Supporting Information

for

Selective inhibition of CBX6—a methyllysine reader protein in the polycomb family

Table of Contents

Synthesis **2**
Compound Characterization Data **3**
Protein Expression and Purification **12**
Fluorescence Polarization Methods **12**
Fluorescence Polarization Data **14**
SPR Methods **18**
SPR Data **19**
MD Methods and Data **21**

A) Percent Identities (similarities)

Protein	CBX2	CBX4	CBX6	CBX7	CBX8
CBX2	100	68 (88)	63 (89)	65 (88)	64 (88)
CBX4	68 (88)	100	69 (86)	68 (90)	71 (86)
CBX6	63 (89)	69 (86)	100	66 (79)	81 (97)
CBX7	65 (88)	68 (90)	66 (79)	100	69 (88)
CBX8	64 (88)	71 (86)	81 (97)	69 (88)	100

B)

```
CBX2 |12-70| FAAECILSKR LRKGKLEYLV KWRGWSSKHN SWEPEENILD PRLLLAFQKK-----EHEKE 55
CBX4 |11-69| FAVESIEKKR IRKGRVEYLV KWRGWSPKYN TWEPEENILD PRLLIAFQNR-----ERQEQ 55
CBX6 |11-69| FAAESIIKRR IRKGRIEYLV KWKGWAIKYS TWEPEENILD SRLIAAFEQK-----ERERE 55
CBX7 |11-69| FAVESIRKKR VRKGKVEYLV KWKGWPPKYS TWEPEEHILD PRLVMAYEEK-----EERDR 55
CBX8 |11-69| FAAEALLKRR IRKGRMEYLV KWKGWSQKYS TWEPEENILD ARLLAAFEER-----EREME 55
```

Figure S1. Similarity of CBX proteins. A) percent identities and similarity scores (in red) of CBX chromodomains, as determined by EMBOSS matcher software pairwise alignment, B) sequence alignments that highlight the aromatic cage residues and residues lining the bottom of the (–2) pocket which is divergent within the family, as determined by ClustalW2 alignment.

Synthesis

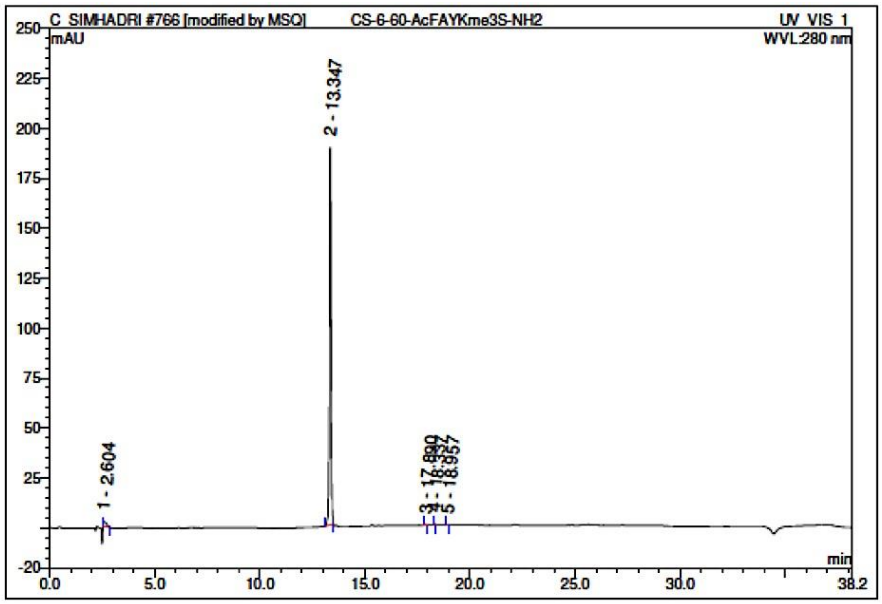
Fmoc-Lys(Me)$_3$-OH was purchased from GL Biochem. All other natural and un-natural amino acids, and coupling agents were purchased from ChemImpex. Compounds were synthesized on Rink amide resin using CEM Liberty microwave synthesizer standard Fmoc protocols using HBTU as coupling agent.

For side-chain functionalization with fluorescein isothiocyanate (FITC), a side-chain Mtt-protected lysine was used. After complete synthesis on resin, treatment with 2:2:96 TFA/triisopropylsilane/CH_2Cl_2 for 20 minutes (x4) was used to cleave the Mtt protecting group selectively. The resin was filtered and washed with CH_2Cl_2 (3 x 5 mL) and DMF (3 x 5 mL). FITC (5 equiv) was added to a 10 mL solution of 1:1 DMF:pyridine. This solution was added to the resin and allowed to stir at ambient temperature overnight. The solution was filtered and the resin washed with DMF (3 x 5 mL) and CH_2Cl_2 (3 x 5 mL) and air-dried. The product was cleaved from resin with 10 mL of 95:2.5:2.5 TFA/H_2O/triisopropylsilane for 2.5 h. The solution was then concentrated in vacuo and added to cold diethyl ether to yield a crude yellow precipitate that was collected by centrifugation. Purifications by preparative HPLC were carried out using a Phenomenex Luna, 5 µm, C-18 column, 250 x 21.20 mm column. Peptides were characterized using LC-MS and ESI-MS and purity was determined to be >95% by analytical LC-MS. Retention times reported arise from analytical traces done on a Thermo Scientific C-18 column, 5 µm, 4.6 mm x 250 mm, flow rate of 1.5 mL/min, gradient running from 90:10 water (0.1 % TFA) and MeCN (0.1 % TFA) to 10:90 water (0.1 % TFA) and MeCN (0.1 % TFA) over 30-38 min.

For side-chain functionalization with biotin, a side-chain Mtt-protected lysine was used. After synthesis on resin, the Mtt protecting groups were cleaved as described above. (+)-Biotin N-hydroxysuccinimide ester (5 equiv) was added to a 10 mL solution of DMF with DIPEA (10 equiv). This solution was then added to the resin and allowed to stir at ambient temperature overnight. Cleavage and purification is the same as described above.

Compound Characterization Data

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount	Type
1	2.60	n.a.	2.295	0.410	2.26	n.a.	BMB*
2	13.35	n.a.	189.021	17.665	97.42	n.a.	BMB*
3	17.89	n.a.	0.338	0.024	0.13	n.a.	BMB*
4	18.34	n.a.	0.337	0.030	0.16	n.a.	BMB*
5	18.96	n.a.	0.049	0.003	0.02	n.a.	BMB*
Total:			192.039	18.132	100.00	0.000	

B)

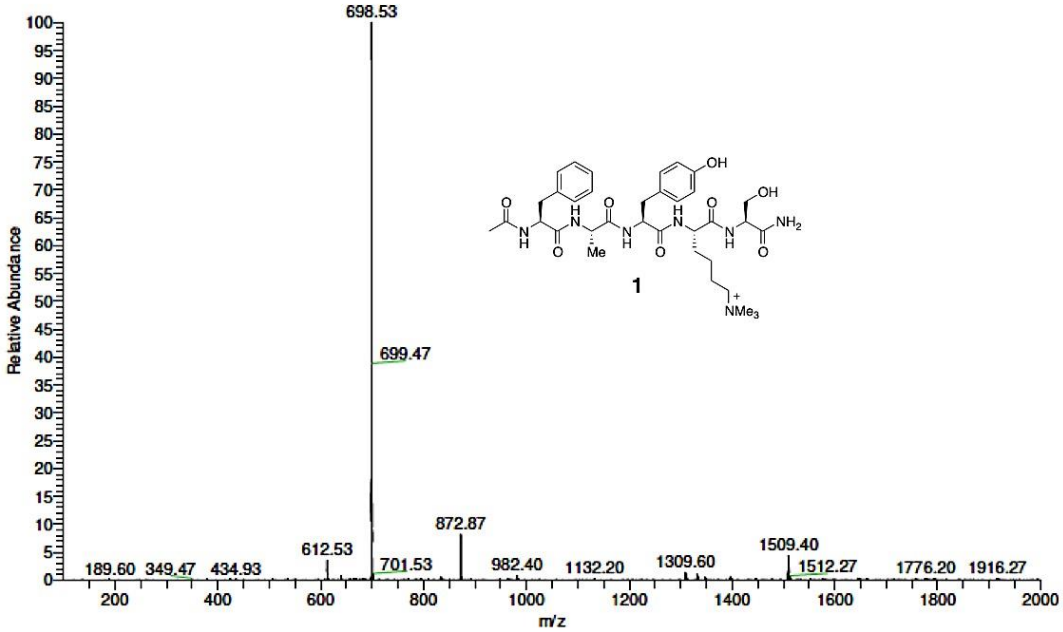
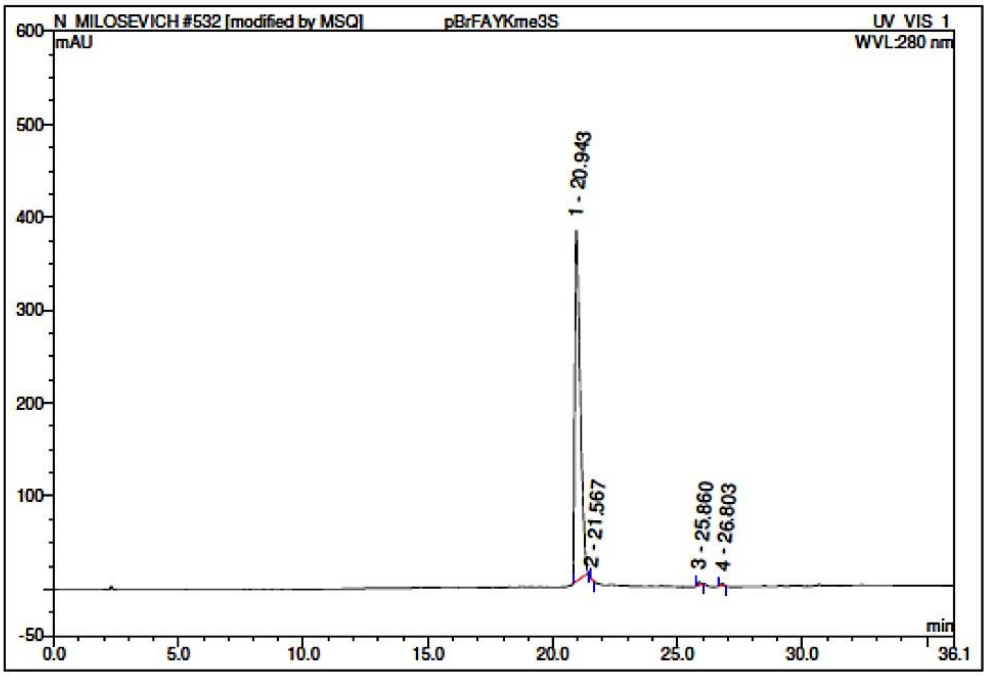


Figure S2. Characterization data for compound **1**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{35}H_{52}N_7O_8^+$: 698.39; found: 698.53.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	20.94	n.a.	377.219	94.395	99.05	n.a.	BMB*
2	21.57	n.a.	0.474	0.023	0.02	n.a.	BMB*
3	25.86	n.a.	3.386	0.466	0.49	n.a.	BMB*
4	26.80	n.a.	2.205	0.420	0.44	n.a.	BMB*
Total:			383.285	95.304	100.00	0.000	

