Pan-specific and partially selective dye-labeled peptidic inhibitors of the polycomb paralog proteins

Natalia Milosevich[a], James McFarlane[a], Michael C. Gignac[a], Janessa Li[a], Tyler M. Brown[a], Chelsea R. Wilson[a], Lindsey Devorkin[b], Caitlin S. Croft[a], Rebecca Hof[a], Irina Paci[a], Julian J. Lum[b], Fraser Hof[a]*

[a]Department of Chemistry, University of Victoria, Victoria BC, V8W 3V6, Canada
[b]BC Cancer – Trev and Joyce Deeley Research Centre, Victoria BC, V8R 6V5, Canada
*Corresponding author: Fraser Hof. Email: fhof@uvic.ca

ABSTRACT: Epigenetic regulation of gene expression is in part controlled by post-translational modifications on histone proteins. Histone methylation is a key epigenetic mark that controls gene transcription and repression. There are five human polycomb paralog proteins (Cbx2/4/6/7/8) which use their chromodomains to recognize trimethylated lysine 27 on Histone 3 (H3K27me3). Recognition of the methyllysine side chain is achieved through multiple cation-pi interactions within an 'aromatic cage' motif. Despite high structural similarity within the chromodomains of this protein family, they each have unique functional roles and are linked to different cancers. Selective inhibition of different CBX proteins is highly desirable for both fundamental studies and potential therapeutic applications. We will report on a series of peptidic inhibitors that selectively target certain polycomb paralogs. We have identified peptidic scaffolds with sub-micromolar potency, and will report examples that are pan-specific and that are partially selective for individual members within the family. These results highlight important structure-activity relationships that allow for selectivity to be achieved through interactions outside of the methyllysine binding aromatic cage motif.

INTRODUCTION

Post-translational modifications on histones control the functions of chromatin through the actions of various epigenetic protein complexes.[1] Methyllysine reader proteins bind to post-translationally methylated lysine residues via an aromatic cage motif.[2] The five human polycomb paralog proteins (CBX2/4/6/7/8) recognize trimethyllysine residues on histone 3 and each participate in the multi-protein Polycomb Repressive Complex 1 (PRC1).[3] PRC1 serves to activate or silence genes by altering accessibility and compaction of chromatin.[4]

Each CBX protein has unique functional roles and displays distinct activities in different stages of cancer and in different tissues.[5-8] To better understand the biology of these proteins and

to test their potential as drug targets, chemical tools are needed to understand the phenotypes generated by inhibition. Significant progress has been made in understanding the biological impacts of inhibiting CBX7,[9,10] but comparatively much less is understood about the other CBX proteins.[11]

Typical approaches to generating small-molecule inhibitors have proven very challenging for CBX proteins. Early efforts in our group at virtual and small molecule screening for CBX7 did not yield potent inhibitors (unpublished results). The challenges associated with screening small molecules against the CBX proteins have also been reported by others.[9,11,12]

We have previously reported a peptide-driven approach to identify a series of sub-micromolar inhibitors targeting CBX4/CBX7 and CBX6.[12-14] Potent peptidic inhibitors of CBX4/CBX7 have also been identified by the Frye group and have shown activity in cellular based studies.[9,15] The first small molecule inhibitors of any CBX protein also targeted CBX7.[10,16] Recent work on new approaches to target the CBX proteins has identified inhibitors of CBX7 and CBX8 using a DNA-encoded library.[17]

No selective inhibitors have been reported for the other CBX proteins. Relatively little is known about the many biological roles of CBX2/4/6/8, and nothing is known about the impacts of chemical inhibition of any of these proteins. The delay in progress is because of the many challenges in targeting the CBX proteins. CBX chromodomains bind to their native substrates with weak affinity,[3] undergo an induced fit mode of binding,[9] and are between 79-98% similar within the family (Figure 1, Table 1).

Significant structural differences exist between the HP1 homologs (CBX1/3/5) and the polycomb paralogs (CBX2/4/6/7),[3] but the differences within the polycomb paralog family are subtle. The polycomb paralog proteins bind Kme3 through cation-pi interactions in their conserved aromatic cage (Figure 1B, C). The aromatic cage is nearly identical within the family, but differences exist in the protein structure in the nearby beta-groove (Figure 1B). Discovery and exploitation of these structural differences outside of the methyllysine binding motif are necessary for the development of selective inhibitors.

Our initial reports of CBX inhibitors first established the pentapeptide sequence Ac-FALKme3S and related analogs as potent inhibitors of CBX7 (Table 1).[12,13] A co-crystal structure of the Ac-FAYKme3S complex with CBX7 shows the peptide ligand forming several key binding interactions in the peptide-binding groove (Figure 1). The N-terminus of the ligand

sits in the protein's beta-groove, and the ligand's (–2) Ala residue points into a small hydrophobic pocket called the (–2) pocket. The hydrophobic clasp of the CBX proteins, made up of residues Val10 and Leu49, fold and clamp around the ligand (Figure 1A). The Kme3 group of the ligand forms several cation-pi interactions in the protein's aromatic cage, and the C-terminal Ser residue of the ligand peptide hydrogen bonds with the carboxylate side chain of Glu43 (Figure 1B).

We have previously reported a CBX6 ligand that contains a valine residue at the (–2) position relative to the ligand's Kme3 residue.[14] This substitution exploited differences within the family of the size of the protein's (–2) hydrophobic pocket. In this work, we report novel SAR of peptidic ligands that explore multiple regions of the CBX-ligand binding interface.

Table 1. Binding affinities for previously reported peptidic ligands with CBX7 and CBX8.[12]

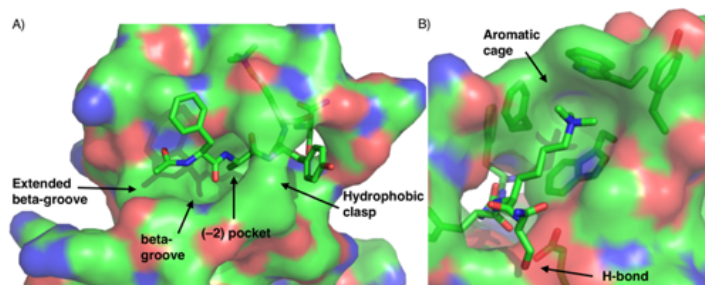|	IC_{50} values (µM) determined by CBX7-H3K27me3 disruption	K_d values (µM) determined by ITC	
Ligand	CBX7	CBX7	CBX8
Ac-FALKme3S	11 ± 0.4	2 ± 0.2	14.2 ± 2
Ac-FAYKme3S	6 ± 0.4	2 ± 0.1	12 ± 1.4
pBr-FALKme3S	5 ± 0.3	0.3 ± 0.05	5 ± 0.3

Figure 1. Key regions of the CBX proteins responsible for binding. A and B) Co-crystal structure of Ac-FAYKme3S in complex with CBX7 (pdb: 4MN3). Key structural features of the proteins binding pocket are labeled with black arrows.

The goal of this work was to further study the structural determinants of recognition, so that we can better target each family member. A secondary goal was to create novel dye-labeled inhibitors as chemical tools that would allow for new biochemical and biophysical studies of the CBX proteins. To this end, we synthesized a small library of peptidomimetic compounds and tested each compound with a panel of CBX proteins. The peptides synthesized are labeled with

the fluorescent dye fluorescein isothiocyanate (FITC). Labeled inhibitors were used in for multiple forms of testing, including direct fluorescence polarization assays (to determine affinity), microarray testing (to determine selectivity), and live cell imaging (to determine uptake and cellular localization). Each of these studies is unprecedented for the majority of methyllysine reader proteins in the CBX family.

Results

Synthesis of peptides

Peptides were synthesized using standard Fmoc solid-phase peptide synthesis protocols. Peptides contained either a beta-alanine residue at the N-terminus or a Lys(Mtt) residue at the C-terminus to allow dye labeling while still on resin. Fmoc-beta alanine was deprotected using standard protocols and the peptide was then reacted with FITC to produce compounds **1** and **2**. Peptides containing a C-terminal Lys-(Mtt) residue were selectively deprotected under mildly acidic conditions, followed by a reaction with FITC to give compounds **3**, **4**, **8-11** (Scheme 1).
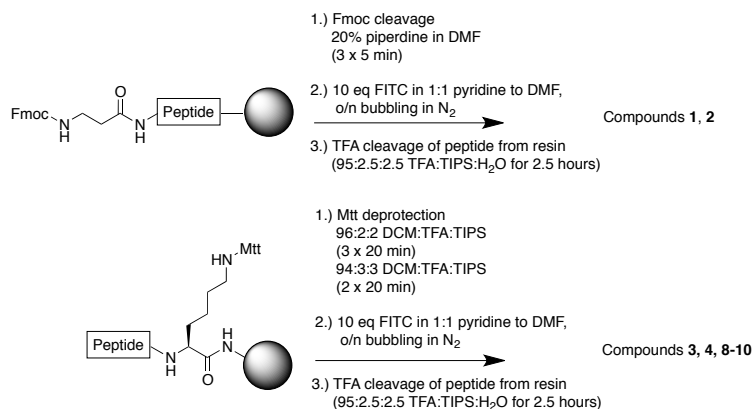
Scheme 1. Synthesis of peptidic compounds 1-4 and 8-10.

Fluorescence polarization-driven studies to understand polycomb paralog SAR

Dye labeling CBX7 inhibitor and N-terminal modifications

A dye-labeled analog of our previously identified CBX7 inhibitor FALKme3S was synthesized with a N-terminal beta-alanine residue (compound **1**) for covalent attachment to FITC. Compound **1** was screened against all CBX polycomb paralogs and the HP1 homolog CBX1. **1** displayed potent binding to all CBX proteins tested, near equipotent binding to CBX2/4/6/7 and weaker binding to CBX1 (an HP1 paralog) and CBX8 (Figure 2A). The

previously reported K_d values of Ac-FALKme3S with CBX7 and CBX8 are similar and within the same magnitude of those reported for the dye-labeled analog **1** (Table 1, Figure 2A).
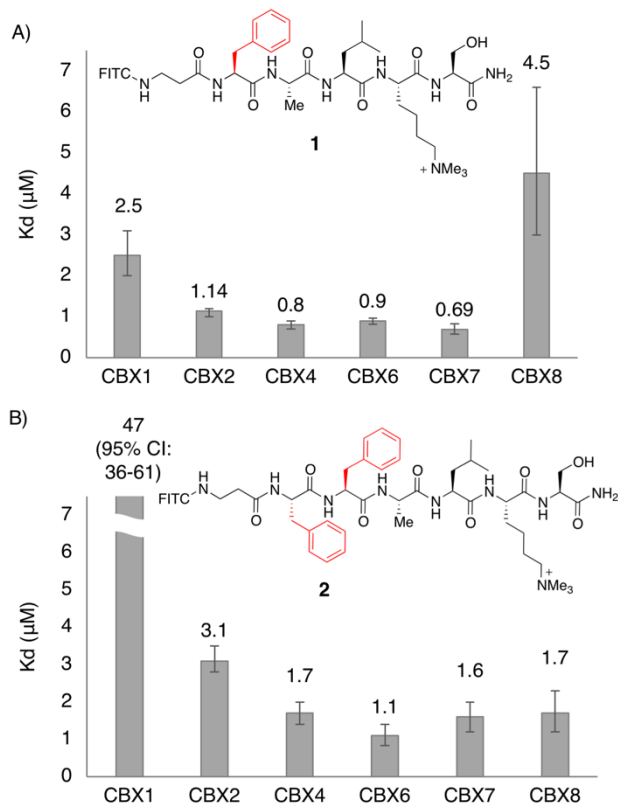

Figure 2: Peptidic inhibitors for CBX proteins and corresponding dissociation constants for CBX1/2/4/6/7/8. A) Binding affinities and chemical structure of compound **1**, B) binding affinities and chemical structure of compound **2**. Error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

The addition of a second phenylalanine to the N-terminus of **1** resulted in an increase in binding to CBX8 (2.6-fold) and decrease in binding to CBX7 (2.3-fold) (Figure 2B, **2**). Compound **2** showed a 19-fold decrease in binding to CBX1, 2 to 3-fold decrease in binding to CBX2/4 and no significant change in binding to CBX6.

Salt bridge interactions between ligands and CBX6 and CBX8

Another key structural difference between CBX7/CBX8 is the presence of a Arg9 residue in the protein near the C-terminus of the ligand interacts (Figure 3A). We predicted that an anionic substitution at the (+2) position would improve binding to CBX6/8 but not CBX7. Where the N-terminus is free of dye modification, we used the *p*-bromobenzamide end-capping group that we had previously shown to provide a boost in potency (pBr-FALKme3S, Table 1). Two analogous compounds were synthesized containing either a Leu residue at the (+2) position

(compound **3**) or a Glu residue (compound **4**) (Figure 3). Compound **3** was a potent inhibitor of CBX6 and CBX7 with K_d values of 78 and 11 nM. Addition of a Glu residue in the (+2) position provided compound **4** (previously reported).[14] The (+2) Glu gave a small increase in binding to CBX6 (1.6-fold) and CBX8 (1.4-fold), no change in binding to CBX7 and a decrease in binding to CBX1 (2.7-fold) (Figure 3C, D).
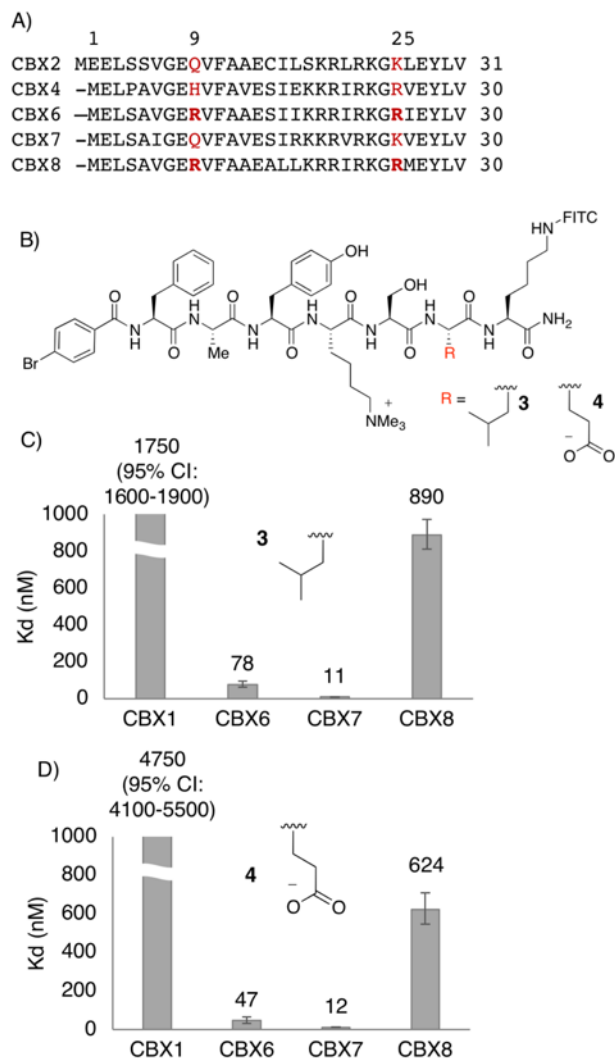

Figure 3: Interactions between anionic ligand substituent and cationic protein residue in CBX6/CBX8. A) Sequence alignments highlighting residues predicted to interact with anionic ligand substituents determined using ClustalW2 alignment, B) Chemical structure of compounds **3** and **4**, B) and C) Binding affinities of **3** and **4** to CBX1/6/7/8. K_d values are reported in nM and error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

To investigate further the role of the Glu residue and to understand the effects of the dye on binding, we synthesized analogues of compounds **3** and **4** lacking the C-terminal lysine and

FITC label. A competitive FP assay was used to determine the IC_{50} values of the unlabeled compounds **5**, **6** and **7** (Figure 4). Compound **4** was used as the dye-labeled probe in the competitive FP assay because of its good solubility and low K_d values for all CBX proteins. Compound **5**, lacking a residue at the (+2) position, displayed binding to CBX7 with 5- and 30-fold selectivity over CBX6 and 8 respectively. The addition of a Leu residue at the (+2) position (**6**) did not significantly change binding to any of the CBX proteins tested. Compound **7** with a (+2) Glu residue gave a 1.7- and 1.4-fold increases in binding to CBX6 and 8, respectively, and a 1.6-fold decrease in binding to CBX7. The trends in binding affinities observed with the dye-labeled compounds **3** and **4** are consistent with the trends seen for the unlabeled compounds **6** and **7**. Each of these sets of binding data suggest that the ligands' (+2) Glu residue favors binding to CBX6 and CBX8 over CBX7.
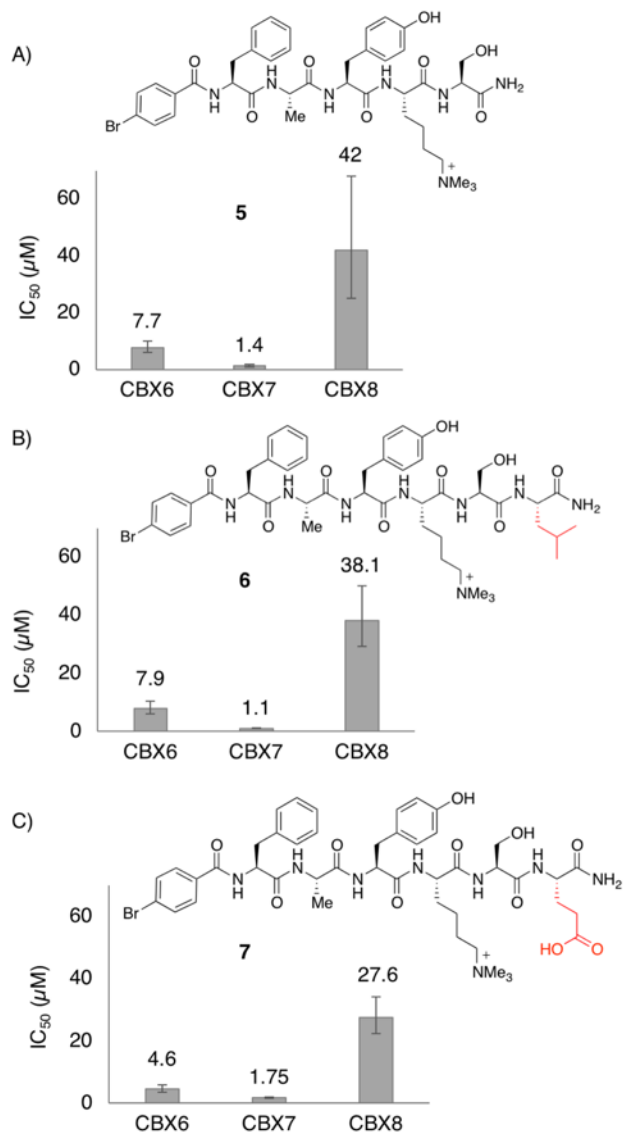
Figure 4: IC_{50} values of unlabeled ligands **5-7** for CBX6/7/8 determined by competitive FP. A) Binding affinities and chemical structure of compound **5**, B) compound **6**, C) compound **7**. Error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

Docking and MD simulations of compound 6 and 7

We sought to further examine the importance of the salt bridge interaction to binding of compounds **6** and **7**, using an accelerated sampling molecular dynamics technique developed in our group.[18] The SLICE method is a hybrid, iterative stochastic deterministic methodology that fuses AutoDock Vina[19] and the Amber16 molecular dynamics suite.[20] The method allows for fast structural identification of binding complexes of highly flexible hosts with peptidomimetic ligands. Binding of compounds **6** and **7** with CBX8 was investigated using SLICE.

8

Two 1-ns SLICE iterations were used to generate partially-induced-fit host configurations of CBX8. A third docking round of the ligands on the resulting host configurations provided starting points for a 100 ns molecular dynamics production run. After the two SLICE iterations, only the glutamate-containing compound **7** was observed to bind in the presumed correct orientation with the trimethyllysine placed in the aromatic cage. Template docking for compound **6**, using bound configurations of compound **7**, was performed to generate initial coordinates for the final production molecular dynamics of bound compound **6**. The simulation was required to examine the behavior of a leucine in the electropositive region of CBX8 depicted in Figure 5.
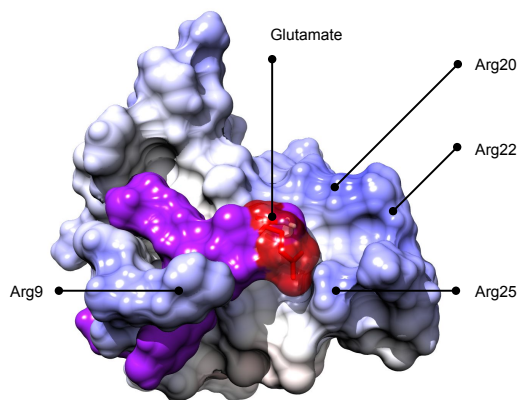
Figure 5. Final SLICE structure of compound **7** with CBX8. The initial docked pose of compound **7** (purple) glutamate (red) predicts a salt bridge with Arg25 instead of the proposed Arg9. The electrostatic potential surface area (APBS[21]) of CBX8 shows the entire region adjacent to the aromatic cage as an electropositive environment (blue surface) consisting of several potentially competing arginine residues.

