HHS Public Access

Author manuscript

Cell Chem Biol. Author manuscript; available in PMC 2022 April 24.

Published in final edited form as:

Cell Chem Biol. 2022 April 21; 29(4): 555–571.e11. doi:10.1016/j.chembiol.2021.10.003.

Reprogramming CBX8-PRC1 function with a positive allosteric modulator

Junghyun L. Suh[1], **Daniel Bsteh**[2,12], **Bryce Hart**[1,12], **Yibo Si**[2], **Tyler M. Weaver**[3], **Carina Pribitzer**[4], **Roy Lau**[5], **Shivani Soni**[6], **Heather Ogana**[7], **Justin M. Rectenwald**[1], **Jacqueline L. Norris**[1], **Stephanie H. Cholensky**[1], **Cari Sagum**[8], **Jessica D. Umana**[1], **Dongxu Li**[9,10], **Brian Hardy**[1], **Mark T. Bedford**[8], **Shannon M. Mumenthaler**[5,6], **Heinz-Josef Lenz**[6], **Yong-Mi Kim**[7], **Gang Greg Wang**[9,10], **Ken H. Pearce**[1], **Lindsey I. James**[1,9], **Dmitri B. Kireev**[1], **Catherine A. Musselman**[3,11], **Stephen V. Frye**[1,9,*], **Oliver Bell**[2,4,13,*]

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

[2]Department of Biochemistry and Molecular Medicine, Keck School of Medicine of the University of Southern California, Los Angeles, CA 90033, USA

[3]University of Iowa, Department of Biochemistry, Iowa City, IA 52242, USA

[4]Institute of Molecular Biotechnology of the Austrian Academy of Sciences (IMBA), Vienna Biocenter (VBC), Dr. Bohr-Gasse 3, 1030 Vienna, Austria

[5]Lawrence J. Ellison Institute for Transformative Medicine, University of Southern California, Los Angeles, CA 90033, USA

[6]Division of Medical Oncology, Norris Comprehensive Cancer Center, Keck School of Medicine, University of Southern California, Los Angeles, CA 90033, USA

[7]Department of Pediatrics, Children's Hospital Los Angeles, Keck School of Medicine of the University of Southern California, Los Angeles, CA 90027, USA

[8]Department of Epigenetics and Molecular Carcinogenesis, University of Texas MD Anderson Cancer Center, Smithville, TX 78957, USA

[*]Correspondence: svfrye@email.unc.edu (S.V.F.), oliver.bell@med.usc.edu (O.B.).

AUTHOR CONTRIBUTIONS

Conceptualization, J.L.S., L.I.J., S.V.F., and O.B.; formal analysis, D.B., D.B.K., T.M.W., and J.L.S.; investigation, J.L.S., D.B., Y.S., T.M.W., B. Hart, C.P., R.L., S.S., H.O., J.M.R., J.L.N., S.H.C., C.S., J.D.U., and B. Hardy; writing – original draft, J.L.S., D.B., T.M.W., C.A.M., D.B.K., O.B., and S.V.F.; writing – review & editing, J.L.S., D.B., B. Hart, M.T.B., H.-J.L., S.M.M., J.-M.K., G.G.W., K.H.P., D.B.K., L.I.J., C.A.M., S.V.F., and O.B.; resources, M.T.B., H.-J.L., S.M.M., J.-M.K., G.G.W., K.H.P., D.B.K., L.I.J., C.A.M., S.V.F., and O.B.; visualization, J.L.S., D.B., B. Hart, T.M.W., D.B.K., and O.B.; supervision, M.T.B., H.-J.L., S.M.M., J.-M.K., G.G.W., K.H.P., D.B.K., L.I.J., C.A.M., S.V.F., and O.B.; project administration, O.B. and S.V.F.; funding acquisition, L.I.J., O.B., M.T.B., and S.V.F.

SUPPLEMENTAL INFORMATION

Supplemental information can be found online at https://doi.org/10.1016/j.chembiol.2021.10.003.

DECLARATION OF INTERESTS

The authors declare no competing interests.

[9]Lineberger Comprehensive Cancer Center, UNC School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[10]Department of Biochemistry and Biophysics, UNC School of Medicine, University of North Carolina at Chapel Hill, Chapel Hill, NC 27599, USA

[11]University of Colorado Anschutz Medical Campus, Department of Biochemistry and Molecular Genetics, Aurora, CO 80045, USA

[12]These authors contributed equally

[13]Lead contact

SUMMARY

Canonical targeting of Polycomb repressive complex 1 (PRC1) to repress developmental genes is mediated by cell-type-specific, paralogous chromobox (CBX) proteins (CBX2, 4, 6, 7, and 8). Based on their central role in silencing and their dysregulation associated with human disease including cancer, CBX proteins are attractive targets for small-molecule chemical probe development. Here, we have used a quantitative and target-specific cellular assay to discover a potent positive allosteric modulator (PAM) of CBX8. The PAM activity of UNC7040 antagonizes H3K27me3 binding by CBX8 while increasing interactions with nucleic acids. We show that treatment with UNC7040 leads to efficient and selective eviction of CBX8-containing PRC1 from chromatin, loss of silencing, and reduced proliferation across different cancer cell lines. Our discovery and characterization of UNC7040 not only reveals the most cellularly potent CBX8-specific chemical probe to date, but also corroborates a mechanism of Polycomb regulation by non-specific CBX nucleotide binding activity.

In brief

Suh et al. describe the discovery of UNC7040, a potent, cellularly active positive allosteric modulator of CBX8. In addition to blocking H3K27me3 binding, UNC7040 enhances CBX8 interaction with nucleic acids leading to efficient canonical PRC1 eviction and activation of Polycomb target genes.

INTRODUCTION

Precise inheritance of distinct gene expression programs during cell division is essential for cellular differentiation and development of multicellular organisms. Epigenetic mechanisms involving posttranslational histone modifications promote heritable transmission of gene expression (Moazed, 2011; Reinberg and Vales, 2018). Important epigenetic regulators include Polycomb group (PcG) proteins, which assemble into large multi-subunit complexes that mediate chromatin modifications to enforce transcriptional gene silencing (Aranda et al., 2015; Schuettengruber et al., 2017; Simon and Kingston, 2013). Mutations in PcG protein-encoding genes are frequently associated with malignancies, underscoring the importance of the Polycomb pathway in maintenance of cell identity (Chan and Morey, 2019; Laugesen and Helin, 2014). The two major complexes, Polycomb repressive complex 1 (PRC1) and PRC2, harbor distinct writing and reading activities of histone modifications.

PRC1 complexes are defined by an E3 ligase activity catalyzing monoubiquitination of histone H2A at lysine 119 (H2AK119ub1) (Cao et al., 2005; de Napoles et al., 2004; McGinty et al., 2014; Wang et al., 2004a), while PRC2 complexes generate mono-, di-, or tri-methylation of histone H3 at lysine 27 (H3K27me1, 2, 3) (Cao et al., 2005; Czermin et al., 2002; Müller et al., 2002). PRC1 can be further subdivided into canonical and variant complexes (cPRC1 and vPRC1, respectively), based on incorporation of chromobox domain-containing (CBX) proteins. In contrast to sequence-dependent vPRC1 targeting, cPRC1 complexes rely on H3K27me3 binding via CBX proteins facilitating recruitment to PRC2 target loci (Cao et al., 2002; Min et al., 2003; Wang et al., 2004b).

Mammalian genomes encode five Polycomb-associated CBX proteins (CBX2, 4, 6, 7, and 8) that are thought to contribute unique and non-overlapping functions to PRC1 target selectivity and activity during development and in disease (Chan and Morey, 2019; Laugesen and Helin, 2014). In embryonic and adult stem cells, cPRC1 is mostly associated with CBX7, which promotes self-renewal by silencing differentiation-specific genes including CBX2, 4, and 8. Conversely, CBX2, 4, and 8 are expressed during differentiation and repress genes involved in stem cell maintenance (Creppe et al., 2014; Morey et al., 2013). CBX7 and CBX8 have been linked to cancer through PRC1-dependent transcriptional repression of the well-known tumor-suppressor locus Cdkn2a (also known as Ink4a/Arf) (Dietrich et al., 2012; Gil et al., 2004; Scott et al., 2007; Tan et al., 2011; Yap et al., 2010). In addition, there is evidence in breast cancer and leukemia that CBX8 exerts oncogenic activity through non-canonical functions independent of PRC1, which are poorly defined (Chung et al., 2016; Tan et al., 2011).

Polycomb's direct role in many disease states (Chan and Morey, 2019; Koppens and van Lohuizen, 2016; Pasini and Di Croce, 2016) has inspired the development of chemical probes to explore Polycomb biology, and consequential drug discovery opportunities leading to PRC2-directed inhibitors entering clinical trials (He et al., 2017; Knutson et al., 2013, 2014; Qi et al., 2017). PRC1 chemical probe development has not yet led to clinical studies, but we and others have worked toward potent ligands for the readers of PRC1, CBX chromodomains (CDs) (Lamb et al., 2019; Milosevich et al., 2016; Ren et al., 2015, 2016; Simhadri et al., 2014; Stuckey et al., 2016a, 2016b; Wang et al., 2020).

We recently reported a potent cellular positive allosteric modulator (PAM), UNC4976, which binds to CBX7 competitively with H3K27me3 peptides, while increasing affinity for oligonucleotides (Lamb et al., 2019). Both DNA and methyl-lysine (Kme) binding are also critical for chromatin association of CBX8 (Connelly et al., 2019). Yet, how these targeting activities contribute to CBX8 functions in development and disease remains largely unclear. Herein, we report the discovery and characterization of UNC7040, a potent, cellularly active PAM specific for CBX8. UNC7040 efficiently impairs PRC1 targeting and transcriptional repression by disrupting CBX8 interaction with H3K27me3. Notably, the PAM activity of UNC7040 increases CBX8's affinity for nucleic acids and enhances cellular efficacy to provide a powerful tool to dissect its canonical and non-canonical roles in gene regulation. Together, our results further establish the importance of allostery in CBX cellular pharmacology and activity in disease model systems dependent upon CBX8.

RESULTS

Structure-based design of CBX8 selective compounds

During the development of UNC3866, a peptidomimetic cellular probe that selectively targets CBX4/7 CDs, we observed that the replacement of an alanine methyl group by an ethyl group within UNC3866 modestly decreased the affinity toward CBX7 (~6-fold), while slightly increasing its affinity toward CBX8 (Figure 1A, UNC4939) (Stuckey, 2016; Stuckey et al., 2016a). Moreover, replacement of the alanine residue with an isopropyl group resulted in a compound (UNC4030) that was a 100-fold less-potent antagonist of CBX7 and a 2-fold more-potent antagonist of CBX8 when compared with UNC3866. In addition, we recently developed a PAM of CBX7, UNC4976, which contains a norbornyl, methyl substitution at the lysine mimetic position instead of the diethyl group present in UNC3866, which binds to both CBX7 and CBX8 to a modestly greater extent *in vitro* and additionally possesses much better cellular efficacy than UNC3866 (Figure 1A, UNC4976) (Lamb et al., 2019). Based on these results, we hypothesized that the combination of a more sterically demanding group at the alanine position of UNC3866 with lysine mimetics that could induce positive allostery with nucleotide binding, as seen with UNC4976 (Connelly et al., 2018; Lamb et al., 2019), could result in a potent and selective PAM for CBX8.

Recent efforts to develop CBX7 peptidomimetic antagonists have also provided structural insights into the slight differences between the CBX CDs, CBX2, 4, 7, and 8 (Milosevich et al., 2016; Stuckey et al., 2016a, 2016b). Structural analysis of UNC3866 bound to CBX7 (PDB: 5EPJ) and CBX8 (PDB: 5EQ0) provides a clear rationale for the potency differences between UNC3866, UNC4939, and UNC4030 (Figure 1A), wherein the co-crystal structures revealed that CBX7 possesses a much smaller hydrophobic pocket where the side chain of the alanine resides as compared with CBX8 (Figure 1B). Accordingly, we decided to take advantage of CBX8's larger binding pocket and other structural differences in our overall structure-activity relationship (SAR) (Figures 1C–1E) strategy as depicted in Figure 1F. Initially, we introduced six bulkier and structurally diverse functional groups at the alanine position to investigate the optimum substituent size to occupy this pocket (Figure 1F, red color code). In addition to the alanine pocket being a key factor for CBX8 selectivity versus CBX7, the UNC3866 co-crystal structure revealed that the aromatic cage of CBX8 is more expansive than that of CBX7 (Figure 1C). Therefore, introduction of larger functional groups than the methyl, norbornyl group at the lysine residue in UNC4976 could increase the selectivity and potency for CBX8 (Figure 1F, magenta color code). In addition, a limitation of UNC4976 is its diminished water solubility due to the hydrophobic norbornyl group at the lysine mimetic position. Therefore, given that the CBX8 potency and selectivity SAR pointed to even further increases in hydrophobicity, it was apparent that adding a solubilizing group would be necessary to enable biological evaluation while avoiding biologically incompatible solvents, such as >1% DMSO. The N-terminal *tert*-butyl benzoyl residue (N-cap) of UNC3866 is solvent accessible and there are several proximal hydrogen bond donors (HBDs) in CBX8 (Figure 1D). Therefore, we explored introduction of heteroatoms and protonatable groups at this position to improve solubility and perhaps potency (Figure 1F, blue color code). The side chain of leucine in the UNC3866 co-crystal structure with CBX8 is also solvent accessible, and we hypothesized that installation of an

HBD at this position could create a favorable interaction with E8 to increase the binding affinity and/or solubility of CBX8 ligands (Figure 1F, green color code). We pursued an iterative strategy of rounds of synthesis and testing followed by rounds of mixing and matching of optimal substituents, rather than a fully combinatorial approach that would result in synthesis of many mismatched substituents (i.e., combinations of all hydrophobic residues or all polar residues).