B)

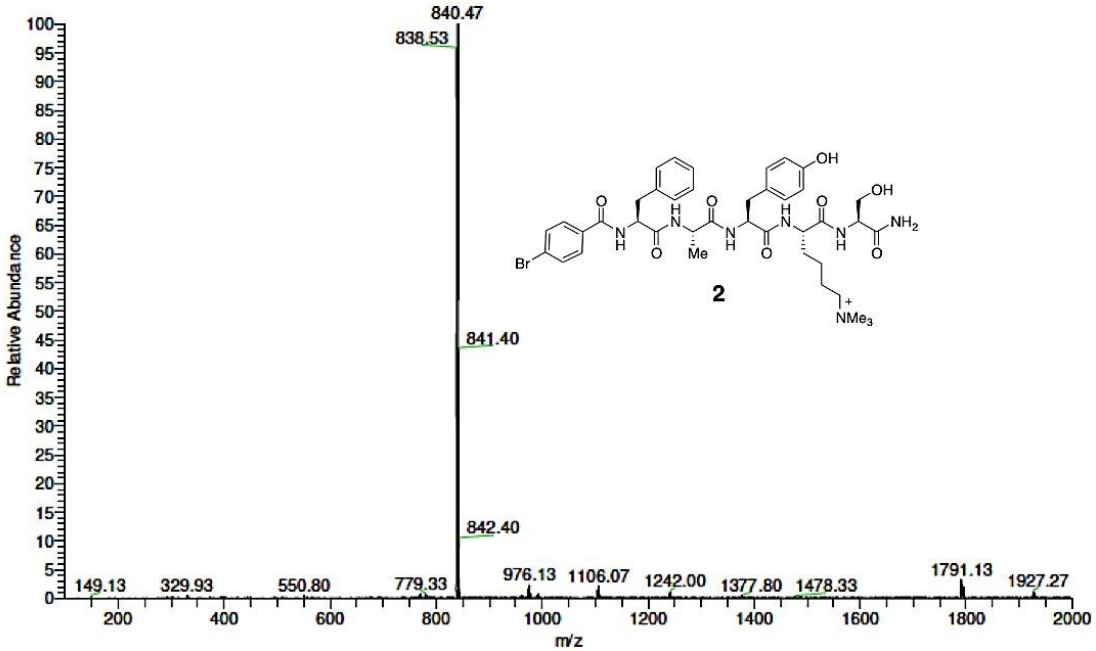
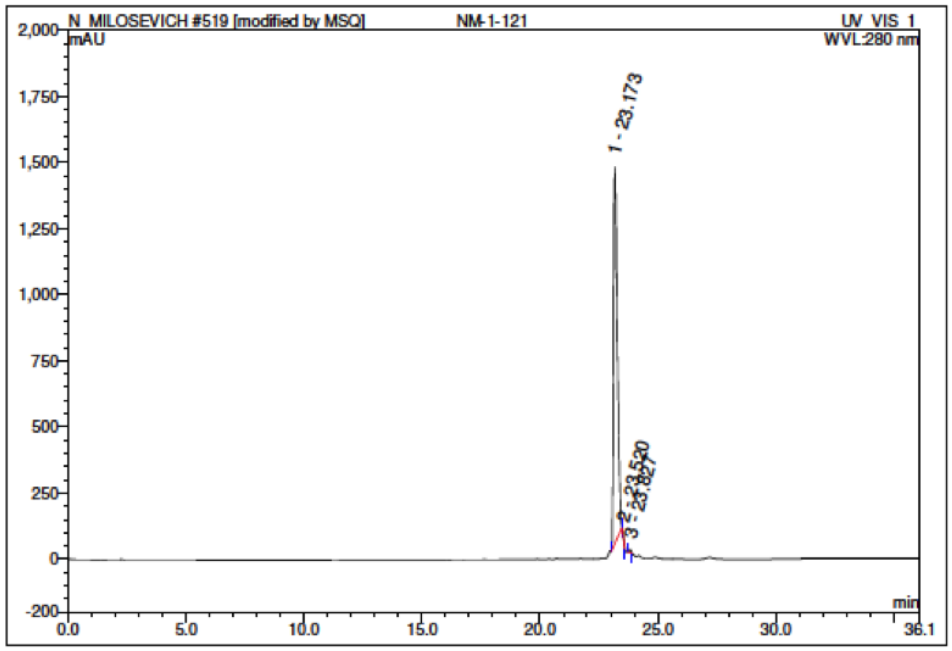


Figure S3. Characterization data for compound **2**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{40}H_{53}BrN_7O_8^+$: 838.31; found: 838.53.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	23.17	n.a.	1424.024	264.891	99.50	n.a.	BMb*
2	23.52	n.a.	5.588	0.269	0.10	n.a.	bMB*
3	23.83	n.a.	11.773	1.050	0.39	n.a.	BMB*
Total:			1441.385	266.210	100.00	0.000	

B)

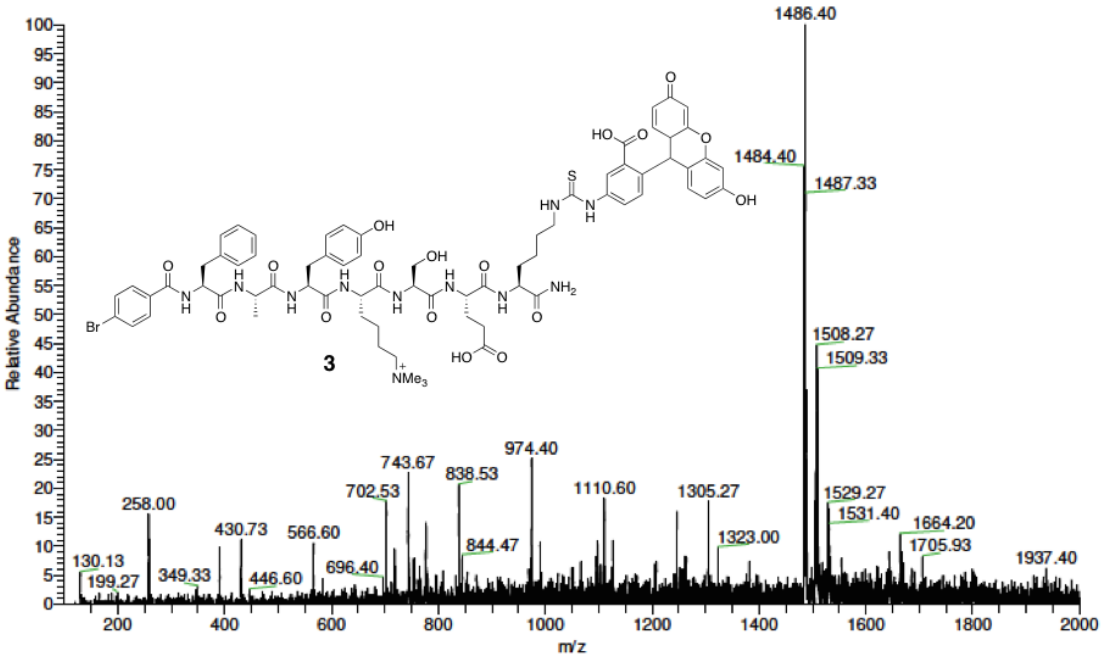
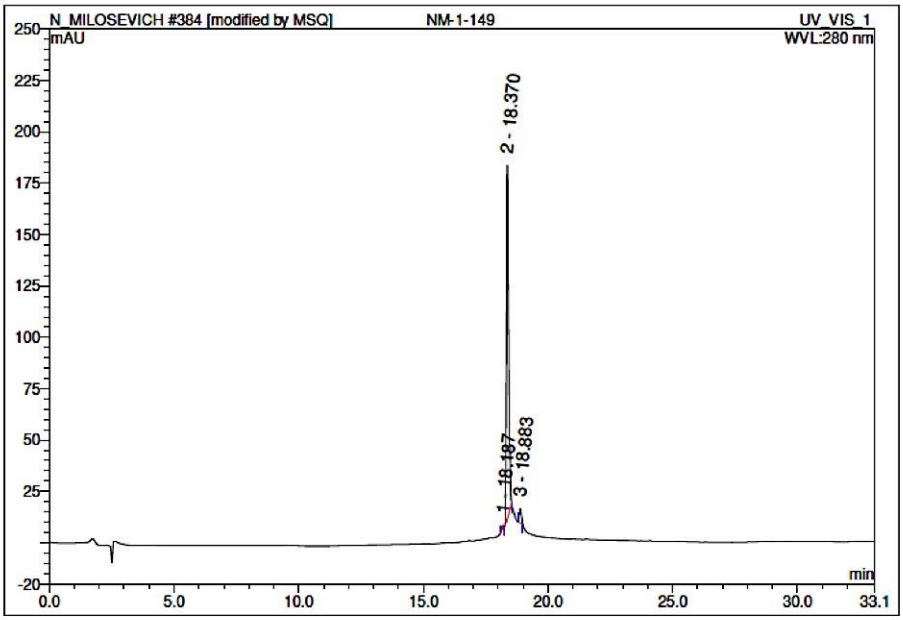


Figure S4. Characterization data for compound **3**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{72}H_{85}BrN_{11}O_{17}S^{+}$: 1486.50; found: 1486.40.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	18.19	n.a.	1.691	0.170	0.90	n.a.	BMB*
2	18.37	n.a.	171.650	18.025	95.33	n.a.	BMB*
3	18.88	n.a.	7.016	0.713	3.77	n.a.	BMB*
Total:			180.358	18.908	100.00	0.000	

B)