The SLICE application to compounds **6** and **7** consistently showed a preference of CBX8 for **7**. However, the docked initial structure for the final production MD run presented a salt bridge of the Glu residue with Arg25, instead of the assumed Arg9 (Figures 3A and 6B). This is a consequence of two factors: (i) The protein surfaces of CBX6/8 in the area surrounding the aromatic cage present significant positive charge, for which both Arg9 and Arg25 are jointly responsible;[3] (ii) the docking procedure is insufficient to tease out effectively the binding differences between Arg9 and Arg25 in the complex solvation and thermal environment in which the binding process takes place. However, the Arg25 salt bridge dissociated in the initial stages of the production MD run.
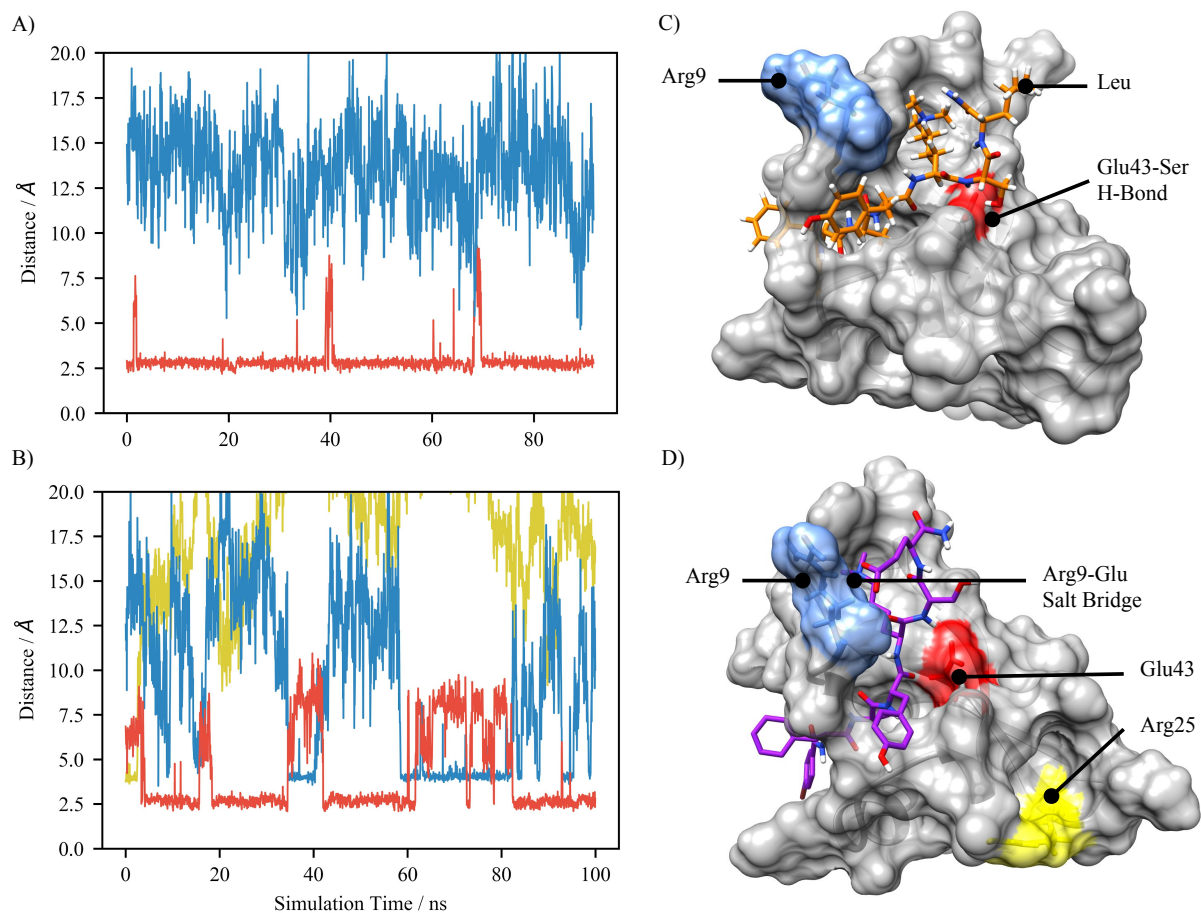
Figure 5. Final SLICE structure of compound **7** with CBX8. The initial docked pose of compound **7** (purple) glutamate (red) predicts a salt bridge with Arg25 instead of the proposed Arg9. The electrostatic potential surface area (APBS[21]) of CBX8 shows the entire region adjacent to the aromatic cage as an electropositive environment (blue surface) consisting of several potentially competing arginine residues.

The SLICE application to compounds **6** and **7** consistently showed a preference of CBX8 for **7**. However, the docked initial structure for the final production MD run presented a salt bridge of the Glu residue with Arg25, instead of the assumed Arg9 (Figures 3A and 6B). This is a consequence of two factors: (i) The protein surfaces of CBX6/8 in the area surrounding the aromatic cage present significant positive charge, for which both Arg9 and Arg25 are jointly responsible;[3] (ii) the docking procedure is insufficient to tease out effectively the binding differences between Arg9 and Arg25 in the complex solvation and thermal environment in which the binding process takes place. However, the Arg25 salt bridge dissociated in the initial stages of the production MD run.

Figure 6. Highlighted interactions of CBX8 with compounds **6** and **7**. A) The evolution of the distance between compound **6** leucine δ-carbon and CBX8 Arg9 guanidinium carbon through the 100 ns simulation is shown in blue. The compound **6** serine OH – Glu43 δ-carbon distance is shown in red, highlighting the stability of the hydrogen bond in this complex. In (B), distances shown are for: compound **7** glutamate δ-carbon – Arg25 guanidinium carbon (yellow), glutamate δ-carbon – Arg9 guanidinium carbon (blue), and compound **7** serine OH – Glu43 δ-carbon (red). C) MD Snapshot of compound **6** (stick representation) with the relevant residues on CBX8 highlighted in red and blue. D) MD Snapshot of compound **7** (stick representation) with CBX8 (with the relevant residues highlighted in blue, red and yellow).

In fact, monitoring the 100 ns trajectories for interaction distances showed the docked Arg25 salt bridge to be unstable compared to the interaction with Arg9. The latter was maintained for nearly half of the simulation, as shown in Figure 6B. An interesting feature arising from the simulation is the hydrogen bond between the compound **7** serine and Glu43 on CBX8. The bond shows a reverse correlation with the formation of the Arg9 salt bridge: both interactions cannot exist at the same time. The penalty of losing a hydrogen bond likely reduces the increased binding activity expected by the addition of newly formed salt bridge for

11

compound **7**. On the other hand, compound **6** maintained its Ser-Glu43 hydrogen bond throughout the simulation, as its leucine residue showed no preference of interaction in the region adjacent to the aromatic cage. Hydrophobic clasp distances, (–2) pocket placement, and aromatic cage interactions were stable for both compounds throughout the simulations and suggest that the key differences between the binding of **6** and **7** are the local interactions of the Leu/Glu substitutions (see Figure 7).
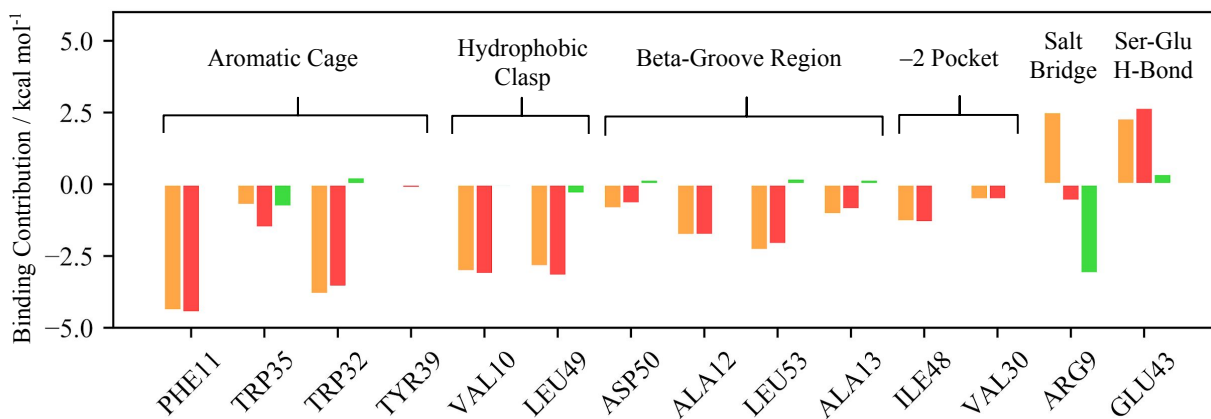
Figure 7. MMPBSA per-residue binding free energies for CBX8 residues in the interaction region. Orange and red bars show energy values for the interaction with compound **6** (E**6**) and **7** (E**7**), respectively. The relative stabilization contribution of compound **7**, (E**7**-E**6**), is presented in green. Residues that did not contribute to binding for either compounds **6** and **7** were not included.

Binding energy contributions

To explore the contributions of competing interactions to the binding energy, we calculated the per-residue free energies of binding using MMPBSA.py.[22] Coordinates for the calculation for host/ligand/complex were all taken from the 100 ns trajectory. This single trajectory approach ignores the induced-fit energy penalty of the host, which is likely significant, based on our experience with apo-protein simulations of CBX8. However, this contribution is also likely the same for **6** and **7**, given their structural similarities, and can be neglected for the purpose of this comparison.

Further energy decomposition, specifically the electrostatic and solvation energies of the residues involved in the Arg8 salt-bridge and surface Glu43 hydrogen bond (Table 2), support our structural findings regarding the competition between salt bridge and hydrogen bond formation for compound **7**. Overall, the hydrogen bond interaction for compound **7** is weaker than that of compound **6**. This is an overall result that arises from significantly more stable

electrostatic contributions for compound **6**, due to residence time, balanced out by a proportionally large solvent exclusion effect.

On the other hand, the salt bridge interaction favors compound **7**, as initially designed: The interactions in the salt bridge region are slightly destabilizing for compound **6**, whereas the intermittent compound **7** Glu – Arg9 interaction provides an overall stabilizing effect. Note that the values listed in Table 2 are overall per-residue values, and not pairwise energies.

The difference of the total MMPBSA binding free energies for Compounds **6** and **7**, using the Poisson Boltzmann solvent, were found to be 3.79 kcal/mol in favor of compound **7**.

Table 2. MMPBSA.py Per-Residue Energies for Selected Residues in Salt Bridge and Adjacent Glu43 Hydrogen Bond

		$E_{el}{}^{\alpha}$	$E_{sol}{}^{\beta}$	$E_{tot}{}^{\gamma}$	$E_{int}{}^{\varepsilon}$
CBX8 / compound 6					
Hydrogen Bond	Glu43 (CBX8)	−37.32	40.61	2.31	−4.16
	Ser1 (compound **6**)	−13.01	7.07	−6.47	
Salt Bridge Region	Arg8 (CBX8)	33.13	−29.38	2.52	2.05
	Leu2 (compound **6**)	−0.49	0.65	−0.47	
CBX8 / compound 7					
Hydrogen Bond	Glu43 (CBX8)	−16.08	19.71	2.68	−2.11
	Ser1 (compound **7**)	−9.97	5.71	−4.79	
Salt Bridge Region	Arg8 (CBX8)	−22.58	23.66	−0.58	−2.81
	Glu2 (compound **7**)	−68.53	66.89	−2.23	
	Glu2 (Compound **7**)	−68.53	66.89	−2.23	

Electrostatic potential$^{\alpha}$, PBSA solvation energy$^{\beta}$, total per-residue energy$^{\gamma}$, and combined total energies$^{\varepsilon}$ (sum of E_{tot} for the two residues) are presented for the interacting residues on CBX8 and the two ligands. All energies reported in kcal/mol.

Ligand substitutions at the (–1) and (–2) position

Returning to synthesis and testing of new ligands, we sought to explore substitutions of the (–2) residue within the scaffold of **4.** An additional difference between CBX7 and CBX8 is the hydrophobic (–2) pocket. The (–2) pocket is different sizes in CBX7 and CBX8. We sought to exploit this difference by adding a larger alkyl substituent at the (–2) position. Within the scaffold of compound **4**, the methyl substituent at the (–2) position was replaced with an ethyl group to produce compound **8** (Figure 8A).[14] This subtle change weakened binding to all

partners except CBX8. Significantly weaker binding was observed for **8** with CBX1 (31-fold), between a 2- to 5-fold decrease in binding for CBX2/4/6/7 and no change in binding to CBX8.
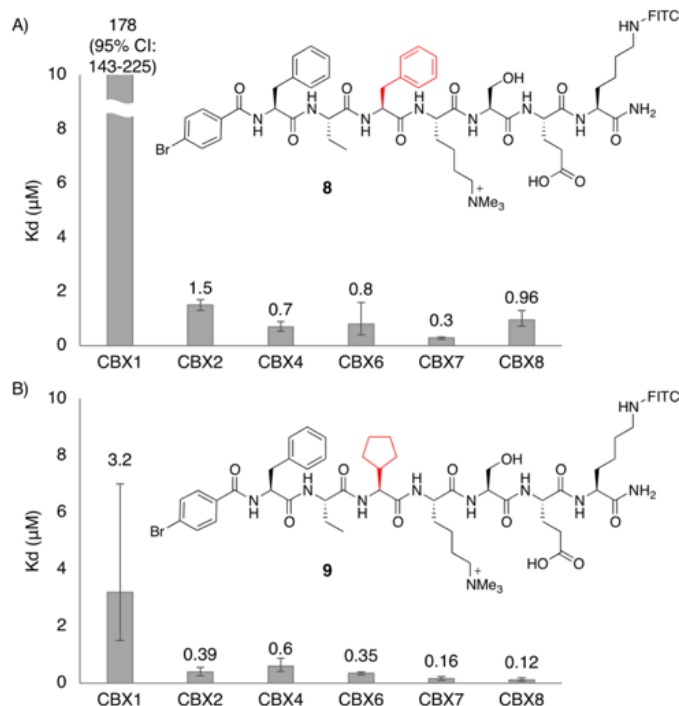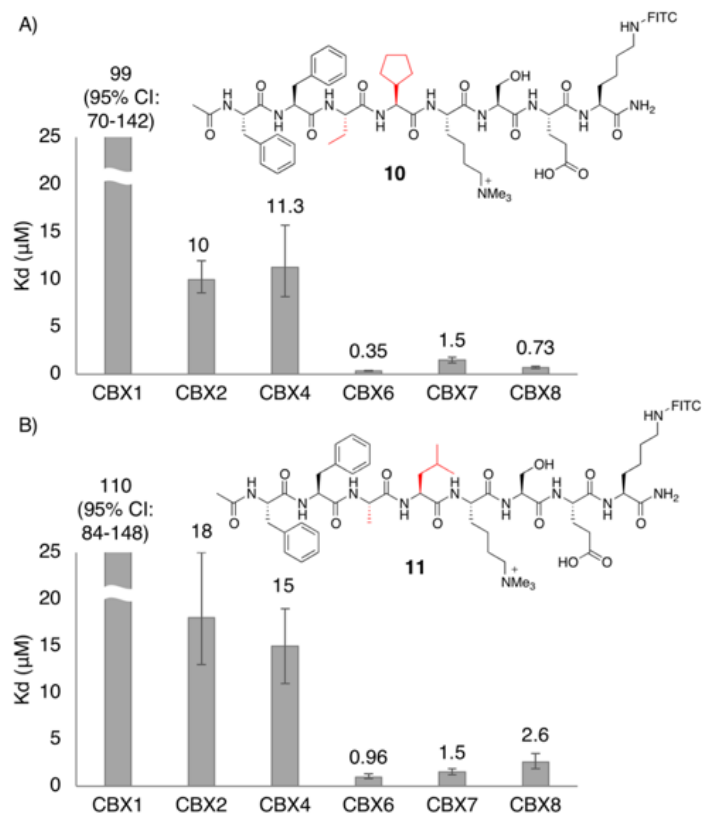

Figure 8: Peptidic inhibitors **8** and **9** and corresponding dissociation constants for CBX1/2/4/6/7/8. A) Binding affinities and chemical structure of compound **8**, B) and compound **8**. Error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

We next replaced the Phe residue with a cyclopentyl moiety at the (−1) position to give compound **9** (Figure 8B). This swap produced the most potent CBX8 inhibitor reported to date with a K_d value of 120 nM. In addition to improving potency for CBX8, the cyclopentyl side chain in the (−1) position increased binding to all CBX proteins tested. Compound **9** is 26-fold selective for CBX8 over CBX1, 3-fold selective over CBX2/6, and 5-fold selective over CBX4. The difference in affinity of **9** for CBX7 and CBX8 is not significant.

We hoped to push further toward CBX8 selectivity by adding a second Phe residue to **9**, as in compound **2**. To this end, we synthesized compound **10** containing an acetylated N-capping Phe residue in combination with the (−2) ethyl group and (−1) cyclopentyl side chain (Figure 9A). **10** is the first compound in this series with stronger affinity to CBX8 compared to CBX7. **10** is most potent for CBX6, and is 4 and 2-fold selective for CBX6 and 8 over CBX7. In respect to CBX8, **10** is between 14-135x selective over CBX1/2/4.

Replacing the (–2) and (–1) substituents of **10** to Ala-Leu residues known to be favored by CBX1/2/4/7 did not significantly change binding to these proteins, but did decrease binding to CBX6 and 8 by a factor of 2.7 and 3.5 (Figure 9B). Compounds **10** and **11** are selective for CBX6/7/8 over CBX1/2/4.

Selective inhibition of CBX7 over CBX4 has not yet been reported. The CBX4 chromodomain is the most similar to CBX7 (similarity score of 90%),[3,14] both bind the native histone substrate with similar affinity,[3] and almost all CBX7 ligands reported to date have had similar affinities for CBX4.[9,10,12,13] Interestingly, both **10** and **11** show significantly weaker binding to CBX4 compared to CBX7. Future efforts on selective inhibition of CBX7 may benefit from extended engagement of the peptide-binding groove.

Figure 9: Peptidic inhibitors **10** and **11** and corresponding dissociation constants for CBX1/2/4/6/7/8. A) Binding affinities and chemical structure of compound **10**, B) and compound **11**. Error bars are reported as asymmetric 95% confidence intervals from experiments done in duplicate for CBX1/2/4/6/7 and triplicate for CBX8.

Selectivity studies using a methyl reader protein microarray

Using a microarray to test our fluorescently tagged peptides provides both validation of our FP results and more diverse knowledge on selectivity beyond the CBX family of proteins.

We utilized a protein microarray made up of 98 different recombinant human methyl reader proteins (including all CBX proteins) arrayed in duplicate.[23] Initial testing of tetramethylrhodamine isothiocyanate (TRITC) dye-labeled inhibitors produced the expected binding trends but with a high degree of background fluorescence. To prevent this, a biotinylated analog of **3** was synthesized (compound **12**, Figure 10). The microarray was incubated with the probe, and the binding of the probe was imaged using a fluorescent streptavidin reagent.

The inhibitor tested in the microarray showed excellent selectivity for CBX proteins over a broad selection of other methyl readers. Some off-target binding to the chromodomain Y like (CDYL) proteins was observed,[24] along with weak off-target binding to the chromodomain-containing mortality factor 4-like protein (MORF4L1 or MRG15). Compound **12** showed the highest selectivity for CBX4 and 7, with weaker binding observed to CBX2/6/8 (Figure 10).
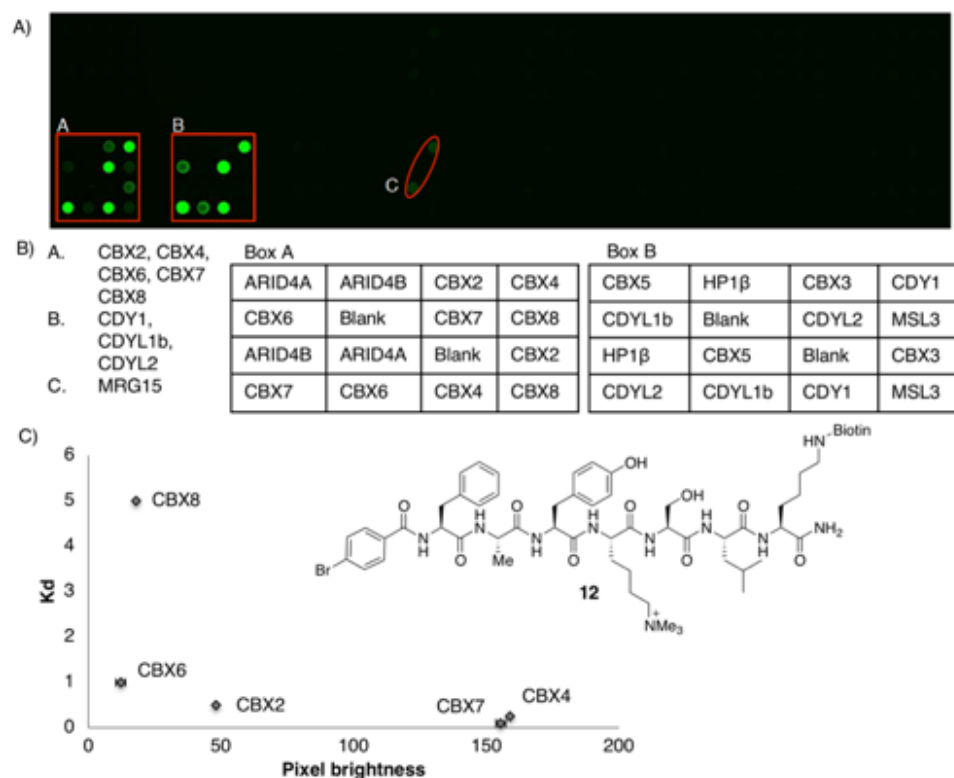
Figure 10: Protein microarray made up of 98 human methyl reader proteins shows peptidic inhibitors are selective against a broad set of methyl reader targets. Proteins were coated onto the wells in each plate with each square representing a family of proteins. A) Protein microarray with compound **12**. B) Legend of proteins in box A and B. Box A contains polycomb paralogs, box B contains HP1 homologs and CDYL proteins, and box C highlights weak off target binding to MRG domains. The full map of the protein microarray is available in Supporting Figure S26. C) K_d values of compound **3** (dye labeled analog of **12**) with CBX proteins plotted against brightness of each spot from protein microarray (quantified using ImageJ) and chemical structure of biotin-tagged compound **12**.