We sequentially investigated the effect of replacing each position: "the Ala pocket," "the N-cap," "the Leu position," and "Lys mimetics." We synthesized compounds using conventional solid-phase peptide synthesis with Rink amide resin and tested their affinity to CBX7/8-CDs using a time-resolved fluorescence energy transfer (TR-FRET) assay (Rectenwald et al., 2019). As expected, CBX8 could accommodate larger substituents in the Ala pocket relative to CBX7 (Table S1, more detailed analysis is described in Data S1), and compound **2**, which has valine at this position, was a potent antagonist for CBX8 with 21-fold selectivity over CBX7. Interestingly, the introduction of just one heteroatom at the N-cap position resulted in full water solubility at 10 mM. Even though there were small differences in CBX8 potency between N-cap variations (Table S1), we decided to retain a *tert*-butyl substitution at this position and rely on introduction of a solubilizing group at the leucine position. Accordingly, compound **2** analogs varying at the leucine position were synthesized (Table S1). Similar to the N-cap modifications, adding one or more heteroatoms at this position improved solubility. Morpholino substitution as in compound **14** resulted in high water solubility and increased CBX8 selectivity to 58-fold. Finally, we evaluated the effect of incorporating a group bulkier than methyl at the lysine mimetic position while maintaining the norbornyl substitution from UNC4976, which had been shown to result in greatly enhanced cellular efficacy (Lamb et al., 2019). We combined the optimal substituents defined above: an isopropyl (**2**) or cyclobutyl (**5**) group at the alanine position and a morpholino group at the leucine position (Table S1) and synthesized eight compounds that contained ethyl, propyl, or isobutyl substitution on the norbornyl-lysine mimetic (Table S1). Incorporation of a larger alkyl group did not result in a significant enhancement in CBX8 *in vitro* potency and selectivity compared with compounds containing a methyl group (**14** and **19**). However, as discussed below, variation at this position had a dramatic effect on cellular activity.

Cellular screen of new antagonists of CBX8

Previous studies describing PAMs of CBX7 revealed that UNC4976 (Figure 1A) demonstrates 14-fold enhanced cellular potency compared with UNC3866 (Figure 1A, top compound), despite almost identical *in vitro* thermodynamic and kinetic affinity profiles, as well as cell permeability (Lamb et al., 2019; Moussa et al., 2019; Stuckey et al., 2016a, 2016b). Based on this finding, we evaluated our compounds in a CBX8-specific cell-based reporter assay to determine if cellular potency trends differ substantially from *in vitro* results. We engineered a mouse embryonic stem cell (mESC) line that harbors an artificial DNA binding array composed of 12xZFHD1, 4xGAL4 UAS, and 7xTetO sites upstream of a PGK promoter controlling the expression of puromycin resistance and a GFP gene (Figure 2A). In addition, we expressed CBX8 fused to the Tet repressor domain (TetR), enabling recruitment and assembly of functional cPRC1 at the TetO DNA binding sites (Figures S1A

and S1B). After the nucleation event, additional PRC1 and PRC2 complexes are recruited to sites along the gene body through interactions of the CBX8 CDs with trimethylated histone H3, lysine 27 (H3K27me3), which results in formation of a repressive Polycomb chromatin domain similar to endogenous targets and transcriptional silencing of the GFP gene, as measured by a decrease in GFP signal using flow cytometry (Figures 2B, S1B, and S1C). Notably, GFP reporter gene repression is reversible upon doxycycline (DOX)-dependent release of the TetR fusion protein (Figure S1D). Thus, while ectopic targeting of CBX8 can initiate formation of a repressive Polycomb chromatin domain in mESCs, in contrast to CBX7 it fails to promote propagation of silencing in the absence of the initial stimulus (Moussa et al., 2019). Having established this reversible CBX8 reporter, we reasoned that addition of CBX8 antagonists would inhibit binding of PRC1 and de-repress the GFP reporter gene.

A DMSO tolerance test was performed using a CellTiter-Glo cell viability assay in the mESC line utilized in our CBX8 GFP reporter assay. We found that treatment with up to 1.25% DMSO resulted in no cellular toxicity (Figure S1E). Accordingly, we assessed the cellular efficacy of our CBX8 antagonists discussed above with final assay concentrations of 1% DMSO. After 48 h exposure of test compounds in the CBX8 reporter mESC line, we determined the change of GFP levels compared with 1% DMSO control by flow cytometry. Results organized as discussed above for our *in vitro* SAR are depicted in Figures S2A–S2D.

Consistent with CBX8 potency trends in our TR-FRET results, compounds with alanine position substituents isopropyl (**2**) or cyclopropyl (**4**) showed the largest reactivation of GFP signals among the seven compounds, followed by cyclobutyl (**5**) and cyclopentyl (**6**) groups, while the methyl (**1**), isobutyl (**3**), and *tert*-butyl (**7**) all demonstrated minimal effect on GFP reactivation (Figure S2A). Next, we examined N-cap modifications. Although there were no significant *in vitro* potency differences between these four N-cap variations (compounds **2**, **8**, **9**, and **10**), cellular data showed that the *tert*-butyl moiety from the parent compounds, UNC3866 and UNC4976, was the most potent modification (Figure S2B). For the Leu position modifications, ornithine substitution (**13**) at this position provided the highest GFP reactivation compared with morpholino-containing (**12** and **14**) or leucine-containing (**2**) compounds at the highest tested concentration (99 μM) (Figure S2C), consistent with formation of peptide-rich domains that can transiently disrupt the membrane or promote peptide translocation into the cell (Duchardt et al., 2007; Herce et al., 2009; Tunnemann et al., 2006). Finally, lysine mimetic modification had a dramatic effect on cellular efficacy (Figure S2D, Table S1), wherein compounds with isobutyl groups **18** and **22** demonstrated almost fully reactivated GFP levels in contrast to the other compounds in the series, which did not reactivate GFP by greater than 50%. This striking result showed that increased steric bulk at this position greatly enhanced cellular efficacy despite having little effect on *in vitro* potency.

Since the ornithine substitution (**13**, Figure S2C) at the leucine position and the isobutyl group at the lysine mimetic position (**18**, Figure S2D) showed the greatest cellular efficacy, we incorporated these modifications into one molecule and synthesized two additional compounds that mix the best substituents at each position (Figure 2C, compounds **23** and **24**). The four best compounds from the previous SAR study and these additional

compounds, **23** and **24**, were then tested in the CBX8 cellular assay to compare the effect of combined optimal modifications. Interestingly, isobutyl-containing compounds **18** and **22** displayed 20- to 30-fold enhanced cellular potency compared with the propyl-containing compounds, **17** and **21**. We hypothesized that this could be due to enhanced allosteric influence of the isobutyl compounds on CBX8 nucleic acid binding, similar to the allostery we observed with UNC4976 versus UNC3866 with CBX7 (Lamb et al., 2019). An allosteric mechanism is plausible based on a recent nuclear magnetic resonance (NMR) study of CBX8, which revealed that trimethylated H3K27 binds to one face of CBX8 while DNA binds to the other face with the aromatic cage where our lysine mimics bind sandwiched in between (Connelly et al., 2018). However, before seeking other explanations for the dramatic difference in cell potency mediated by addition of a single methyl group to convert the propyl substituent in **21** to an isobutyl group in **22**, we examined the cell permeability of our compounds, as this could also be a significant variable influencing cell potency.

Influence of modifications on cellular permeability

An important factor to consider is how cell permeability changes with each ligand modification as this can influence cellular efficacy. To address this question, we utilized the chloroalkane penetration assay (CAPA) (Peraro et al., 2017, 2018) to quantitatively measure cell permeability. CAPA is especially useful in assessing compounds of modest permeability that cannot be rank ordered by assays, such as Caco-2 or PAMPA, because of inadequate dynamic range at the lower limits of detection (Foley et al., 2020). To systematically compare the effect of each functional group modification on cell permeability, we synthesized nine compounds containing a chloroalkane tag as depicted in Table S2.

Assessment of CT-modified compounds by CAPA revealed that incorporation of a cyclobutyl group at the alanine position had little effect on the permeability of the compound compared with the parent compound (Table S2, **25** versus **26**, and Figure S2E). For the N-cap position the *tert*-butyl was the most permeable, which correlates with the potency trend observed in the CBX8 reporter assay (Table S2, **26** and **28**, and Figure S2F). However, based on their CP_{50} values, the greatly diminished cellular activity for the compounds containing a hydroxy (**27**) or morpholino (**28**) functional group at the N-cap position are not fully attributable to permeability (or *in vitro* affinity; Table S1)) and there could be other factors beyond cell permeability that influence their cellular activity. At the leucine position, the leucine functional group increased permeability relative to compounds containing a positively charged side chain (Table S2, **26** versus **29**–**30**, and Figure S2G). Interestingly, ornithine-containing compound (**30**) demonstrates a very steep CP_{50} curve above a certain concentration (between 11 and 33 μM), similar to our data in the CBX8 reporter assay (Figure S2C) and consistent with the aforementioned transient plasma membrane permeabilization (Duchardt et al., 2007; Herce et al., 2009; Tunnemann et al., 2006). Critically, variations at the lysine mimetic position demonstrated no differences in cell permeability (Table S2, **30**–**33**, and Figure S2H). Therefore, although permeability data explain the increased cell efficacy observed between compounds with leucine position modifications, neither cell permeability nor *in vitro* affinity can explain the significant difference in cellular efficacy of compounds **21** and **22**, which differ only at the lysine mimetic position.

NMR studies of CBX8 with 21 (UNC6779), 22 (UNC6782), and DNA

The ability for the isobutyl lysine functionalized compounds to more effectively disrupt PRC1 activity in cells as compared with unbranched alkyl substitution (Figure S2D), and the similar permeability (Figure S2H) and *in vitro* profile of these compounds (Table S1), encouraged us to investigate the underlying mechanism for this difference by exploring the interaction of **21** and **22** with CBX8 and DNA utilizing NMR spectroscopy. ^{1}H-^{15}N-HSQC titrations were performed by collecting spectra of the ^{15}N-CBX8-CD in the apo state and upon addition of **21** or **22**. Addition of increasing concentrations of **21** and **22** led to significant chemical shift perturbations (CSPs) in the CBX8-CD spectrum consistent with an interaction between the inhibitors and CBX8-CD (Figure S3A). In addition, several residues disappeared upon inhibitor binding, which is consistent with these regions undergoing a conformational exchange on an intermediate timescale. Interestingly, several minor peaks appear only in the **22** titration, which are not seen in the **21** titration, indicating that the CDs can adopt a minor population structural state when bound to **22** but not **21**. As a distinct set of peaks is seen, this indicates that the minor population is either a fully stable separate population or that the major and minor populations are in slow exchange between each other on the NMR timescale.

To further investigate the structural basis for the interaction of **21** and **22** with CBX8-CD, a normalized change in chemical shift from the apo to inhibitor-bound states was calculated using a set of previously published assignments for the CBX8-CD (Connelly et al., 2018). Importantly, this initial analysis is only for the major state population of the **22**-bound CBX8-CD. Residues with significant CSPs upon addition of **21** and **22** map largely to the aromatic cage region and the N-terminal portion of the αA helix, while residues that broaden beyond detection are in the N-terminal portion of the β1 strand, which is known to undergo a conformational change upon binding H3K27me3 (Figures S3B and S3C). Notably, the CSPs are in good agreement with previous data for the addition of the canonical H3K27me3 ligand (Figures S3B and S3C). The residues with CSPs (14/14) and residues that broaden beyond detection (4/5) in the titrations of **21** and **22** titrations are almost identical, indicating the major population CDs bind to the two inhibitors largely in a similar manner.