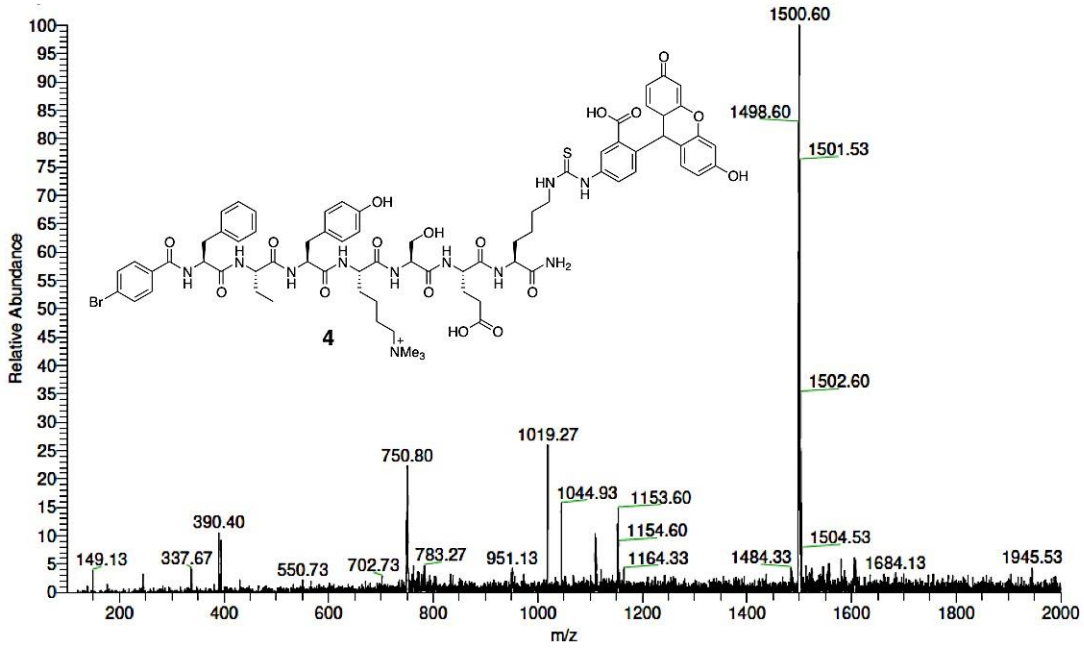
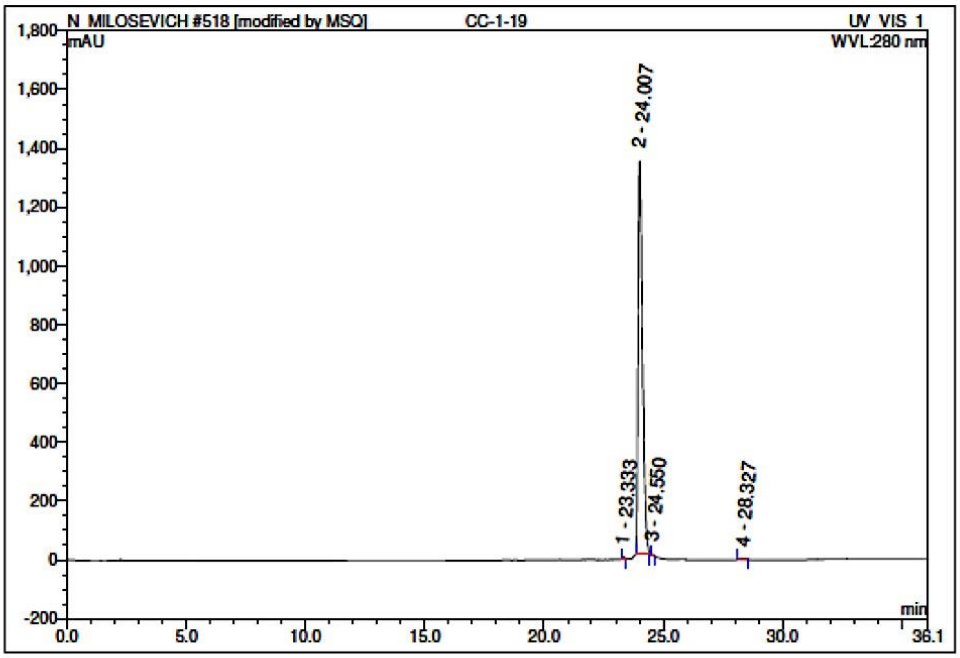


Figure S5. Characterization data for compound **4**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{73}H_{87}BrN_{11}O_{17}S^{+}$: 1500.52; found: 1500.60.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	23.33	n.a.	5.786	0.448	0.16	n.a.	BMB*
2	24.01	n.a.	1335.288	274.794	99.45	n.a.	BM *
3	24.55	n.a.	5.077	0.436	0.16	n.a.	BMB*
4	28.33	n.a.	3.197	0.637	0.23	n.a.	BMB*
Total:			1349.347	276.315	100.00	0.000	

B)

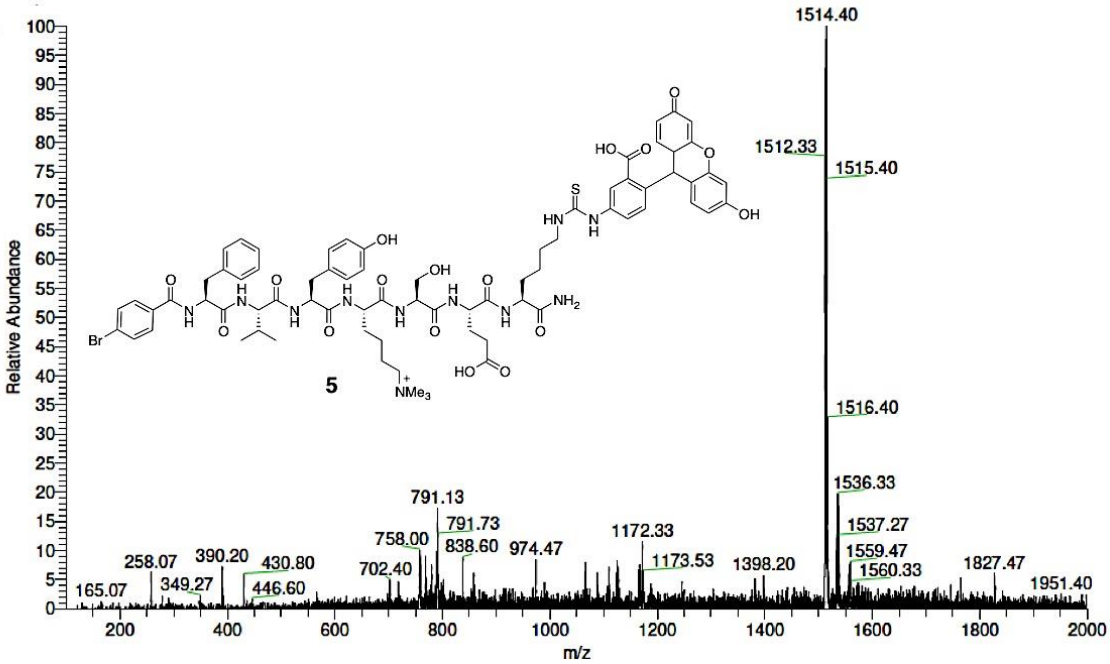
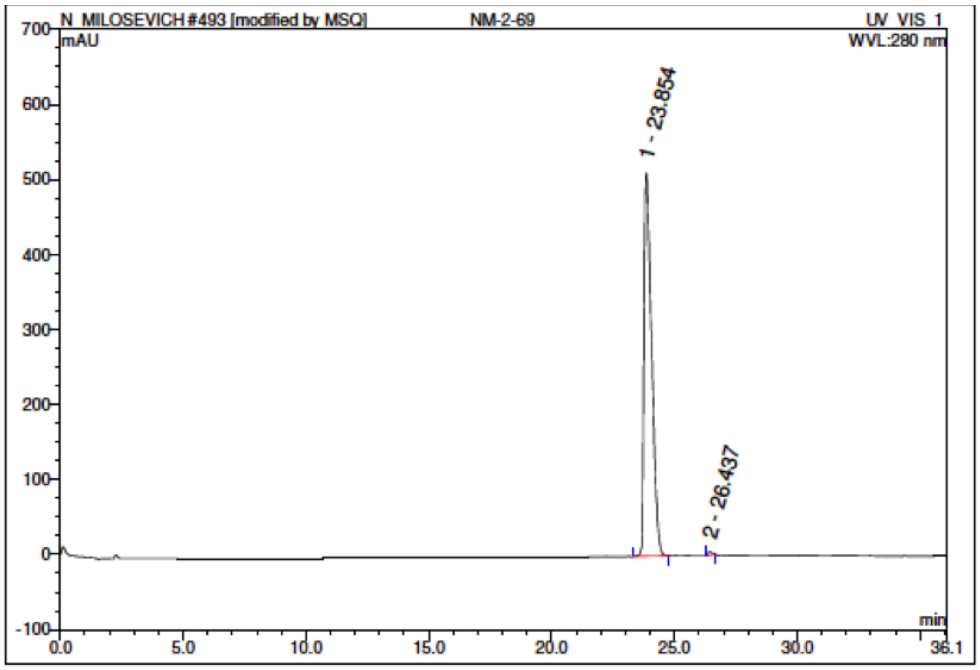


Figure S6. Characterization data for compound **5**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{74}H_{89}BrN_{11}O_{17}S^{+}$: 1514.53; found: 1514.40.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	23.85	n.a.	510.549	182.338	99.58	n.a.	BMB
2	26.44	n.a.	4.092	0.765	0.42	n.a.	BMB*
Total:			514.641	183.103	100.00	0.000	

B)

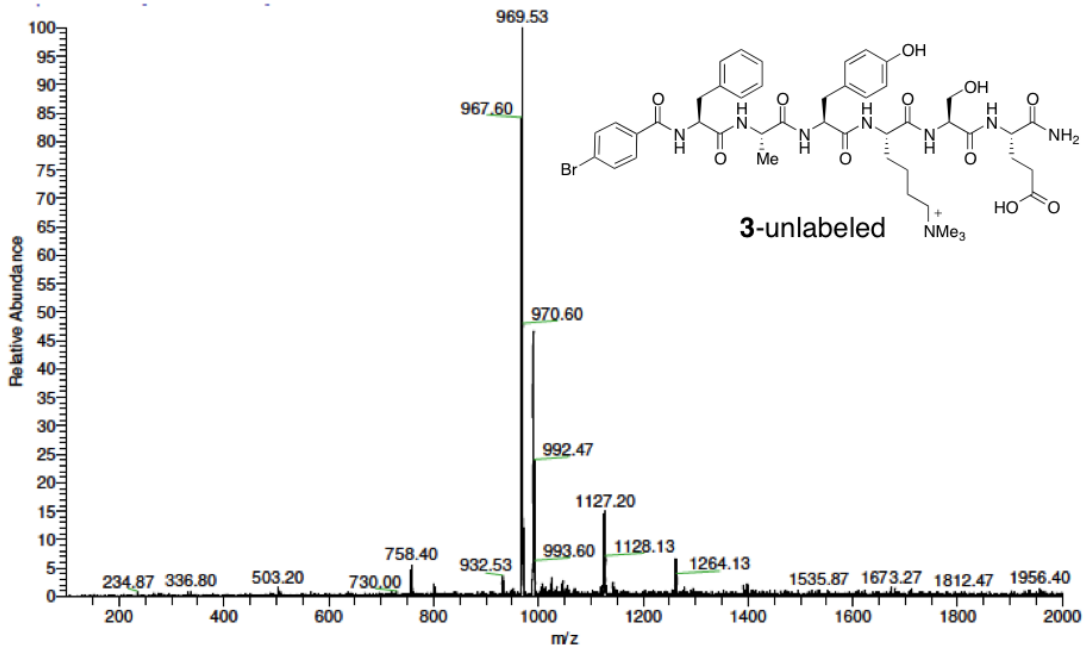
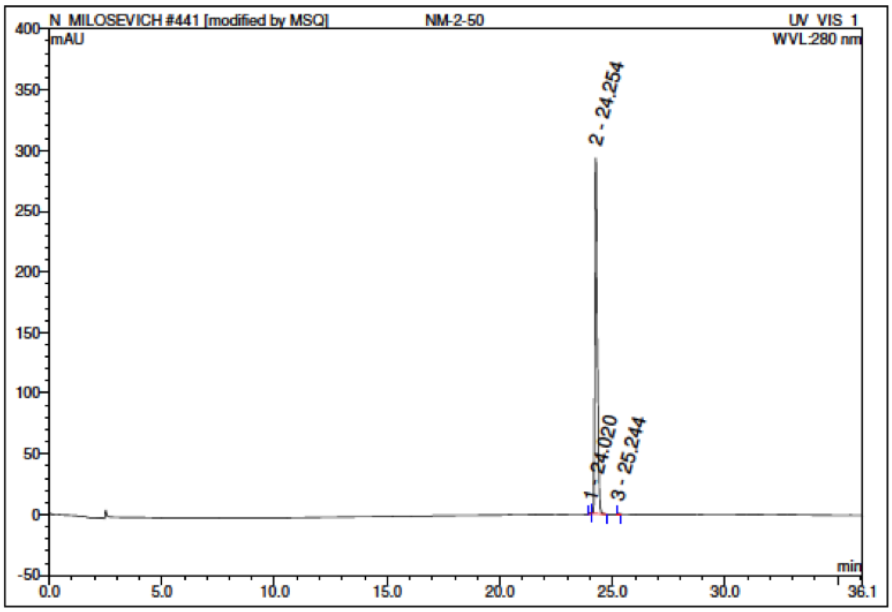


