the side chain nitrogen atom of acceptor UB residue 63 and the carbonyl group on G76 of on the donor UB was plotted as a function of simulation time.

Python scripts were also created to measure the time of specific geometries created between the thioester bond and position 63 C4/C5 residues in the complex simulations. The angle of the acceptor side chain relative to the active site were measured by the torsion angle between the acceptor UB Cα and side-chain amine and UBE2N Cβ and S. At every frame of the complex simulations, this angle was recorded using a MDAnalysis-utilizing script along with the atomic distance between the amine and the cysteine. These angle/distance combinations were then filtered by two different types of binning. Combinations were categorized by 3 Å distance bins that ranged from 2-20 Å. 12 30-degree bins were also used to further classify the combinations.

Supplementary Note references:

1 El Oualid, F. *et al.* Chemical synthesis of ubiquitin, ubiquitin-based probes, and diubiquitin. *Angew Chem Int Ed Engl* **49**, 10149-10153 (2010).

2 Liu, M. *et al.* Improved WATERGATE Pulse Sequences for Solvent Suppression in NMR Spectroscopy. *Journal of Magnetic Resonance* **132**, 125 (1998).

3 Piotto, M., Saudek, V. & Sklenar, V. Gradient-tailored excitation for single-quantum NMR spectroscopy of aqueous solutions. *J Biomol NMR* **2**, 661-665, doi:10.1007/BF02192855 (1992).

4 Sklenar, V., Piotto, M., Leppik, R. & Saudek, V. Gradient-tailored water suppression for 1H-15N HSQC experiments optimized to retain full sensitivity. *Journal of Magnetic Resonance* **102**, 241-245 (1993).

5 Vranken, W. F. *et al.* The CCPN data model for NMR spectroscopy: development of a software pipeline. *Proteins* **59**, 687-696, doi:10.1002/prot.20449 (2005).

6 Leman, J. K. *et al.* Macromolecular modeling and design in Rosetta: recent methods and frameworks. *Nat Methods* **17**, 665-680 (2020).

7 Renfrew, P. D., Choi, E. J., Bonneau, R. & Kuhlman, B. Incorporation of Noncanonical Amino Acids into Rosetta and Use in Computational Protein-Peptide Interface Design. *PLoS One* **7**, e32637 (2012).
8 Hill, S., Harrison, J. S., Lewis, S. M., Kuhlman, B. & Kleiger, G. Mechanism of Lysine 48 Selectivity during Polyubiquitin Chain Formation by the Ube2R1/2 Ubiquitin-Conjugating Enzyme. *Mol Cell Biol* **36**, 1720-1732 (2016).
9 Amaro, R., Daliwal, B. & Luthey-Schulten, Z. Parameterizing a novel residue. (2014).
10 Jones, W. M., Davis, A. G., Wilson, R. H., Elliott, K. L. & Sumner, I. A conserved asparagine in a ubiquitin-conjugating enzyme positions the substrate for nucleophilic attack. *J Comput Chem* **40**, 1969-1977 (2019).
11 Phillips, J. C. *et al.* Scalable molecular dynamics with NAMD. *J Comput Chem* **26**, 1781-1802 (2005).
12 Brooks, B. R. *et al.* CHARMM: the biomolecular simulation program. *J Comput Chem* **30**, 1545-1614 (2009).
13 R. J. Gowers *et al.* MDAnalysis: A Python package for the rapid analysis of molecular dynamics simulations. . *Proceedings of the 15th Python in Science Conference*, 98-105 (2016).
14 Liu, P., Agrafiotis, D. K. & Theobald, D. L. Fast determination of the optimal rotational matrix for macromolecular superpositions. *J Comput Chem* **31**, 1561-1563 (2010).
15 Shapovalov, M. V. & Dunbrack, R. L., Jr. A smoothed backbone-dependent rotamer library for proteins derived from adaptive kernel density estimates and regressions. *Structure* **19**, 844-858 (2011).