To assess whether the CBX8-CDs can interact with **21** or **22** and DNA simultaneously, ^{1}H-^{15}N-HSQC spectra were collected on the ^{15}N-CBX8-CDs in the apo state, inhibitor-bound state, and inhibitor plus DNA-bound states. Addition of an 11bp DNA with the CBX8-CDs pre-bound to either **21** or **22** led to significant CSPs in the CBX8-CD spectrum indicating binding (Figure 3A). Comparison of the CBX8-CD spectrum bound to DNA only, inhibitor only, and inhibitor + DNA reveal thats the bound states are unique from each other consistent with the formation of a ternary complex with addition of inhibitor + DNA (Weaver et al., 2018). Residues with significant CSPs upon addition of the 11bp DNA to the CDs pre-bound to either **21** and **22** (major population) map largely to the C-terminal portion of the β1 strand and αA helix consistent with the previously determined 11bp DNA binding site identified in the absence of inhibitor (Figures 3B and 3C). Comparison of the residues with significant CSPs upon addition of DNA reveal 11 residues that are identical between **21** and **22** binding, which indicates a relatively similar binding mode. However, a small subset of CSPs upon addition of DNA to the inhibitor-bound CDs were different,

these included S36 and Y39 for **21**-bound and K33, E43, L49, A51, and A56 for **22**-bound, suggesting there are subtle differences in the structural mechanism of DNA binding for each ligand. Notably, for the **22**-bound CBX8-CDs, CSPs were also seen in the subset of peaks corresponding to the minor population, indicating that it also associates with DNA. Although the structure of this minor population is currently unknown, it is clearly active with respect to DNA binding.

Taken together, NMR data indicate that both **21** and **22** bind to the CBX8-CDs and form a DNA-ternary complex. For the majority of resonances, the chemical shifts upon association with **21** or **22** alone lie along a near linear trajectory toward the tertiary complex chemical shifts, supporting that these compounds stabilize a DNA binding competent conformation of the CBX8-CDs. However, by analyzing the major state-bound populations of **21** and **22**, we could not identify significant differences in the inhibitor binding mode. The presence of a minor state-bound population of the CBX8-CDs with **22** could be the key to the differences we detected for these two inhibitors in a cellular context, and we decided to explore this further using molecular dynamics (MD).

MD of CBX8 with 21 (UNC6779) and 22 (UNC6782)

We applied MD simulations to explore how the ligands (**21** and **22**) affect the conformational ensembles of CBX8 and analyzed the respective MD trajectories to infer a structural mechanism by which **22** may allosterically enhance the affinity of CBX8 for DNA/RNA. In this study, a total of ~50 μs of MD simulations on systems including the CBX8 CDs in complex with, respectively, **21** and **22**, as well as the CBX8 CDs alone were performed. Structural snapshots, one per 40 ps, were extracted from the MD trajectories, aligned and subjected to a cluster analysis. The analysis was performed in such a way that each cluster contained closely related protein folds, within ~1 Å of root-mean-square deviation. Hence, a centroid of each cluster approximates a distinct conformation within the protein's conformational ensemble. Of particular interest were clusters that predominantly consisted of the snapshots featuring either **21** or **22**. Indeed, such clusters can be associated with ligand-induced conformations of the ensemble. As hypothesized, we observed that both **21** and **22**, each in its unique way, alter the conformational ensemble of the CBX8 CDs. Of the total of 395 conformations identified in all three simulated systems, 94 and 40 were induced by **21** or **22**, respectively (Figures 4A and 4B), which suggests a significantly higher conformational mobility of the protein-bound **21**. These ligand-induced conformations were observed during 24% and 31% of time, respectively, for **21**- and **22**-bound CBX8. The existence of such ligand-induced conformations supports the idea that the enhanced binding of the **22**-CBX8 complex to DNA/RNA might be due to the compound's ability to induce "DNA/RNA-friendly" CD conformations. Some of these ligand-specific conformations could represent the "minor state" inferred from the NMR data for **22**. Moreover, the clustering data suggest a higher average stability of the conformations induced by **22** (40 conformations account for 31% of the simulation time, compared with 94 conformations/24% time for **21**). This observed stability is consistent with the NMR spectra showing a slower exchange between the major and minor populations for **22**. A 31% share of such conformations in the ensemble of the **22**-bound CBX8 seems sufficient to make a measurable difference in binding. Moreover, it should be kept in mind

that this share may dynamically change in the presence of DNA/RNA since the latter would capture and preserve these DNA/RNA-friendly CBX8 conformations as soon as they are spontaneously produced, shifting the overall conformational equilibrium in favor of these species. While we expected ligand-induced conformations to show more focused changes around the aromatic cage, as the two compounds only differ in the methyl-lysine mimetic that is expected to bind in this region (Stuckey et al., 2016a), compound-induced conformations instead reflected broader, CD-wide shifts. This suggests that, although the propyl (**21**) to isobutyl (**22**) change is modest in the context of the entire peptidomimetic scaffold, **22** has the ability to drastically alter the conformation of the CBX8 CD even outside the aromatic cage and thereby allosterically enhance DNA/RNA binding.

We then investigated a possible structural mode of the ligand-induced interaction of CBX8 CDs with the DNA double helix. To this end, three sets of 10 MD snapshots each were selected at random from structural clusters predominantly containing either ligand-bound (**21** or **22**) or ligand-free CBX8 CDs. All 30 structures were then submitted to automated protein-DNA docking simulations by the high ambiguity driven protein-protein DOCKing (HADDOCK) algorithm (van Zundert et al., 2016). The resulting HADDOCK scores were in the range between −70 and −110 kcal/mol, which is typical of a small-size protein. Remarkably, HADDOCK scores are consistent with the experimental data, i.e., cluster-weighted averages for CBX8:**22** and CBX7:**21** are, respectively, −81 ± 7 and −73 ± 5 kcal/mol. The top-ranked docking poses show a large contact surface area between the protein and DNA (Figures 4C and 4D) with the K18-G24 β strand binding deep into the major groove. Both ligand-bound CDs share significant similarities in the way they bind to DNA. In particular, the protein-DNA interaction implicates residues R19, R20, R22, and R60, which have been previously identified as important for the interaction with DNA (Figure 4C) (Connelly et al., 2018). However, **21**- and **22**-bound CBX8 displayed significant local differences in engaging the DNA double helix. In particular, only **22**-bound protein obtains binding modes with all four essential DNA binding residues (R19, R20, R22, and R60) simultaneously interacting with DNA (Figures 4B and 4C). Moreover, the latter binding modes were adopted by the most populated ligand-induced conformation clusters wherein the cluster-weighted averages of the number of essential residues simultaneously bound to DNA were 2.4 ± 0.3 and 3.1 ± 0.3 for, respectively, **21-** and **22**-bound CBX8 CDs. Overall, the combination of MD clustering and HADDOCK docking data suggests that modification of the Kme mimetic significantly affects the dynamics of both the K18-G24 loop and the aromatic cage, which in turn are implicated in CBX8:DNA binding, thus providing a clear structural rationale for the positive allosteric effect of **22**.

Final cellular probe optimization, negative control design, and characterization

Considering CBX8 affinity, selectivity against CBX7, cellular efficacy, and cell permeability, we selected compound **22** as the basis for further studies. Although they have similar profiles, we selected the cyclobutyl functional group (**22**) over the isobutyl (**21**) at the alanine position because unnatural amino acids incorporated into peptidomimetics are known to be beneficial for metabolic stability (Blaskovich, 2016; Qvit et al., 2017). While for synthetic ease, our SAR studies and initial mechanistic studies were carried out with C-terminal amides, C-terminal methyl esters are known to show higher cellular efficacy

due to improved permeability (Lamb et al., 2019; Stuckey et al., 2016a). Consequently, compound **34**, which has a *tert*-butyl benzoyl at the N-cap position, cyclobutyl glycine at the alanine position, morpholino ornithine at the leucine position, norbornyl, isobutyl lysine at the lysine mimetic position, and a methyl ester at the C terminus, was synthesized as our final CBX8 cellular probe (UNC7040, Figure 5A, top). UNC7040 was profiled in the CBX8 reporter assay and the CellTiter-Glo viability assay, which resulted in an EC_{50} of 0.84 ± 0.11 μM (n = 9) and no cytotoxicity up to 100 μM, respectively (Figures 5B and 5C). The cellular potency enhancement of UNC7040 versus **22** (EC_{50} = 2.8 ± 0.5 μM) was as expected for replacing a primary amide with a methyl ester (Lamb et al., 2019).

We then set out to develop an appropriate negative control compound for UNC7040. The availability of a negative control is important to enable cellular and *in vivo* studies to be carried out with a matched pair of compounds that are as similar as possible in their structure, physical properties, and off-target profiles, while differing greatly in their on-target activity. Based on prior SAR, which showed that epimers at the leucine position did not bind to CBX CDs (Stuckey et al., 2019), we substituted L-cyclobutyl glycine for D-cyclobutyl glycine resulting in compound **35** (Figure 5A, bottom). As expected, compound **35** displayed a negligible effect on de-repression of GFP in the CBX8 reporter assay (Figure 5B). However, unexpectedly, this compound possessed much higher cellular toxicity (DC_{50} = 6.0 ± 1.4 μM) than UNC7040 (Figure 5C) and was therefore unsuitable as negative control. Alternatively, an analog of the active compound UNC7040 was synthesized with a methylated Nα in the cyclobutyl glycine residue (Figure 5A, compound **36**). Since the incorporation of a methyl group at this position disrupts a key hydrogen bond with Leu49 in CBX CDs (Stuckey et al., 2016a), as expected, **36** showed no cellular activity in the CBX8 reporter assay (Figure 5B). However, once again, **36** was more toxic than UNC7040 with a DC_{50} of 3.6 ± 1.8 μM. As an alternative approach, we incorporated an Nα-methylated valine at the alanine position (**37**), but **37** was again more toxic than UNC7040 (Figure 5C, DC_{50} = 4.4 ± 0.41 μM) but inactive as expected in the CBX8 reporter assay (Figure 5B). Finally, we focused our search for a nontoxic negative control on the lysine mimetic position. As depicted in Figure 4, close structural analogs, **19** and **22**, that differ only in bearing a methyl versus isobutyl group, respectively, in the Kme mimic differ dramatically in their cellular activity (**22**, EC_{50} = 2.8 ± 0.49 μM whereas **19**, EC_{50} > 100 μM). Based on this result, we synthesized compound **38** (UNC7263), which possesses no cellular activity in the CBX8 reporter assay as well as no cytotoxicity up to 100 μM (Figures 5A–5C). Thus, even though the *in vitro* potency for CBX8 in our TR-FRET competition assay is identical for UNC7040 versus UNC7263 (Figure 6A), and their cell permeability is expected to be identical (Figure S2H, compare **30** and **33**), the PAM activity of UNC7040 results in potent cellular activity that is attenuated by at least 100-fold in UNC7263. With final probe, UNC7040 and negative control, UNC7263 in hand, we proceeded with more thorough profiling of selectivity and biological activity.

In vitro selectivity profiling of UNC7040

Using a protein domain microarray, we evaluated the specificity of our CBX8 PAM UNC7040, in a broad and unbiased fashion against a binding domain array containing 274 purified Kme readers, including 33 CDs, 43 Tudor domains, >100 PHDs, and representatives

from several additional domains, such as PWWP, BAH, ELM2, and HORMA domains, and Ank repeats. This protein microarray contains GST fusions of these domains immobilized on a glass slide (see Figure S4A and B for list of the domains). The biotinylated analog of UNC7040 (**39**) bound strongly to CBX2 and CBX8 CDs, and also interacted less strongly with CBX4, CBX6, and CBX7, and weakly to the CD of CDYL2 (Figure S4A). As expected, **39** did not display binding to CBX1, CBX3, or CBX5, which are CBX CDs of human heterochromatin protein 1 paralogs. This qualitative data focused our quantitative selectivity profiling on CBX domains.

To quantitate the selectivity of UNC7040 we utilized a TR-FRET assay with CBX2/4/6/5/7/8, CDYL2, and the MPP8 CD, which was recently established (Rectenwald et al., 2019). The IC_{50} trends generated from TR-FRET results were generally consistent with the intensity of the microarray data (Figure 6A, UNC7040). UNC7040 is more than ~5-fold selective for CBX8 against CBX2, CBX4, and CBX6, and 22-fold selective for CBX8 over CBX7 via TR-FRET analysis. We also tested compound **35**, which represents an *in vitro* negative control for UNC7040, and UNC7263, which is the cellular negative control for UNC7040. As expected, **35** did not bind to the CDs in this study at concentrations up to 100 μM (Figure 6A, **35**) and UNC7263, which still has the potential to show *in vitro* binding to these proteins, displayed a similar affinity profile to UNC7040 (Figure 6A, UNC7263). The only significant difference between UNC7040 and UNC7263 is the binding affinity to CDYL2, with IC_{50} values of ~100 μM and 0.96 ± 0.14 μM, respectively. Based on the UNC3866-bound crystal structures of CBX8 and CDYL2, the aromatic cage of CDYL2 is more constrained compared with that of CBX8 because of the side chain of Glu36 in CDYL2 facing into the aromatic cage while the side chain of Tyr29 in CBX8 faces out from the aromatic cage (Figure 6B). Therefore, UNC7263 would more favorably bind to CDYL2 compared with UNC7040, which has a bulkier side chain at the lysine mimetic position.