To ensure this data was in agreement with K_d values we collected from our FP assays, ImageJ software was used to quantify the brightness of each microarray spot and this was plotted against the K_d values for analogous dye-labeled compound of **12** (Figure 10C). In general, it is clear that proteins with weak K_d values also have weak pixel intensity showing that the microarray data is in qualitative agreement with our solution-phase results. CBX6 shows significantly less pixel brightness than would be expected from its K_d for this probe ligand, suggesting that microarrayed CBX6 does not accurately represent solution activity.

Cell-based data

We sought to determine the ability of the dye-labeled peptidic agents to be used in cell-based studies. Our efforts to study the inhibitors in cells included live cell imaging, flow cytometry, and MTT-based viability studies.

Cell-based studies showed varying degrees of cellular uptake and minimal changes cell cycle distribution as well as viability. Live cell imaging and immunofluorescence confocal imaging using TOV21G ovarian carcinoma cells and PC3 prostate cancer cells, treated with compounds **2** and **4** did not show compounds entering the cytoplasm or nucleus. Inhibitors were seen in characteristic punctate dots in the cytoplasm at concentrations above 10 μM, suggesting endosomal entrapment (data not shown). Flow cytometry experiments with TOV21G cells treated with compound **2** showed uptake of the inhibitor, supported by the formation of a population of fluorescent cells in each case (Figure S1). In the presence of the inhibitor, there was no observable change in the distribution of cells across Go, G1, S or G2/M (Table S1). This is also consistent with compounds being taken up in endosomes, but not able to escape to cytoplasm or nucleus in order to have a biological effect. We also carried out MTT assays, used to measure metabolic activity and cell viability, with TOV21G cells treated with inhibitor **4**, and unlabeled analogs **6** and **7**. A slight decrease in cell viability was seen for the cells treated with **4**, however we did not observe a dose-response for the unlabelled analogs **6** and **7** (Figure S2). We conclude that these compounds are unfortunately not suitable for cell-based activity studies.

Discussion

Potent inhibitors were developed that are either pan-specific or partially selective within the polycomb paralog family. While none of the inhibitors developed were selective for a single

CBX protein, we have identified several compounds with sub-micromolar affinity including two inhibitors that bind CBX6/7 with IC_{50} values < 100 nM (**3**, **4**).

Several compounds displayed class selectivity, with preferential binding to CBX6/7/8 over CBX1/2/4. We report inhibitors with a novel affinity for CBX6/8 over CBX7, which is unusual because CBX7 almost always gives higher affinity binding to peptidic and small molecule ligands.[9,10,12,13,15-17] This includes the most potent CBX8 inhibitor to date (**9**, 120 nM) and an inhibitor that is 2-fold selective for CBX8 over CBX7 (**10**). Our work also produced the first peptidic inhibitor that is 10-fold selective for CBX7 over the highly similar chromodomain of CBX4 (compound **10**).

We have identified areas of the peptide-binding groove that are different within the family of CBX proteins and form distinct interactions with the peptide ligands reported. Substitutions at the (−1) position effect the hydrophobic clasp of the CBX proteins and alter binding affinities differentially within the family. Addition of a cyclopentyl moiety in the (−1) position increases binding to all CBX members with the greatest increase observed for binding to CBX8. The addition of two Phe residues at the N-terminus of the ligand diminishes binding to CBX1/2/4 to give inhibitors that are selective for CBX6/7/8 (**9**, **10**). We predict that future efforts targeting the extended beta-groove, (−1) and (−2) position of the protein will aid in the discovery of selective inhibitors.

Substitutions at the (+2) position of the ligand that participate in salt-bridge interactions alter the binding to CBX6 and CBX8. Our prediction that a Glu residue at the (+2) position could improve binding affinity to CBX6 and 8 by interacting with Arg9 was partially correct. We did observe favorable salt-bridge interactions with a ligand (+2) Glu interacting with the protein residues Arg9 and Arg25 in MD simulations with CBX8. However, this interaction destabilized a key hydrogen bond between the (+1) serine in the ligand with Glu43 in CBX8.

The inhibitors developed are highly selective for the CBX polycomb proteins over other methyllysine reader proteins. The protein microarray studies with **12** show the inhibitors to be highly selective for the CBX polycomb paralogs over many other methyllysine readers. Potent off-target binding was observed with the CDY proteins. The peptidomimetic inhibitor UNC3866 targeting CBX4/CBX7, was discovered to have off target binding to the CDY proteins.[9] The authors followed up this work by repurposing the scaffold to develop a combinatorial peptide library resulting in the discovery of potent inhibitors of CDYL1/CDYL2.[24] Future optimization

of the inhibitors is needed to differentiate binding between the two highly similar families of chromodomains.

While the dye-labeled agents gave access to new biochemical assays, our efforts to use them in cell-based studies were unsuccessful. Several earlier CBX inhibitors have shown poor cellular activities due to low permeability, and the addition of FITC is unlikely to have improved this situation. Poor solubility of the peptidic inhibitors was another challenge and limited our ability to test higher concentrations of the inhibitors. Future efforts to use the reported inhibitors in cell-based studies will require alternative delivery strategies.

Conclusion

The goals of this work were to study the structural determinants of recognition for the CBX proteins and to create dye-labeled inhibitors as tools for biochemical and biophysical studies of the CBX proteins. We have successfully created potent inhibitors for each CBX polycomb paralog protein. The inhibitors reported are useful tools for biochemical assays and for future competitive based screens for the discovery of new ligands.

The SAR learned from this work provides new insights into the structure and molecular recognition properties of these proteins outside of the previously well explored aromatic cage binding pocket. This SAR lays the foundation for creating highly selective and cell-permeable chemical tools to study the role of CBX proteins in epigenetic regulation, which we will report in due course.

ASSOCIATED CONTENT

Supporting Information

Supplementary Figures and experimental methods are available below.

AUTHOR INFORMATION

Corresponding Author

* Fraser Hof. Email: fhof@uvic.ca

Author Contributions

The manuscript was written by N.M with revisions provided by F.H. N.M carried out the experimental design, synthesis, and characterization of peptides and binding affinities. J.M wrote

and conceived of the modeling and MD simulations with input for revisions provided by I.P. C.C and C.W assisted with peptide synthesis. M.G, T.B and R.B grew and purified all CBX proteins. N.M, M.G and T.B carried out the FP assays. J.L and L.D conducted the cell-based studies. All authors have given approval to the final version of the manuscript.

ABBREVIATIONS

Ac, acetyl; arg, arginine; CBX, chromobox; CDY, chromodomain Y, chromodomain, chromatin organization modifier; DNA, deoxyribonucleic acid; FITC, fluorescein isothiocyanate; FP, Fmoc, fluorenylmethoxyloxycarbonyl; FP, fluorescence polarization; Glu, glutamic acid; H3, histone 3; H3K9me3, trimethylated lysine 9 on histone 3; H3K27me3, trimethylated lysine 27 on histone 3; HP1, heterochromatin protein; IC_{50}, inhibitory concentration that reduces effect by 50 percent; K, lysine; Kme3, trimethyllysine; K_d, dissociation constant; Leu, leucine; MeCN, acetonitrile; MD, molecular dynamics; MORF4L1, mortality factor 4-like protein.

REFERENCES

1.   Rothbart SB, Strahl BD. Interpreting the language of histone and DNA modifications. *Biochim Biophys Acta.* 2014;1839(8):627-643.

2.   Musselman CA, Khorasanizadeh S, Kutateladze TG. Towards understanding methyllysine readout. *Biochim Biophys Acta.* 2014;1839(8):686-693.

3.   Kaustov L, Ouyang H, Amaya M, et al. Recognition and specificity determinants of the human cbx chromodomains. *J Biol Chem.* 2011;286(1):521-529.

4.   Aranda S, Mas G, Di Croce L. Regulation of gene transcription by Polycomb proteins. *Science advances.* 2015;1(11):e1500737.

5.   Richly H, Aloia L, Di Croce L. Roles of the Polycomb group proteins in stem cells and cancer. *Cell Death Dis.* 2011;2:e204.

6.   Vincenz C, Kerppola TK. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proc Natl Acad Sci U S A.* 2008;105(43):16572-16577.

7.   Ma RG, Zhang Y, Sun TT, Cheng B. Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *Journal of Zhejiang University Science B.* 2014;15(5):412-428.

8.   Koppens M, van Lohuizen M. Context-dependent actions of Polycomb repressors in cancer. *Oncogene.* 2015;35:1341.

9.   Stuckey JI, Dickson BM, Cheng N, et al. A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. *Nat Chem Biol.* 2016;12(3):180-187.

10.   Ren C, Morohashi K, Plotnikov AN, et al. Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem Biol.* 2015;22(2):161-168.

11.   Milosevich N, Warmerdam Z, Hof F. Structural aspects of small-molecule inhibition of methyllysine reader proteins. *Future Med Chem.* 2016;8(13):1681-1702.

12.   Simhadri C, Daze KD, Douglas SF, et al. Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J Med Chem.* 2014;57(7):2874-2883.

13.   Simhadri C, Gignac MC, Anderson CJ, et al. Structure–Activity Relationships of Cbx7 Inhibitors, Including Selectivity Studies against Other Cbx Proteins. *ACS Omega.* 2016;1(4):541-551.

14.   Milosevich N, Gignac MC, McFarlane J, et al. Selective Inhibition of CBX6: A Methyllysine Reader Protein in the Polycomb Family. *ACS Med Chem Lett.* 2016;7(2):139-144.

15.   Stuckey JI, Simpson C, Norris-Drouin JL, et al. Structure–Activity Relationships and Kinetic Studies of Peptidic Antagonists of CBX Chromodomains. *J Med Chem.* 2016;59(19):8913-8923.

16.   Ren C, Smith SG, Yap KL, et al. Structure-Guided Discovery of Selective Antagonists for the Chromodomain of Polycomb Repressive Protein CBX7. *ACS Med Chem Lett.* 2016;7(6):601-605.

17.   Denton KE, Wang S, Gignac MC, et al. Robustness of In Vitro Selection Assays of DNA-Encoded Peptidomimetic Ligands to CBX7 and CBX8. *SLAS Discov.* 2018;23(5):417-428.

18.   James M, Katherine K, Irina P. *Accelerated Molecular Dynamics for Structural Prediction in Protein/Peptide Binding: The SLICE Method.* 2019.

19.   Trott O, Olson AJ. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J Comput Chem.* 2010;31(2):455-461.

20.   *AMBER 2018* [computer program]. San Francisco: University of California; 2018.

21.   Jurrus E, Engel D, Star K, et al. Improvements to the APBS biomolecular solvation software suite. *Protein Sci.* 2018;27(1):112-128.

22.   Miller BR, McGee TD, Swails JM, Homeyer N, Gohlke H, Roitberg AE. MMPBSA.py: An Efficient Program for End-State Free Energy Calculations. *Journal of Chemical Theory and Computation.* 2012;8(9):3314-3321.

23.   Kim J, Daniel J, Espejo A, et al. Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* 2006;7(4):397-403.

24.   Barnash KD, Lamb KN, Stuckey JI, et al. Chromodomain Ligand Optimization via Target-Class Directed Combinatorial Repurposing. *ACS Chemical Biology.* 2016;11(9):2475-2483.

Supplementary information for:

Pan-specific and partially selective dye-labeled peptidic inhibitors of the polycomb paralog proteins

Natalia Milosevich[a], James McFarlane[a], Michael C. Gignac[a], Janessa Li[a], Tyler M. Brown[a], Chelsea R. Wilson[a], Lindsay Devorkin[b], Caitlin S. Croft[a], Rebecca Hof[a], Irina Paci[a], Julian J. Lum[b] and Fraser Hof[a]*

[a]Department of Chemistry, University of Victoria, Victoria BC, V8W 3V6, Canada
[b]BC Cancer Agency – Trev and Joyce Deeley Research Centre, Victoria BC, V8R 6V5, Canada
*Email: fhof@uvic.ca

Table of Contents

---

*Fraser Hof. e-mail: fhof@uvic.ca

Supplementary Figures S1-S2 and Supplementary Table S1

Figure S1: Uptake efficiency and PI based cell cycle analysis for TOV21G cells treated with compound **2** show uptake of inhibitor with no change to cell cycle distribution. A) Uptake efficiency based on FITC intensity. Samples treated with equivalent volume of vehicle control (DMSO - light blue trace compared to purple trace) show increased fluorescent signal arising from 50 μM treatment of compound **2**. B) Cell cycle distribution (G1, S and G2) based on PI intensity. Samples treated with equivalent volume of vehicle control (DMSO - light blue trace compared to purple trace) show no change in cell cycle distribution. Samples were run on Becton Dickinson FACS Calibur™ and analyzed using FlowJo 7.6.5 software.

Table S1: Cell cycle distribution (G1, S, G2) of TOV21G cells remained constant with varying concentrations of compound **2** (measured from Figure S1).

Concentration of 2	G1 (%)	S (%)	G2 (%)
0 μM (no DMSO)	49	14	21
0 μM (0.25% DMSO)	66	14	12
5 μM (0.02% DMSO)	60	13	16
10 μM (0.05% DMSO)	65	14	12
25 μM (0.1% DMSO)	66	15	11
50 μM (0.25% DMSO)	68	12	12
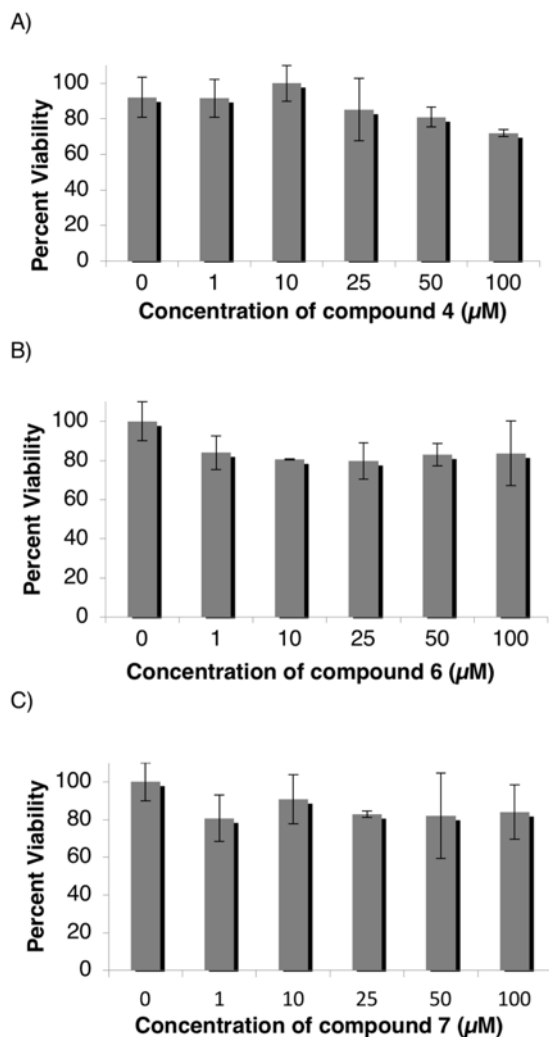
Figure S2: Cell viability determined using an MTT assay carried out with TOV21G cells treated with compounds **4**, **6** and **7**. A) Percent viability of TOV21G cells treated with **4**, B) Percent viability of TOV21G cells treated with **6**, C) Percent viability of TOV21G cells treated with **7**.

General methods - 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.

Side-chain functionalization with fluorescein isothiocyanate (FITC)

Side-chain functionalization with fluorescein isothiocyanate (FITC) was achieved by incorporation of either a Fmoc-beta-Alanine or a side-chain Mtt-protected lysine (see Figure 3). For functionalization of Fmoc-beta-Alanine, Fmoc was removed following standard protocols after synthesis was completed on resin. A solution of 20% piperdine in DMF (10 mL) was added

to the resin and bubbled with N_2 for 5 minutes (repeated 3×). The resin was washed 3× with DMF between each reaction. FITC (5 equiv) was added to a 10 mL solution of 1:1 DMF and pyridine. This solution was added to the resin and bubbled with N_2 at ambient temperature overnight.

For functionalization of side-chain Mtt-protected lysine, the Mtt protecting group was cleaved after complete synthesis on resin. Selective cleavage was achieved with treatment of 2:2:96 TFA/triisopropylsilane/CH_2Cl_2 for 20 minutes (×4). The resin was filtered and washed with CH_2Cl_2 (3 × 5 mL) and DMF (3 × 5 mL). FITC (5 equiv) was added to a 10 mL solution of 1:1 DMF and pyridine. This solution was added to the resin and bubbled with N_2 at ambient temperature overnight.

Following the functionalization with FITC, all peptide/resin solutions were filtered and the resin washed with DMF (3 × 5 mL) and CH_2Cl_2 (3 × 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 × 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 × 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 bubbled with N_2 at ambient temperature overnight. Cleavage and purification is the same as described above.

