Supporting Information

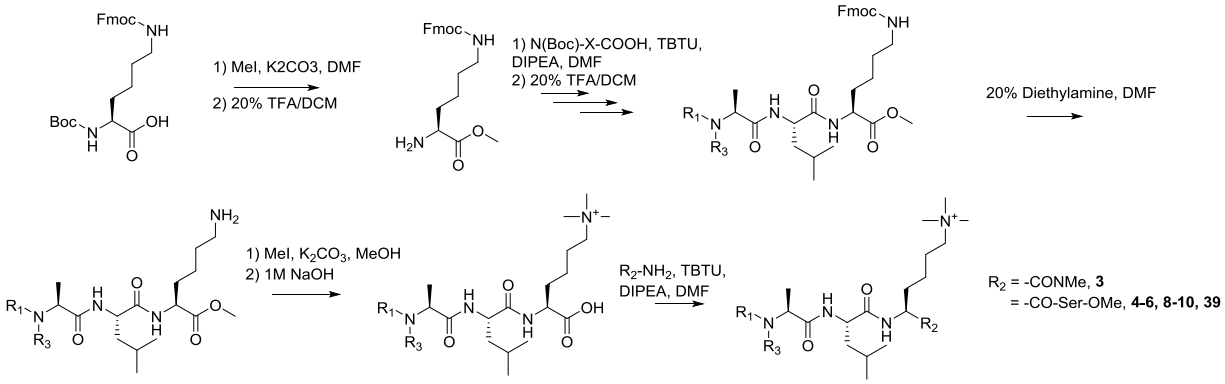
Structure-activity relationships and kinetic studies of peptidic antagonists of CBX chromodomains

Jacob I. Stuckey, Catherine Simpson, Jacqueline L. Norris-Drouin, Stephanie H. Cholensky, Junghyun Lee, Ryan Pasca, Nancy Cheng, Bradley M. Dickson†, Kenneth H. Pearce, Stephen V. Frye, Lindsey I. James**

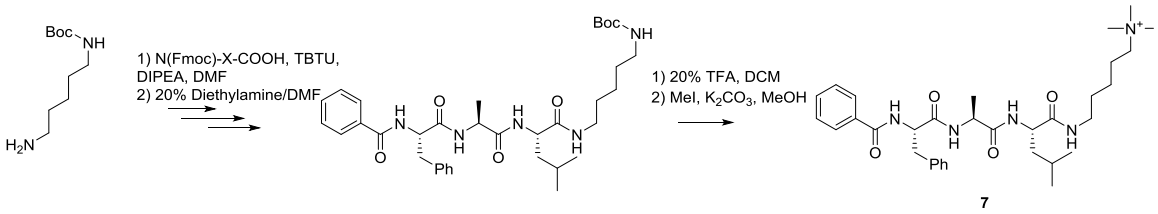
Center for Integrative Chemical Biology and Drug Discovery, Division of Chemical Biology and Medicinal Chemistry, UNC Eshelman School of Pharmacy, University of North Carolina at Chapel Hill, Chapel Hill, North Carolina 27599, USA.

Table of Contents

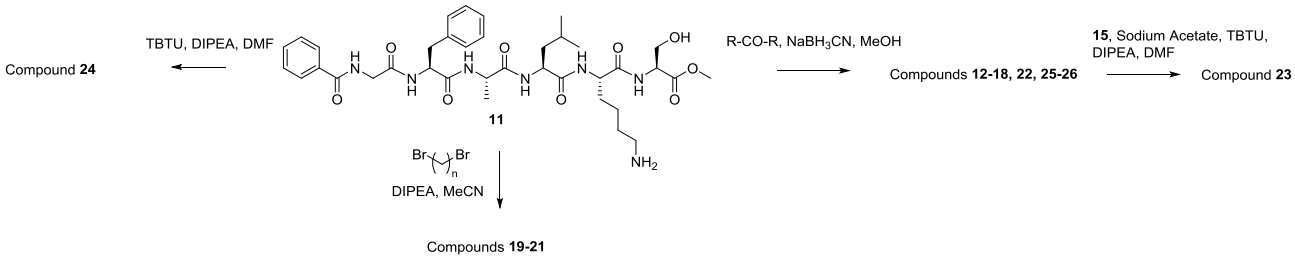
Synthetic Scheme 1.



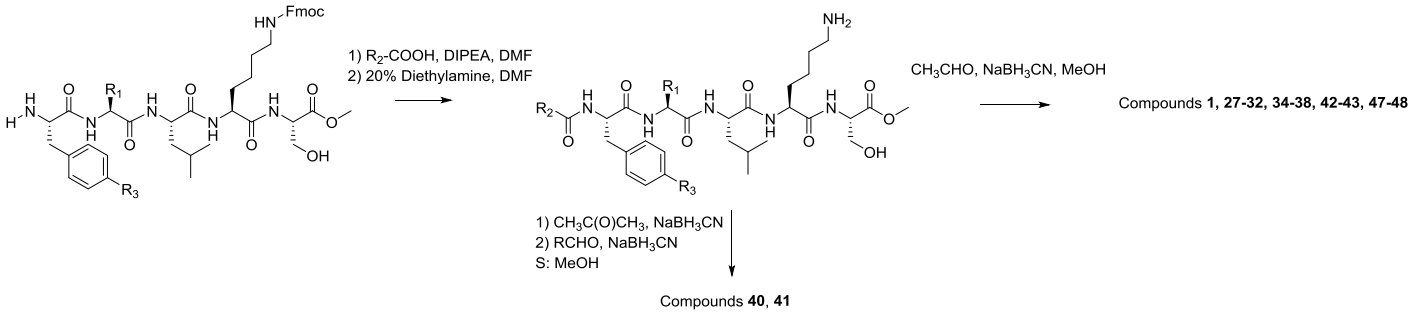
Synthetic Scheme 2.