In addition to the TR-FRET assay, we utilized surface plasmon resonance (SPR) to measure K_d values of compound UNC7040 versus different CDs including CBX2/4/6/8 and CDYL2. To achieve this, we used compound **39** and immobilized this compound on a NeutrAvidin chip. The trends in K_d generated from SPR were consistent with the TR-FRET data except for CBX2, which showed a slightly higher K_d (Figure 6A). Compound **39** is more than 5-fold selective for CBX8 against CBX4, 43-fold selective for CBX8 over CBX6, and more than 100-fold selective over CBX2 and CDYL2. Taken together, *in vitro* binding data suggest that compound UNC7040 possessed at least 5-fold *in vitro* selectivity within CBX CDs and excellent selectivity against other Kme binding domains.

Cellular selectivity profiling of UNC7040

We next sought to assess the selectivity of UNC7040 against other CDs in a cellular context. To specifically test its activity against CBX7-dependent gene silencing we utilized a previously reported CBX7 reporter assay which involves CBX7 tethering via ZFHD1 recapitulating Polycomb silencing of a GFP reporter gene in mESCs (Lamb et al., 2019). Here, as a positive control, UNC4976 was tested along with UNC7040 and, consistent with our published data, UNC4976 had an EC_{50} of 3.0 ± 0.40 μM (n = 9). On the contrary, UNC7040 demonstrated no cellular activity in the CBX7 reporter cell line (Figure S4C).

To determine the impact of UNC7040 treatment on chromatin binding of paralogous CBX proteins we used chromatin immunoprecipitation (ChIP)-qPCR comparing changes in occupancy of CBX2, CBX6, CBX7, and CBX8 at endogenous Polycomb target genes in mESCs. Compound treatment was limited to 6 h to exclude indirect effects on CBX protein binding due to transcriptional changes. As expected, addition of 20 μM UNC7040 strongly reduced CBX8 occupancy relative to untreated TetR-CBX8-expressing reporter mESCs (Figure 6C). In contrast, chromatin binding of CBX2, CBX6, and CBX7 remained unaffected by UNC7040 treatment. CBX8 is normally not expressed in mESCs (Figure S1). To exclude the possibility that ectopic CBX8 expression masks cross-reactivity of UNC7040 against other CDs, we evaluated CBX2, CBX6, and CBX7 binding in response to 20 μM UNC7040 treatment of wild-type mESCs (Figure S4D). In the absence of CBX8, UNC7040 had minimal effects on chromatin binding of other Polycomb-associated CBX paralogs. Together, these results extend our *in vitro* data demonstrating that UNC7040 is a selective PAM against CBX8 in the context of native chromatin regulation in cells.

Comparison of UNC7040, UNC7042, and UNC7263 effects on CBX8 binding to DNA by TR-FRET

Our NMR and MD exploration of the influence of Kme mimetics containing an isobutyl-norbornyl (**21**) versus a propyl-norbornyl (**22**) in the context of primary amides provided qualitative support of a differential effect on nucleic acid binding. To gain a quantitative readout we sought to develop an *in vitro* assay for ternary complex formation examining our final methyl ester-containing probe and negative controls. We developed a TR-FRET ternary complex assay to examine whether UNC7040 and UNC7263 could stabilize interactions between CBX8 and biotin-tagged double-stranded DNA (biotin-dsDNA). CBX8 CD titration was performed with a constant 10 nM concentration of biotin-dsDNA, in the presence or absence of saturating amounts (5 μM) of compounds. Treatment with UNC7040 strongly enhanced the CBX8-DNA interaction (Figure 6D) and, in comparison, UNC7263 improved this interaction to a much weaker degree. Importantly, *in vitro* negative control UNC7042 had no effect on modulating CBX8 binding to DNA. Using the same approach, we also tested the binding of mutant CBX8 R19A to DNA (Figure 6E). This substitution abrogates the interaction of CBX8 with DNA. Notably, unlike UNC7263, UNC7040 rescued a significant amount of binding between CBX8 R19A and DNA. Taken together with our qualitative NMR data and MD simulations of the primary amides (**21** and **22**, above), these data provide quantitative evidence that supports the hypothesis that UNC7040 possesses an improved ability to stabilize the binding of CBX8 to DNA relative to UNC7263 and this difference translates to significant differences in cellular efficacy in multiple systems, as predicted by our CBX8 reporter assay (Figure 5).

Activity of UNC7040 in clinically relevant cancer cell lines

To assess the activity of UNC7040 in CBX8-dependent, clinically relevant cancer models, we interrogated the impact of treatment in cellular models of diffuse large B cell lymphoma (DLBCL) and colorectal cancer (CRC). Lymphomagenesis results from aberrant transcriptional repression of cell-cycle check point and B cell maturation genes by the Polycomb pathway (Beguelin et al., 2016). Specifically, silencing is mediated by PRC1 containing BCOR and CBX8. By binding H3K27me3, CBX8 stabilizes BCOR

interaction with BCL6 and facilitates robust transcriptional silencing (Beguelin et al., 2016). Importantly, CBX8 RNAi knockdown mimics the antiproliferative effects of EZH2 and BCL6 inhibition in DLBCL cells, arguing that CBX8 presents a potential therapeutic target in B cell lymphoma. We examined proliferation of SUDHL4 cells in response to UNC7040 or UNC7263 treatment. To benchmark against proliferation defects by PRC2 inhibition, we also treated SUDHL4 cells with small molecules targeting EZH1/2 (UNC1999) or EED (EED226). As expected, small-molecule-mediated PRC2 inactivation, but not the respective control compounds (UNC2400, UNC5679), dramatically reduced DLBCL cell proliferation (Figure 7A). Importantly, SUDHL4 cell proliferation was also impaired when treated with 40 μM of UNC7040, consistent with CBX8 acting downstream of PRC2. In comparison, proliferation was unaffected at 40 μM of UNC7263. However, antiproliferative activity of UNC7263 was observed at an increased dosage of 100 μM, consistent with our TR-FRET ternary complex data demonstrating that this compound is a weak PAM of CBX8 binding to nucleic acids when compared with UNC7040. Finally, we found that cell proliferation was further reduced by combined treatment of UNC7040 and EED226, suggesting that allosteric PRC1 and PRC2 antagonism has additive effects on DLBCL cell growth (Figure 7B).

To relate the proliferation defect in response to UNC7040 to changes in PRC1 occupancy, we treated SUDHL4 cells with the compound for 6 h and used calibrated ChIP coupled to next-generation sequencing (cChIP-seq). In addition, we profiled gene expression changes by RNA-seq after 6 days. At first, we used cChIP-seq of CBX8, RING1B, and H3K27me3 in untreated cells to determine the genome-wide distribution of PRC1 and PRC2 targets. We also mapped histone H3 lysine 27 acetylation (H3K27ac), an important enhancer mark, to discriminate active and inactive gene regulatory regions (Rada-Iglesias et al., 2011). We detected 1,735 peaks with significant RING1B enrichment identifying cPRC1 and vPRC1 binding sites in SUDHL4 cells (Figures 7C and 7D). Surprisingly, the majority of RING1B peaks colocalized with H3K27ac at active gene regulatory regions. Only a fraction of RING1B overlapped with H3K27me3 suggesting that most PRC1 targeting involves PRC2-independent mechanisms (groups 3–10, vPRC1 targets). CBX8 was preferentially enriched at H3K27me3 targets, consistent with its proposed role in cPRC1 recruitment (groups 1 and 2, cPRC1 targets) (Figure 7D). H3K27me3 peaks were located predominantly at intronic and distal intergenic regions, suggesting that PRC2 signals for CBX8 recruitment to repress enhancers in DLBCL cells (Figure 7E). Short treatment with UNC7040 strongly reduced CBX8 and RING1B occupancy at H3K27me3-modified regions (groups 1 and 2), whereas binding remained largely unaffected at most vPRC1 target sites (groups 3–8) (Figures 7C, 7D, and 7F). Surprisingly, at vPRC1 targets in groups 9 and 10, CBX8 binding was modestly but significantly increased upon UNC7040 treatment (Figures 7D and 7F). By comparison, UNC7263 caused less cPRC1 displacement, consistent with its limited effect on SUDHL4 cell proliferation and diminished activity in the TR-FRET ternary complex assay and our CBX8 reporter assay (Figures 5D, 5B, and 7A). These results demonstrate that UNC7040 efficiently disrupts H3K27me3 recognition and canonical CBX8 function in cancer cells. It is currently unclear if genomic CBX8 redistribution to vPRC1 target sites is linked to PAM activity enhancing nucleic acid binding.

Transcriptome analysis by RNA-seq revealed that UNC7040 triggered robust, differential expression of 481 genes (p_{adj} = 0.05, LFC = ±0.5; Figures 7G, and S5A). The majority of

genes with altered mRNA levels were upregulated in line with loss of Polycomb repression by UNC7040. UNC7263 treatment had negligible impact on SUDHL4 gene regulation. Consistent with reduced proliferation of SUDHL4 cells, UNC7040-induced genes included members of the tumor necrosis factor superfamily and regulators involved in NF-κB signaling important for B cell maturation (Mackay and Schneider, 2009) (Figure 7G). Being enriched at distal intergenic and intronic regions marked by H3K27me3, CBX8-PRC1 might control B cell maturation genes by repressing the activity of regulatory enhancers. To predict genes under control of CBX8-dependent enhancers, we employed Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010). GREAT analysis of UNC7040-sensitive CBX8 peaks revealed enrichment of genes with functional annotation in B cell proliferation supporting the notion that CBX8 directs cPRC1-dependent silencing to enhancers of B cell maturation genes and this control is disrupted by UNC7040 (Figure 7H).

CRC presents another emerging model for CBX8-dependent control of cell proliferation, but the underlying mechanism remains largely unexplored (Zhang et al., 2019). We selected a panel of established CRC cell lines, confirmed CBX8 expression, and evaluated changes in proliferation after 72 h of treatment with either vehicle control or with increasing concentrations of UNC7040 or UNC7263 (Figures S5B and S5C). UNC7040 treatment reduced LoVo cell proliferation at concentrations above 1.85 μM, indicating CBX8 dependency of this CRC cell line (Figure S5C). The antiproliferative effects were specific to CBX8 inhibition since treatment with CBX7 antagonist UNC4976 or UNC7263 displayed cell counts similar to vehicle control (Figures S5C and S5D). UNC7040 treatment also impaired spheroid formation of LoVo cells but not HCT116 cells, corroborating the cell-line-specific effect (Figure S5E).

Similar to DLBCL cells, cChIP-seq profiling revealed strong reduction in CBX8 and RING1B occupancy at H3K27me3 targets in LoVo cells, indicating efficient cPRC1 displacement in response to UNC7040 (Figures S5F and S5G). Despite substantial cPRC1 loss, UNC7040 triggered only minor transcriptional changes, suggesting that additional mechanisms, such as vPRC1, might act redundantly to maintain gene silencing in CRC cells (Figures S5H and S5I). Nevertheless, we found that expression of *KRT20*, a differentiation marker gene, was upregulated, whereas LGR5, a stem cell marker gene, was downregulated by UNC7040, consistent with CBX8 dependence of LoVo cell proliferation and self-renewal (Shimokawa et al., 2017). In conclusion, epigenomic and transcriptomic profiling in two distinct cancer cell lines demonstrated potent activity of UNC7040 to displace CBX8-containing cPRC1 and impair Polycomb-dependent gene silencing.

DISCUSSION

Here, we took advantage of a synthetic cellular reporter assay of CBX8-dependent repression and an iterative, structure-guided approach to design a potent allosteric modulator, UNC7040, and a matched weak allosteric negative control, UNC7263. Strikingly, while these compounds have similar *in vitro* affinity for CBX8, their cellular activity differs dramatically. This difference can be explained by UNC7040's enhanced ability to stabilize non-specific binding of CBX8 to nucleic acids as a PAM.