Compound Characterization Data

A)

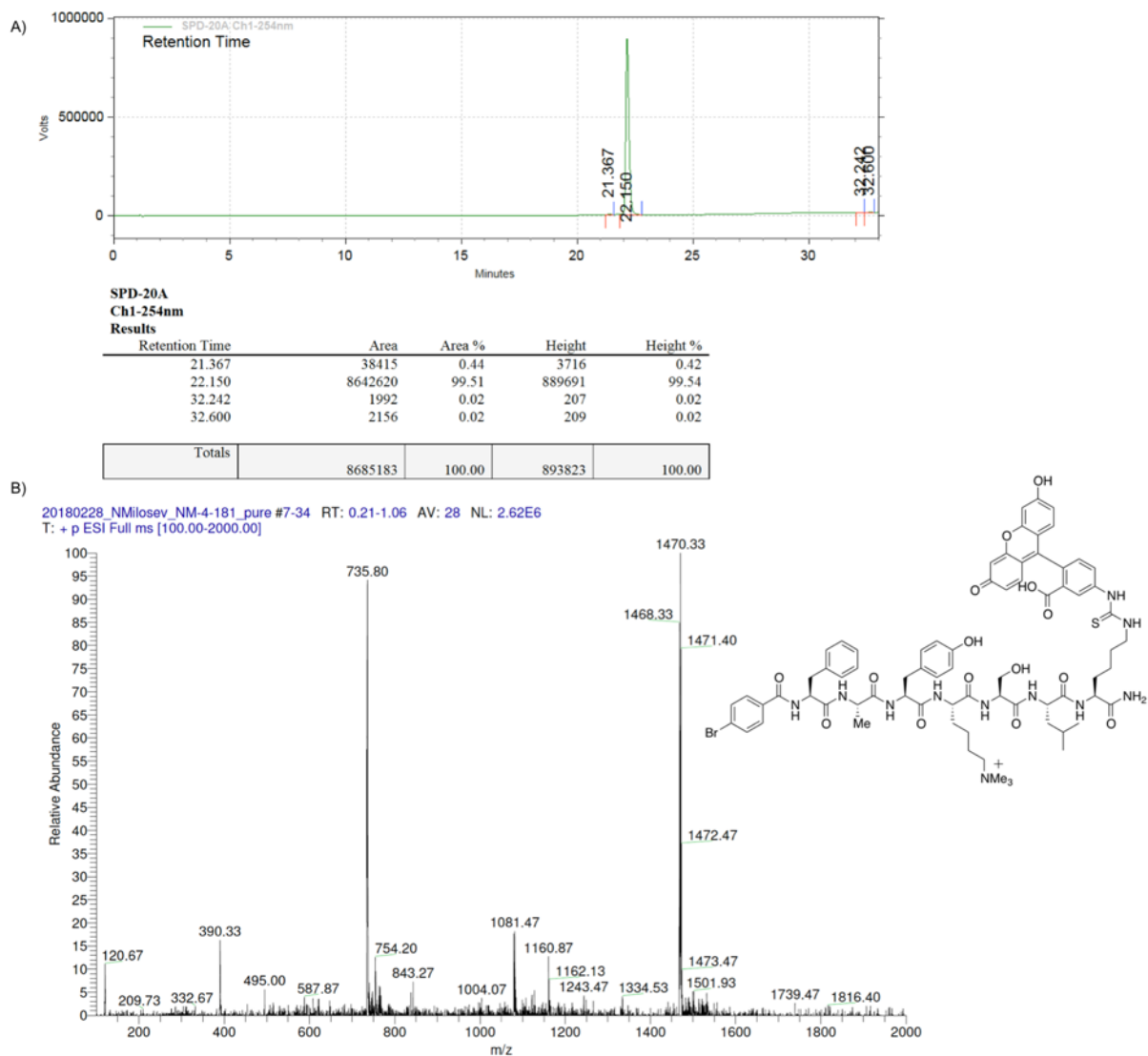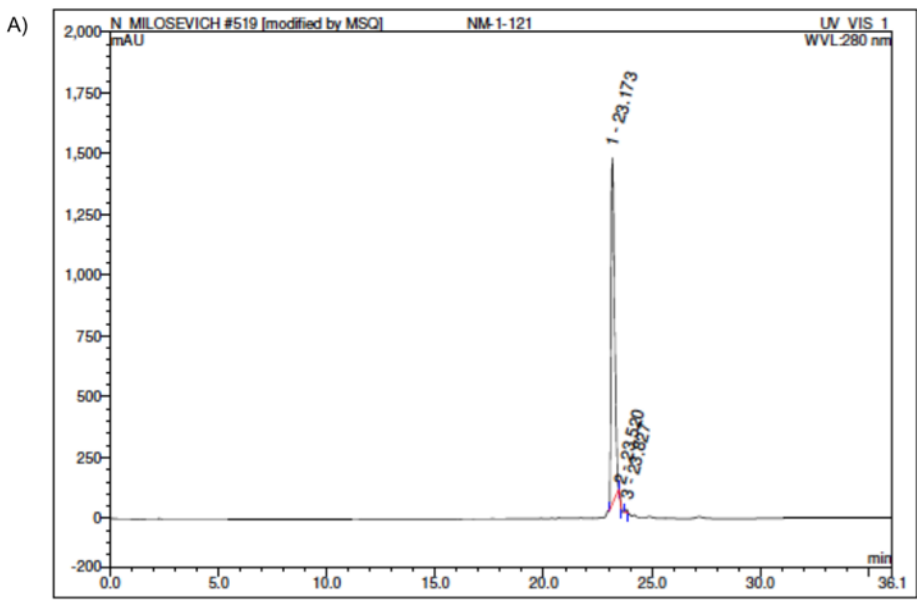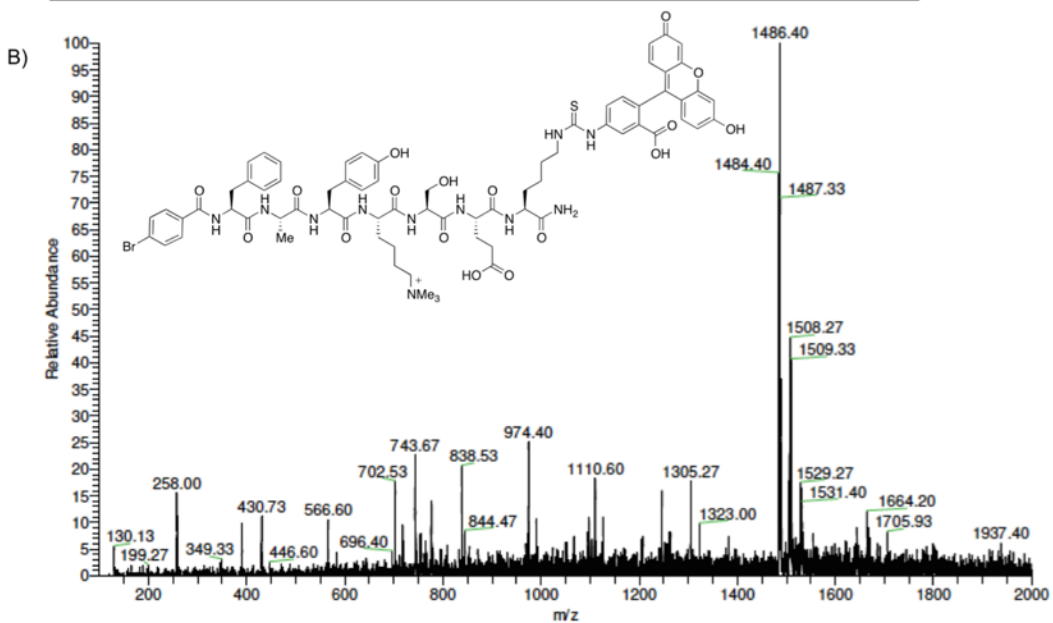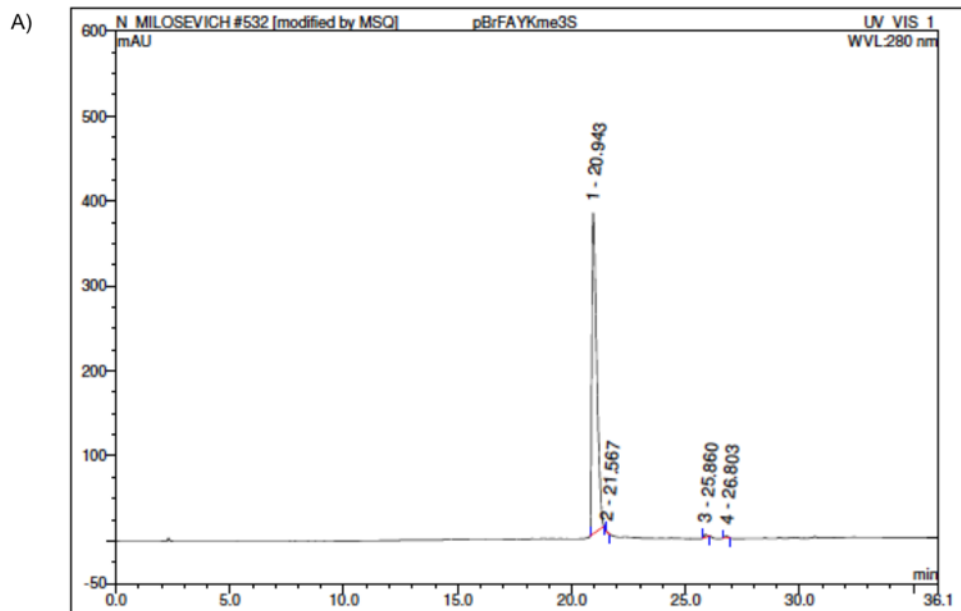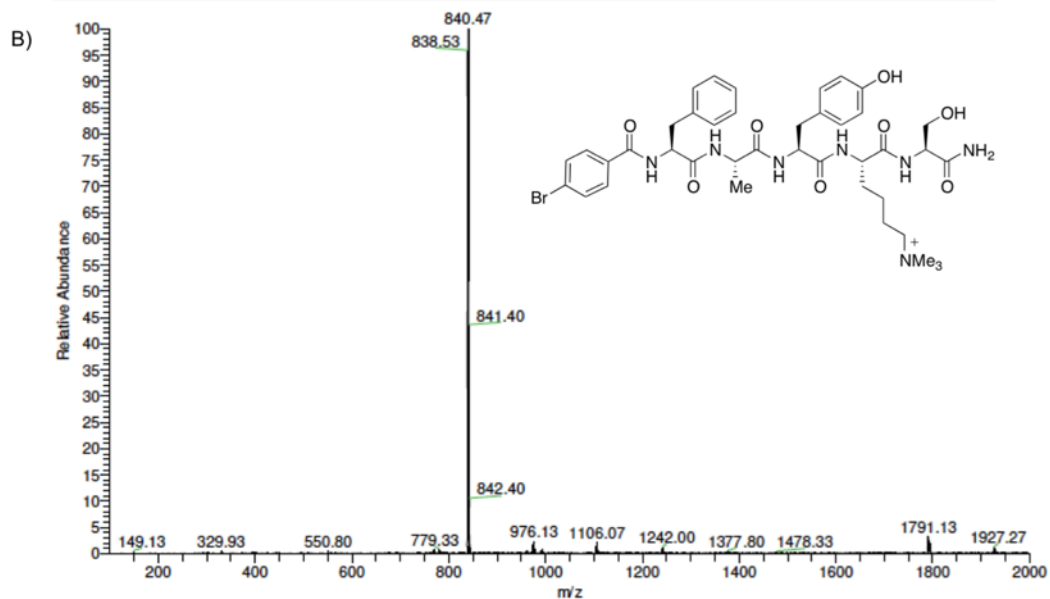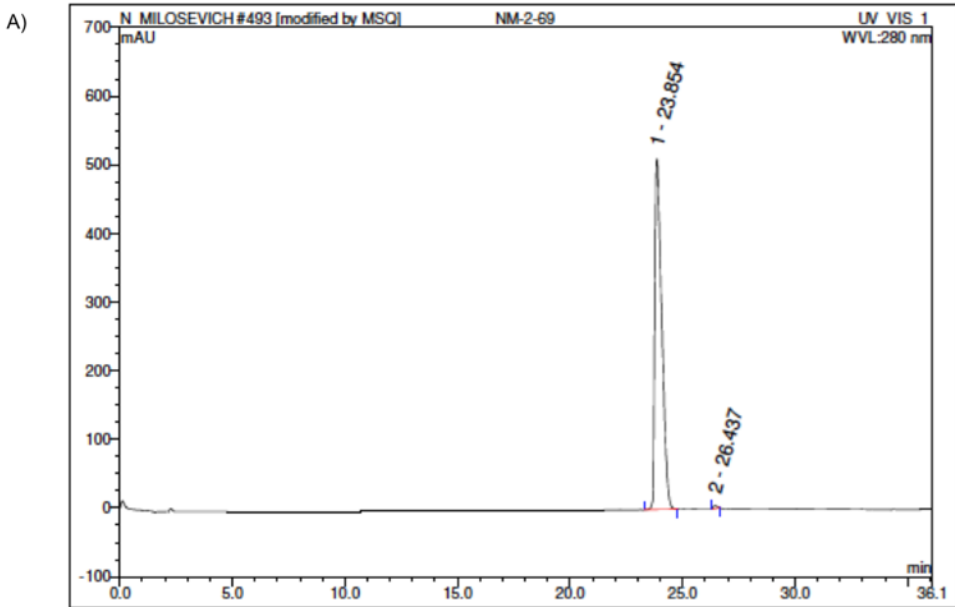| No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	16.464	0.471	5.219	0.35	0.39
2	16.790	134.480	1326.013	99.65	99.61
Total:		134.950	1331.232	100.00	100.00

B)
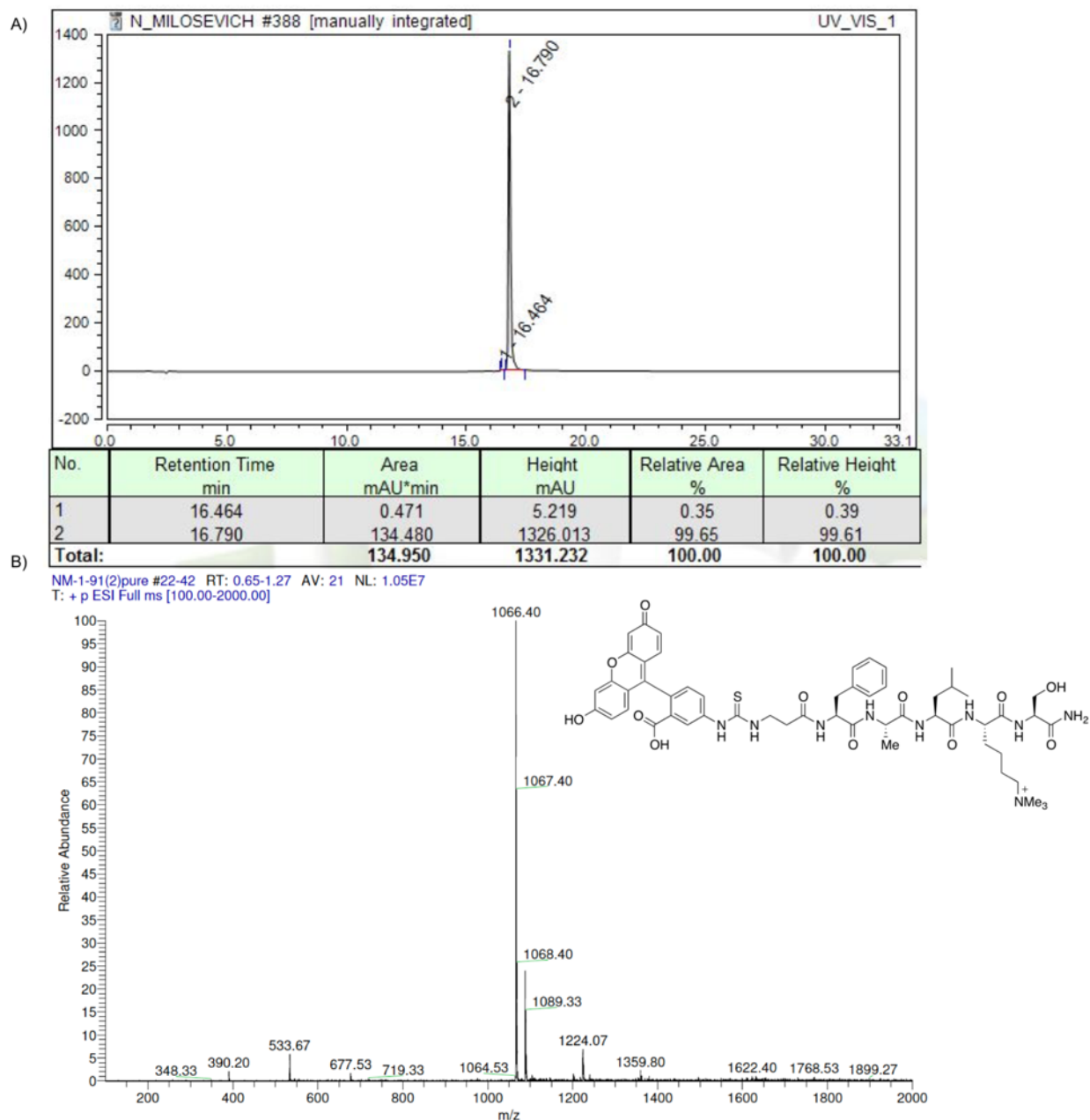
Figure S3. Characterization data for compound **1** A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{54}H_{68}N_9O_{12}S$: 1066.47; found: 1066.40.

A)

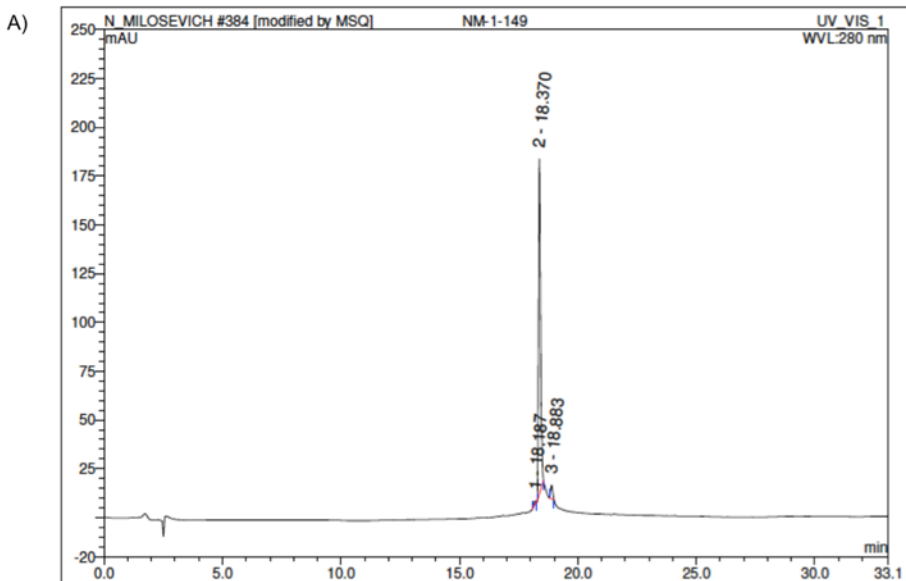| No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	16.537	0.150	1.856	0.98	1.17
2	17.234	0.057	0.640	0.37	0.40
3	17.947	0.093	0.910	0.61	0.57
4	18.440	15.038	155.370	98.05	97.85
Total:		15.337	158.777	100.00	100.00

B)

Figure S4. Characterization data for compound **2**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{63}H_{77}N_{10}O_{13}S$: 1213.54; found: 1213.60.

A)

SPD-20A
Ch1-254nm
Results

Retention Time	Area	Area %	Height	Height %
21.367	38415	0.44	3716	0.42
22.150	8642620	99.51	889691	99.54
32.242	1992	0.02	207	0.02
32.600	2156	0.02	209	0.02
Totals				
	8685183	100.00	893823	100.00

B)

20180228_NMilosev_NM-4-181_pure #7-34 RT: 0.21-1.06 AV: 28 NL: 2.62E6
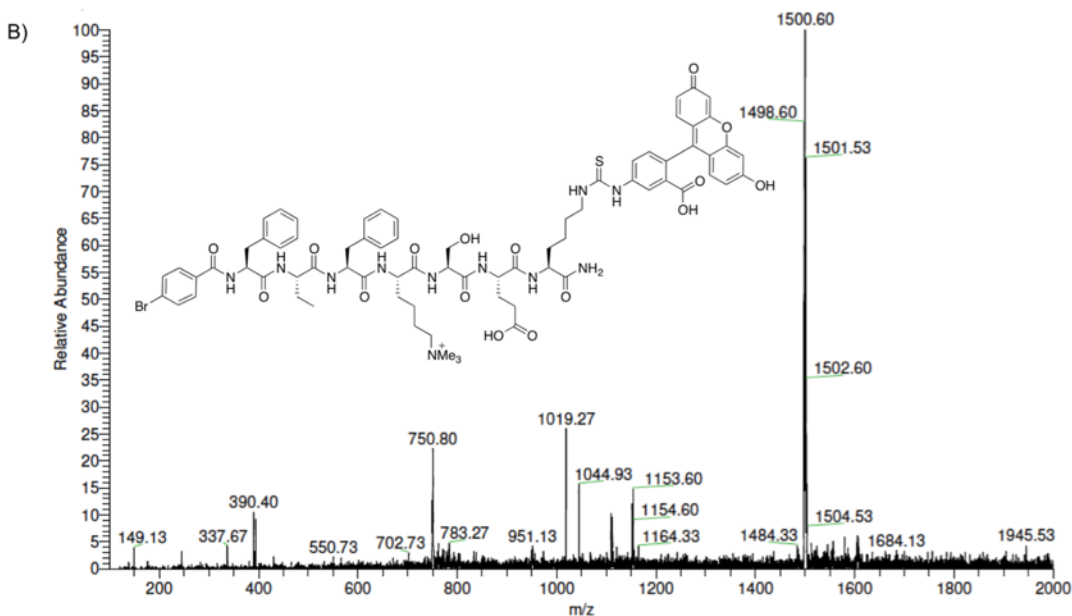T: + p ESI Full ms [100.00-2000.00]