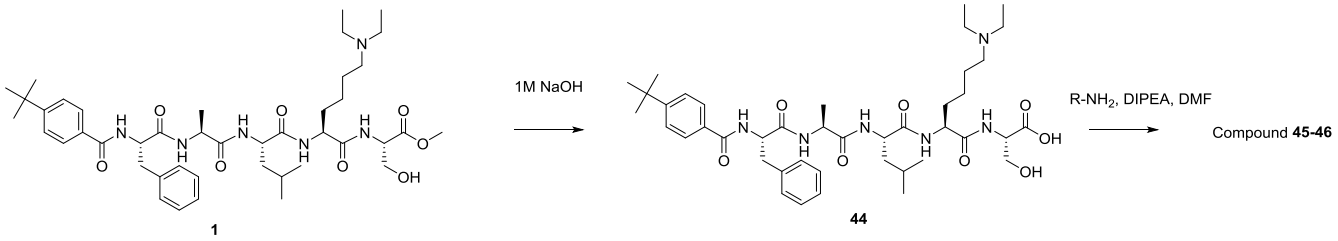
Synthetic Scheme 3.



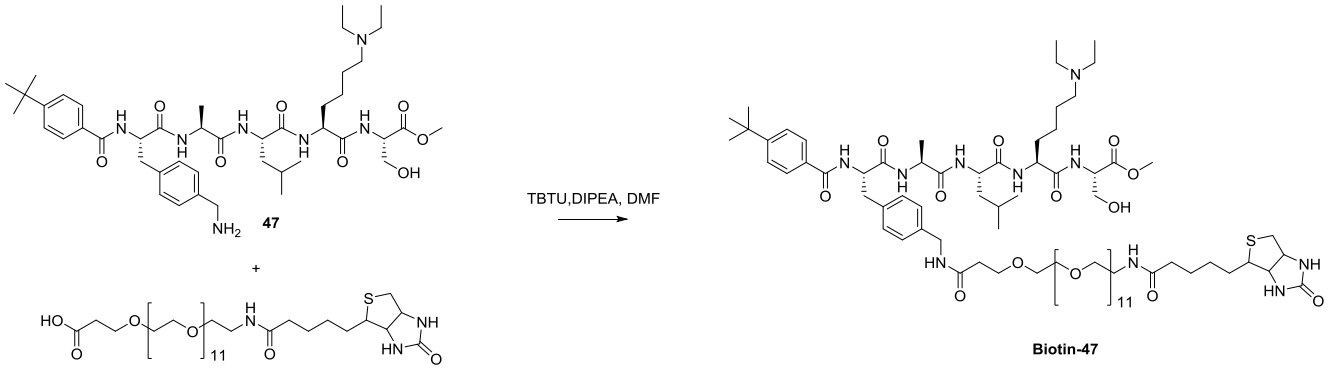
Synthetic Scheme 4.

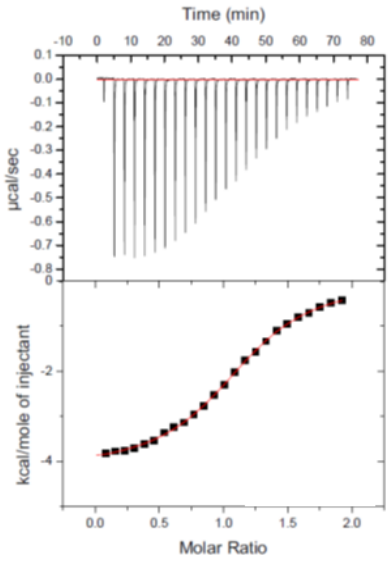


Synthetic Scheme 5.