Application of UNC7040 in two distinct cancer types demonstrated rapid and efficient displacement of CBX8-containing PRC1 from chromatin marked with H3K27me3. In contrast, CBX8 and RING1B remained largely unchanged at non-canonical targets. By blocking interaction with H3K27me3, UNC7040 offers a powerful new molecular tool to investigate the distinct functions of CBX8 in Polycomb-dependent and -independent gene regulation. The combinatorial nature of PRC1 assembly and its partial redundancy have hampered functional dissection of PRC1 subunits using classical genetic approaches. Adding to a chemical biology tool kit, including small-molecule-induced degraders of PcG proteins (Dobrinić et al., 2020; Hsu et al., 2020; Ma et al., 2020; Potjewyd et al., 2020; Zepeda-Martinez et al., 2020), UNC7040 provides an opportunity for selective, acute, and reversible inactivation of CBX8 reader activity to distinguish its role in cPRC1 regulation. Furthermore, a CBX8-specific PAM chemical probe is highly relevant for translational studies to explore the potential of CBX8 as a target for cancer therapy. Within the framework of the Structural Genomics Consortium chemical probes effort (Müller et al., 2018), we will make UNC7040 available to the expert biomedical community for use in *in vitro* and cellular disease models.

STAR★METHODS

RESOURCE AVAILABILITY

Lead contact—Requests for further information or reagents should be directed to the lead contact and corresponding author, Oliver Bell (oliver.bell@med.usc.edu)

Materials availability—Compounds, plasmids and cell lines generated in this study will be made available on reasonable request and with Material Transfer Agreement.

Data and code availability—Genomic data sets generated in this study have been deposited at NCBI under GEO accession number: GSE182592 and are publicly available as of the date of publication. This paper does not report original code. Any additional information required to reanalyze the data reported in this paper is available from the lead contact upon request.

EXPERIMENTAL MODEL AND SUBJECT DETAILS

Generation of polycomb in-vivo assay in mouse embryonic stem cell (mESC) and culture conditions—CBX8 reporter mESCs with 12xZFHD1, 4xGAL4 and 7xTETO DNA binding sites upstream of a Puromycin-GFP reporter gene were generated previously described dual reporter mESCs (Moussa et al., 2019) by recombinase mediated cassette exchange. mESCs were cultivated without feeders in high-glucose-DMEM (Sigma, D6429) supplemented with 13.5% fetal bovine serum (Gibco), 10 mM HEPES pH 7.4 (Corning, 25–060-CI), 2 mM GlutaMAX (Gibco, 35050–061), 1 mM Sodium Pyruvate (Gibco, 11360–070), 1% Penicillin/Streptomycin (Sigma, P0781), 1X non-essential amino acids (Gibco, 11140–050), 50 mM β-mercaptoethanol (Gibco, 21985–023) and recombinant LIF, and incubated at 37°C and 5% CO_2. mESCs were passaged every 48 hours by trypsinization in 0.25% Trypsin-EDTA (1X) (Gibco, 25200–056) and seeding of 2.0×10^6 cells on a 10 cm tissue culture plate (Genesee Scientific, #25–202).

Culturing of cancer cell lines—Human DLBCL cells, SUDHL4 (ATCC #CRL-2957), were cultured in RPMI-1640 complemented with 10% fetal bovine serum (Gibco) and 1% penicillin/streptomycin. The human colorectal cell lines HCT116 and HT-29 (ATCC #CCL-247 and #HTB-38, respectively) were cultured in McCoy's 5A medium supplemented with 10% fetal bovine serum (Gemini) and 1% penicillin/streptomycin (Gemini). Caco2 (ATCC #HTB-37) were maintained in EMEM supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. LoVo (ATCC #CCL-229) were maintained in F-12K medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin. The DLD-1 cell line (gift from Dr. Yun at Baylor College of Medicine) was cultured in McCoy's 5A medium supplemented with 10% fetal bovine serum and 1% penicillin/streptomycin.

Bacterial strains—DH5α and Stabl3 competent cells were purchased from Thermo Fisher Scientific and used for plasmid transformation and propagation based on manufacturer's instructions.

METHOD DETAILS

Protein expression and purification

Expression constructs: The chromodomains of CBX2 (residues 9–66 of NP_005180), CBX4 (residues 8–65 of NP_003646), CBX6 (residues 8–65 of NP_055107), CBX7 (residues 8–62 of NP_783640) and CDYL2 (residues 1–75 of NP_689555) were expressed with C-terminal His-tags in pET30 expression vectors. The chromodomain of CBX5 (residues 18–75 of NP_036429) and MPP8 (residues 55–116 of NP_059990) were expressed with a N-terminal His-tag in a pET28 expression vector. The chromodomain of CBX8 (residues 8–61 of NP_065700) was expressed with either a N-terminal His-tag in a pET28 expression vector, a N-terminal His-tag in a pET30 expression vector, or a N-terminal GST-tag in a pGEX derived expression vector.

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

His-tagged proteins were purified by re-suspending thawed cell pellets in 30 ml of lysis buffer (50 mM sodium phosphate pH 7.2, 50 mM NaCl, 30 mM imidazole, 1× EDTA free protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)) per liter of culture. Cells were lysed on ice by sonication with a Branson Digital 450 Sonifier (Branson Ultrasonics, Danbury, CT) at 40% amplitude for 12 cycles with each cycle consisting of a 20 s pulse followed by a 40 s rest. The cell lysate was clarified by centrifugation and loaded onto a HisTrap FF column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 10 column volumes of binding buffer (50 mM sodium phosphate, pH 7.2, 500 mM NaCl, 30mM imidazole) using an AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was washed with 15 column volumes of binding buffer and protein was eluted in a linear gradient to 100% elution buffer (50 mM sodium phosphate, pH 7.2, 500 mM NaCl, 500

mM imidazole) over 20 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2 ml in Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 75 prep grade column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 1.2 column volumes of sizing buffer (25 mM Tris, pH 7.5, 250 mM NaCl, 2 mM DTT, 5% glycerol) using an ATKA Purifier (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2 ml/min collecting 3-ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Protein was exchanged into a buffer containing 25 mM Tris, pH 7.5, 150 mM NaCl, 2 mM β-mercaptoethanol before use in ITC.

GST-tagged CBX8 was purified by re-suspending thawed cell pellets in 30 ml of lysis buffer (1× PBS, 5 mM DTT, 1× EDTA free protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)) per liter of culture. Cells were lysed on ice by sonication as described for His-tagged proteins. Clarified cell lysate was loaded onto a GSTrap FF column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 10 column volumes of binding buffer (1× PBS, 5mM DTT) using a AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was washed with 10 column volumes of binding buffer and protein was eluted in 100% elution buffer (50 mM Tris, pH 7.5, 150 mM NaCl, 10 mM reduced glutathione) over 10 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2 ml in Amicon Ultra-15 concentrators, 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 1.2 column volumes of sizing buffer (25 mM Tris, pH 7.5, 500 mM NaCl, 2 mM DTT, 5% glycerol) using an ATKA FPLC (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2 ml/min collecting 3 mL fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

Affinity tag removal: The N-terminal GST-tag was removed from CBX8 proteins by TEV protease cleavage according to manufacturer's recommendations (Sigma-Aldrich, St. Louis, MO). Briefly, purified protein was incubated with TEV protease at a final concentration of 50 units TEV protease per milligram tagged protein for 16 h at 4°C. The cleavage reaction was loaded onto a HiLoad 26/60 Superdex 75 to separate tag free CBX8 from GST and any protein that still retained the GST-tag. Size exclusion was performed as described above. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure tag free CBX8 protein were pooled and concentrated using Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Protein was exchanged into a buffer consisting of 20mM MES pH 6.5, 250mM NaCl, 1mM DTT.

TR-FRET assay—TR-FRET assay was performed as described in Rectenwald et al. (Rectenwald et al., 2019). A stock solution of 10X Kme reader buffer (200 mM Tris pH 7.5, 1500 mM NaCl, and 0.5% Tween 20) was prepared, 0.2 μm filtered, stored at room temperature, and was used throughout. Assays were completed using freshly made Kme reader buffer containing 20 mM Tris pH 7.5, 150 mM NaCl, 0.05% Tween 20, and 2 mM dithiothreitol (DTT). White, low-volume, flat-bottom, non-binding, 384-well microplates (Greiner, #784904) were used for assay development and screening with a total assay volume of 10 μL. 384-well, V-bottom polypropylene plates (Greiner, #781280) were used for compound serial dilutions and for transfer of assay mixtures. For compounds with stock solutions in water, serial dilutions were made using Kme reader buffer. For compounds stored in DMSO, serial dilutions were made using DMSO. Following addition of all assay components, plates were sealed with clear covers, gently mixed on a tabletop shaker for 1 minute, centrifuged at 1000 × g for 2 minutes, and allowed to equilibrate in a dark space for one hour before reading. Measurements were taken on an EnVision® 2103 Multilabel Plate Reader (Perkin Elmer) using an excitation filter at 320 nm and emission filters at 615 nm and 665 nm. 615 nm and 650 nm emission signals were measured simultaneously using a dual mirror at D400/D630. TR-FRET output signal was expressed as emission ratios of acceptor/donor (665/615 nm) counts. Percent inhibition was calculated on a scale of 0% (i.e., activity with DMSO vehicle only) to 100% (100 μM UNC3866) using full column controls on each plate. The interquartile mean of control wells was used to calculate Z' values. For dose-response curves, data was fit with a four-parameter nonlinear regression analysis using GraphPad Prism 7.0 or ScreenAble software to obtain IC_{50} values.

For testing compounds in higher throughput, 384-well assay ready plates were prepared in standard plate format: columns 1 and 2 were used for low signal controls (100% inhibition with competitor compound), columns 23 and 24 were used for high signal controls (DMSO only), and columns 3–22 were used for 25 μM single-dose test compounds. First, controls were added to a mother plate where columns 1 and 2 were filled with 10 mM stock of UNC3866 in DMSO and columns 23 and 24 were filled with DMSO. Test compounds were dispensed across the mother plate at 100X (10 mM) concentration in columns 3–22 using a TECAN Freedom EVO liquid handling workstation. Using a TTP Labtech Mosquito® HTS liquid handling instrument, assay ready plates were stamped by stamping 100 nL of control compound into columns 1 and 2, 25 nL of compounds from the mother plate into columns 3–22, and 25 nL of DMSO into columns 23 and 24. Protein, biotinylated tracer ligand, and the TR-FRET reagents were added together and gently mixed by pipetting and rocking. 10 μL was then added to each well of an assay ready plate using a Multidrop Combi (ThermoFisher). Percent inhibition was calculated on a scale of 0% (i.e., activity with DMSO vehicle only) to 100% (100 μM UNC3866) from the full column controls on each plate.

CBX8-GFP reporter assay—CBX8-GFP reporter containing mESCs were cultivated without feeders in high-glucose-DMEM (Sigma, D6429) supplemented with 13.5% fetal bovine serum (Gibco), 10 mM HEPES pH 7.4 (Corning, 25–060-CI), 2 mM GlutaMAX (Gibco, 35050–061), 1 mM Sodium Pyruvate (Gibco, 11360–070), 1% Penicillin/Streptomycin (Sigma, P0781), 1X non-essential amino acids (Gibco, 11140–050),

50 mM β-mercaptoethanol (Gibco, 21985–023) and recombinant LIF, and incubated at 37°C and 5% CO_2. mESCs were passaged every 48 hours by trypsinization in 0.25% Trypsin-EDTA (1X) (Gibco, 25200–056) and seeding of 2.0×10^6 cells on a 10 cm tissue culture plate (Genesee Scientific, #25–202).

For the screening, compounds were prepared at 10 mM top concentration from 10 mM DMSO stocks as a 10-point and 2-fold dilution series. 0.5 mL of each serial-dilution was then added to a 384-well assay plate (Corning, 3764) in triplicate using a TTP Labtech Mosquito HTS liquid handling instrument. mESCs were trypsinized in 0.25% Trypsin-EDTA, counted on a BioRad TC20 cell counter, and diluted to a density of 2,000 cells/50 mL. 50 mL of cell suspension per well was then plated on top of previously added compound stocks to achieve a final 1X compound concentration + 0.99% DMSO, and assay plates were incubated for 48 hours at 37°C and 5% CO_2. For testing the methyl ester version compounds, compounds were prepared at 10X concentration from 10 mM DMSO stocks as a 15-point, 1.5-fold dilution series, diluted into mESC media + 0.6% DMSO. 5 μL of each 10X stock was then added to a 384-well assay plate (Corning, 3764) in triplicate. mESCs were trypsinized in 0.25% Trypsin-EDTA, counted on a BioRad TC20 cell counter, and diluted to a density of 2,000 cells/45 μL. 45 μL of cell suspension per well was then plated on top of previously added 10X compound stocks to achieve a final 1X compound concentration + 0.06% DMSO, and assay plates were incubated for 48 hours at 37°C and 5% CO_2.

After 48 hours, cells were washed once in 50 mL of 1X PBS and trypsinized with 12.5 mL of clear 0.25% Trypsin-EDTA (5X) (Gibco, 15400–054) per well. Cells were incubated at 37°C and 5% CO_2 for 15–20 min to ensure in complete dislodging of cells from the assay plate. Trypsin was then quenched with 12.5 mL of 50% FBS in 1X PBS. Flow cytometry was completed on an Attune NxT equipped with Attune NxT v3.1 acquisition software. Live, single cells were gated for GFP expression and data analysis was completed with FlowJo and GraphPad Prism 8 software.