Figure S5. Characterization data for compound **3**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{73}H_{89}BrN_{11}O_{15}S^{+}$: 1470.54; found: 1470.33.
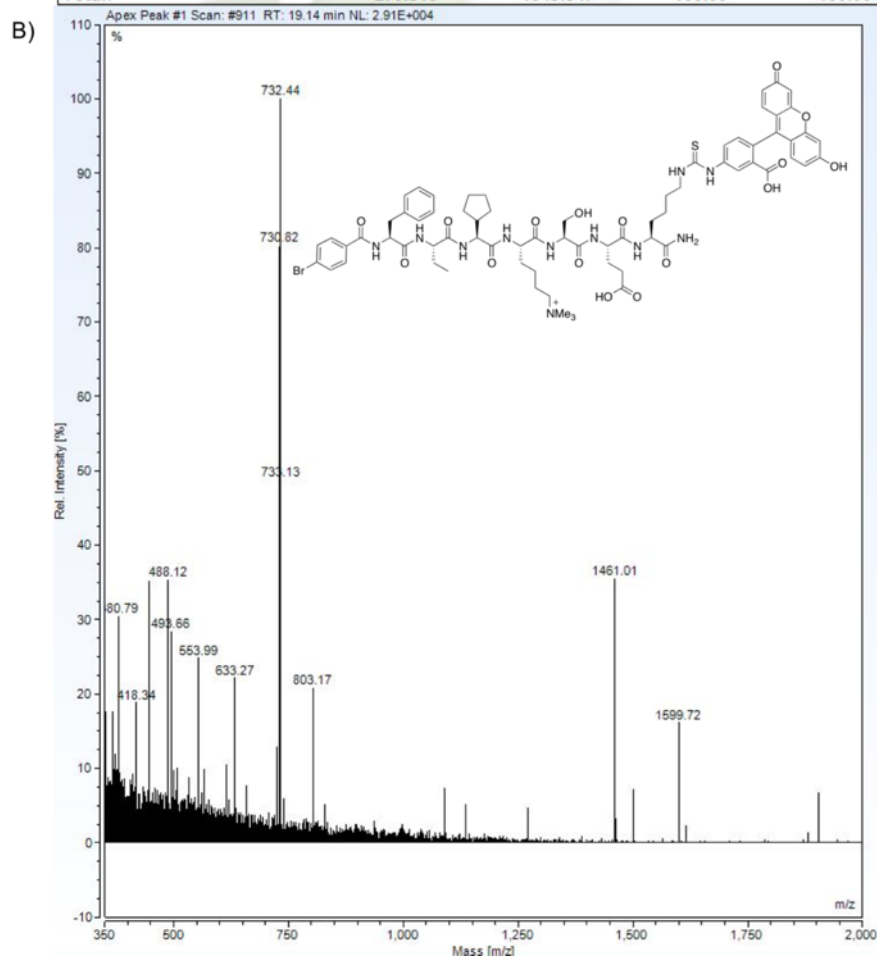
Figure S6. Characterization data for compound **4**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{72}H_{85}BrN_{11}O_{17}S^+$: 1486.50; found: 1486.40.
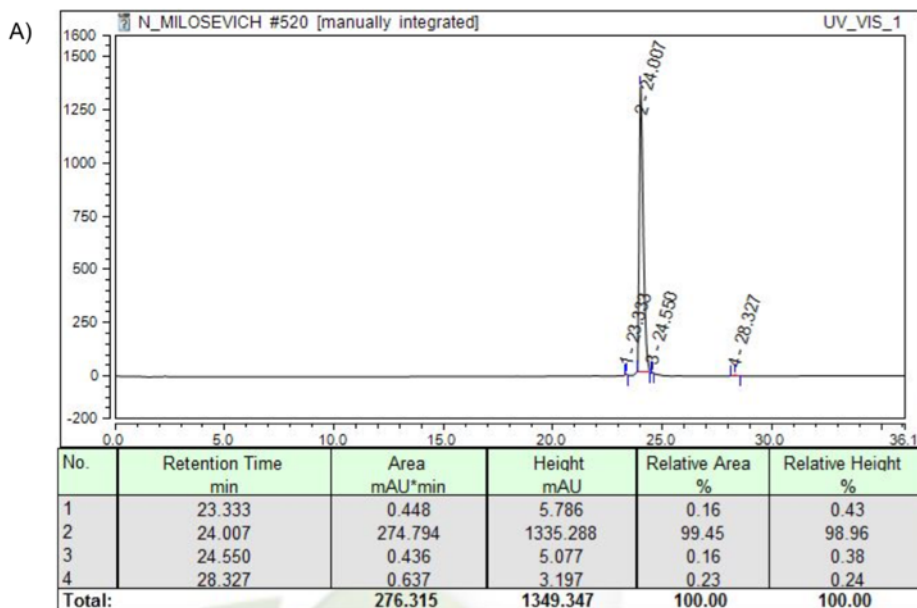
Figure S7. Characterization data for compound **5**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{40}H_{53}BrN_7O^+$: 838.31; found: 838.53.
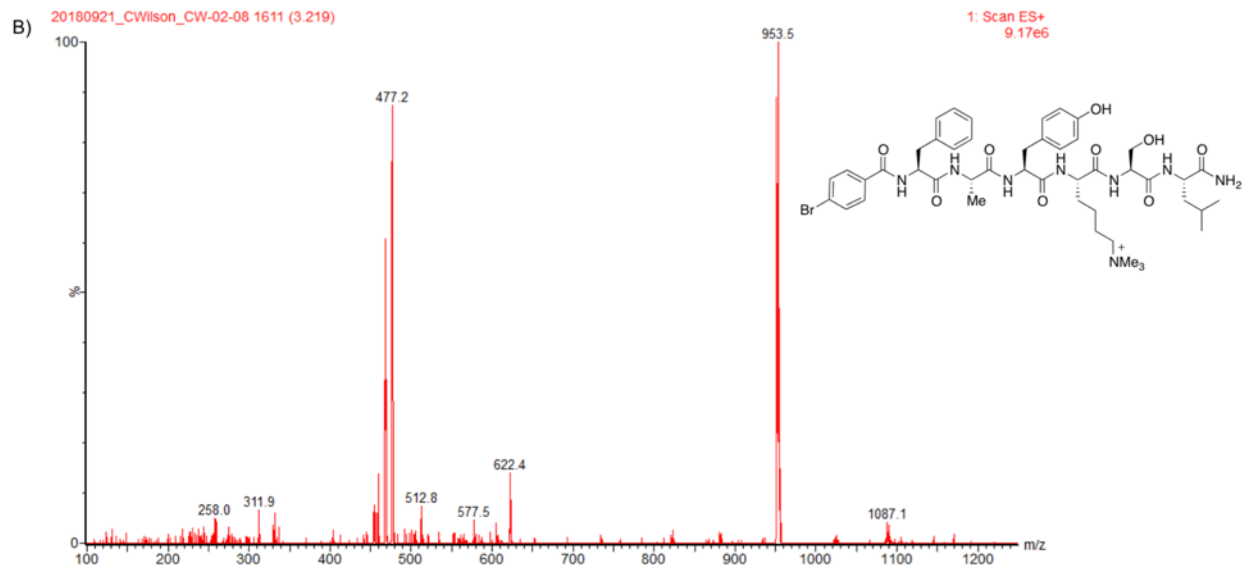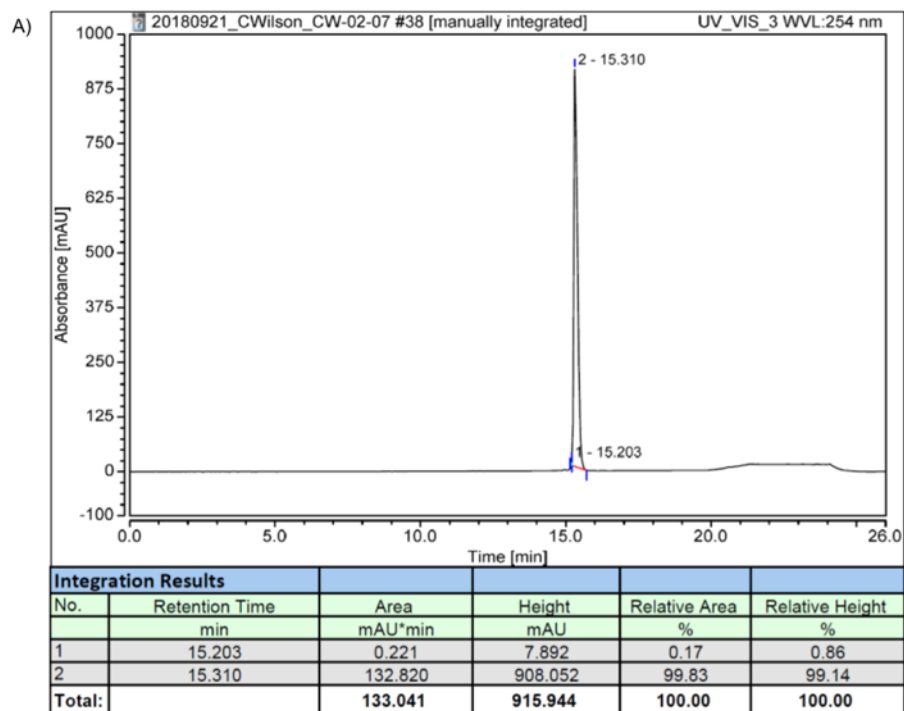
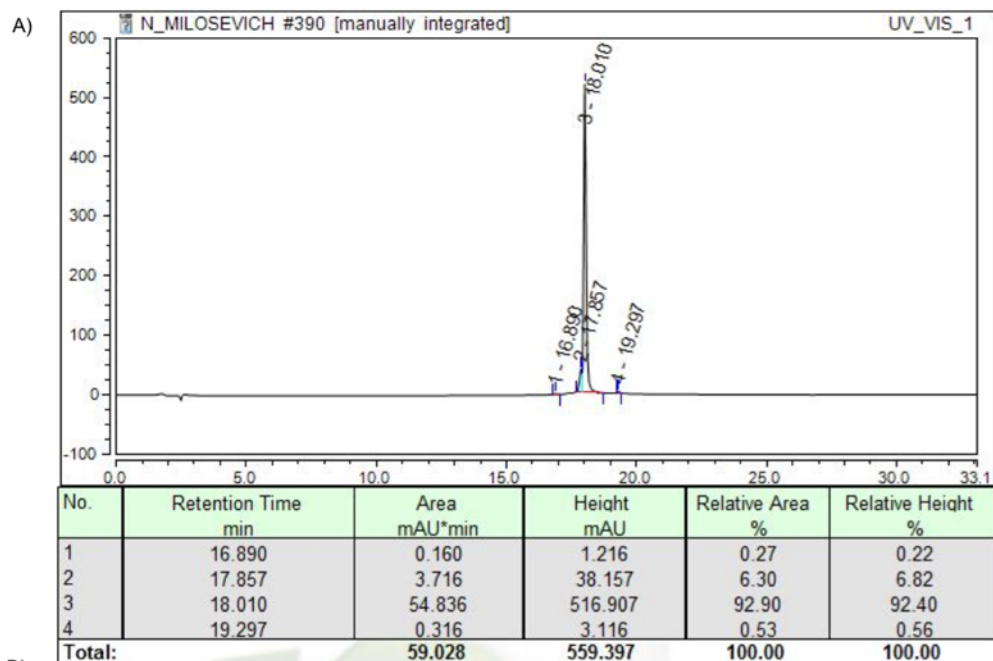**Figure S8.** Characterization data for compound **6**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{46}H_{64}BrN_8O_9^+$: 953.40; found: 953.50.
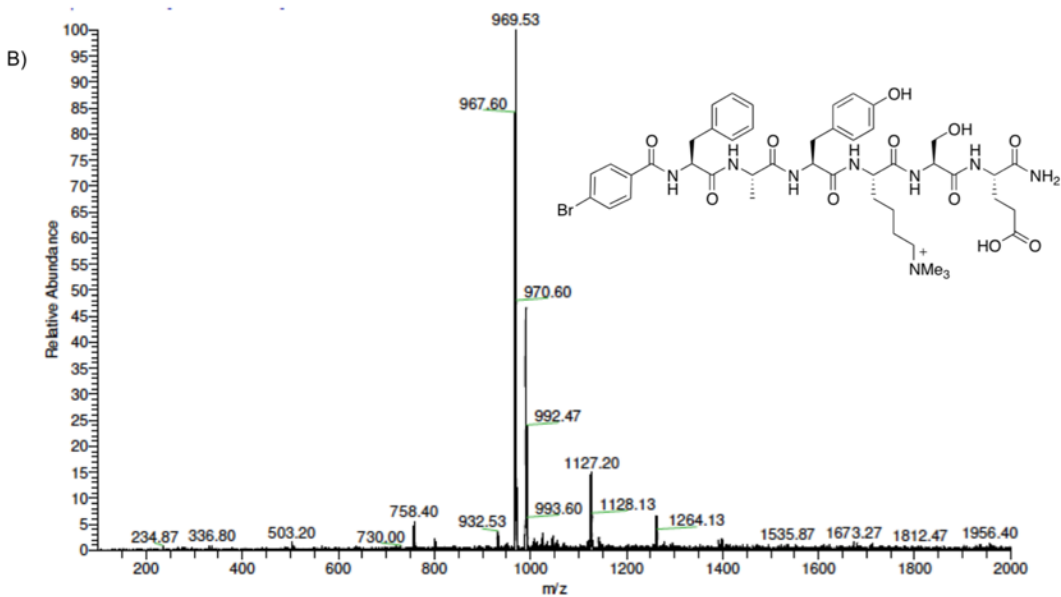
Figure S9. Characterization data for compound **7**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [m/z] calcd. for $C_{45}H_{60}BrN_8O_{11}^+$: 967.36; found: 967.60.
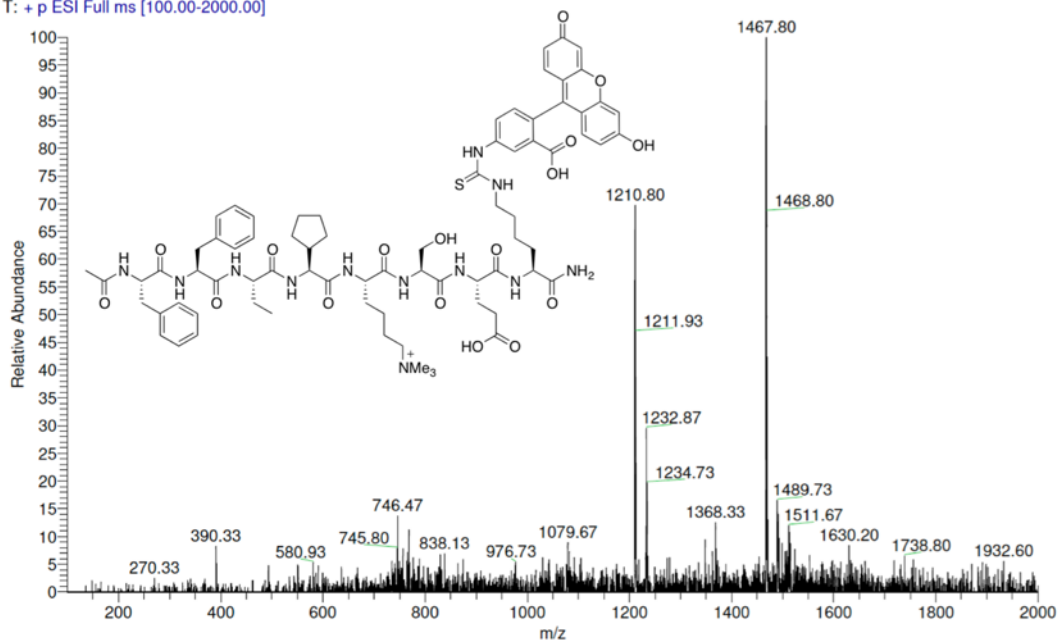
Figure S10. Characterization data for compound **8**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. For $C_{73}H_{87}BrN_{11}O_{17}S^+$: 1500.52; found: 1500.60.

Figure S11. Characterization data for compound **9**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [m/z] calcd. For $C_{71}H_{89}BrN_{11}O_{16}S^{+}$: 732.27, 1464.54; found: 732.44, 1461.01.

Figure S12. Characterization data for compound **10**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{75}H_{95}N_{12}O_{17}S^+$: 1469.68; found: 1468.80.
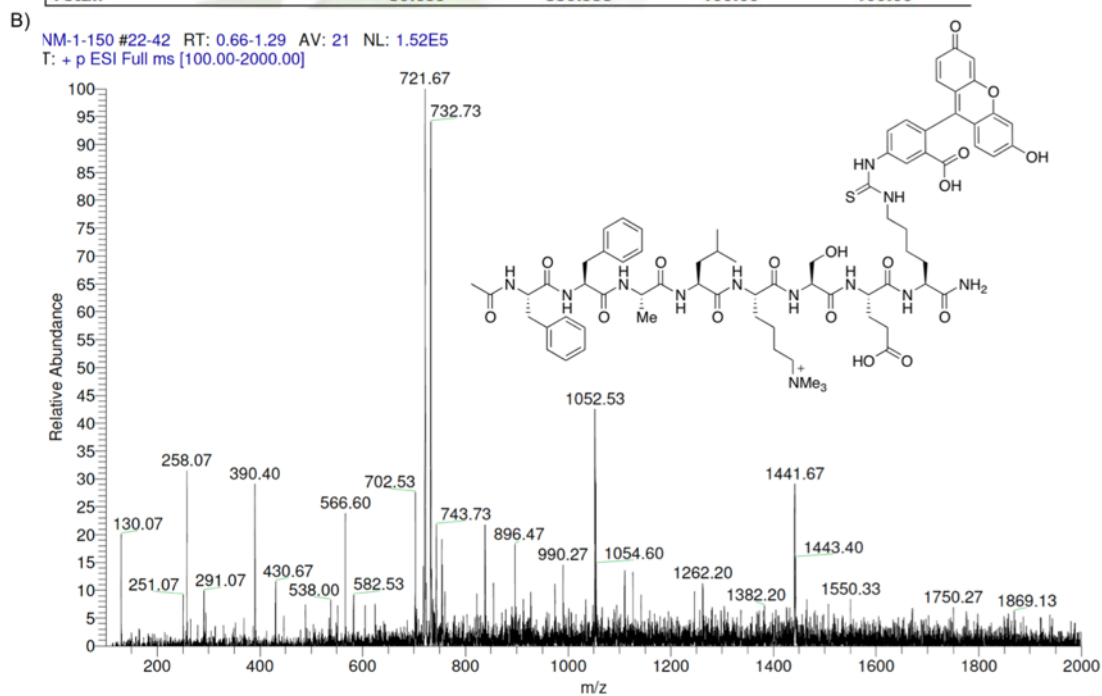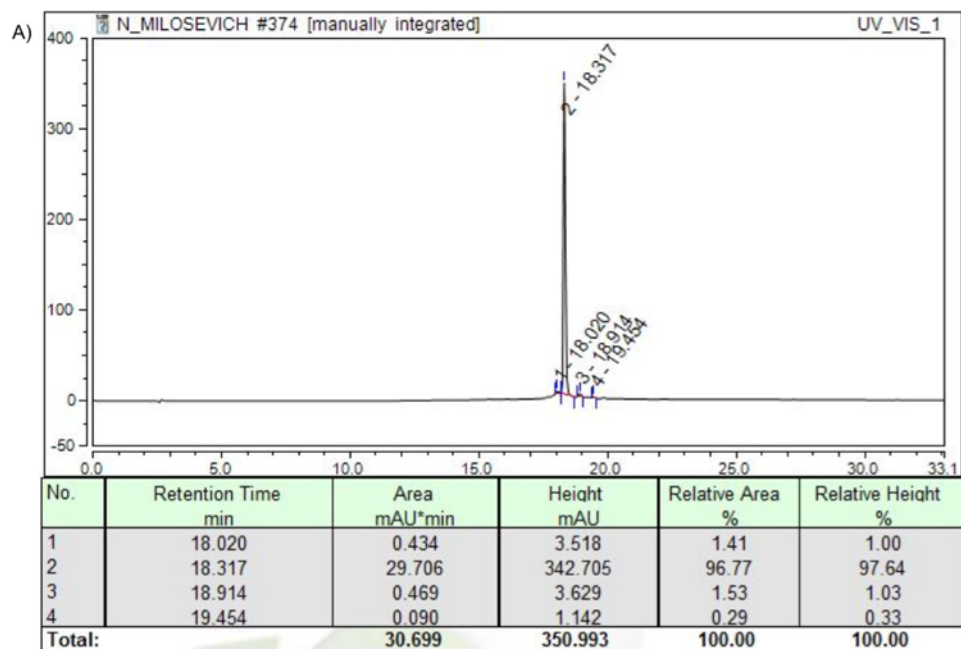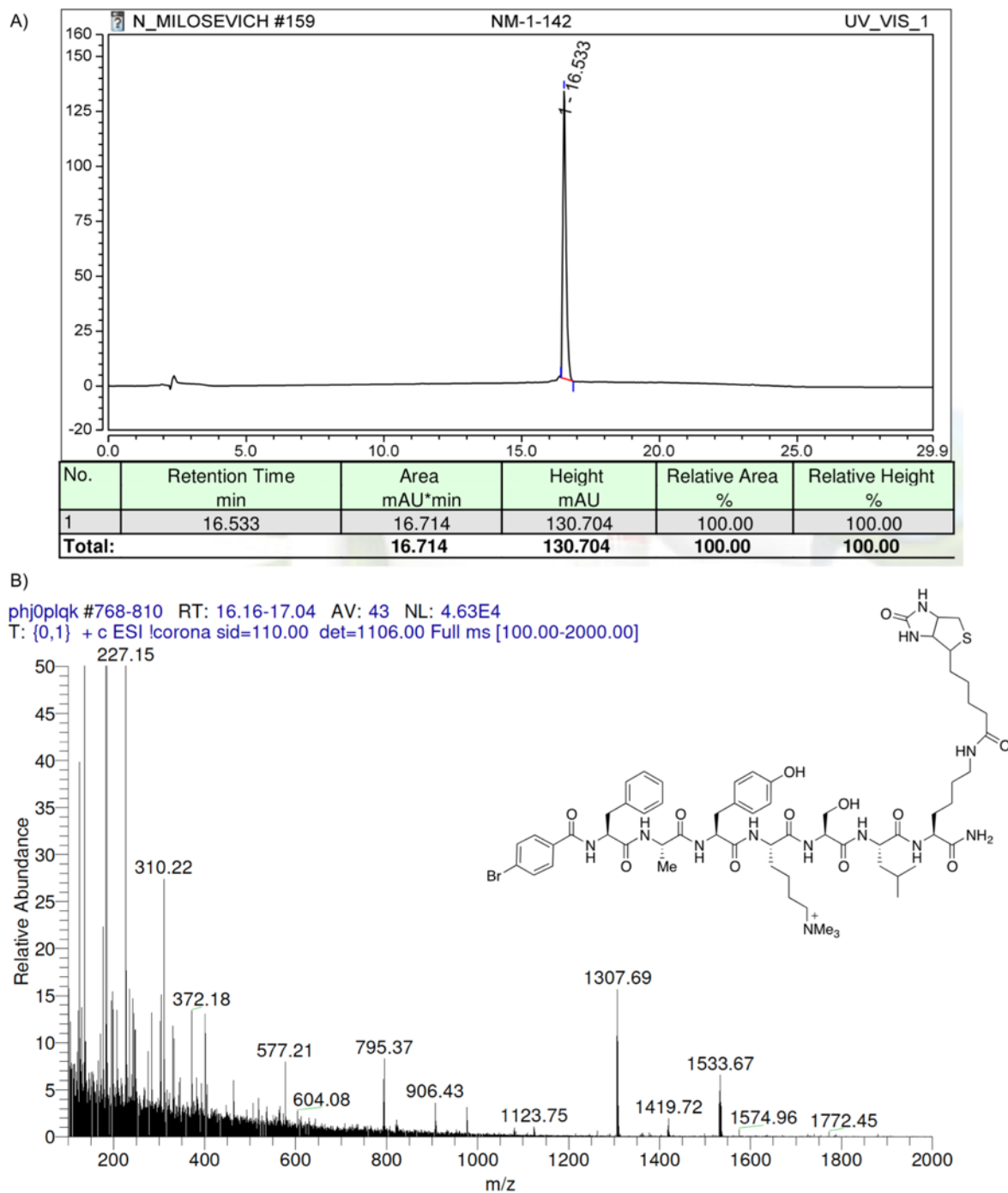
Figure S13. Characterization data for compound **11**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{73}H_{95}N_{12}O_{17}S^+$: 721.84, 1443.67; found: 721.67, 1443.40.