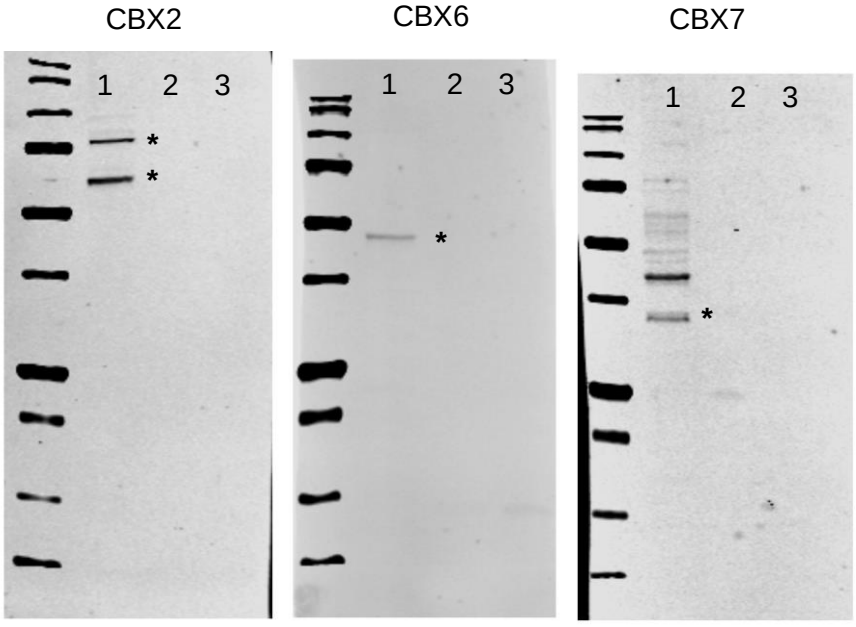


Synthetic Scheme 6

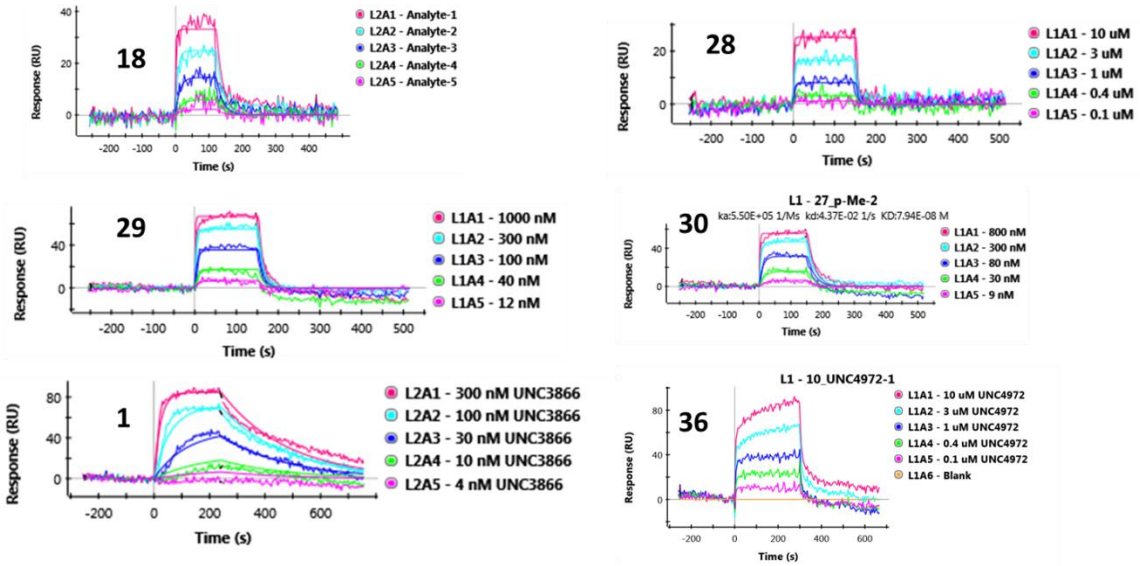




Supplementary Figure 1. Representative ITC curve of compound **6** binding to CBX7 (K_d = 10.3 ± 0.9 µM).



Supplementary Figure 2. Western blots from pulldown Experiments with biotin-47. Target protein bands are marked with an *. For CBX2, two isoforms are known to exist, resulting in the presence of two CBX2 bands. For each experiment, lane **1** = input, lane **2** = biotin-47 pull-down and lane **3** = biotin-47 pull-down + 100 µM UNC3866 (**1**).



Supplementary Figure 3. SPR experiments with CBX7. Representative curves for each compound are shown. Compound numbers are indicated on the respective sensorgrams.

Protein Expression and Purification

All expression constructs were transformed into Rosetta BL21(DE3)pLysS competent cells (Novagen, EMD Chemicals, San Diego, CA). Protein expression was induced by growing cells at 37°C with shaking until the OD600 reached ~0.6-0.8 at which time the temperature was lowered to 18°C and expression was induced by adding 0.5mM IPTG and continuing shaking overnight. Cells were harvested by centrifugation and pellets were stored at -80°C.

His-tagged proteins were purified by re-suspending thawed cell pellets in 30ml of lysis buffer (50mM sodium phosphate pH 7.2, 50mM NaCl, 30mM imidazole, 1X EDTA free protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)) per liter of culture. Cells were lysed on ice by sonication with a Branson Digital 450 Sonifier (Branson Ultrasonics, Danbury, CT) at 40% amplitude for 12 cycles with each cycle consisting of a 20 second pulse followed by a 40 second rest. The cell lysate was clarified by centrifugation and loaded onto a HisTrap FF column (GE Healthcare, Piscataway, NJ) that had been preequilibrated with 10 column volumes of binding buffer (50mM sodium phosphate pH 7.2, 500mM NaCl, 30mM imidazole) using an AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was washed with 15 column volumes of binding buffer and protein was eluted in a linear gradient to 100% elution buffer (50mM sodium phosphate pH 7.2, 500mM NaCl, 500mM imidazole) over 20 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2ml in Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare, Piscataway, NJ) that had been preequilibrated with 1.2 column volumes of sizing buffer (25mM Tris pH 7.5, 250mM NaCl, 2mM DTT, 5% glycerol) using an ATKA Purifier (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2ml/min collecting 3ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

GST-tagged proteins were purified by re-suspending thawed cell pellets in 30ml of lysis buffer (1xPBS, 5mM DTT, 1X EDTA free protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)) per liter of culture. Cells were lysed on ice by sonication as described for His-tagged proteins. Clarified cell lysate was loaded onto a GSTrap FF column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 10 column volumes of binding buffer (1xPBS, 5mM DTT) using a AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was

washed with 10 column volumes of binding buffer and protein was eluted in 100% elution buffer (50mM Tris pH 7.5, 150mM NaCl, 10mM reduced glutathione) over 10 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2ml in Amicon Ultra-15 concentrators, 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare, Piscataway, NJ) that had been preequilibrated with 1.2 column volumes of sizing buffer (25mM Tris pH 7.5, 250mM NaCl, 2mM DTT, 5% glycerol) using an ATKA FPLC (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2ml/min collecting 3ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

Supplementary Table 1. Protein-bait ligand concentrations used for AlphaScreen® assays.

Protein (final assay concentration)	Bait ligand (final assay concentration)
CBX4 (10 nM)	UNC4195 (**2**) (10 nM)
CBX7 (15 nM)	UNC4195 (**2**) (15 nM)
CBX8 (125 nM)	UNC4195 (**2**) (15 nM)
CDYL2 (8 nM)	UNC4195 (**2**) (8.0 nM)
CBX5 (31 nM)	ARTKQTARK(Me3)STGGKAPRKQL-K(Biotin)-NH2 (65 nM)

ITC Experiments

All ITC measurements were recorded at 25 °C with an AutoITC200 microcalorimeter (MicroCal Inc., MA). All protein and compound stock samples were in the target buffer (25 mM Tris-HCl, pH 8, 150 mM NaCl, and 2 mM β-mercaptoethanol), and then diluted in the same buffer to achieve the desired concentrations: 100 µM protein and 1.0 mM compound. The concentration of the protein stock solution was established using the Edelhoch method, whereas compound stock solutions were prepared based on mass. A typical experiment included a single 0.2 µl compound injection into a 200 µl cell filled with protein, followed by 26 subsequent 1.5 µl injections of compound. Injections were performed with a spacing of 180 seconds and a reference power of 8 µcal/sec. Control experiments were performed titrating each compound into buffer under identical conditions to determine the heat signals, if any, that arise from diluting the compound. If applicable, the heats of dilution generated were then subtracted from the

protein-compound binding curves. The initial data point was routinely deleted. The titration data was analyzed using Origin Software (MicroCal Inc., USA) by non-linear least squares, fitting the heats of binding as a function of the compound:protein ratio to a one site binding model.