Cell toxicity assay—The effect of DMSO and methyl ester version compounds (compound **34**–**38**) on cell viability was determined using a CellTiter-GloTM ATP detection system (Promega #7573). For testing DMSO tolerance, ten point, 1:0.8 and 1:0.5 dilution curves of DMSO starting 5 μM final concentration in mECS media were prepared. For testing compound tolerance, ten point, 1:2 dilution curves of compounds starting at 100 μM final concentration were diluted to 5X final concentration in mESC media. Then, 5 μL were added to 384-well white, clear bottom tissue culture plates (Corning #3707) with a Multimek automated liquid handling device (Nanoscreen, Charleston, SC). mESCs were trypsinized in 0.25% Trypsin-EDTA, counted on a BioRad TC20 cell counter, and diluted to a density of 5,000 cells/20 μL. 20 μL of cell suspension per well was then plated on top of previously added 5X DMSO or compound stocks to achieve a final 1X DMSO or compound concentration, and assay plates were incubated for 48 hours at 37°C and 5% CO_2. After 48 hours, cells were lysed with 25 μL of CellTiter-GloTM reagent. Luminescence was read on an Envision platereader (Perkin Elmer) after 15 minutes at room temperature in dim light.

TR-FRET CBX8/dsDNA/compound ternary complex assay—All protein, biotin-dsDNA, and compound stocks were diluted to desired final concentrations in TR-FRET assay buffer (20 mM Tris-HCl, pH 7.5, 10 mM NaCl, 0.01% Tween-20, 2 mM DTT) and all experiments were conducted in white, low-volume, flat-bottom, non-binding, 384-well assay plates (Grenier, No: 784904). For CBX8 CD titration experiments, an eleven-point, two-fold dilution series was generated starting at 1.25 μM CBX8 CD. CBX8 CD was incubated with 5 μM compound, 10 nM biotin-dsDNA, 50 nM Lance *Ultra* U*Light* anti-6X-histidine antibody, and 2 nM Lance Eu-W8044 steptavidin conjugate at a final reaction volume of 10 μL (0.05% DMSO). Following addition of all assay components, plates were sealed with clear covers, gently mixed on a tabletop shaker for 1 minute, centrifuged at 1000 × g for 2 minutes, and allowed to equilibrate in a dark space for 1 hour before reading. Measurements were taken on an EnVision® 2103 Multilabel Plate Reader (Perkin Elmer) using an excitation filter at 320 nm and emission filters at 615 nm and 665 nm. 615 nm and 650 nm emission signals were measured simultaneously using a dual mirror at D400/D630. TR-FRET output signal was expressed as emission ratios of acceptor/donor (665/615 nm) counts. Data was plotted using GraphPad Prism 9.0.

NMR experiments

Protein expression and purification for NMR studies: The recombinant CBX8-CD was expressed in BL21 (DE3) (New England Biolabs) *Escherichia coli* cells. Cells were grown in LB-medium or M9-minimal media supplemented with ^{15}N-NH_4Cl or ^{15}N-NH_4Cl and ^{13}C-glucose. For unlabeled protein, cells were grown shaking at 215 rpm at 37°C until an OD ~1.0 was reached and induced with 1 mM IPTG for 16–18 hr overnight. For isotopically-enriched protein, cells were grown in LB-medium until an OD ~1.0, spun down at 4000 rpm for 10 minutes, and resuspended in M9-medium (4 l LB cells per 1 l M9) supplemented with either ^{15}N-NH_4Cl or ^{15}N-NH_4Cl/^{13}C-glucose. The cells were allowed to recover in M9 media for 1 hr shaking at 18°C and induced with 1.0 mM IPTG for 16–18 hr overnight. Cells were subsequently collected by centrifugation at 6000 rpm for 20 minutes, frozen in N_2(l) and stored at −80°C.

For purification, cells were resuspended in a buffer containing 100 mM NaCl, 25 mM Tris (pH 7.5) supplemented with DNase I and lysozyme. Cells were then lysed using an Emulsiflex homogenizer (Avestin) or by sonication, and lysate cleared by centrifugation at 15 000 rpm for 1 hr at 4°C. The soluble supernatant was incubated with glutathione agarose resin (ThermoFisher Scientific) rotating at 4°C for 1 hr. The GST-tagged CBX8 CD was washed extensively with a high salt buffer containing 1 M NaCl and 25 mM Tris (pH 7.5) before elution with a buffer containing 150 mM NaCl, 25 mM Tris (pH 7.5) and 50 mM reduced glutathione. The GST-CBX8 CD was concentrated using a 3000 MWCO centrifugation filter unit to 2 ml and cleaved with TEV protease for 3 hr at 25°C. The cleaved CBX8 CD was further purified using a combination of cation-exchange and size exclusion chromatography (Superdex 75, GE Healthcare Life Sciences). For NMR studies, ^{15}N-CBX8 CD and ^{15}N/^{13}C-CBX8 CD were used in a final NMR buffer containing 100 mM NaCl and 40 mM phosphate buffer (pH 6.8). For EMSAs, the unlabeled CBX8 CD was used in a final buffer containing 25 mM phosphate buffer (pH 6.8), 25 mM NaCl, 1 mM EDTA, 1 mM DTT.

NMR experiments: Titrations experiments were performed by collecting ^{15}N-HSQC spectra on 0.05 mM ^{15}N-CBX8-CD in the apo state and upon addition of increasing concentrations of ligand. Titrations experiments were performed at 25°C on a Bruker Avance II 800 MHz NMR spectrometer equipped with a 5 mm triple resonance cryoprobe. The data was processed using NMRPipe (Delaglio et al., 1995) and further analysis performed using CcpNmr (Vranken et al., 2005). To determine residues with CSPs, the following equation was used:

$$\Delta\delta = (\Delta\delta H)^2 + (0.20\Delta\delta N)^2$$

A resonance was considered significantly perturbed if the $\Delta\delta$ value was greater than the average plus 0.5 standard deviations after trimming 10% of residues with the largest $\Delta\delta$ value.

CAPA assay—HeLa cells stably expressing the HaloTag-GFP-Mito construct were provided by the Kritzer lab (Peraro et al., 2017, 2018). Cells were cultured in high-glucose-DMEM (Sigma, D6429) supplemented with 10% fetal bovine serum, 1% Penicillin/Streptomycin (Sigma, P0781) and 1 mg/mL Puromycin (InvivoGen, ant-pr) and incubated at 37°C and 5% CO_2. Cells were passaged every 48–72 hours by trypsinization in 0.25% Trypsin-EDTA (1X) (Gibco, 25200–056) and seeding of 3.0×10^6 cells on a T75 tissue culture plate.

GFP-HaloTag HeLa cells were seeded in a 384-well assay plate (Corning, 3764) at a density of 5,000 cells/well on the day before the experiment and allowed to adhere overnight at 37°C and 5% CO_2. A 100X mother plate of compound dilutions was prepared in a separate 384-well plate, and 10-points and 3-fold dilution series were generated using 10 mM DMSO stock. Compound-free DMSO control wells were also included in the mother plate to be used as no-pulse (100%) and no-pulse/no-chase (0%) controls. On the day of the experiment, a daughter plate of compound dilutions and control samples was prepared by transferring 0.5 μL of each well from the mother plate using a TTP Labtech Mosquito HTS liquid handling instrument. Then, the wells were diluted with 49.5 μL of HeLa media to ensure a final DMSO concentration of 1%. The media of the 384-well assay plate containing cells was aspirated and 50 μL of the compound and control solutions from the daughter plate were added to each well. The assay plate was then incubated at 37°C and 5% CO_2 for 4 hours. Media was removed, and cells were washed once with phenol-red free Opti-MEM (1X) (Gibco, 11058–021) and incubated at 37°C and 5% CO_2 for 30 minutes. Media was again removed and replenished with 30 mL phenol-red free Opti-MEM supplemented with 5 mM CT-TAMRA (Promega, G8251), except for no CT-TAMRA control wells, which were replenished with 30 mL phenol-red free Opti-MEM alone. Cells were then incubated at 37°C and 5% CO_2 for 30 minutes. Media was removed and cells were washed a final time with phenol-red free Opti-MEM, this time supplemented with 10% FBS + 1% Penicillin/Streptomycin, and incubated at 37°C and 5% CO_2 for 30 minutes. Media was removed, and cells were washed once in 50 mL of 1X PBS and trypsinized with 12.5 mL of clear 0.25% Trypsin-EDTA (5X) (Gibco, 15400–054) per well. Cells were incubated at 37°C and 5% CO_2 for 15–20 minutes to ensure in complete dislodging of cells from the assay

plate. Trypsin was then quenched with 12.5 mL of 50% FBS in 1X PBS. Flow cytometry was completed on an IntelliCyt iQue Screener PLUS equipped with ForeCyt acquisition software. Live, single cells were gated first for GFP expression, and GFP positive cells were then analyzed for mean fluorescence intensity of CT-TAMRA dye by double normalization to a no dye sample (0% red signal) and dye only sample (100% red signal). Data analysis was completed with FlowJo and GraphPad Prism 8 software.

Surface plasmon resonance experiments—SPR experiments were performed on a BioRad ProteOn XPR36 Interaction Array System. All compound stock solutions were diluted to desired final concentrations in SPR Buffer (20 mM Tris-HCl, pH 7.0, 150 mM NaCl, 0.005% Tween-20), and protein stock solutions were diluted into SPR Buffer supplemented with 1 mg/mL BSA. Biotinylated derivative of UNC7040, compound 39 (UNC7045) was made up as 150 nM stocks in SPR buffer and immobilized at a flow rate of 30 μL/min and a contact time of 60 s onto a NeutrAvidin-containing ProteOn NLC sensor chip. Following a 30 min buffer blank in which SPR Buffer was switched to buffer supplemented with 1 mg/mL BSA, proteins (CBX2, CBX4, CBX6, CBX8, CDYL2) were flowed at a rate of 50 μL/min with a contact time of 200 s and a dissociation time of 800 s. Regeneration of the sensor chip in 0.1% SDS/5mM NaOH was completed between each protein sample at a flow rate of 30 μL/min for 120 s. Double referencing subtraction was done with buffer and protein blank channels to account for nonspecific binding to the sensor chip. Data were fit to a two-state binding model in which ka and kd parameters were fit as grouped, ka2, kd2, and RI parameters were fit locally, and all other parameters were fit globally.

Protein domain microarray analysis—The protein domain microarray was conducted by the Protein Array and Analysis Core (PAAC) at MD Anderson Cancer Center. A comprehensive library of human protein domains that potentially read methyllysine marks was cloned into a pGEX vector by Biomatik (Cambridge, Canada) using gene synthesis to best optimize the open reading frames for bacterial expression. Escherichia coli was used to express the protein domains as GST fusions, which were purified using glutathione-Sepharose beads. The recombinant domains were arrayed onto nitrocellulose-coated glass slides (Oncyte®Avid slides, Grace Bio-Labs, Bend, OR), using an Aushon 2470 pin microarrayer, as previously described (Espejo et al., 2002). Biotinylated compounds were pre- conjugated to streptavidin-Cy3 to generate fluorescent probes. Following probing, fluorescent signal was detected with a GenePix 4200A Microarray Scanner (Molecular Devices).

Molecular dynamics and docking—Molecular dynamics (MD) simulations for all three systems ($CBX8_{CD}$, $CBX8_{CD}$:**21**, and $CBX8_{CD}$:**22**) were performed using the Gromacs 2018.2 simulation package with CHARMM22 protein force field (Vanommeslaeghe et al., 2010). The 3D structures of $CBX8_{CD}$ in complex with **21** and **22** were built by analogy from the crystal structure of CBX8 in complex with a chemically similar ligand UNC3866 (PDB: 5EQ0) (Stuckey et al., 2016a) using the Maestro modeling suite (release 2016–2, Schrödinger, LLC: New York, NY). Both above structures served as starting points for MD simulations. CHARMM22 force field parameters for **21** and **22** were generated by

Swissparam (Zoete et al., 2011). End caps were added to both termini of each protein. The protein complex was minimized in vacuum using steepest decent algorithm for 5,000 steps or until the maximum force of 1,000 $kJ*mol^{-1}*nm^{-1}$ was reached. The molecular systems were then solvated in TIP3P water (Mark and Nilsson, 2001), counterions were added to ensure the systems' electric neutrality, and NaCl ions (0.15 M) were added by randomly replacing certain water molecules in order to mimic physiological conditions. An energy minimization with solvent was then performed, followed by a two-step equilibration: 5 ns in NVT ensemble at 310 K using the modified Berendsen thermostat and 5 ns in NPT ensemble at 1 atm (and 310 K) using the Parinello-Rahman pressure coupling (Nosé and Klein, 1983). All simulations were conducted using the Leapfrog integrator in periodic boundary conditions. The particle mesh Ewald algorithm (Essmann et al., 1995) controlled the long-range electrostatic interactions. Bonds involving hydrogen atoms were constrained using the linear constraint solver algorithm (LINCS) (Hess et al., 1997). The production simulations were performed in NVT ensemble. For each of the three systems, three independent ~5 μs MD simulations were run. Molecular visualizations were produced using Maestro [Schrodinger, LLC]. MD trajectories were clustered and analyzed by means of the Pipeline Pilot data processing environment (v. 18.1.100.11, BIOVIA, 3dsbiovia.com). The input data (sets of the protein's atomic coordinates and the backbone φ and ψ angles) were generated from the MD trajectories using custom Pipeline Pilot scripts (protein structures were centered and aligned using the Gromacs trjconv tool). The clustering technique used was *k*-means with Euclidian distance metrics. The cluster aggregation criteria were chosen so that root mean square distances (RMSD) between the cluster members would be on the order of 1Å.