Figure S7. Characterization data for compound **3**-unlabeled. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{45}H_{60}BrN_8O_{11}^+$: 967.36; found: 967.60.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	24.02	n.a.	0.664	0.058	0.16	n.a.	BMb*
2	24.25	n.a.	292.917	37.070	99.78	n.a.	bMB*
3	25.24	n.a.	0.384	0.024	0.06	n.a.	BMB*
Total:			293.965	37.152	100.00	0.000	

B)

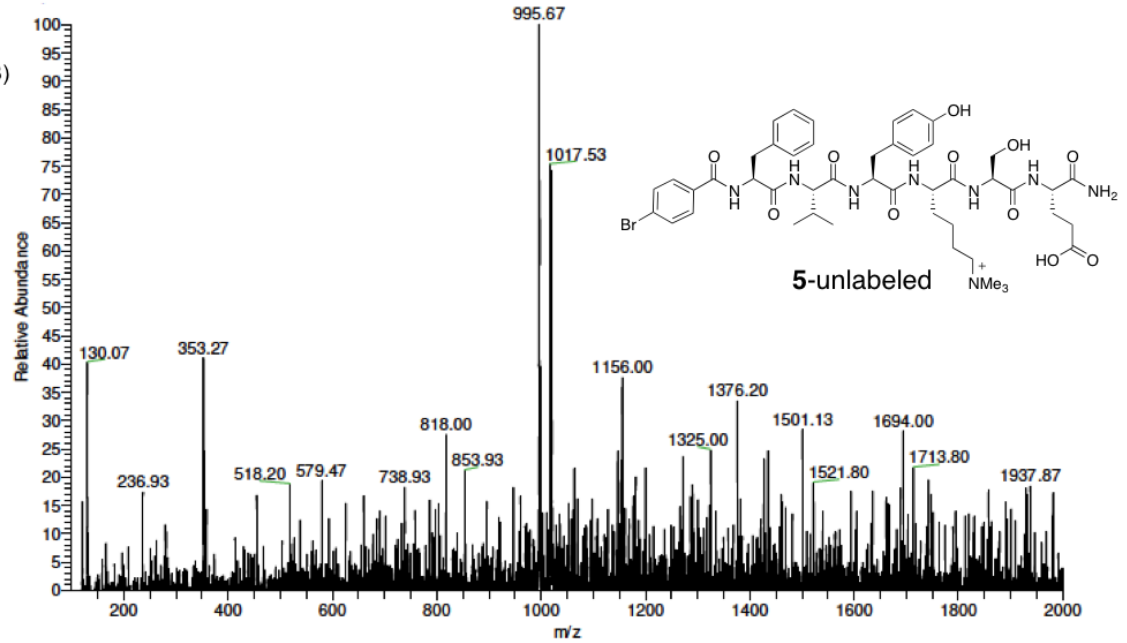
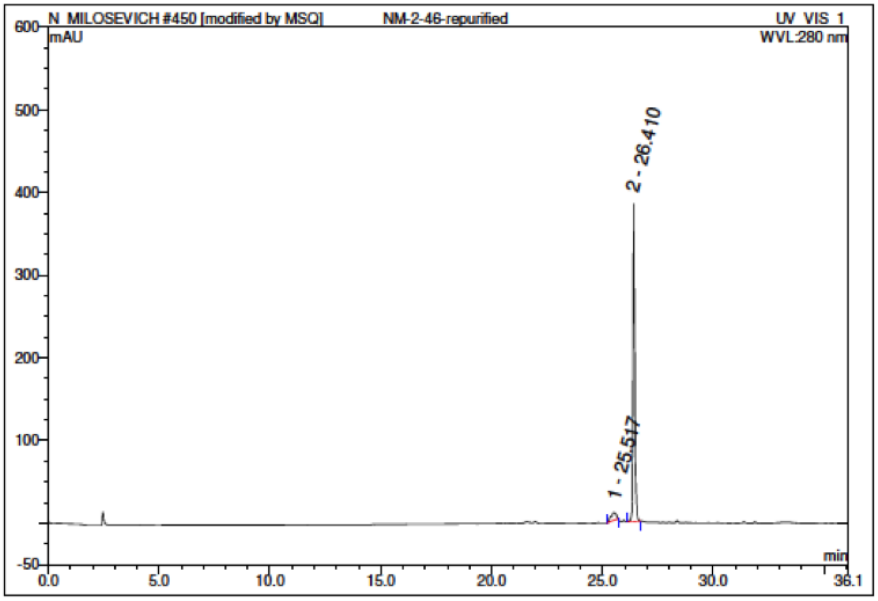


Figure S8. Characterization data for compound **5**-unlabeled. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{47}H_{64}BrN_8O_{11}^+$: 995.39; found: 995.67.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	25.52	n.a.	9.467	2.478	6.63	n.a.	BMB*
2	26.41	n.a.	384.248	34.916	93.37	n.a.	BMB
Total:			393.715	37.394	100.00	0.000	

B)

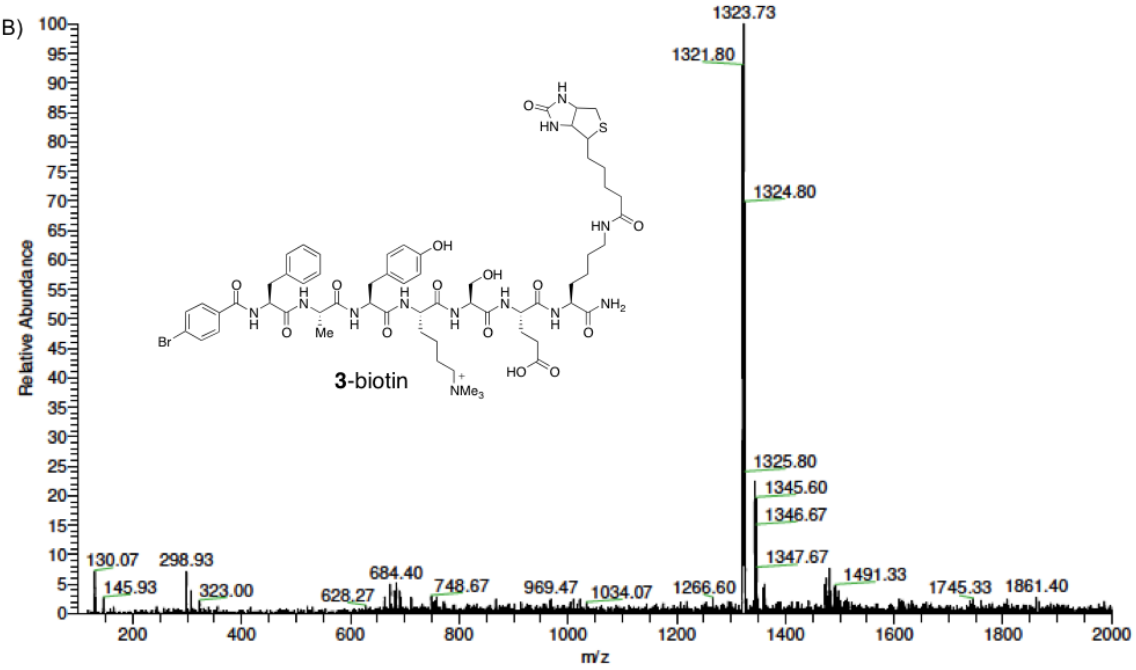
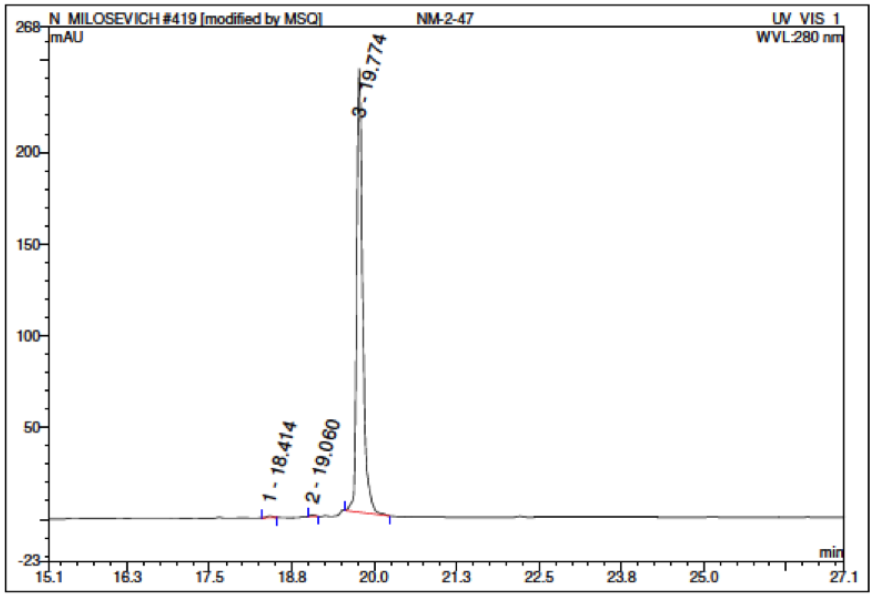


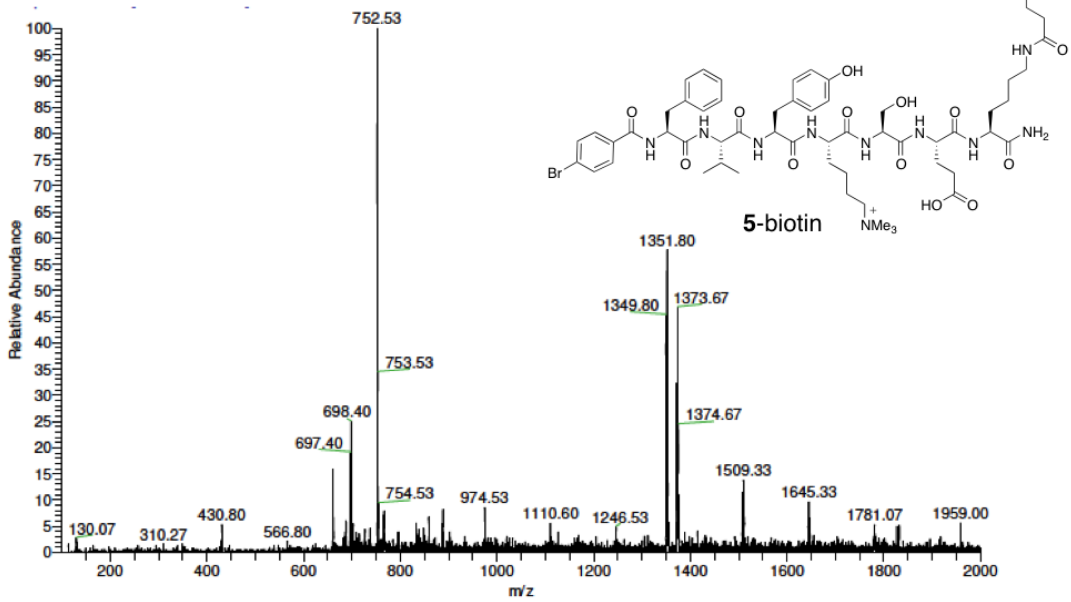
Figure S9. Characterization data for compound **3**-biotin. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{61}H_{86}BrN_{12}O_{14}S^+$: 1321.53; found:1321.80.