Cell culture and lysis

PC3 cells were obtained from ATCC® (CRL-1435™) through the UNC Lineberger Tissue Culture Facility. Cells were cultured using GIBCO® DMEM/F12 (Ham), [+] L-Glutamine, and [+] 15 mM HEPES media. Cells were trypsinized using 0.25% trypsin. Lysis was performed using Cytobuster™ protein extraction reagent supplemented with protease inhibitors and Benzonase® (used at 25 U/mL). Samples were incubated at 37°C for 10 minutes, followed by incubation at RT on a rotator for 20 minutes. The samples were spun down and the supernatant collected and transferred to a clean Eppendorf tube. Protein concentrations were quantified using the Bradford protein assay.

General chemistry procedures

Analytical LCMS data for all compounds were acquired using an Agilent 6110 Series system with the UV detector set to 220 nm. Samples were injected (<10 µL) onto an Agilent Eclipse Plus 4.6 × 50 mm, 1.8 µm, C18 column at room temperature. Mobile phases A (H_2O + 0.1% acetic acid) and B (MeOH + 0.1% acetic acid) were used with a linear gradient from 10% to 100% B in 5.0 min, followed by a flush at 100% B for another 2 minutes with a flow rate of 1.0 mL/min. Mass spectra (MS) data were acquired in positive ion mode using an Agilent 6110 single quadrupole mass spectrometer with an electrospray ionization (ESI) source. Reverse phase column chromatography was performed with a Teledyne Isco CombiFlash®R_f 200 using C18 RediSep®R_f Gold columns with the UV detector set to 220 nm and 254 nm. Mobile phases of A (H_2O + 0.1% TFA) and B (MeOH or MeCN) were used with default column gradients. Preparative HPLC was performed using an Agilent Prep 1200 series with the UV detector set to 220 nm and 254 nm. Samples were injected onto a Phenomenex Luna 250 × 30 mm, 5 µm, C18 column at room temperature. Mobile phases of A (H_2O + 0.1% TFA) and B (MeOH or MeCN) were used with a flow rate of 40 mL/min. A general gradient of 0-15 minutes increasing from 10 to 100% B, followed by a 100% B flush for another 5 minutes. Small variations in this purification method were made as needed to achieve ideal separation for each compound. Analytical LCMS (at 220 nm) was used to establish the purity of targeted compounds. All

compounds that were evaluated in biochemical and biophysical assays had >95% purity as determined by LCMS.

Amide couplings followed by Boc removal

To a solution of 0.1 M carboxylic acid (1.0 eq) in DMF was added 1.2 equivalents of amine, 1.3 equivalents of TBTU and 2.2 equivalents of DIPEA. The reaction was stirred at room temperature for 30 min to overnight. The mixture was diluted with EtOAc (100 mL) and washed 3X with brine (100 mL). The organic layer was then dried over Na_2SO_4, filtered and concentrated by rotary evaporation. The resulting residue was dissolved in a mixture of 20% TFA in DCM. The solution was stirred at room temperature for 30 minutes to overnight. The mixture was concentrated by rotary evaporation and purified by reverse-phase column chromatography ((Water + 0.1% TFA)/Methanol) to yield the TFA salt of the target compound. Typical yields ranged from 60-90%.

Amide couplings followed by FMOC-removal

To a solution of 0.1 M carboxylic acid (1.0 eq) in DMF was added 1.2 equivalents of amine, 1.3 equivalents of TBTU and 2.2 equivalents of DIPEA. The reaction was stirred at room temperature for 30 min to overnight. The mixture was diluted with EtOAc (100 mL) and washed 3X with brine (100 mL). The organic layer was then dried over Na_2SO_4, filtered and concentrated by rotary evaporation. The resulting residue was dissolved in a mixture of 20% diethylamine in DMF. The solution was stirred at room temperature for 30 minutes to overnight. The mixture was concentrated by rotary evaporation and purified by reverse-phase column chromatography ((Water + 0.1% TFA)/Methanol) to yield the TFA salt of the target compound. Typical yields ranged from 40-70%.

Reductive aminations with aldehydes

The lysine containing peptide (1.0 eq) was dissolved in methanol at a concentration of 0.1 M. Sodium cyanoborohydride (3.0 eq) and aldehyde (8.0 eq.) were added and the solution was stirred at RT until the amine starting material was consumed. The mixture was then concentrated and purified by reverse phase column chromatography ((H_20 + 0.1% TFA)/MeOH) to yield the target compound as a TFA salt. Typical yields ranged from 60-90%.

Reductive aminations with ketones

The lysine containing peptide (1.0 eq) was dissolved in methanol at a concentration of 0.1 M. Sodium cyanoborohydride (3.0 eq) and aldehyde (8.0 eq.) were added and the solution was stirred at 50°C for 2 hours until the amine starting material was consumed. The mixture was then concentrated and purified by reverse phase column chromatography ((H_20 + 0.1% TFA)/MeOH) to yield the target compound as a TFA salt. Typical yields ranged from 40-70%.