Protein-DNA docking calculations were performed using the HADDOCK web service (van Zundert et al., 2016; Wassenaar et al., 2012). Twelve protein structures were selected for docking from the MD trajectories of the two simulated systems ($CBX8_{CD}$:**21** and $CBX8_{CD}$:**22**; 6 structures per systems). All selected structures were centroids of the most populated conformational clusters. The 3D structure of the DNA double helix (35 base pairs) was generated using the Discovery Studio 4.0 modeling suite (www.3dsbiovia.com). A set of default HADDOCK parameters was used for all docking simulations. The parameter file and all input and output HADDOCK files are available upon request.

DLBCL cell proliferation assay—SU-DHL-4 cells were seeded at a cell density of 1×10^5 cells/mL in a 24 well plate (Costar®, 3526), with 0.5 mL of cells per well. The vehicle control treated cells were dosed with 0.4% water or DMSO. Every 48 hours, the cells were counted on an automated Bio-Rad TC20™ cell counter with Trypan blue (Abcam, ab233465) and cell counting slides (1450015) to give the cell count (cells/mL) and cell viability (%). At each time point, the cells were split back to the original seeding density with fresh media, and the cells were re-dosed with compound or vehicle. The percent cell proliferation of cells treated with compound was calculated based on the total cell number expressed as split-adjusted viable cells, relative to the vehicle control at the same time point. The data is reported as the average of between 4 to 6 replicates ± standard deviation.

CC cell proliferation assays—HCT116, Caco2, LOVO and DLD-1 were seeded in 96-well Cell Carrier plates (Perkin Elmer) at 1500, 2000, 1200 and 1000 cells per well, respectively. The day after seeding (D0), cells were treated with various concentrations of UNC7263, UNC7040, UNC4976 or vehicle. Live and dead cell counts were measured daily over a 3-day period using the Operetta High Content Screening system (PerkinElmer). Prior to imaging, cells were stained with 5 μg/ml of Hoechst 33342 (nuclear dye) (Life Technologies) and 5 μg/ml of propidium iodine (vital dye) (Life Technologies) for 30 minutes. Cells were then segmented based upon the nuclear dye using the Harmony software (PerkinElmer). Propidium iodide intensity levels were calculated, and cells were classified as 'dead' if their intensity was above the established threshold.

CC colony assay formation—To test the effect of CBX 8 inhibitors on the cell proliferation, we performed colony assays. For colony assay, HCT116 or LOVO cells were plated at the density of 500 cells per well in six well plates. After 24hours, cells were treated with 10 or 20 mM UNC7040 or UNC7263. Depending on cell lines, after 1 to 2 weeks colonies were fixed with 100 % methanol for 20 minutes at room temperature, followed by rinsing with water. Next, colonies were stained with 0.5% crystal violet in 25 % methanol for 5 minutes at room temperature. Cells were washed with water to remove excess dye. Plates were inverted and left overnight for drying. Colonies were viewed using bright field microscopy. A non-overlapping cluster of minimum 50 cells were counted as a colony. All the treatments were done in triplicates.

Chromatin immunoprecipitation and quantitative PCR (ChIP-qPCR) and next-generation sequencing (ChIP-seq)—25×10^6 mES cells were collected, washed in once in 1x PBS and crosslinked with formaldehyde at a final concentration of 1 % for 7 min. The crosslinking was stopped on ice and with glycine at final 0.125 M concentration. The crosslinked cells were pelleted by centrifugation for 5 min at 1200g at 4°C. Nuclei were prepared by washes with NP-Rinse buffer 1 (final: 10 mM Tris pH 8.0, 10 mM EDTA pH 8.0, 0.5 mM EGTA, 0.25% Triton X-100) followed by NP-Rinse buffer 2 (final: 10 mM Tris pH 8.0, 1 mM EDTA, 0.5 mM EGTA, 200 mM NaCl). Afterwards the cells were prepared for shearing by sonication by two washes with Covaris shearing buffer (final: 1 mM EDTA pH 8.0, 10 mM Tris-HCl pH 8.0, 0.1% SDS) and resuspension of the nuclei in 0.9 mL Covaris shearing buffer (with 1× protease inhibitors complete mini (Roche)). The nuclei were sonicated utilizing 15 ml Bioruptor tubes (Diagenode, C01020031) with 437.5 mg sonication beads (Diagenode, C03070001) for 6 cycles on a Bioruptor Pico sonicator. Lysates were incubated in 1× IP buffer (final: 50 mM HEPES/KOH pH 7.5, 300 mM NaCl, 1 mM EDTA, 1% Triton X-100, 0.1% DOC, 0.1% SDS), with following antibodies at 4°C on a rotating wheel: H3K27ac (Abcam, ab4729), H3K27me3 (Diagenode, C15410195), Ring1B (Cell Signaling, D22F2), Cbx2 (Abcam, ab80044), Cbx6 (Santa Cruz Biotechnology, sc-3930400), Cbx7 (Abcam, ab21873) and Cbx8 (Bethyl Laboratories, A300–882A). ChIPs were washed 5x with 1x IP buffer (final: 50 mM HEPES/KOH pH 7.5, 300 mM NaCl, I mM EDTA, 1% Triton-X100, 0.1% DOC, 0.1% SDS), or 1.5× IP buffer for H3K27me3, followed by 3x with DOC buffer (10 mM Tris pH 8, 0.25 mM LiCl, 1 mM EDTA, 0.5% NP40, 0.5% DOC) and 1x with TE (+50 mM NaCl).

qPCR analysis—The PCIA extracted IP DNA was precipitated and quantified using a homemade EvaGreen based qPCR mix on a CFX Connect Real-Time PCR Detection System (BioRad). qPCR primers are listed in the Oligonucleotide Section of the Key Resources Table.

cChIP-seq library preparation—Libraries were prepared using the NEXTflex ChIP-Seq kit (Bio Scientific) following the "No size-selection cleanup" protocol. Each sample of ChIPed DNA was end-repaired and ligated to unique barcoded adaptors to produce individual libraries. Libraries corresponding to samples to be directly compared to each other (e.g. DMSO vs UNC7040) were pooled together and purified using 1 volume of Agencourt AMPure XP (Beckman Coulter). The pooled libraries were eluted with 23 μL of elution buffer (NEXTflex ChIP-Seq kit) and amplified using the KAPA Real-Time Library Amplification Kit (KAPABiosystems) following the kit instructions. Finally, the amplified libraries were size-selected and purified using 0.9x volume of Agencourt AMPure XP (Beckman Coulter). Library quality control including determination of average size and concentration was performed prior to sequencing by commercial Next Generation Sequencing providers. NGS libaries were eventually sequenced as 150 bp paired-end reads on the Illumina HiSeq platform.

cChIP-seq data analysis

Processing and mapping of raw reads: The raw reads of cChIP-seq were mapped to the custom concatenated human (hg38) and spike-in mouse (mm10) genome sequences using bowtie 2 with "–no-mixed" and "no-discordant" options (Langmead and Salzberg, 2012). Subsequently, low quality reads were removed using SAMtools (Li et al., 2009), duplicated reads were discarded with the Picard toolkit (http://broadinstitute.github.io/picard/) and only unique mapped reads were retained for subsequent analysis.

Data visualization: For visualization uniquely mapped human reads were normalized by random subsampling using calibration factors calculated from the corresponding mm10 spike-in reads as described previously (Fursova et al., 2019) using samtools. High correlation between replicates was confirmed using multiBamSummary and plotCorrelation functions from deepTools (Ramirez et al., 2016) before merging for visualization and downstream analysis. Genome coverage tracks (bigWig files) were generated with MACS2′ pileup function (Zhang et al., 2008) and heatmaps and profile plots were generated with deepTools (Ramirez et al., 2016).

Peak calling: Peaks were called on each replicate independently using MACS2 (Zhang et al., 2008) with -broad option and q-value cutoff of 0.1. Only peaks called in both replicates were retained for downstream analysis.

RNA-seq library preparation—For RNA-seq sample preparation, 5–10×10^6 SUDHL-4 or LoVo cells (untreated, UNC7263 or UNC7040 treated in triplicates) were trypsinized and collected by centrifugation for 5 min at 500 g. Subsequently, cell pellets were washed 1× with PBS and collected in 1X DNA/RNA protection reagent (Monarch Total RNA Miniprep Kit, NEB). Cells were lysed and total RNA was isolated following the mammalian

cell protocol inlcuding on-column DNase I treatment. Total RNA samples were submitted to Novogene Co. for quality control and library preparation, applying poly-A enrichment and using the NEBNext Ultra II RNA Library Prep kit (NEB), indexed using NEBNext Multiplex Oligos (NEB). Final libraries were sequenced as 150 bp paired-end reads on the Illumina HiSeq platform.

RNA-seq data analysis—Raw paired RNA-seq reads were aligned to the hg38 genome sequences using STAR-2.6.1c (Dobin et al., 2013). Overlap of aligned reads in bam format with genes was performed using HTSeq count function (Anders et al., 2015) with stranded = no option and the GRCh38 version 101 GTF file. The HTseq count matrix was pre-filtered to remove any genes with less than 10 reads. Differential gene expression analysis was performed using DESeq2 (Love et al., 2014) using “apeglm” method (Zhu et al., 2019) for LFC shrinkage. We applied a threshold of p-adj < 0.05 and fold change >0.5 or −0.5 to consider gene expression changes significant. Log2 fold change values were visualized in volcano plots using R, ggplot2 and GraphPad prism. To assign intergenic cis-regulatory regulatory elements identified by cChIP-seq with associated genes we utilized the Genomic Regions Enrichment of Annotations Tool (GREAT) (McLean et al., 2010).

Co-immunoprecipitation—Whole cell protein extract from CBX8 reporter mESCs (45 × 10^6) was obtained by lysis in 500 µl lysis buffer (final: 20 mM Tris-HCl pH 7.5, 150 mM NaCl, 2 mM $MgCl_2$, 10% Glycerol, 1 mM DTT, 1 mM PMSF, 0.2% NP-40, 1× Complete Mini protease inhibitor). Lysate was homogenized by sonication in a Diagenode Bioruptor Pico followed by rotation for 3 h at 4°C. After 30 min centrifugation at 4°C protein concentration of the lysate was determined by Bradford assay (Biorad). In parallel, 30 µl Protein G coupled Dynabeads (Thermo Fisher Scientific) were prepared for each IP reaction as follows: 3x wash in lysis buffer, incubation with 1.5 µg FLAG antibody (Sigma Aldrich, F1804) for 3 h at 4°C, 1× wash in lysis buffer and finally resuspension in 30 µl of lysis buffer. For each IP, 30 µl of pre-bound Dynabeads were incubated with 1 mg protein extract in a final volume of 500 µl overnight at 4°C. Finally, beads were washed four times with lysis buffer and proteins were eluted at 95°C in SDS sample buffer and analyzed by Western blot.

Co-immunoprecipitation samples and corresponding 2% Input samples were run on Novex Life Technology NuPAGE 4–12% Bis-Tris gels in Invitrogen NuPAGE MES SDS Running Buffer and transferred on a Merck Chemicals Immobilon-FL Membrane (PVDF 0.45 µm). The membrane was blocked (5% non-fat dry milk in 1× PBS, 0.1% Tween 20) and incubated in 5% non-fat dry milk in 1× PBS and 0.1% Tween 20 with the primary antibodies (PHC1 Cell Signaling #13505 1:1000; Rybp Cell Signaling D8J7W 1:1000; SUZ12 Cell Signaling (D39F6) 1:1000; Cbx7 Merck Millipore 07–981 1:1000; Cbx8 Bethyl Laboratories A300–882A 1:1000; Ring1B Cell Signaling D22F2 1:1000). Finally, the membrane was incubated with corresponding secondary IRDye 800CW Goat anti-Rabbit IgG (H+L) (LICOR) or IRDye 680RD Goat anti-Mouse IgG (H+L) (LICOR) and imaged by an Odyssey CLx Near-Infrared Imaging System (LICOR).