A)



No.	Ret.Time min	Peak Name	Height mAU	Area mAU*min	Rel.Area %	Amount µl	Type
1	18.41	n.a.	0.869	0.078	0.30	n.a.	BMB*
2	19.06	n.a.	0.846	0.068	0.26	n.a.	BMB*
3	19.77	n.a.	241.698	25.641	99.43	n.a.	BMB
Total:			243.413	25.787	100.00	0.000	

B)



5-biotin

Figure S10. Characterization data for compound **5**-biotin. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{63}H_{90}BrN_{12}O_{14}S^+$: 1349.56; found:1349.80.

Protein Expression and Purification

Protein Expression

Addgene plasmids 25245 (CBX1), 25158 (CBX2), 25237 (CBX4), 25296 (CBX6), 25241 (CBX7) and a CBX8 plasmid donated directly by C. Arrowsmith (Structural Genomics Consortium, Toronto, Canada) were used to transform competent BL21 CodonPlus RIL *Eschericia* coli cells (Stratagene). The chromodomains of CBX1 (20-73), CBX2 (8-62), CBX4 (8-65), CBX6 (8-65), CBX7 (8-62) and CBX8 (8-61) were overexpressed as N-terminal His6-tagged proteins by culturing *E.coli* at 37°C in 2XYT media to OD ~1.8, dropping the temperature to 15°C for 30-60 minutes to 1 hour, inducing with 1 mM IPTG (Life Technologies) and culturing overnight (~16 hours) before pelleting cells, re-suspending in chromodomain-specific binding buffer plus protease inhibitor (Calbiochem), and freezing until required for purification.

Protein Purification

E. coli pellets were thawed overnight and sequentially lysed via a 30 minute incubation with a CHAPS/dH_2O solution (Biobasic) and sonication. After centrifugation to remove cell debris, the purification procedure involved two chromatographic steps. First, an affinity chromatographic step on a nickel-nitrilotriacetic acid chelating column (Qiagen) followed by a gel filtration step using an FPLC outfitted with a Hiload 16/600 Superdex 75 pg size exclusion column (GE). The FPLC step performed a buffer exchange into our ‘minimal’ fluorescence polarization (FP) buffer consisting of 20 mM Tris HCl, 250 mM NaCl, and 0.01% Tween-20. Purified protein was then concentrated to the desired volume using Amicon centrifugal filter units, whereupon PMSF, DTT and benzamidine were added to a 1 mM final concentration (Thus creating ‘full’ FP buffer). Proteins were flash frozen and stored at -80°C while purity was assessed by SDS page. It is worth noting that the His tags on CBX7 and CBX8 were cleaved off overnight after the nickel column step using our in-house TEV protease.

Generation of CBX7 Point Mutant

A plasmid encoding the N-terminal His6-tagged chromodomain of CBX7 (amino acids 8-62) was obtained from Addgene (supplier Cheryl Arrowsmith). The chromodomain was then mutated using a Quickchange II XL Site-Directed Mutagenesis Kit (Stratagene) to create the V13A mutant.

Fluorescence Polarization Methods

Competitive FP analysis of **1**, **2**, **3**-unlabeled, and **5**-unlabeled binding to CBX7 was used to determine IC_{50} values as previously reported in reference 33, except using the new FITC-peptide **3** as the competitive binding probe (Figures S11, S12). The experiments on **3**-unlabeled and **5**-unlabeled were added in response to a reviewer’s concern, in order to confirm that the FITC tag on compound **5** is not somehow responsible for its observed CBX6 selectivity. The data, although qualitative, do confirm this by showing that **3**-unlabeled actually has 2.8-fold preference for CBX7 that is reversed for **5**-unlabeled, which has 7.3-fold preference for CBX6 (Figure S12). In spite of these results agreeing with our conclusion that **5** is selective for CBX6 because of its isopropyl substituent, we emphasize that competitive FP in such systems provides IC_{50} values that depend strongly on the intrinsically different

affinities between each protein and the FITC-peptide probe. Direct FP and/or SPR experiments (see below) don't rely on a third binding partner as a reporter of the interaction, and are therefore better measures of protein-ligand K_d values when comparing data from a panel of different proteins.

Direct FP was done by titration of CBX proteins into FITC-labelled probes **3-5**. Binding assays were performed in black 384 well plates with optical bottoms at a total volume of 50 µl per well. A solution containing 100 nM of the labelled probe and a varying concentration of the CBX protein in question (dissolved in full FP buffer) was titrated into wells containing constant concentrations (100 nM) of labelled probe in full FP buffer. 100 µl of the CBX-containing solution was added to well 19. Wells 2-18 contained 50 µl of the probe-only solution. 50 µl of the well 19 solution was added to well 18 and mixed. These dilutions continued well 3, whereupon 50 µl was discarded after mixing, leaving the total volume of all wells at 50 µl. Well 1 contained only full FP buffer and well 2 contained only the labelled probe in full FP buffer. The maximum concentrations used for each CBX protein varied (120 µM – 600 µM) based on availability and solubility of the protein at high concentration, as well as the expected strength of the probe being tested. Assays were performed in triplicate or duplicate depending on CBX protein availability (noted in raw data). After removing any bubbles with a needle, plates were incubated in the dark at RT for 15 minutes and then read with a SpectraMax M5 plate reader (Molecular Devices). Raw data for perpendicular (P) and parallel (S) light was collected with SoftMax Pro software using a fluorescence polarization endpoint read. 100 reads were made and averaged per well with 5 second automix before reading and medium PMT sensitivity. Excitation was 450 nm for FITC and emission was 530 nm (with a 515 nm cutoff).

Fluorescence Polarization Data Analysis

Average P and S values (minus the average values for buffer control wells) were determined for sets of replicates using Microsoft Excel and millipolarization units (mP) were calculated using the formula ((S-P)/(S+P))*1000. Data were then plotted using XLFit curve fitting add-in for Excel (Microsoft). The mathematical model for the fit was described by the equation (C+((BLmax*x)/(Kd+x))). For certain assays, some of the data points (0 to 5 points) representing the highest CBX concentrations had to be excluded due to abnormally high mP values, likely arising from protein aggregation.

Fluorescence Polarization Data

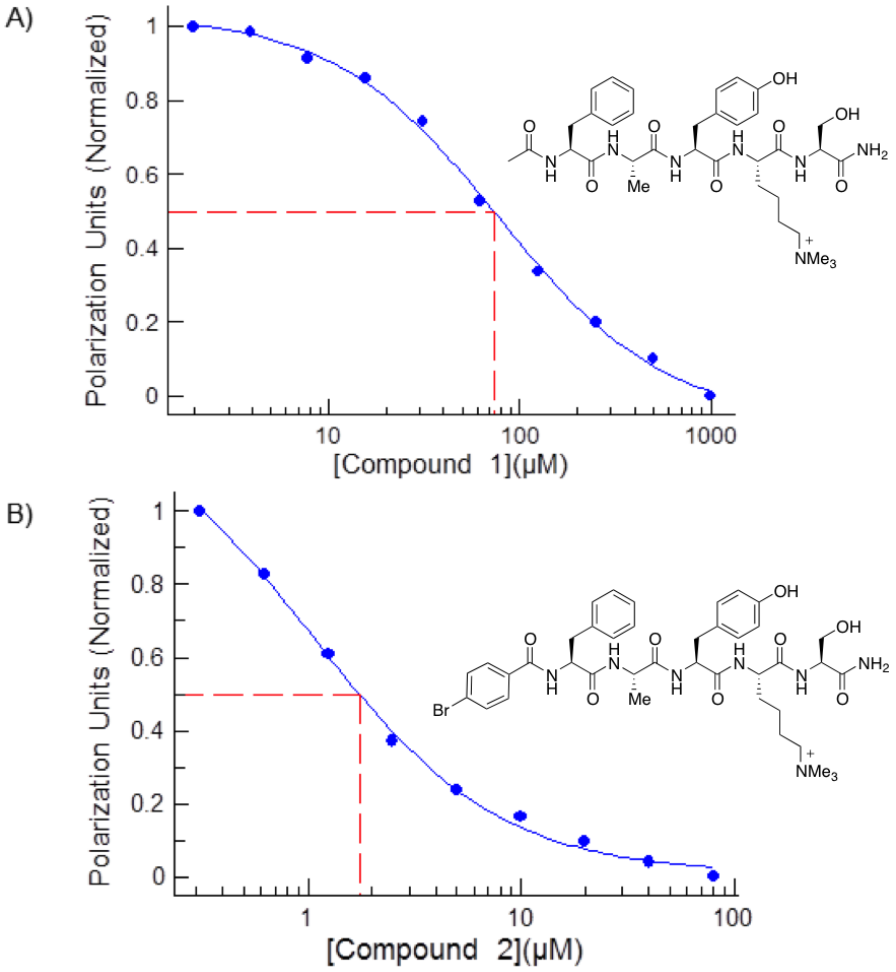


Figure S11. Normalized competitive fluorescence polarization for compound **1** and **2** with CBX7 using compound **3** as fluorescent probe. A) Compound **1** with CBX7 (IC_{50} = 73 +/– 7.3 μM), B) Compound **2** with CBX7 (IC_{50} = 1.7 +/– 0.18 μM).