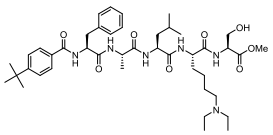
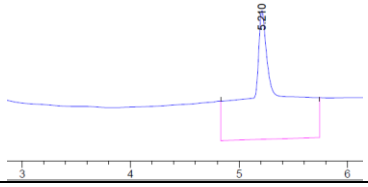
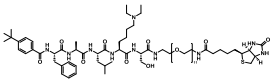
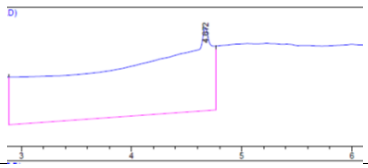
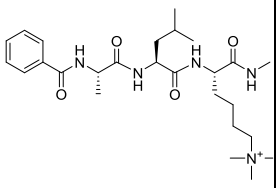
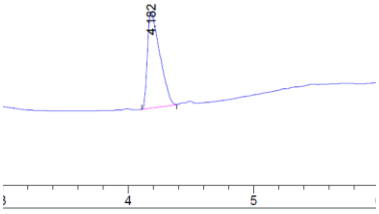
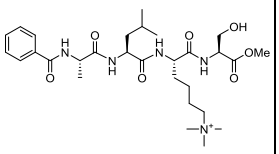
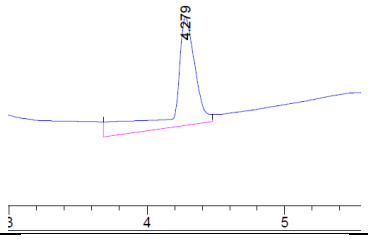
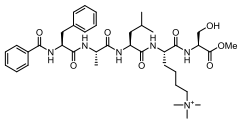
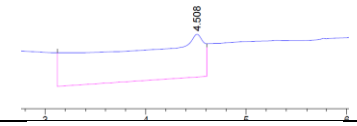
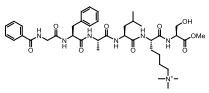
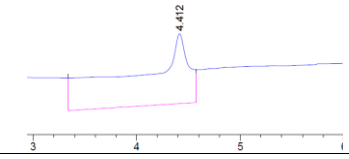
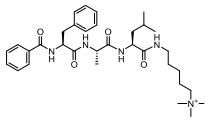
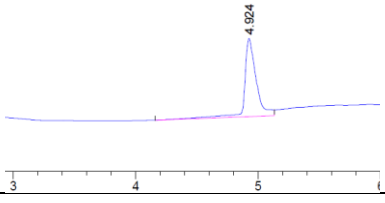
Cyclic alkylations

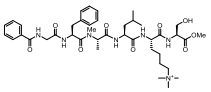
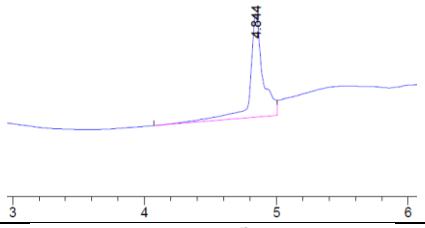
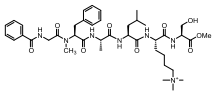
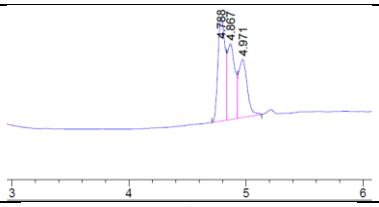
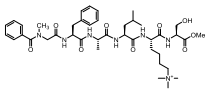
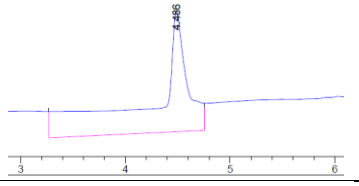
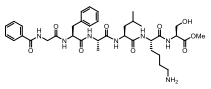
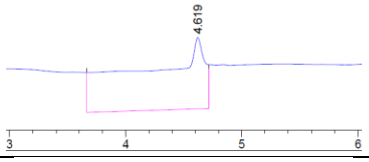
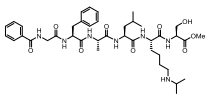
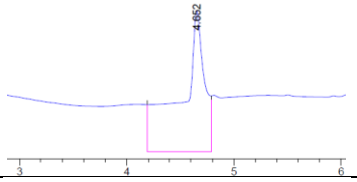
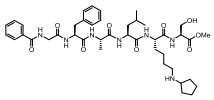
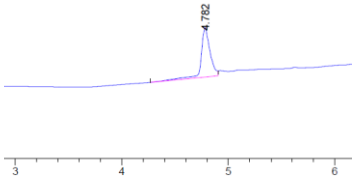
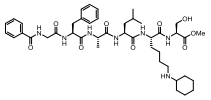
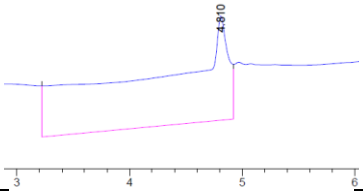
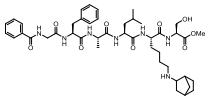
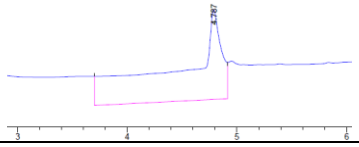
The lysine containing peptide (1.0 eq) was dissolved in MeCN at a final concentration of 0.1 M. Dibromoalkane (1.2 eq) and DIPEA (3.1 eq) were added. The solution was heated at 110°C in a microwave reactor for 1 hour. The solution was then concentrated by rotary evaporation and the residue was purified by HPLC ((H_20 + 0.1% TFA)/MeOH) to yield the target compound as a TFA salt. Yields ranged from 25-70%.