Figure S14. Characterization data for compound **12**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{62}H_{90}BrN_{12}O_{12}S^{+}$: 1307.57; found:1307.69.

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$_2$O 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. 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.

Fluorescence Polarization Methods

Direct FP Method

Direct FP was done by titration of CBX proteins into FITC-labelled probes. 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 (PMSF, DTT and benzamidine added to minimal FP buffer to a 1 mM final concentration). 100 μL of the CBX-containing solution was added to well 19. Wells 2-18 contained 50 μL of the peptide 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 peptide 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).

Analysis of Direct FP data

The background of the blank buffer was subtracted from the parallel and perpendicular intensities of emission. Average values for parallel and perpendicular emission intensities (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 ((Parallel – Perpendicular)/(Parallel + Perpendicular))*1000. Data were then plotted using Graphpad Prism 8. The mathematical model for the fit was described by the equation $y = (BLmax*x)/(K_d+x) + NS*x + background$. 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.

Competitive FP Method

Competitive FP analysis of **5-7** binding to CBX6/7/8 was used to determine IC_{50} values as with compound **4** as the competitive binding probe (Figures S20-22). Compound **4** and peptide inhibitors were dissolved in distilled H_2O. The assay was carried out in NUNC black 96-well plates and full FP buffer was used as described above (20 mM Tris-HCl, pH = 8.0, 250 mM NaCl, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF, 0.01% Tween). Probe **4** was at a constant concentration of 100 nM. Each CBX protein was at a constant concentration dependent on the K_d of the protein to probe **4.** CBX1 was used at 10 μM, CBX2 at 0.7 μM, CBX4 at 0.7 μM, CBX6 at 1 μM, CBX7 at 0.7 μM and CBX8 at 4 μM. Inhibitor concentrations varied between 0-500 μM. All wells in the plate were made to final volume of 100 μL. 200 μL of the inhibitor-containing solution was added to well 12. Wells 2-11 contained 100 μL of the CBX and **4** only solution. 100 μL of the well 12 solution was added to well 11 and mixed. These dilutions continued to well 3, whereupon 100 μL was discarded after mixing, leaving the total volume of all wells at 100 μL. Well 1 contained only full FP buffer and well 2 contained only CBX protein and **4** in full FP buffer. Following serial dilution of the inhibitor across the plate, plates were incubated for 15 minutes in the dark and then were read with the SpectraMax M5 plate reader as described above.

Analysis of Competitive FP data

The background of the blank buffer was subtracted from the parallel and perpendicular intensities of emission. Values were graphed using Graphpad Prism 8 and fitted using a sigmoidal curve function from which IC_{50} values were extrapolated. The equation used is described as follows: $y = bottom + (top–bottom) / (1 + 10^{(X–LogIC_{50})})$. Errors are reported as 95% confidence intervals. Experiments were done in duplicate or triplicate and this is indicated in the raw data shown in Figures S16-S26.
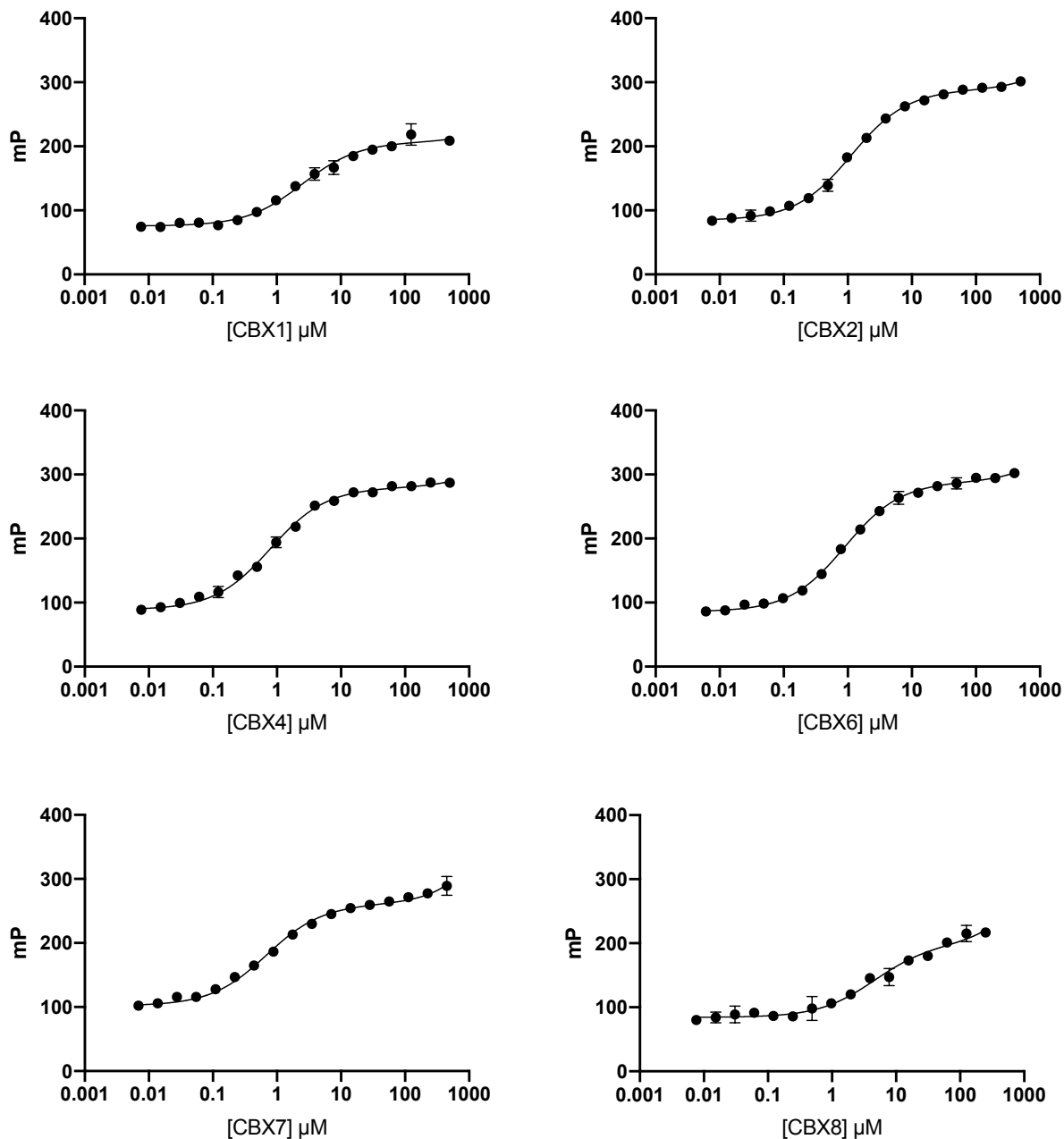
Fluorescence Polarization Data

Figure S15. Direct fluorescence polarization data of compound **1** with CBX1/2/4/6/7/8. CBX1 (K_d = 2.5 µM, 95% CI 2.0 µM to 3.1 µM, R^2 = 0.983), CBX2 (K_d = 1.14 µM, 95% CI 1.0 µM to 1.2 µM, R^2 = 0.997), CBX4 (K_d = 0.79 µM, 95% CI 0.70 µM to 0.90 µM, R^2 = 0.994), CBX6 (K_d = 0.90 µM, 95% CI 0.82 µM to 0.96 µM, R^2 = 0.998), CBX7 (K_d = 0.69 µM, 95% CI 0.57 µM to 0.83 µM, R^2 = 0.993), CBX8 (K_d = 4.5 µM, 95% CI 3.0 µM to 6.8 µM, R^2 = 0.967). Error reported as asymmetrical 95% confidence intervals. Titrations for CBX1/2/4/8 performed in triplicate and CBX6/7 performed in duplicate.
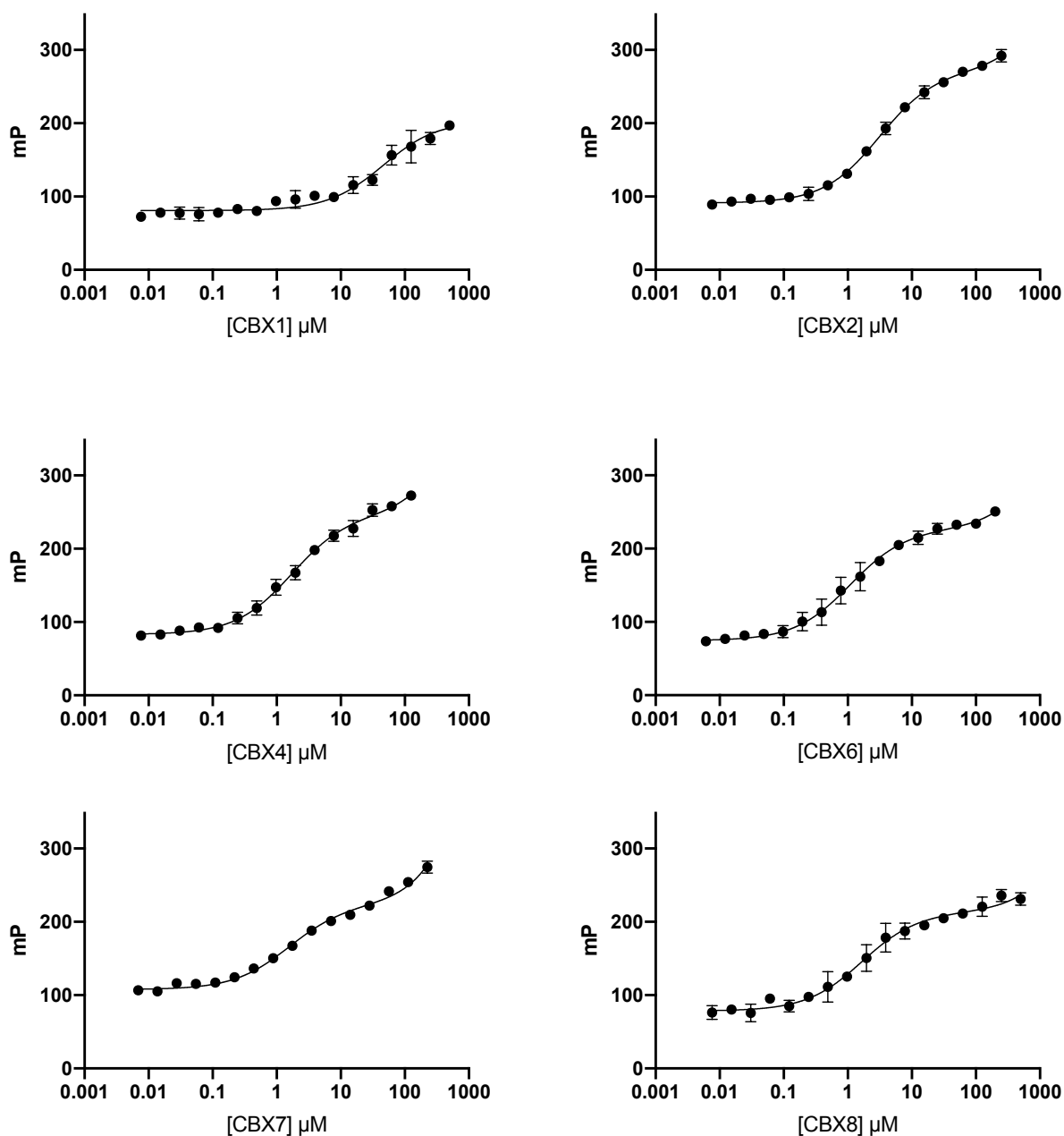
Figure S16. Direct fluorescence polarization data of compound **2** with CBX1/2/4/6/7/8. CBX1 (K_d = 47 µM, 95% CI 36 µM to 61 µM, R^2 = 0.944), CBX2 (K_d = 3.1 µM, 95% CI 2.8 µM to 3.5 µM, R^2 = 0.995), CBX4 (K_d = 1.7 µM, 95% CI 1.4 µM to 2.0 µM, R^2 = 0.991), CBX6 (K_d = 1.1 µM, 95% CI 0.83 µM to 1.4 µM, R^2 = 0.987), CBX7 (K_d = 1.6 µM, 95% CI 1.2 µM to 2.0 µM, R^2 = 0.992), CBX8 (K_d = 1.7 µM, 95% CI 1.2 µM to 2.3 µM, R^2 = 0.977). Error reported as asymmetrical 95% confidence intervals. CBX1/2/4/7 performed in triplicate, and CBX6/8 titrations performed in duplicate.
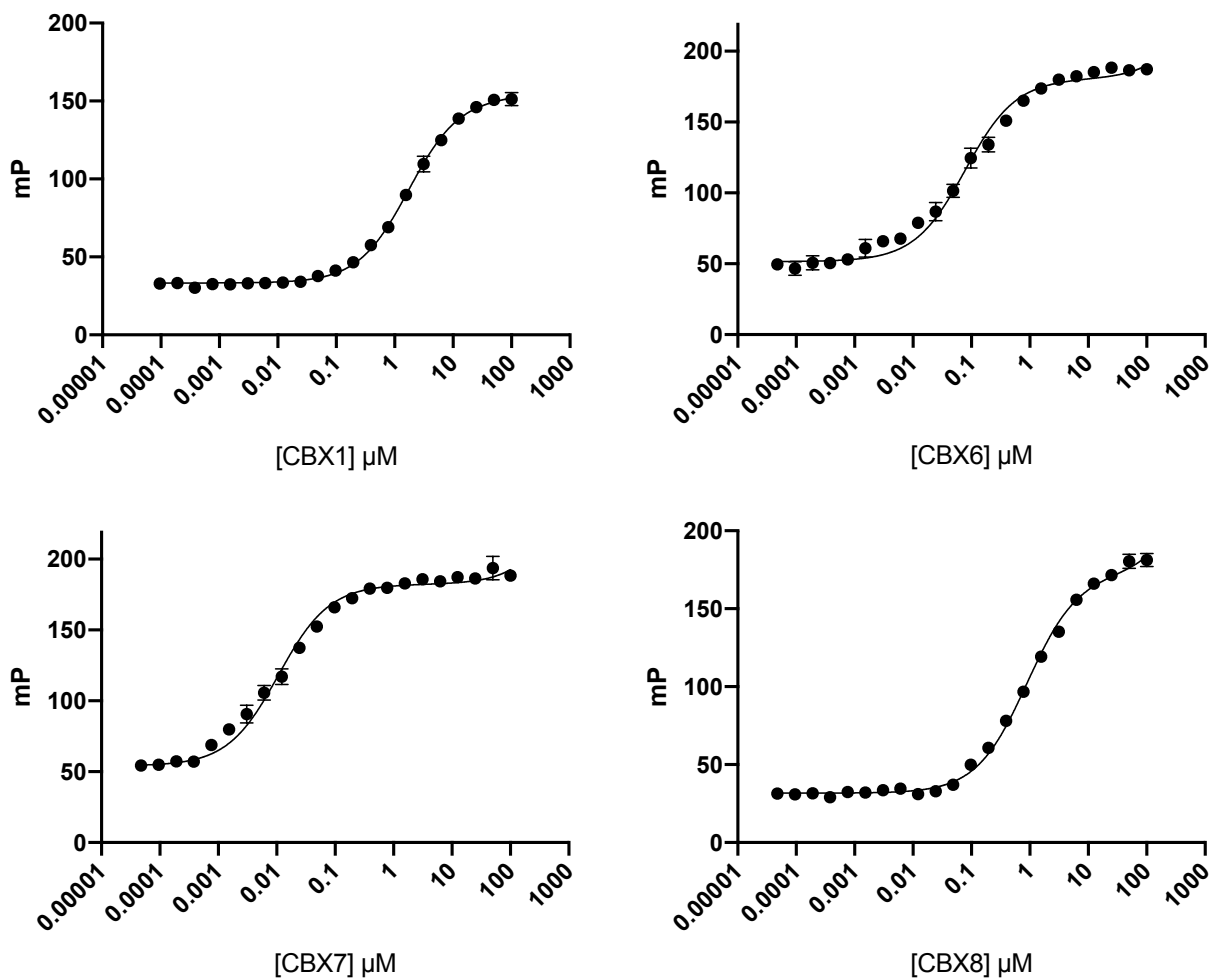
Figure S17. Direct fluorescence polarization data of compound **3** with CBX1/6/7/8 performed in triplicate. CBX1 (K_d = 1.75 µM, 95% CI 1.62 µM to 1.89 µM, R^2 = 0.998), CBX6 (K_d = 0.078 µM, 95% CI 0.063 µM to 0.097 µM, R^2 = 0.984), CBX7 (K_d = 0.011 µM, 95% CI 0.0090 µM to 0.013 µM, R^2 = 0.987), CBX8 (K_d = 0.89 µM, 95% CI 0.81 µM to 0.98 µM, R^2 = 0.997). Error reported as asymmetrical 95% confidence intervals.