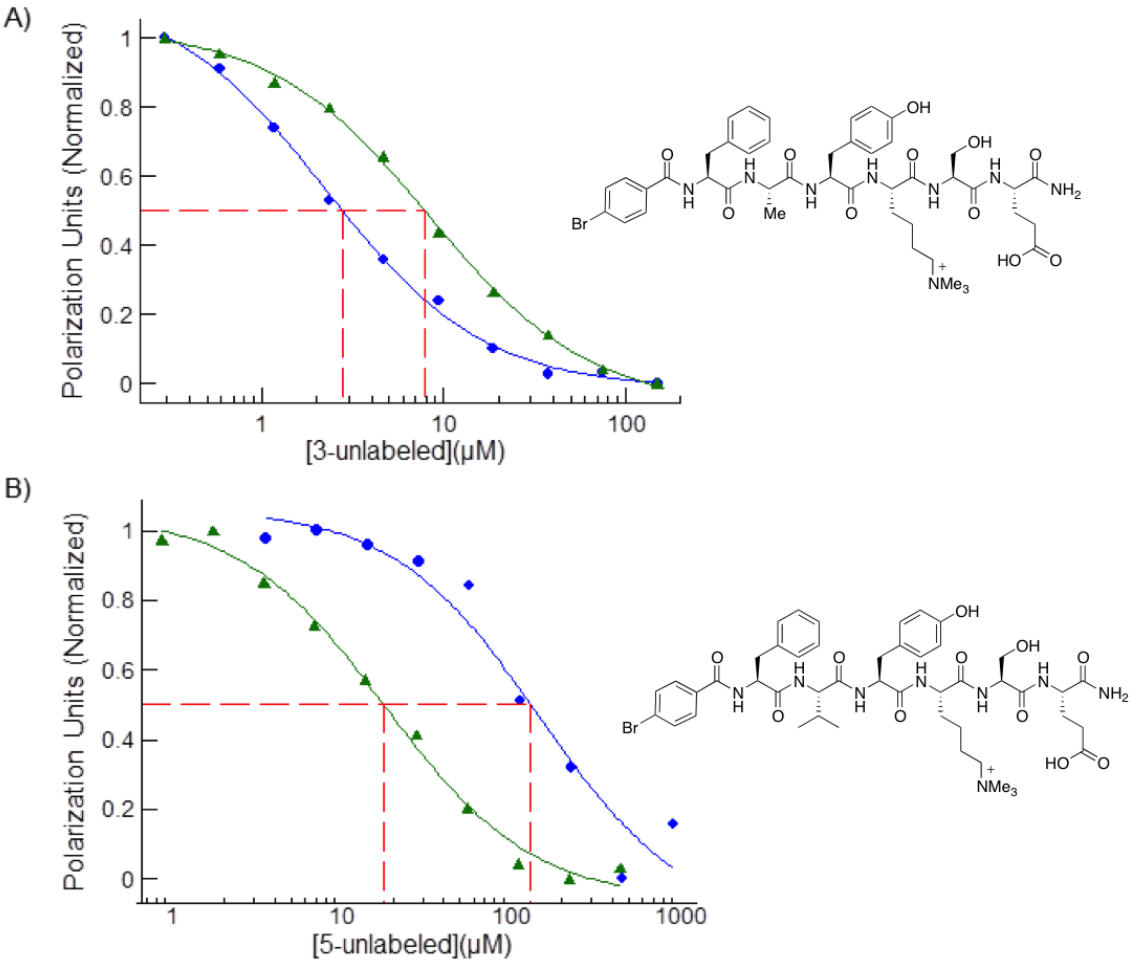


Figure S12. Overlays of normalized competitive fluorescence polarization titrations with **3**-unlabeled and **5**-unlabeled against CBX6 (green triangles) and CBX7 (blue circles). A) **3**-unlabeled studied with CBX6 (IC_{50} = 7.8 μM) and CBX7 (IC_{50} = 2.8 μM). B) **5**-unlabeled studied with CBX6 (IC_{50} = 18 μM) and CBX7 (IC_{50} = 131 μM). Uncertainties in the graphically determined IC_{50} values are ±10%.

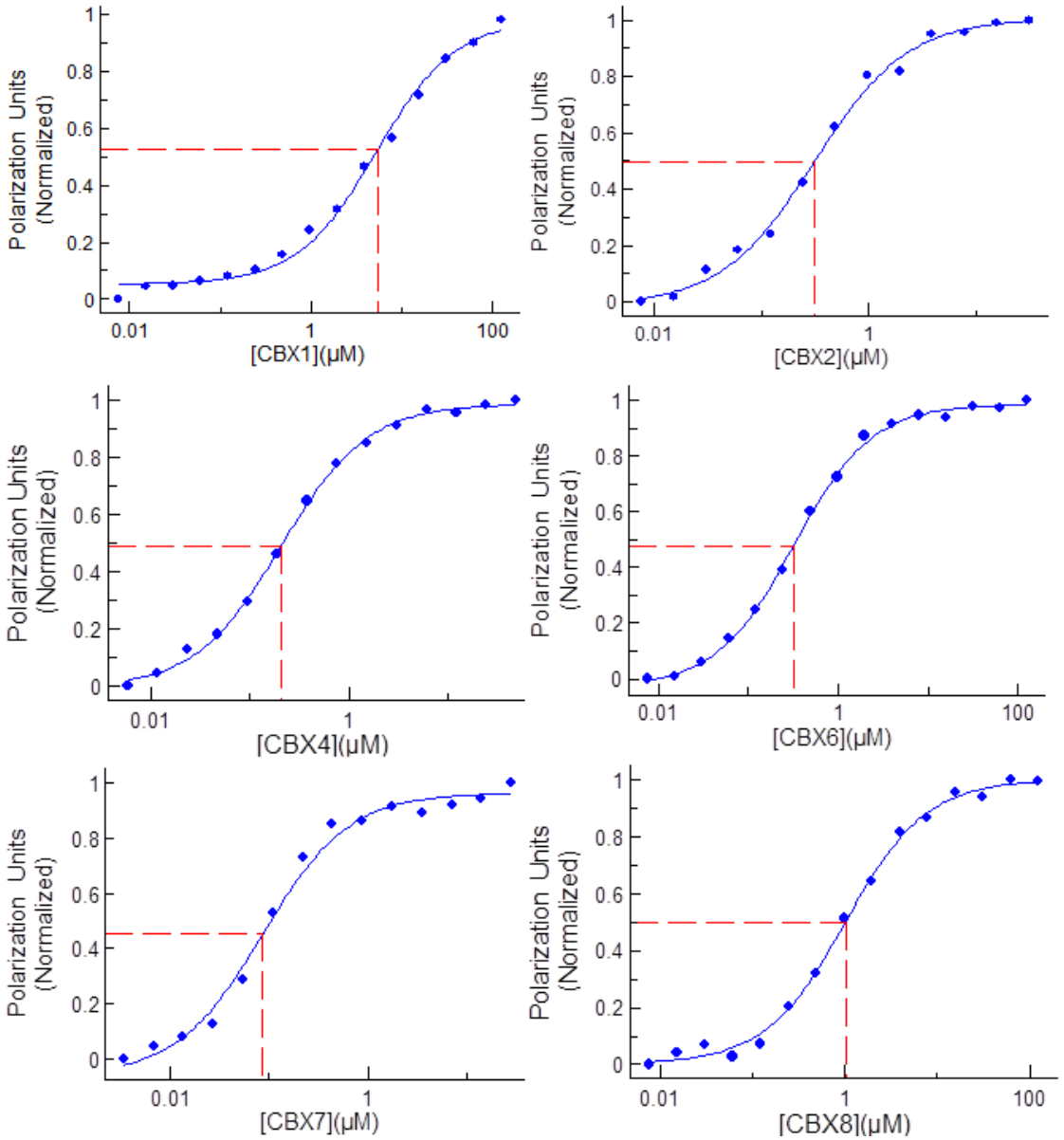


Figure S13. Normalized direct fluorescence polarization data of compound **3** with CBX1/2/4/6/7/8.

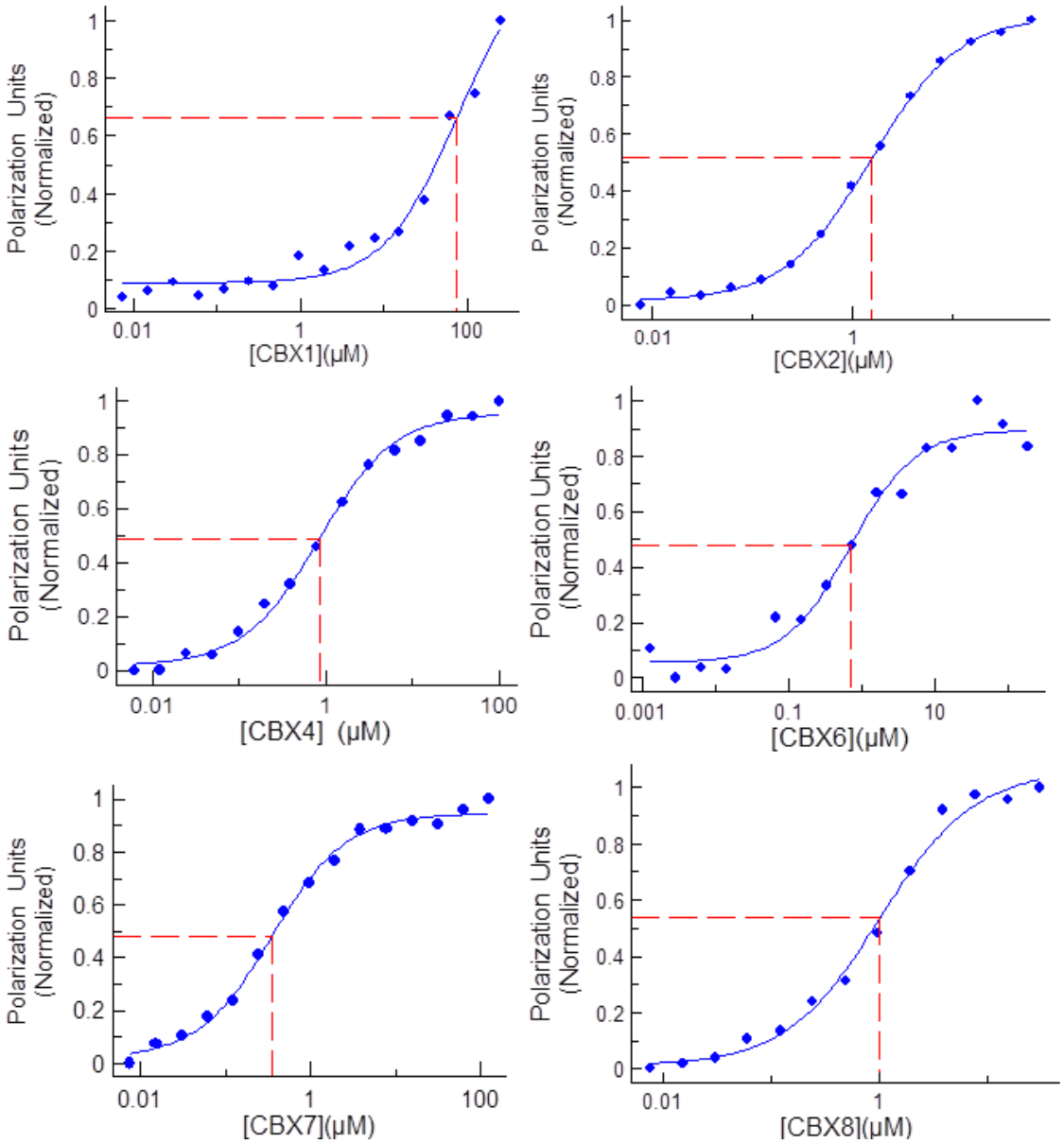


Figure S14. Normalized direct fluorescence polarization data of compound **4** with CBX1/2/4/6/7/8.