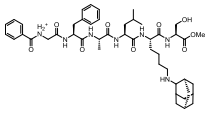
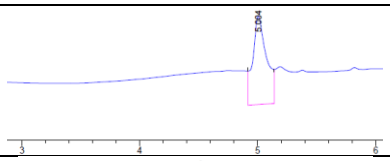
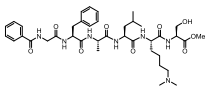
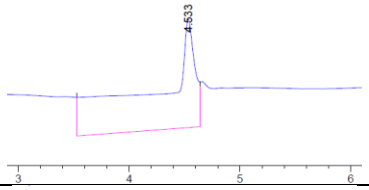
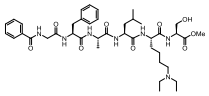
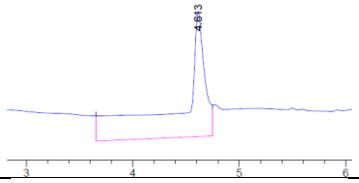
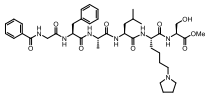
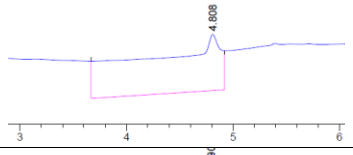
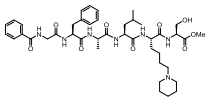
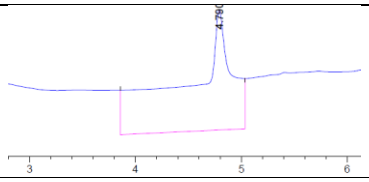
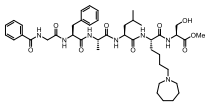
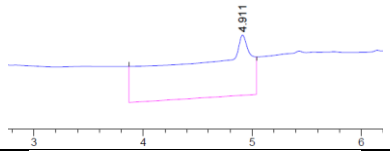
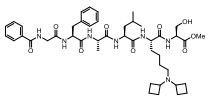
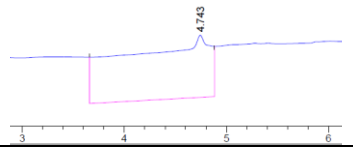
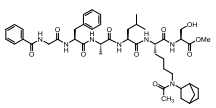
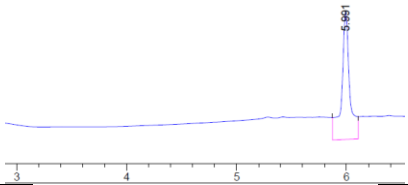
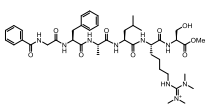
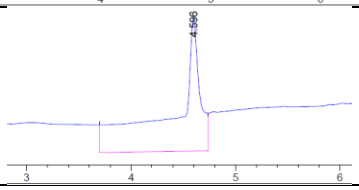
Lysine Trimethylation

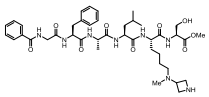
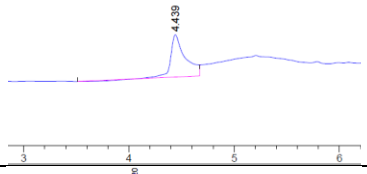
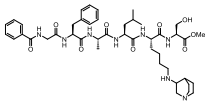
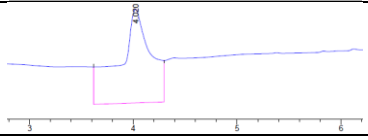
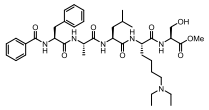
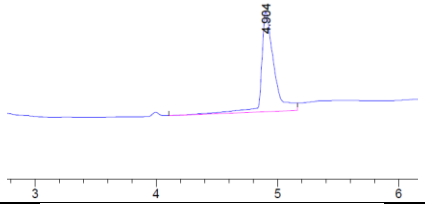
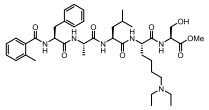
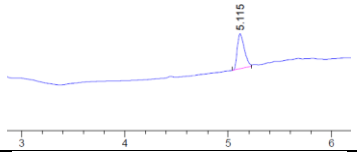
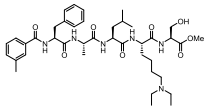
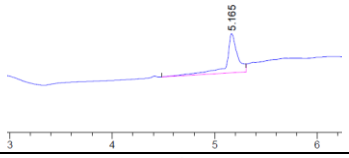
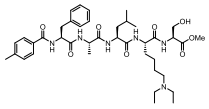
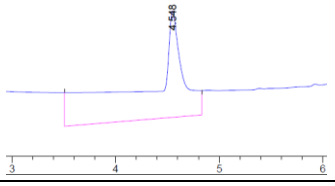
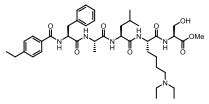
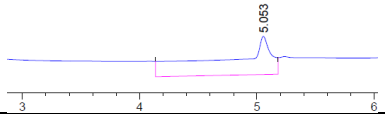
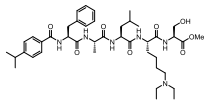
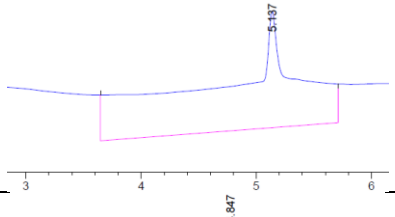
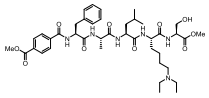
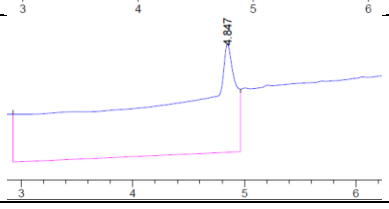
The lysine containing peptide (1.0 eq) was dissolved in MeOH at a final concentration of 0.1 M. Iodomethane (10.0 eq) and K_2CO_3 (6.0 eq) were added and the mixture was stirred overnight at room temperature. The mixture was concentrated by rotary evaporation and dissolved in a solution of 1M NaOH at a concentration of 0.1 M and stirred for 30 minutes at room temperature. The solution was then brought to pH 7 with 37% aq HCl and concentrated by rotary evaporation. The residue was purified by HPLC ((H_20 + 0.1% TFA)/MeCN) to yield the target compound as a TFA salt. Typical yields ranged from 50-80%.