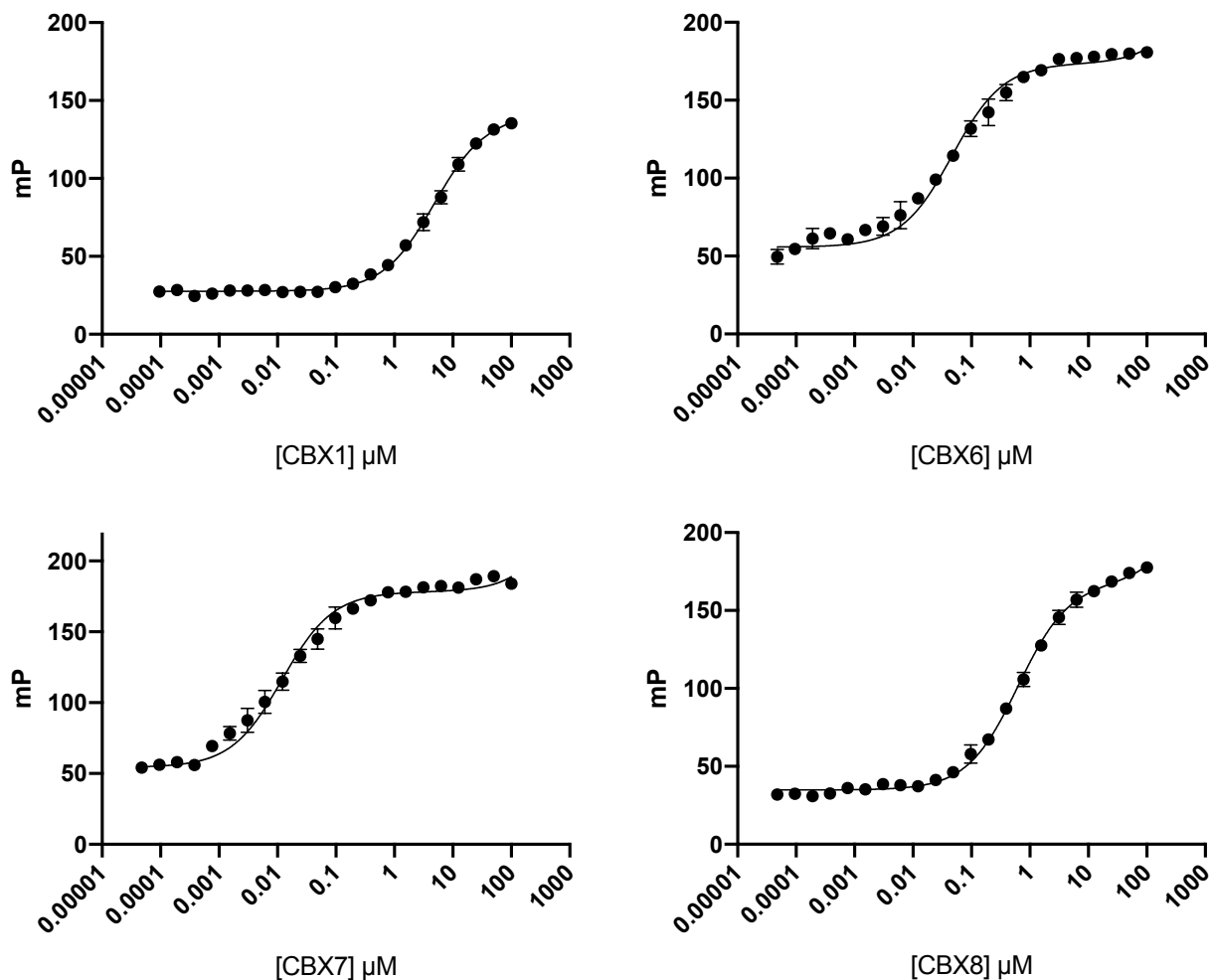**Figure S18**. Direct fluorescence polarization data of compound **4** with CBX1/6/7/8 performed in triplicate. CBX1 (K_d = 4.8 μM, 95% CI 4.1 μM to 5.5 μM, R^2 = 0.944), CBX6 (K_d = 0.047 μM, 95% CI 0.035 μM to 0.062 μM, R^2 = 0.972), CBX7 (K_d = 0.012 μM, 95% CI 0.0097 μM to 0.015 μM, R^2 = 0.982), CBX8 (K_d = 0.62 μM, 95% CI 0.57 μM to 0.68 μM, R^2 = 0.998). Error reported as asymmetrical 95% confidence intervals.
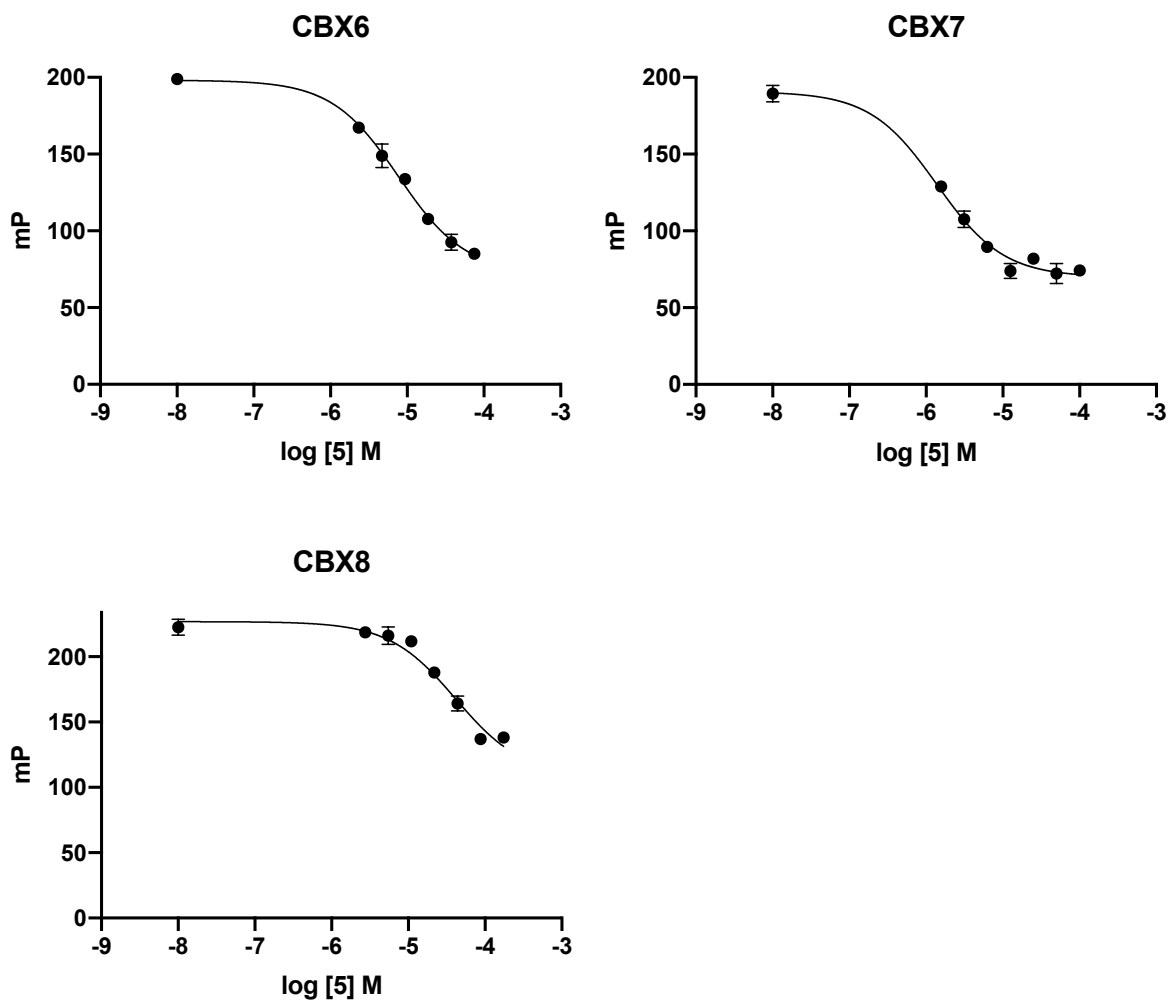
Figure S19. Competitive fluorescence polarization data of compound **5** with CBX6/7/8 performed in duplicate. CBX6 (logIC$_{50}$ = −5.113 M, 95% CI −5.222 to −5.000, R^2 = 0.990), CBX7 (logIC$_{50}$ = −5.862 M, 95% CI −5.983 to −5.746, R^2 = 0.983), CBX8 (logIC$_{50}$ = −4.385 M, 95% CI −4.593 to −4.162, R^2 = 0.967). Error reported as asymmetrical 95% confidence intervals.
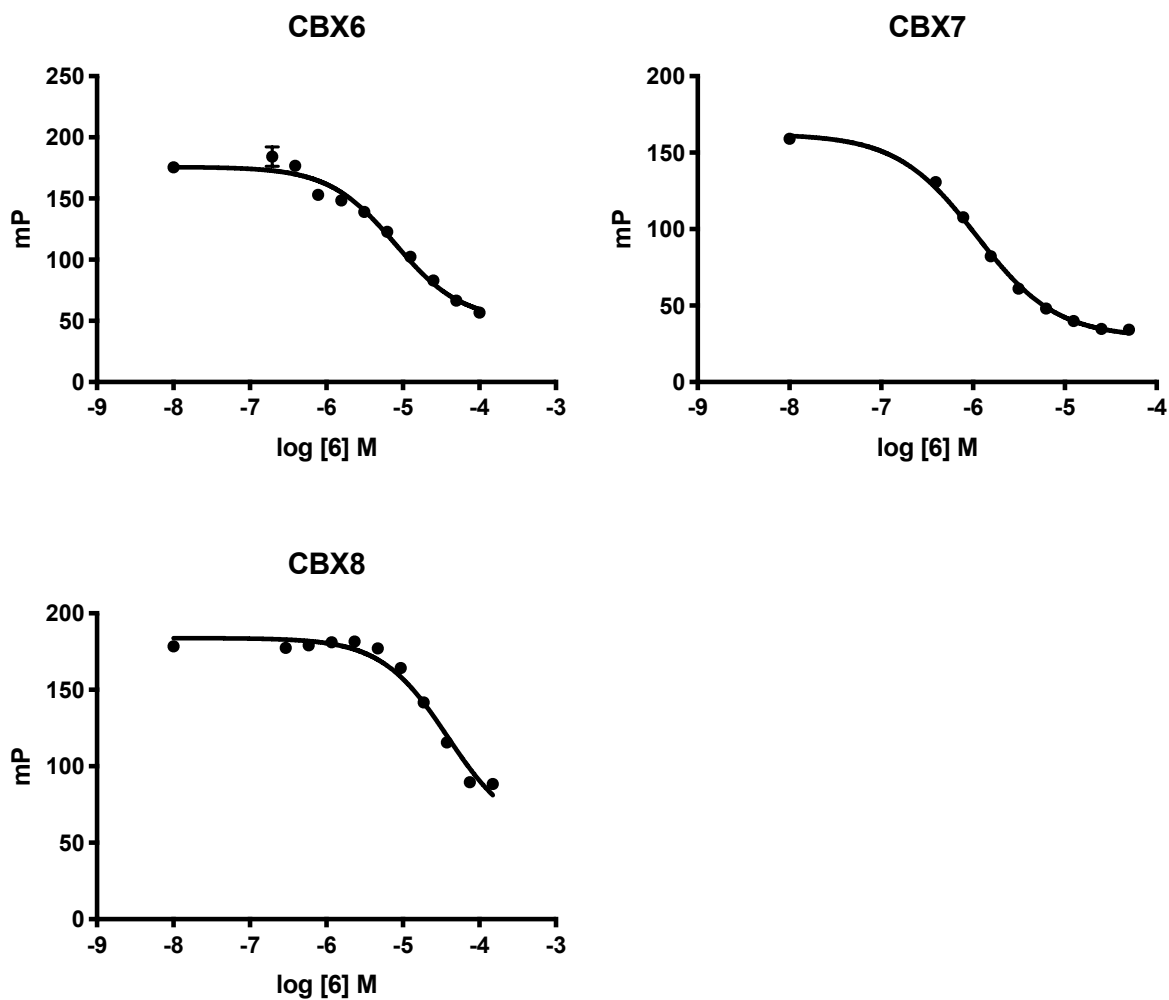
Figure S20. Competitive fluorescence polarization data of compound **6** with CBX6/7/8 performed in triplicate. CBX6 ($logIC_{50} = -5.102$ M, 95% CI -5.219 to -4.985, $R^2 = 0.980$), CBX7 ($logIC_{50} = -5.962$ M, 95% CI -5.996 to -5.982, $R^2 = 0.997$), CBX8 ($logIC_{50} = -4.419$ M, 95% CI -4.533 to -4.300, $R^2 = 0.975$). Error reported as asymmetrical 95% confidence intervals.
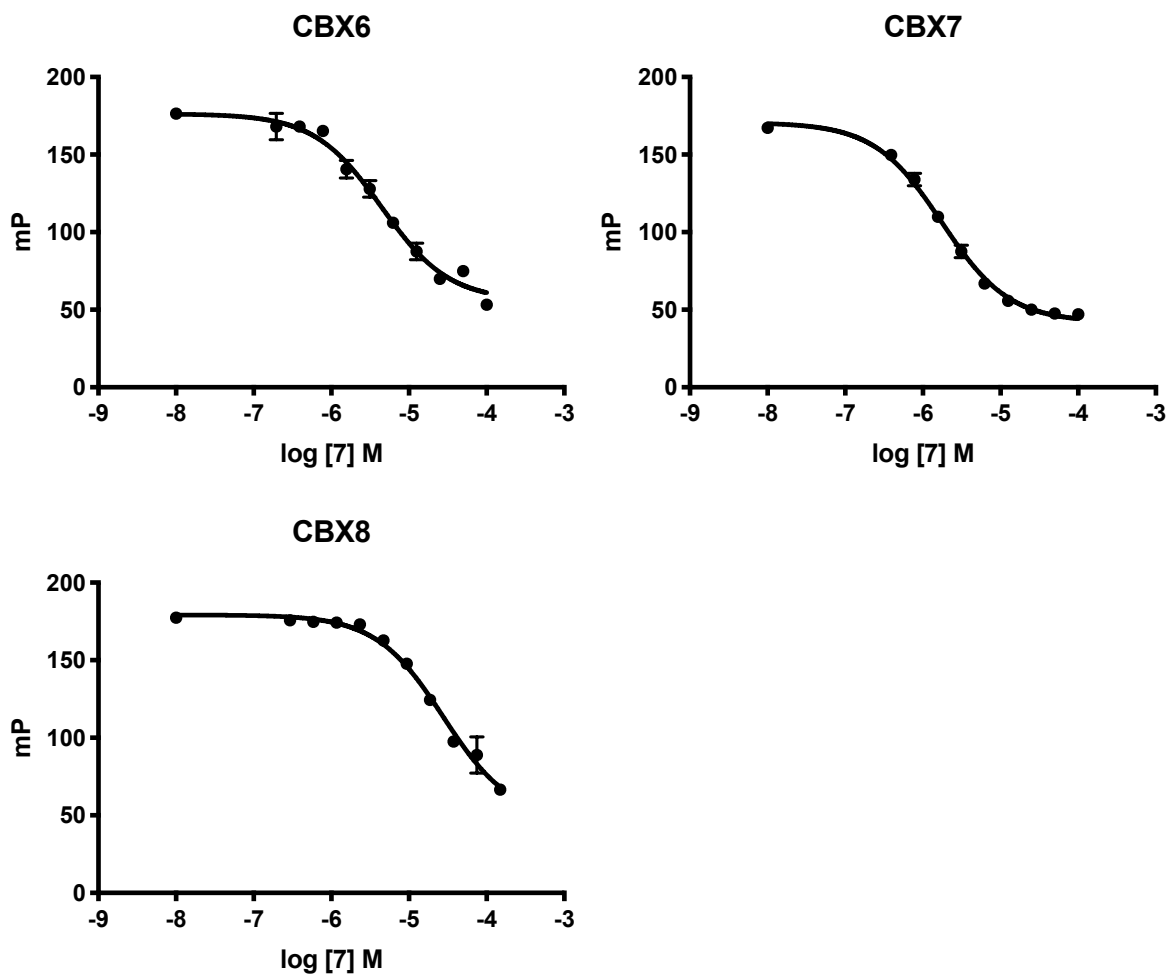
Figure S21. Competitive fluorescence polarization data of compound **7** with CBX6/7/8 performed in triplicate. CBX6 ($\log IC_{50} = -5.338$ M, 95% CI -5.447 to -5.227, $R^2 = 0.979$), CBX7 ($\log IC_{50} = -5.756$ M, 95% CI -5.800 to -5.711, $R^2 = 0.995$), CBX8 ($\log IC_{50} = -4.559$ M, 95% CI -4.648 to -4.467, $R^2 = 0.986$). Error reported as asymmetrical 95% confidence intervals.
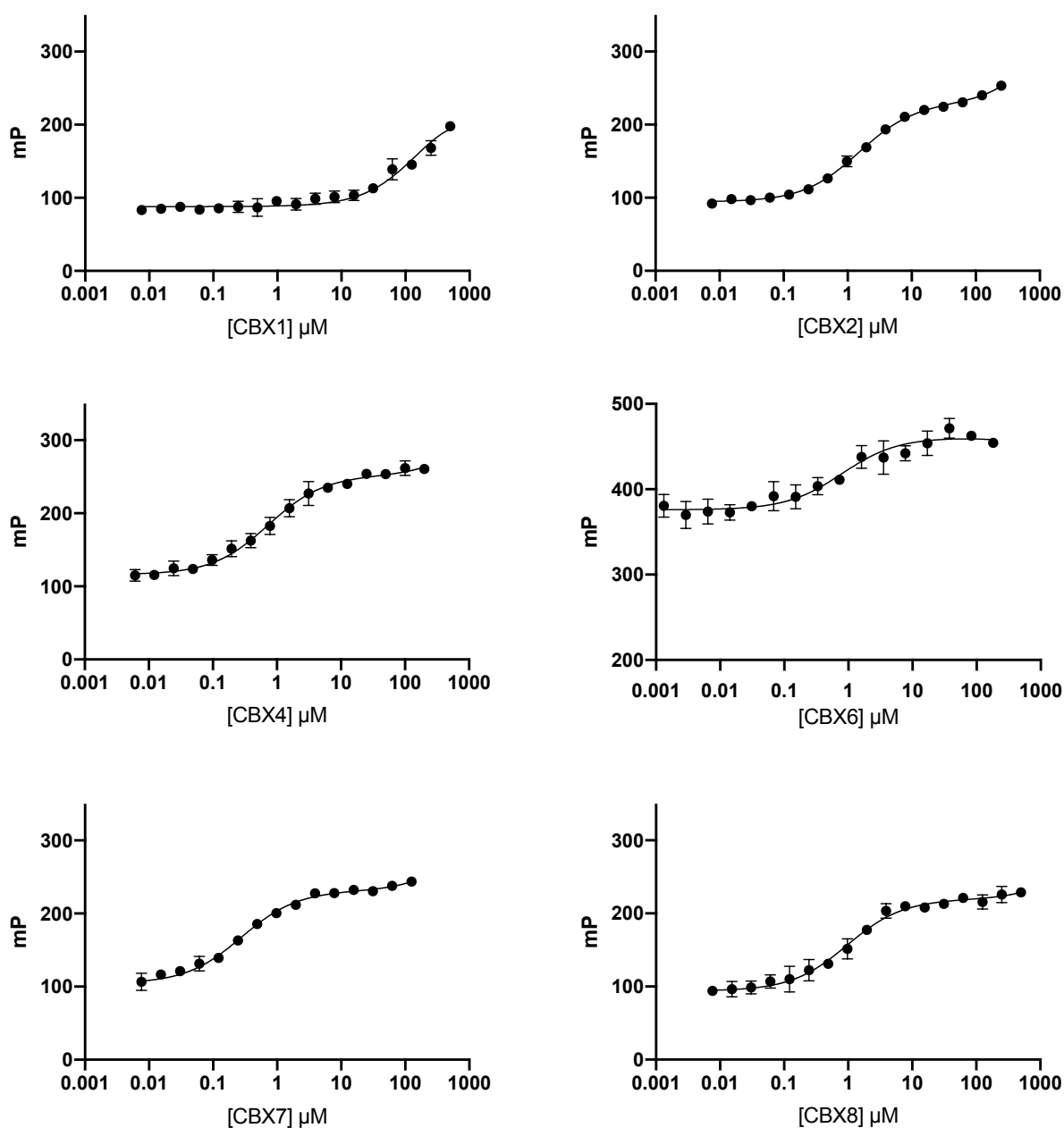
Figure S22. Direct fluorescence polarization data of compound **8** with CBX1/2/4/6/7/8. CBX1 (K_d = 178 μM, 95% CI 143 μM to 225 μM, R^2 = 0.941), CBX2 (K_d = 1.5 μM, 95% CI 1.3 μM to 1.7 μM, R^2 = 0.996), CBX4 (K_d = 0.73 μM, 95% CI 0.57 μM to 0.93 μM, R^2 = 0.981), CBX6 (K_d = 0.80 μM, 95% CI 0.4 μM to 1.6 μM, R^2 = 0.896), CBX7 (K_d = 0.30 μM, 95% CI 0.23 μM to 0.33 μM, R^2 = 0.988), CBX8 (K_d = 0.96 μM, 95% CI 0.73 μM to 1.30 μM, R^2 = 0.980). Error reported as asymmetrical 95% confidence intervals. CBX2/7/8 titrations performed in triplicate. CBX1/2/4/6/7 titrations performed in triplicate and CBX8 titration performed in duplicate.
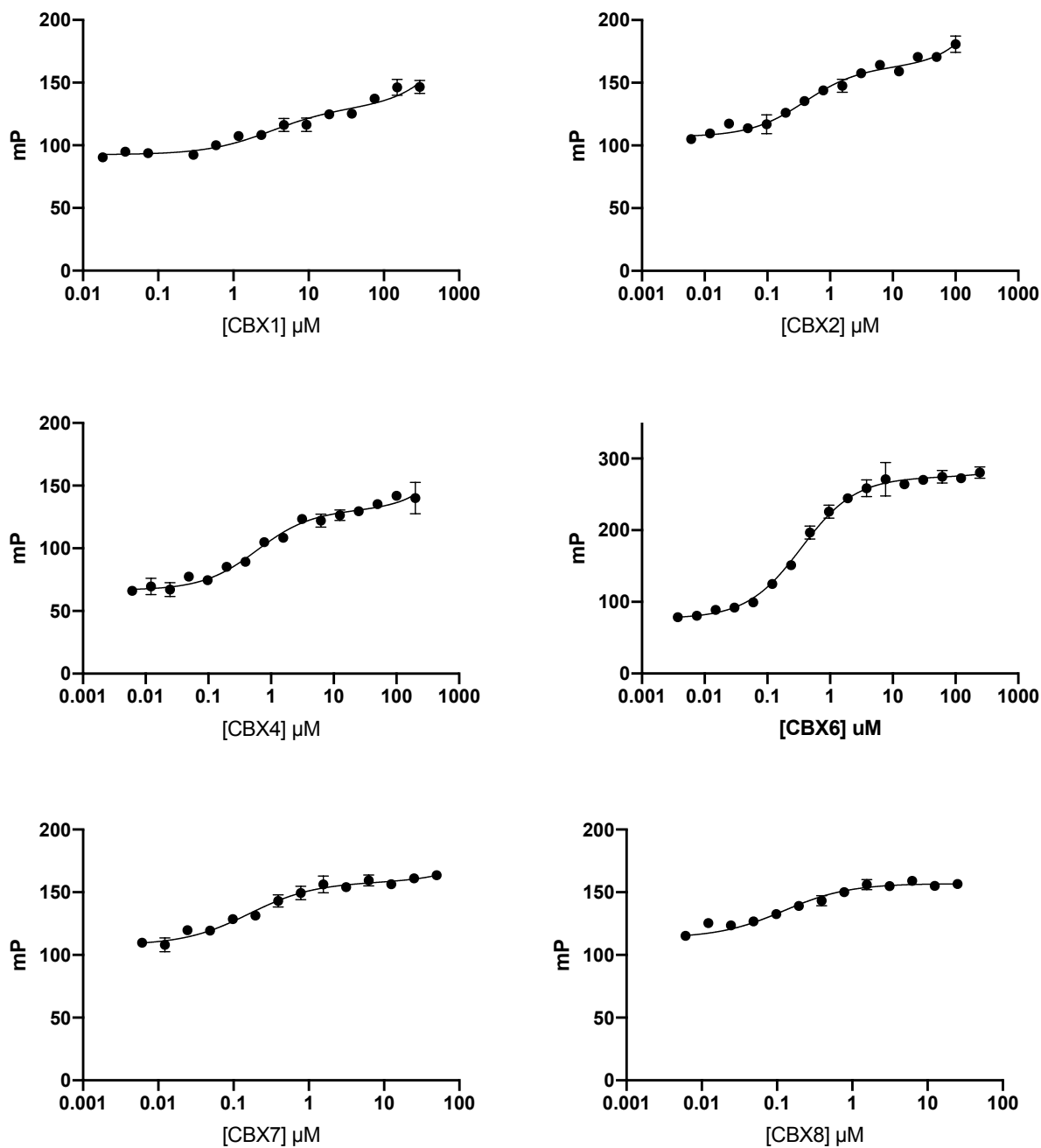
Figure S23. Direct fluorescence polarization data of compound **9** with CBX1/2/4/6/7/8. CBX1 (K_d = 3.2 μM, 95% CI 1.5 μM to 7.0 μM, R^2 = 0.956), CBX2 (K_d = 0.39 μM, 95% CI 0.27 μM to 0.56 μM, R^2 = 0.971), CBX4 (K_d = 0.60 μM, 95% CI 0.41 μM to 0.87 μM, R^2 = 0.974), CBX6 (K_d = 0.35 μM, 95% CI 0.299 μM to 0.401 μM, R^2 = 0.993), CBX7 (K_d = 0.16 μM, 95% CI 0.11 μM to 0.23 μM, R^2 = 0.965), CBX8 (K_d = 0.12 μM, 95% CI 0.08 μM to 0.19 μM, R^2 = 0.960). Error reported as asymmetrical 95% confidence intervals. CBX2/7/8 titrations performed in triplicate. CBX1/4/6 titrations performed in duplicate.