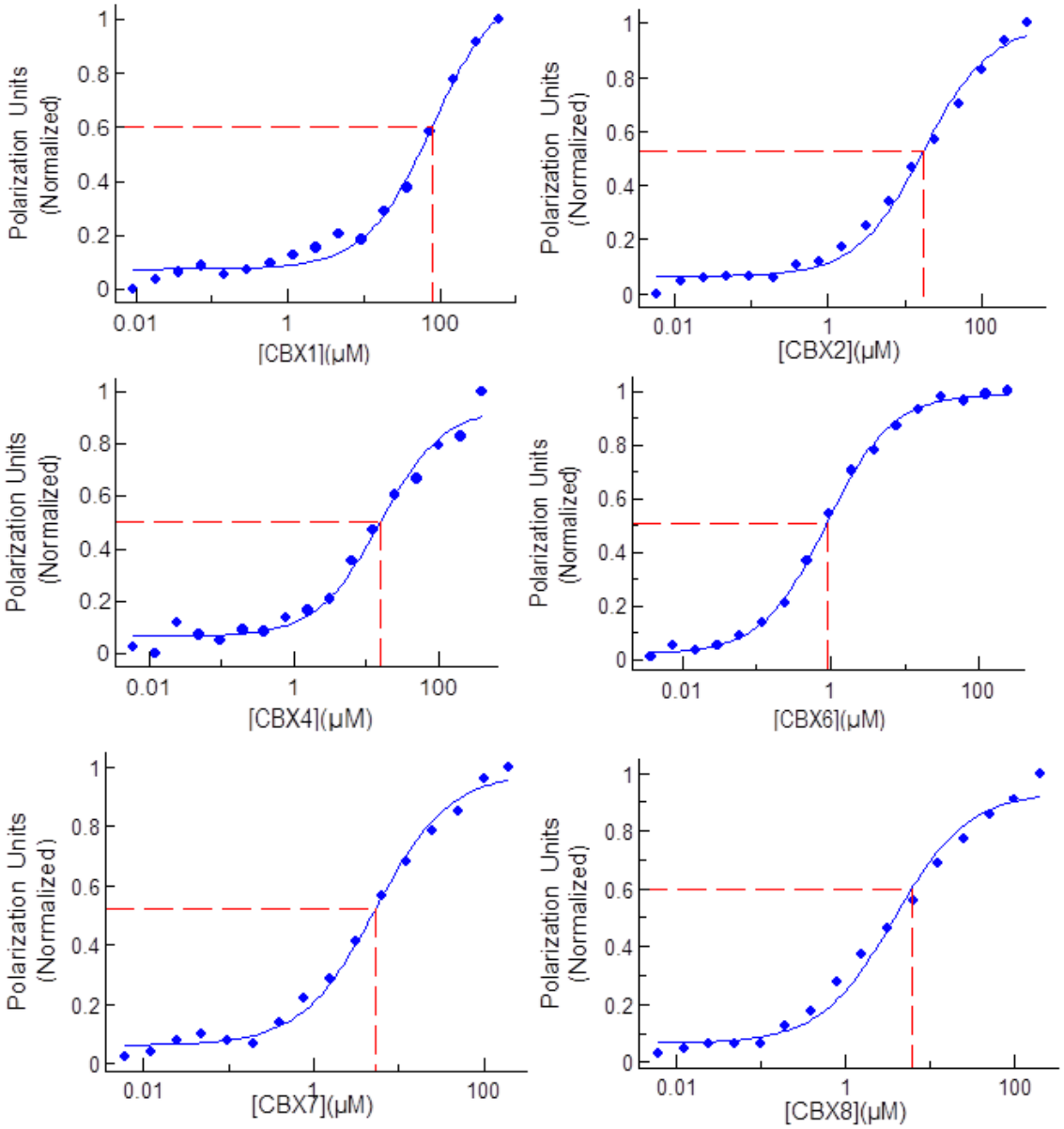


Figure S15. Normalized direct fluorescence polarization data of compound **5** with CBX1/2/4/6/7/8.

SPR Methods

All surface plasmon resonance experiments were performed on a Biacore X100 (GE Healthcare). Research-grade streptavidin (SA) sensor chips were used for ligand immobilization. The biotinylated ligands (**3**-biotin and **5**-biotin) were prepared in HBS-EP+ buffer (10 mM HEPES, 150 mM NaCl, 3mM EDTA, 0.005% P20, pH 7.4) provided by the manufacturer and 10% dimethylsulfoxide. Ligand immobilization in flow cell 2 was carried out by injection at a flow rate of 5 µl/min for 18 minutes for a density of 632 RU and 1313 RU for compound **3**-biotin and **5**-biotin respectively. Flow cell 1 was left blank as a reference. Stock protein solutions were stored in HBS-EP+ running buffer at -80°C and dilution series were prepared as required. Binding data was collected by injecting the analyte at a flow

rate of 30 µl/min over two flow cells at a temperature of 25°C. Association and dissociation of the complex was 3 minutes and 6 minutes respectively for compound **3**-biotin, and 3 minutes and 4.5 minutes for compound **5**-biotin. The flow cell surfaces were regenerated with a 30 second injection of 10 mM glycine-HCl solution pH 2.0 (GE Healthcare). Duplicate injections for each sample, including the blank, were carried out for each protein-ligand interaction. Affinity curves, dissociation constants and sensograms were determined using Biacore X100 Evaluation Software.

SPR Data

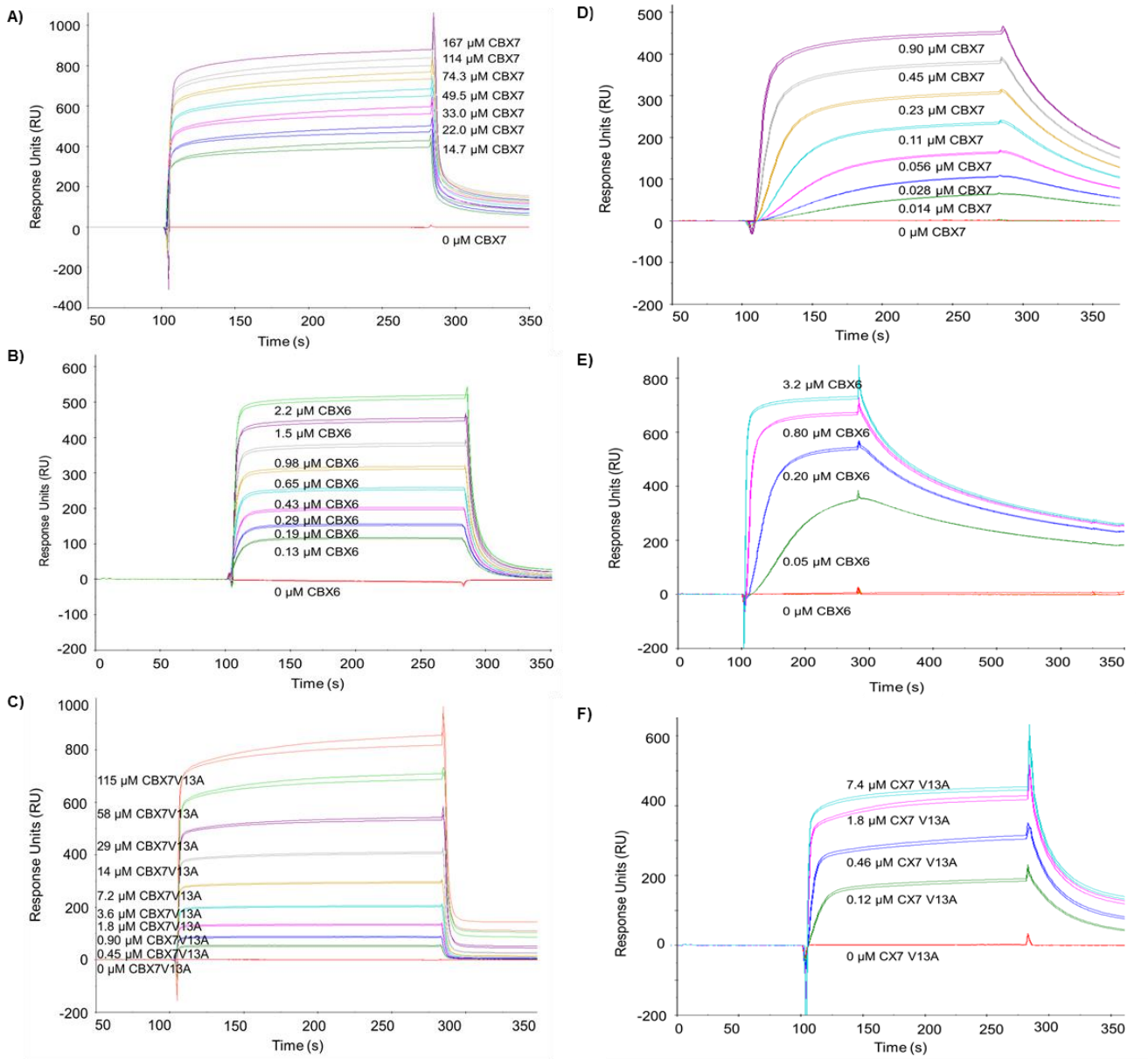


Figure S16. SPR sensorgrams for SA chip functionalized with **3**-biotin or **5**-biotin studied against CBX7, CBX6 and CBX7-V13A. All sensograms were adjusted to baseline = 0 RU. A) **5**-biotin on SA

chip with CBX7, B) **5**-biotin on SA chip with CBX6, C) **5**-biotin on SA chip with CBX7-V13A, D) **3**-biotin on SA chip with CBX7, E) **3**-biotin on SA chip with CBX6, F) **3**-biotin on SA chip with CBX7-V13A.