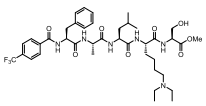
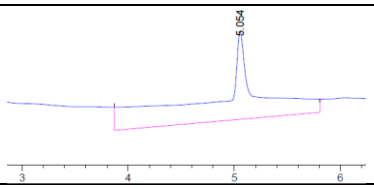
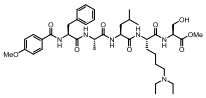
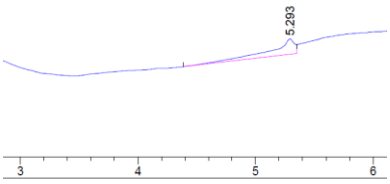
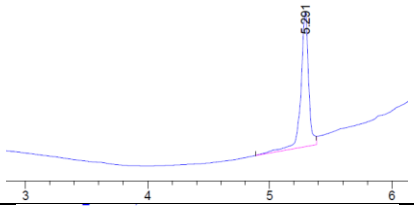
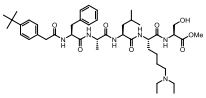
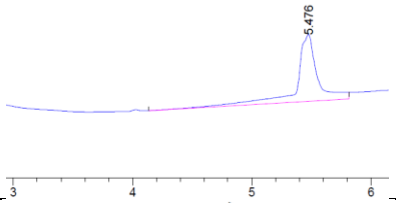
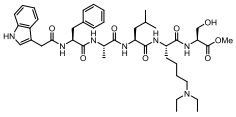
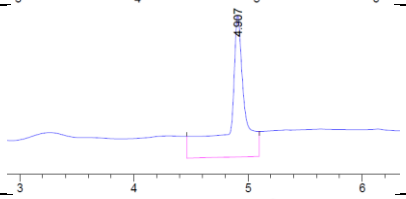
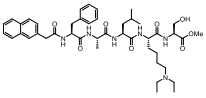
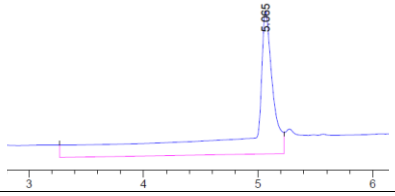
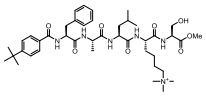
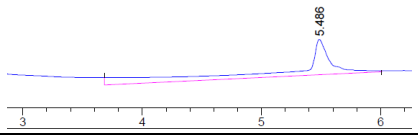
Supplementary Table 2. LC and MS characterization of final compounds.

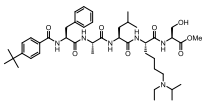
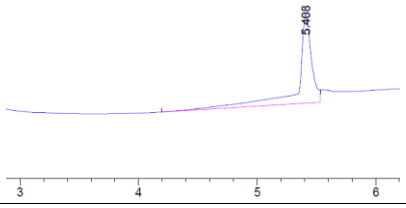
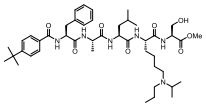
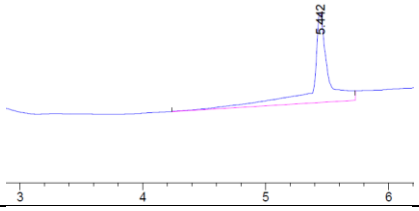
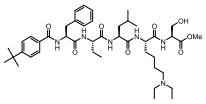
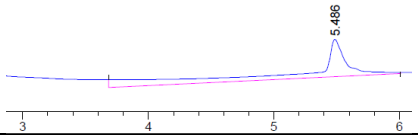
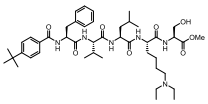
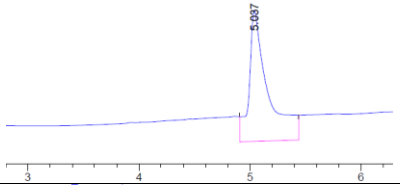
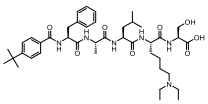
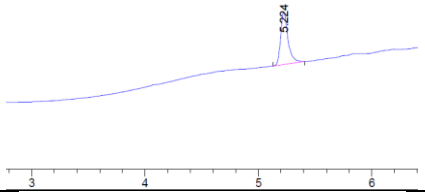
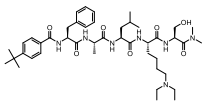
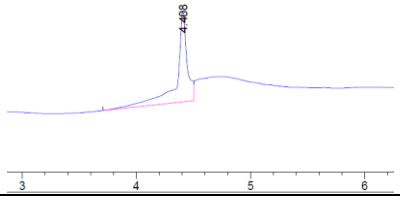
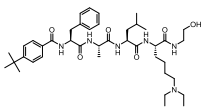
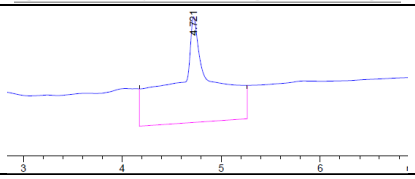
Compound ID	Structure	LCMS UV Trace (220 nm)	Calculated Monoisotopic Mass $[M]^+$ or $[M+H]^+$	Observed Mass(es)
1			795.50	409.35 $[M+H+Na]^{2+}$ 795.50 $[M+H]^+$
2			1532.90	511 $[M+3H]^{3+}$ 767 $[M+2H]^{2+}$
3			490.34	490.35 $[M]^+$
4			578.35	578.40 $[M]^+$
5			725.42	363.30 $[M+H]^{2+}$ 725.40 $[M]^+$
6			782.44	391.80 $[M+H]^{2+}$ 782.45 $[M]^+$
7			580.39	580.40 $[M]^+$