Figure S24. Direct fluorescence polarization data of compound **10** with CBX1/2/4/6/7/8. CBX1 (K_d = 99 μM, 95% CI 70 μM to 142 μM, R^2 = 0.989), CBX2 (K_d = 10 μM, 95% CI 8.6 μM to 12 μM, R^2 = 0.984), CBX4 (K_d = 11.3 μM, 95% CI 8.2 μM to 15.7 μM, R^2 = 0.985), CBX6 (K_d = 0.35 μM, 95% CI 0.30 μM to 0.41 μM, R^2 = 0.993), CBX7 (K_d = 1.5 μM, 95% CI 1.21 μM to 1.78 μM, R^2 = 0.989), CBX8 (K_d = 0.73 μM, 95% CI 0.63 μM to 0.84 μM, R^2 = 0.991). Error reported as asymmetrical 95% confidence intervals. Titrations performed in duplicate.
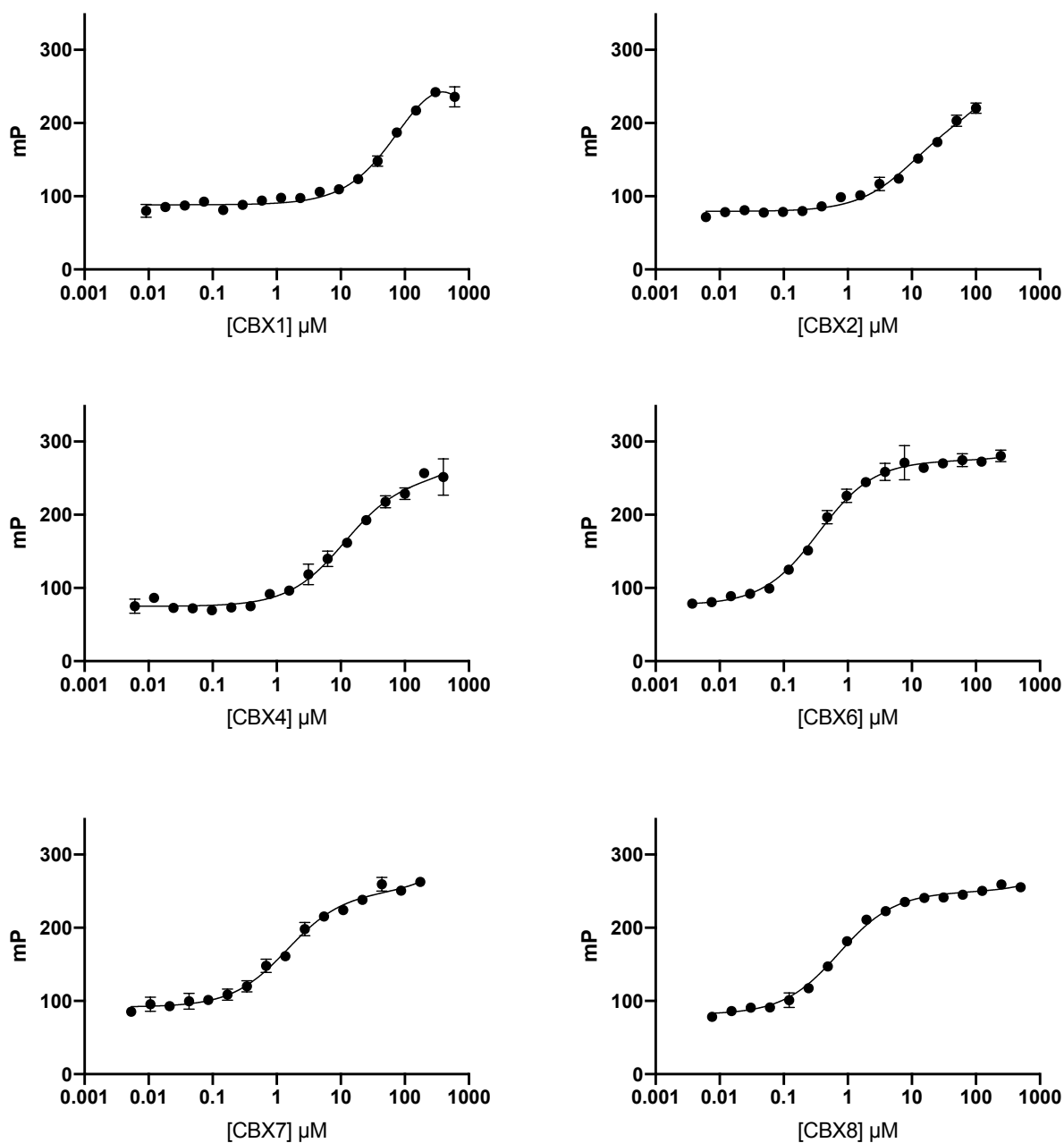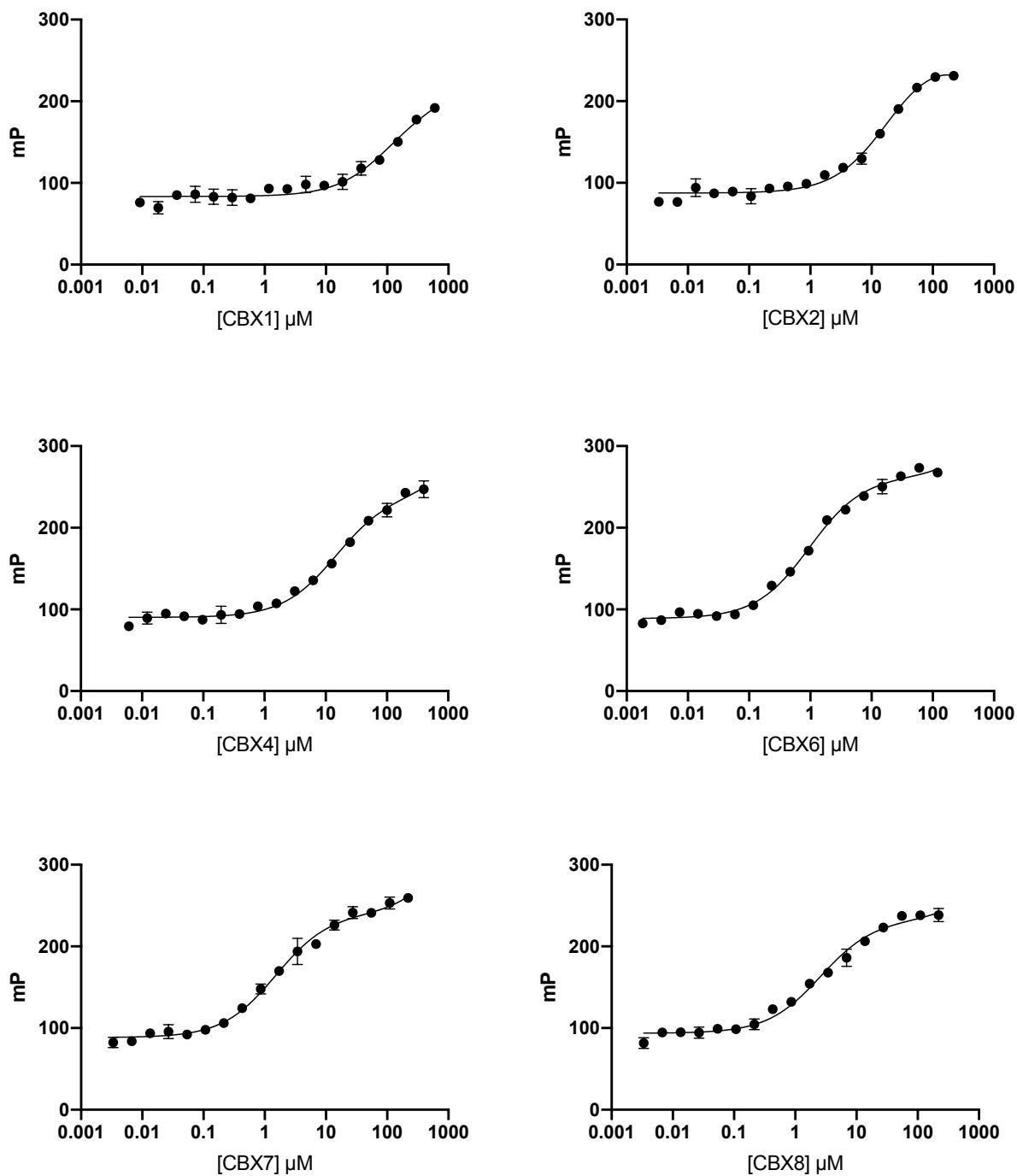
Figure S25. Direct fluorescence polarization data of compound **11** with CBX1/2/4/6/7/8. CBX1 (K_d = 110 μM, 95% CI 84 μM to 148 μM, R^2 = 0.958), CBX2 (K_d = 18 μM, 95% CI 13 μM to 25 μM, R^2 = 0.998), CBX4 (K_d = 15 μM, 95% CI 11 μM to 19 μM, R^2 = 0.991), CBX6 (K_d = 0.96 μM, 95% CI 0.82 μM to 1.3 μM, R^2 = 0.995), CBX7 (K_d = 1.5 μM, 95% CI 1.2 μM to 1.9 μM, R^2 = 0.992), CBX8 (K_d = 2.6 μM, 95% CI 1.9 μM to 3.5 μM, R^2 = 0.987). Error reported as asymmetrical 95% confidence intervals. Titrations performed in duplicate.

Docking and MD Methods

File Preparation

Residues unavailable in the Amber16 ff14SB force field were constructed using the Avogadro molecular visualization tool to include C- and N-terminal methyl caps. These residues include the trimethyllysine and brominated phenyl N-terminal residue. The capped structures were minimized using Gaussian09 at the HF-6-31G* level of theory and the following outputs were used to generate an electrostatic potential map for later fitting.

AmberTools16 residuegen utility was then used to fit the electrostatic potential map and assign force field parameters to the uncapped versions of the residues. The generated prepi file and antechamber generated frcmod file were loaded into tleap along with the Amber ff14SB force field parameters. Connectivity information for both residues was assigned and compounds 6 and 7 were generated using the sequence function to generate PDB files for both compounds.

PDBQT files were generated by first loading the previous PDB files into UCSF Chimera. Applying the built-in AutoDock Vina tool to an arbitrary system generated the ligand PDBQT files used later by our in-house software designed for the exhaustive use of use of AutoDock Vina.

SLICE Method Application

Initial Molecular Docking

Our in-house software combines ligand PDBQTs with an ensemble of MD pre-generated CBX8 host configurations. Each system is assigned a docking box generated by a residue list and a buffer of 5 angstroms outside of the assumed docked location. Using exhaustiveness 7 on 50 host configurations, AutoDock Vina was executed and 10 poses were generated for each configuration, resulting in 500 poses of which the top ten were selected for molecular dynamics.

Table S2. Progression of Top-Docked Scores for Compounds 6 and 7

	Crystal Structure Docking	1st MD Ensemble Docking	2nd MD Ensemble Docking
	/kcal mol^{-1}	/kcal mol^{-1}	/kcal mol^{-1}
Compound 6	–8.0	–8.2	–8.6 *
Compound 7	–8.4	–8.7	–9.2**

* Incorrectly bound

** Pose submitted for 100 ns molecular dynamics

Intermediate and 100 ns Molecular Dynamics

MD simulation starting coordinates and topology files were constructed using tleap. Complexes were initially loaded as PDBs, solvated with approximately 8000 TIP3PBOX waters, and chloride counter-ions were added for charge neutrality. Systems were then minimized using steepest slope

descent, thermally ramped to 300K over 200 ps, and run for either 1 ns or 100 ns with a 2 fs time step, periodic boundary conditions, SHAKE algorithm, and under NPT conditions with an 8 angstrom electrostatic cutoff.

Free Energy Analysis

Ante-MMPBSA.py was used to generate the individual topology files for use in MMPBSA.py from the single 100ns trajectories of compounds 6 and 7 with CBX8. GB radii set to setting 2, and a sampling frequency of 1000 snapshots over 2000 MD frames was set. Per-residue analysis was also done. Both GB and PB energies were calculated. PB and GB energies showed consistent trends with relative binding energies and PB energies were reported.
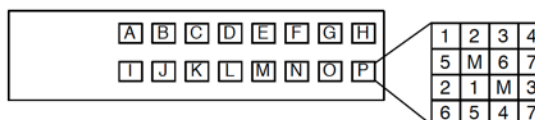
Computational Resources

Each 10-pose iteration for 1 ns trajectories required 1 hour on 280 cores (10 nodes with 2 Intel Xeon E5-2680 v4 processors each). 100 ns trajectories required approximately 120 hours using 28 cores each. Total simulation time for both systems including starting pose generation was approximately 180 hours

Protein Microarray

The methyl domain reader array consists of 96 purified chromatin-associated effector proteins. Currently, 20 of these proteins are known to bind to H3K9me3 and H3K27me3. The array contains 28 chromodomain containing proteins. Each protein is spotted in duplicate onto a nitrocellulose-coated membrane. Visualization of binding is achieved by using streptavidin covalently tagged with a florescent dye.

Methyl Domain Readers Array

TUDOR	TUDOR	TUDOR	TUDOR	TUDOR	TUDOR+PHD	PHD	TUDOR
A1) 53BP1(1-2)	B1) TDRD4-1	C1) TDRD8	D1) AKAP1	E1) SETDB1	F1) JMJD2B	G1) ING2	H1) SGF29
A2) 53BP1(1-2)*	B2) TDRD4-3	C2) TDRD9	D2) LBR	E2) SMN	F2) JMJD2C	G2) TAF3	H2) SPIN1
A3) TDRD1-2	B3) TDRD4-4	C3) TDRD10	D3) PHF1	E3) SMN*	F3) MTF2	G3) TRIM24 PHD+BRD[1]	H3) TDRD2/TDRKH
A4) TDRD1-3	B4) TDRD5	C4) TDRD11	D4) PHF19	E4) SPF30	F4) PHF1*	G4) CHD5(1-2)*	H4) LIN9
A5) TDRD1-4	B5) TDRD6-5	C5) TDRD12-1	D5) PHF20	E5) UHRF1	**TUDOR**	**CHROMO+PHD**	
A6) TDRD2	B6) TDRD6-6	C6) ARID4A	D6) PHF20-2	E6) ZGPAT	F5) Pombe1*	G5) CHD3	
A7) TDRD3	B7) TDRD6-7	C7) ARID4B	D7) PHF20L1	E7) JMJD2A(1-2)	F6) Spindlin1*	G6) CHD4	
					F7) TDRKH[1]	G7) CHD5	

CHROMO	CHROMO	CHROMO	CHROMO	AGENET	IDCL	BROMO*	YEATS
I1) ARID4A	J1) HP1α/CBX5	K1) CHD1	L1) MPP8	M1) FMR1	N1) PCNA	O1) SP140*	P1) YEATS4/GAS41
I2) ARID4B	J2) HP1β	K2) CHD2	L2) MRG15	M2) FXR1	**MBT**	O2) BRD4(1)*	P2) ENL/MLLT1
I3) CBX2	J3) HP1γ/CBX3	K3) CHD6	L3) MYST1/MOF	M3) FXR2	N2) L3MBTL1(1-3)	O3) WDR9(1-2)*	P3) AF9/MLLT3
I4) CBX4	J4) CDY1	K4) CHD7	L4) SMARCC1	**ANK**	N3) L3MBTL1(1-3)*	O4) SMAP*	P4) YEATS2
I5) CBX6	J5) CDYL1b	K5) CHD8	L5) SMARCC2	M4) GLP	N4) L3MBTL3	**HEAT**	
I6) CBX7	J6) CDYL2	K6) CHD9	L6) SUV39H1	M5) IkB FL[2]	**PWWP+**	O5) N-CAPD3(1-4)	
I7) CBX8	J7) MSL3	K7) TIP60	L7) SUV39H2	**BAH**	N5) DNMT3A[2]	O6) N-CAPD3(5-8)	
				M6) ORC1	N6) NSD1 PWWP+PHD*		
					Zf-CW		
					N7) ZCWPW1		

* = Bedford Lab Construct
[1] = Michelle Barton
[2] = Or Gozani

Figure S26: List of the methyl reader domains evaluated in the protein microarray seen in Figure 10. The letter associated with each domain specifies the section of the array containing the protein and the number designates the position of each replicate within that section.

Methods for Cell-based Studies

FACS analysis of Inhibitor Uptake, Cell Viability and Cell Cycle Distribution

TOV21G cells (ATCC® CRL-11730™) seeded at $2.5×10^5$ cells/well in a 6-well plate were adhered overnight at 37°C, 5% CO_2 then treated for 21 hr with 0 μM, 5 μM, 10 μM, 25 μM or 50 μM of compound **2** at 37°C, 5% CO_2. After incubation, cells were washed with media (RPMI-1640 medium containing 25 mM HEPES and L-Glutamine (Thermo Scientific (Hyclone) SH30255.01), supplemented with 10 % fetal bovine serum (FBS)(Life Technologies, 16000044), 1% L-Glutamine (Hyclone SH3003401) and 1% penicillin/streptomycin (Hyclone SV30010). Cells were were rinsed with 1× PBS, trypsinized and collected by centrifuging at 1200 RPM for 5 min at 4°C. Cells were resuspended and washed with 1× PBS, pelleted at 1200 RPM for 5 min at 4°C, resuspended in 1:1000 Fixable Viability Dye eFluor 780 (eBioscience, 65-0865-14) and incubated on ice for 30 min. After incubation, cells were collected by centrifuging at 1400 rpm for 5 min at 4°C and washed twice with ice cold 1× PBS. Cells were simultaneously fixed with 500 µL ice cold 70% ethanol and stained with propidium iodide (PI) (30 µL of 20 mg/mL RNaseA (Invitrogen, 60216-RN) and 26 µL PI (eBioscience, BMS5-FI/100)) for 30 min at room temperature. Cells were centrifuged at 1400 rpm for 8 min at 4°C to remove staining solution, washed with 1× PBS, centrifuged, resuspended with 0.3% FBS in 1× PBS and stored at 4°C until analysis. Samples were filtered (BD Falcon 352235) immediately before running on a Becton Dickinson FACSCalibur™ and analyzed using FlowJo 7.6.5 software.

MTT Assay

TOV21G cells were seeded at 2000 or 3000 cells per well in 96 well culture plates and allowed to adhere overnight at 37°C, 5% CO_2 prior to treatment with inhibitor solutions for 72 hr at 37°C, 5% CO_2. Cell viability was assessed by adding 20 µL of 5 mg/mL MTT in 1x PBS solution (Molecular Probes M6494) to each well and incubating at 37°C, 5% CO_2 for 4 hr. Media was replaced with 100 µL DMSO and absorbance was measured at 560 nm and 670 nm. Cell viability was calculated as follows:

Cell viability = (average A560nm - A670nm of sample) ÷ (average A560nm - A670nm of 0 μM sample) x 100%