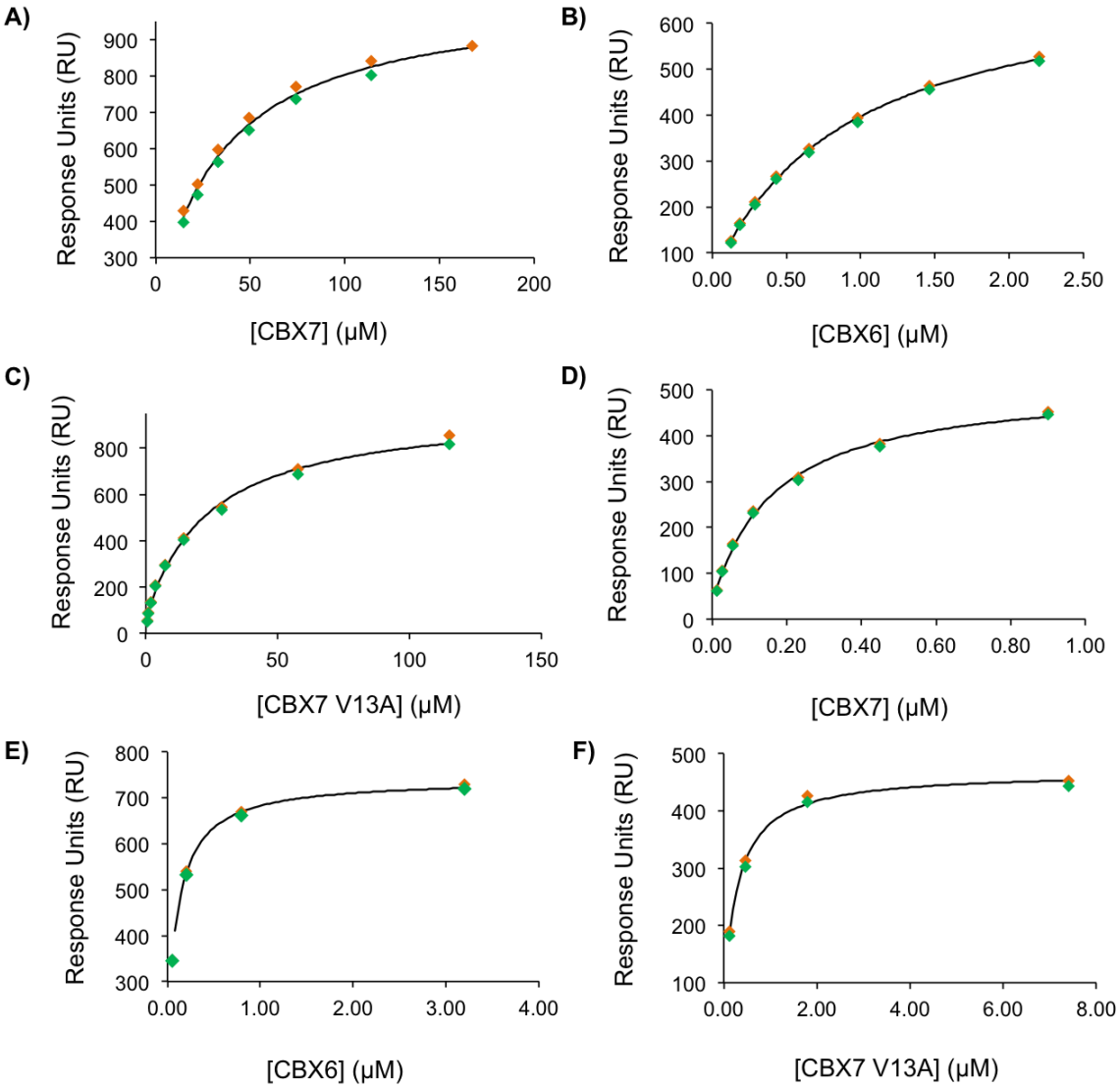


Figure S17. Fitted SPR binding curves for SA chip functionalized with **3**-biotin or **5**-biotin studied with CBX6, CBX7 and CBX7-V13A. A) **5**-biotin on SA chip with CBX7, B) **5**-biotin on SA chip with CBX6, C) **5**-biotin on SA chip with CBX7-V13A, D) **3**-biotin on SA chip with CBX7, E) **3**-biotin on SA chip with CBX6, F) **3**-biotin on SA chip with CBX7-V13A. See main text for K_d values. All experiments were done in duplicate.

MD Methods and Data

Molecular dynamics simulations were run using the AMBER12 software suite employing the ff99SB forcefield. Parameterization of the non-standard residues of the ligand were derived using the RESP charge fitting method (J. Phys. Chem, 1993, 97, 10269-10280). The remaining parameters were taken from the general amber forcefield for small molecules. Systems were solvated in approximately 5000 TIP3PBOX water molecules and neutralized with Cl^- counter ions. A 10 Å electrostatic potential cutoff was used along with the SHAKE algorithm for bonds connected to hydrogen atoms. Prior to production runs, the systems were minimized at 0 K over 10,000 steps, ramped up to 300 K over 50 ps with a time step of 2 fs, then equilibrated to one atmosphere for another 50 ps. Production runs were then ran at 300 K under constant pressure.

Docking of compounds **3** and **5** was done by superposition of the ligand backbone coordinates over those of the crystal structure native peptide H3K27 in CBX6 (See Figure S18). Successful equilibration of the complex was achieved in this scheme, and trajectories of CBX6 and 7, both with compounds **3** and **5**, were run for 50 ns each with a 2 fs time step.

Chronological snapshots of sample trajectory runs are included in MD simulation movies CBX6-5_10ns.mpg and CBX7-5_10ns.mpg, available for download as separate files. Compound **5** valine was stable in the −2 binding pocket of both CBX6 and CBX7, despite the bulkier valine residue at the edge of the pocket in the latter. The zoomed-in region presented in Figure S19 indicates the stability of the Compound **5** valine in the CBX7 binding pocket, at the 25 ns mark.

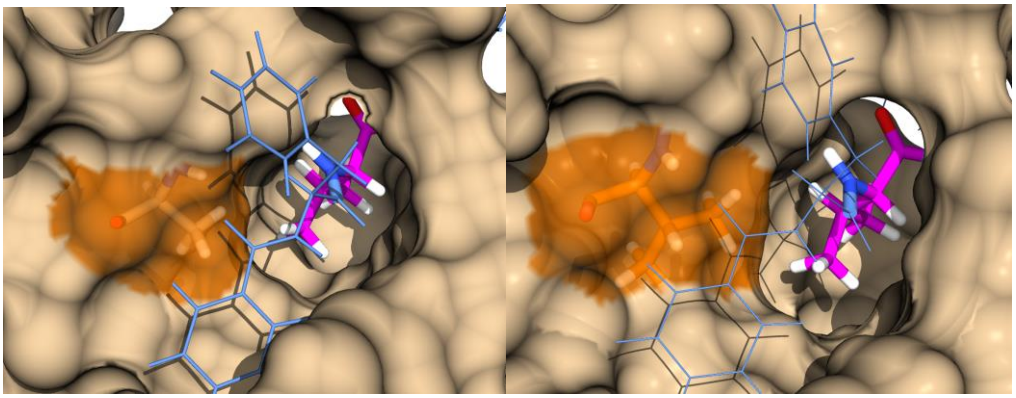


Figure S18. Minimized fit of compound 5 in CBX6 (Left) and CBX7 (Right). Orange highlighted regions indicate the valine (CBX7) and alanine (CBX6) residues on the outside of the hydrophobic clasp. While these residues appear to partially form the rim of the −2 pocket, they have limited impact on the shape at the bottom of the pocket or how much steric interaction with the Compound **5** valine (pink) occurs.

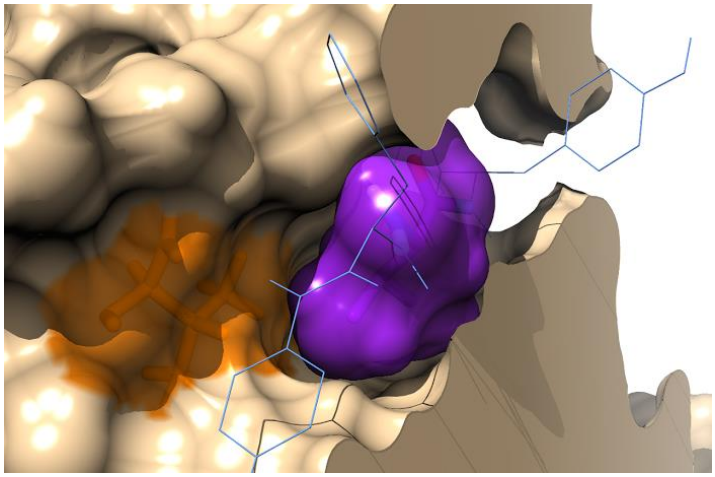


Figure S19. Clipped view of CBX7-Compound **5** interaction after equilibration. Following several nanoseconds of simulation, the Compound **5** valine still sits in the −2 pocket hydrophobic pocket. Clasp residues have been clipped out of the image and a surface area has been applied to the guest residue to show the fit in the bottom of the pocket. The same colour scheme as Figure S18 applies.

However, inclusion of the bulky Compound **5** under the CBX7 hydrophobic clasp induces significant changes in the trimethyllysine recognition site. The opening of the clasp leads to a shift in the aromatic cage phenylalanine, reducing its interaction strength with the trimethyllysine. The distance between the two relevant residues was calculated along the trajectory, and is presented in Figure S20.

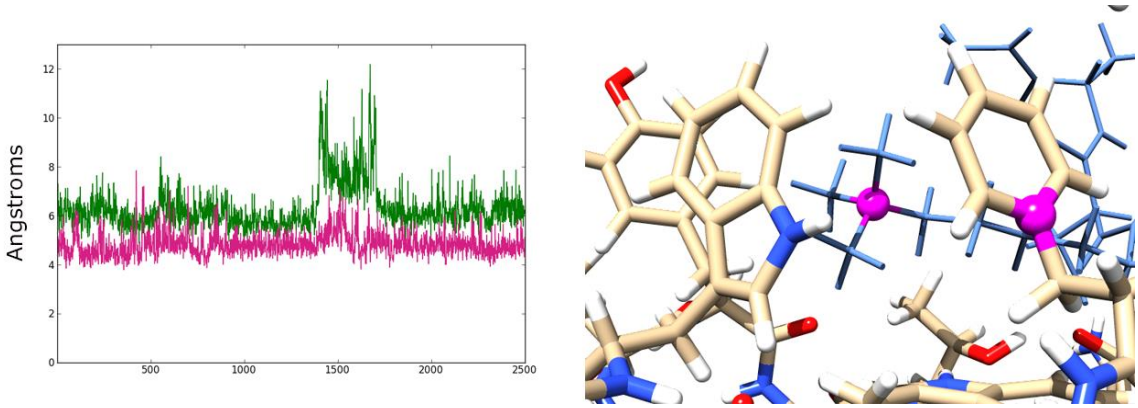


Figure S20. Hydrophobic clasp distance correlations with aromatic cage structure. To show the complex nature of the induced fit in the −2 pocket region, a plot relating the hydrophobic clasp distance to the trimethyllysine interaction shows a correlation between these two binding regions. On the left, the distance plot over the course of 50 ns shows the clasp gap distance (green) compared to the distance between the phenylalanine and trimethyllysine residues on CBX7 and Compound **5**, respectively (pink). The image on the right shows this aromatic cage interaction with the reference atoms coloured in pink as well. From this structural relationship, it is shown that the position of the clasp is tied to the recognition of the trimethyllysine in the aromatic pocket region.

MD simulation movies are available for download as separate files CBX6-5_10ns.mpg and CBX7-5_10ns.mpg. Structures were captured every 0.02 ns and are presented at 15 frames/s. Coloring of the protein is orange=hydrophobic, blue=polar according to the Kyte-Doolittle hydrophobicity scale. The hydrophobic clasp comprised of residues Val10 and Leu49 is the orange bridging feature in the central foreground of each video. The movies are available as Web Enhanced Objects on the ACS Publications website.