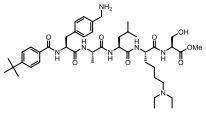
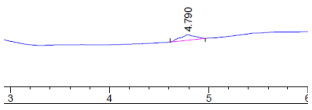
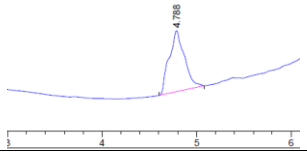
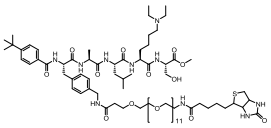
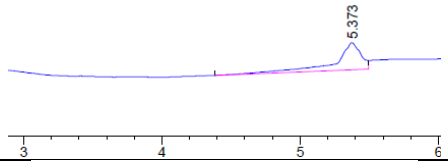
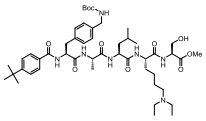
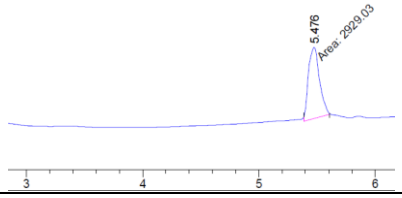
Compound ID	Structure	LCMS UV Trace (220 nm)	Calculated Monoisotopic Mass $[M]^+$ or $[M+H]^+$	Observed Mass(es)
8			796.46	796.45 $[M]^+$
9			796.46	398.80 $[M+H]^{2+}$ 796.40 $[M]^+$
10			796.46	398.80 $[M+H]^{2+}$ 796.40 $[M]^+$
11			740.40	370.80 $[M+2H]^{2+}$ 740.40 $[M+H]^+$ 762.40 $[M+Na]^+$
12			782.44	402.80 $[M+H+Na]^{2+}$ 782.40 $[M+H]^+$
13			808.46	808.45 $[M+H]^+$
14			822.48	411.85 $[M+2H]^{2+}$ 422.80 $[M+H+Na]^{2+}$ 822.45 $[M+H]^+$
15			834.48	417.85 $[M+2H]^{2+}$ 834.50 $[M+H]^+$

Compound ID	Structure	LCMS UV Trace (220 nm)	Calculated Monoisotopic Mass $[M]^+$ or $[M+H]^+$	Observed Mass(es)
16			874.51	448.85 $[M+H+Na]^{2+}$ 874.50 $[M+H]^+$
17			768.43	384.80 $[M+2H]^{2+}$ 768.40 $[M+H]^+$
18			796.46	398.80 $[M+2H]^{2+}$ 409.80 $[M+H+Na]^{2+}$ 796.45 $[M+H]^+$
19			794.44	397.75 $[M+2H]^{2+}$ 408.80 $[M+H+Na]^{2+}$ 794.40 $[M+H]^+$
20			808.46	404.75 $[M+2H]^{2+}$ 415.75 $[M+H+Na]^{2+}$ 808.40 $[M+H]^+$
21			822.48	411.80 $[M+2H]^{2+}$ 422.80 $[M+H+Na]^{2+}$ 822.50 $[M+H]^+$
22			848.49	424.90 $[M+2H]^{2+}$ 435.85 $[M+H+Na]^{2+}$ 848.50 $[M+H]^+$
23			876.49	876.00 $[M+H]^+$ 878.00 $[M+Na]^+$
24			838.48	419.80 $[M+2H]^{2+}$ 430.85 $[M+H+Na]^{2+}$ 838.45 $[M+H]^+$

Compound ID	Structure	LCMS UV Trace (220 nm)	Calculated Monoisotopic Mass $[M]^+$ or $[M+H]^+$	Observed Mass(es)
25			809.46	405.35 $[M+2H]^{2+}$ 416.25 $[M+H+Na]^{2+}$ 809.50 $[M+H]^+$
26			849.49	425.30 $[M+2H]^{2+}$ 849.45 $[M+H]^+$
27			739.44	739.40 $[M+H]^+$
28			753.45	753.45 $[M+H]^+$
29			753.45	377.30 $[M+2H]^{2+}$ 754.45 $[M+H]^+$
30			753.45	388.10 $[M+H+Na]^{2+}$ 753.10 $[M+H]^+$
31			767.47	767.50 $[M+H]^+$
32			781.49	402.30 $[M+H+Na]^{2+}$ 781.50 $[M+H]^+$
33			797.44	797.40 $[M+H]^+$

Compound ID	Structure	LCMS UV Trace (220 nm)	Calculated Monoisotopic Mass $[M]^+$ or $[M+H]^+$	Observed Mass(es)
34			807.43	807.40 $[M+H]^+$
35		220 nm  254 nm 	769.45	385.25 $[M+2H]^{2+}$ 769.45 $[M+H]^+$
36			809.52	809.50 $[M+H]^+$
37			792.47	407.80 $[M+H+Na]^{2+}$ 792.45 $[M+H]^+$
38			803.47	803.50 $[M+H]^+$
39			781.49	391.35 $[M+H]^+$ 781.45 $[M]^+$

Compound ID	Structure	LCMS UV Trace (220 nm)	Calculated Monoisotopic Mass $[M]^+$ or $[M+H]^+$	Observed Mass(es)
40			809.52	405.35 $[M+2H]^{2+}$ 809.50 $[M+H]^+$
41			823.53	823.50 $[M+H]^+$
42			809.52	809.50 $[M+H]^+$
43			823.53	823.15 $[M+H]^+$
44			781.49	391.20 $[M+2H]^{2+}$ 402.15 $[M+H+Na]^{2+}$ 781.10 $[M+H]^+$
45			808.53	404.65 $[M+2H]^{2+}$ 808.20 $[M+H]^+$
46			737.50	737.20 $[M+H]^+$

Compound ID	Structure	LCMS UV Trace (220 nm)	Calculated Monoisotopic Mass $[M]^+$ or $[M+H]^+$	Observed Mass(es)
47		220 nm  254 nm 	824.53	412.85 $[M+2H]^{2+}$
biotin-47			1649.96	550.75 $[M+3H]^{3+}$ 825.50 $[M+2H]^{2+}$ 836.48 $[M+H+Na]^{2+}$
48			924.58	924.55 $[M+H]^+$