General chemistry procedures—Reactions were carried out with commercially available reagents except Fmoc-lysine mimetics and room temperature was generally

22°C. Reverse phase column chromatography was performed with a Teledyne Isco CombiFlash®R_f 200 using C18 RediSep®R_f Gold columns with the UV detector set to 220 nm and 254 nm. Mobile phases of A (H_2O + 0.1% TFA) and B (methanol (MeOH) or acetonitrile (MeCN)) were used with default column gradients. Preparative HPLC was performed using an Agilent Prep 1200 series with the UV detector set to 220 nm and 254 nm. Samples were injected onto either a Phenomenex Luna 250 × 30 mm (5 μm) C18 column or a Phenomenex Luna 75 × 30 mm (5 μm) C18 column at room temperature. Mobile phases of A (H_2O + 0.1% TFA) and B (MeOH or MeCN) were used with a flow rate of 40 mL/min for the large column and 30 mL/min for the small column. A general gradient of 0–22 minutes increasing from 10 to 100% B, followed by a 100% B flush for another 8 minutes. Small variations in this purification method were made as needed to achieve ideal separation for each compound.

Analysis of products—Analytical LCMS (at 220 nm and 254 nm) was used to establish the purity of targeted compounds. All compounds that were evaluated in biochemical and biophysical assays had >95% purity as determined by LCMS. Analytical LCMS data for all compounds were acquired using an Agilent 6110 Series system with the UV detector set to 220 nm and 254 nm. Samples were injected (<10 μL) onto an Agilent Eclipse Plus 4.6 × 50 mm, 1.8 μm, C18 column at room temperature. Mobile phases A (H_2O + 0.1% acetic acid) and B (MeOH or MeCN + 0.1% acetic acid) were used with a linear gradient from 10% to 100% B in 5.0 minutes, followed by a flush at 100% B for another 2 minutes with a flow rate of 1.0 mL/min. Mass spectra (MS) data were acquired in positive ion mode using an Agilent 6110 single quadrupole mass spectrometer with an electrospray ionization (ESI) source. ^{1}H NMR spectra were obtained on a Varian automated 400MR at 400 MHz. Chemical shifts are reported in ppm and coupling constants are reported in Hz with MeOD-d_4 referenced at 3.31 (^{1}H).

Solid-phase peptide synthesis—Peptidomimetic synthesis was conducted on Fmoc Rink amide resin on polystyrene beads (50–100 mg per peptide, Anaspec) and 2-chlorotritylchloride resin (50–400 mg per compound, Chem-Impex).

C-terminal amide-containing compounds *(General Procedure A)*—For C-terminal amides, the Rink amide resin was rinsed once with dichloromethane (DCM), drained, and swollen in DCM for 5 minutes followed by equilibration in N,N-dimethylformamide (DMF) for another 5 minutes. Fmoc protecting group on Fmoc Rink amide resin was first removed in a solution of 2.5% 1,8-diazabicycloundec-7-ene (DBU) and 2.5% pyrrolidine in 8 mL of DMF for 10 minutes. Then, the resin was filtered and washed twice with DMF, methanol, DMF, and DCM (8 mL each) before adding the first amino acid for coupling. Fmoc-protected amino acids (3 eq) were preactivated with HBTU (3 eq), HOAt (3 eq), and DIPEA (10 eq) for 5 minutes with periodic swirling in 5 mL of DMF and 3 mL of DCM. (Fmoc-protected unnatural amino acids (1.2–2 eq) were preactivated with HBTU (1.2–2 eq), HOAt (1.2–2 eq), and DIPEA (5 eq) for 5 minutes with periodic swirling in 5 mL of DMF and 3 mL of DCM.) The solution was then added to the resin and left on a shaker at room temperature for 1 hour (overnight for unnatural amino acids). The resin was filtered and washed twice with DCM, DMF, methanol, and DMF again (8

ml each). Fmoc protecting groups on N-terminal amino acid were removed in a solution of 2.5% DBU and 2.5% pyrrolidine in 8 mL of DMF for 10 minutes. Then, the resin was filtered and washed twice with DMF, methanol, DMF, and DCM (8 mL each) before adding the next amino acid for coupling. Following installation of the capping residue, the resin was rinsed 10 times with DCM. Cleavage cocktail (95% trifluoroacetic acid (TFA), 2.5% triisopropylsilane (TIPS), and 2.5% water) was added to the resin, the mixture was left on the shaker for 2 hours, and the filtrate was collected. Resin was rinsed twice with DCM and filtrates were pooled then concentrated under vacuum. The mixture was dissolved in water, washed with ether, and the peptidomimetics were extracted with water (3 × 10 mL). The aqueous layers were combined then concentrated *in vacuo* and purified via preparative high performance liquid chromatography standard method (10–100% MeCN in H_2O + 0.1% TFA). The solvent was removed *in vacuo* and dissolved in water and lyophilized to yield the desired peptidomimetic products.

C-terminal methyl ester-containing compounds *(General Procedure B)*—For C-terminal methyl ester compounds, 2-chlorotritylchloride resin was swollen DCM for 10 minutes. Fmoc-Ser-OH (5 eq) was loaded onto the resin with DIPEA (10 eq) in DCM (5 mL) for 1 hour. The resin was then treated with MeOH for 15 minutes to block any unreacted loading sites. Next, the resin was rinsed three times with DCM and three times with DMF. All subsequent coupling steps, cleavage, purification, and analysis were conducted as described in the aforementioned synthesis process with the Rink amide resin.

Solution-phase peptide synthesis

N-cap modifications (General Procedure C): The N-cap modifications other than the *tert*-butyl benzoyl group were unstable in the cleavage cocktail (95% TFA, 2.5% TIPS, and 2.5% water). Compounds **8–10** and **27–28** were prepared using a combination of solid and solution phase synthesis. The peptides were synthesized on the beads up to the phenylaniline residue, and they were cleaved from the beads using the previously mentioned cleavage cocktail. After purification, the 5-mer peptides (1 eq) were mixed with the respective preactivated N-terminal acids (N-terminal acid (3 eq), HBTU (3 eq), HOAt (3 eq), and DIPEA (10 eq) for 5 minutes) in a round bottom flask, and the mixture was stirred overnight. Volatiles were removed *in vacuo*, and the crude mixture was dissolved in 80% acetonitrile and 20% water (1 mL). HPLC purification followed the procedure described according to the standard method. Fractions were concentrated on a Genevac HT-4X and lyophilized to yield the desired peptidomimetic products.

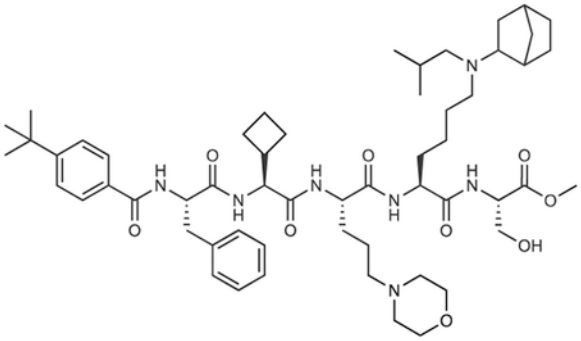
CAPA and biotinylated compounds (General Procedure D): When synthesizing CAPA compounds and a biotinylated compound, the final step is a chloroalkane-tag or Peg_{11}-biotin-tag addition. To a solution of a peptidomimetic compound (1 eq) in DMF (2 mL) was added TBTU (1.1 eq) and DIPEA (1.5 eq). In a separate flask, DIPEA (1.5 eq) was added to a solution of 2-(2-((6-chlorohexyl)oxy)ethoxy)ethan-1-amine 2,2,2-trifluoroacetate (1.2 eq) or *N*-(35-amino-3,6,9,12,15,18,21,24,27,30,33-undecaoxapentatriacontyl)-5-((3a*S*,4*S*,6a*R*)-2-oxohexahydro-1*H*-thieno[3,4-*d*]imidazol-4-yl) pentanamide (1.3 eq) in DMF (2 mL). After 5–10 minutes, the solution was added to the main reaction mixture, and the reaction

was left to stir at room temperature overnight. Volatiles were removed *in vacuo*, and the crude mixture was dissolved in 80% acetonitrile and 20% water (1 mL). HPLC purification followed the procedure described above. Fractions were concentrated on a Genevac HT-4X and lyophilized to yield the desired peptidomimetic products.

Morpholino group introduction (General Procedure E): The morpholino group containing compounds at the leucine position were synthesized on Fmoc Rink amide resin and 2-Chlorotrityl resin. Here, either an ornithine or an aminobutyric acid was incorporated into the peptides at the leucine position according to the standard procedures A or B. The resulting 6-mer peptidomimetic compounds were dissolved in acetonitrile. In a separate round bottom flask, sodium periodate (140 mg, 1 eq, 486 μmol) was dissolved in 5 mL of water. The mixture was then placed in an ice bath at 4°C and (3*R*,4*S*)-tetrahydrofuran-3,4-diol (39 μL, 1 eq, 0.48 mmol) was added, and the reaction mixture was stirred overnight. To the flask of peptidomimetic compound, was added acetic acid (6 eq). In a separate round bottom flask, 10 mL of acetonitrile was added to 2,2'-oxydiacetaldehyde solution (2 eq, 1 mg/mL in water), which was synthesized from (3*R*,4*S*)-tetrahydrofuran-3,4-diol. This solution was sonicated for 30 seconds followed by filtering through a syringe filter. The resulting filtrate was then added to the peptidomimetic compound in acetic acid. After stirring for 10 minutes at room temperature, sodium cyanoborohydride (4 eq) was added and the reaction was stirred overnight at room temperature. Volatiles were removed *in vacuo*, and the crude mixture was dissolved in 80% acetonitrile and 20% water (1 mL). HPLC purification followed the standard method. Fractions were concentrated on a Genevac HT-4X and lyophilized to yield the desired peptidomimetic products.

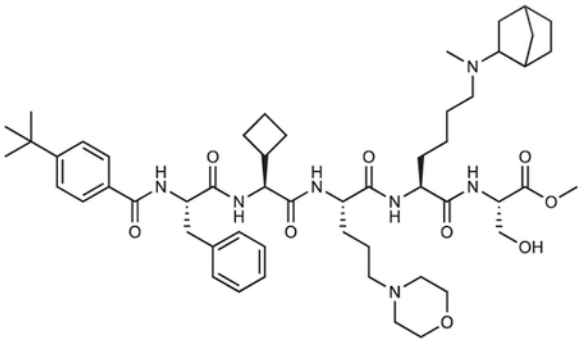
Methyl Ester formation (General Procedure F): Peptidomimetic synthesis was conducted on 2-chlorotritylchloride resin (100–200 mg per compound, Chem-Impex) followed general procedure B and a morpholino group was introduced followed general procedure E. Sulfurous dichloride (100 eq) was added dropwise to an excess of methanol (10 mL) in an ice bath at 4°C. In a separate flask, peptidomimetic compound (C-terminal acid, 1 eq) was dissolved in a small amount of methanol and added dropwise to the flask containing sulfurous dichloride. The reaction mixture was stirred for 20 minutes at 4°C, and then allowed to return to room temperature and stirred overnight. Upon complete conversion to the methyl ester product (monitored by LC/MS), volatiles were removed *in vacuo*, and the crude mixture was dissolved in 80% acetonitrile and 20% water (1 mL). HPLC purification followed the standard method described above. Fractions were concentrated on a Genevac HT-4X and lyophilized to yield the desired peptidomimetic products.

Methyl N^{6}*-(bicyclo[2.2.1]heptan-2-yl)*-N^{2}*-((*S*)-2-((*S*)-2-((*S*)-2-(4-(*tert*-butyl)benzamido)-3-phenylpropanamido)-2-cyclobutylacetamido)-5-morpholinopentanoyl)*-N^{6}*-isobutyl*-L*-lysyl*-L*-serinate (UNC7040, 34).*



Prepared according to general procedure B, E, and F using a 2-chlorotrityl Resin (200 mg) to yield the TFA salt of the desired product as a white solid (23.64 mg, 13.1%). ^{1}H-NMR (MeOD-d_4, 400 MHz) δ 7.65 (d, J = 8.2 Hz, 2H), 7.527.46 (m, 2H), 7.35–7.19 (m, 5H), 4.83–4.76 (m, 1H), 4.54–4.48 (m, 1H), 4.46–4.40 (m, 2H), 4.26–4.23 (m, 1H), 4.01–3.98 (m, 2H), 3.93 (dd, J = 11.3 Hz, 4.8 Hz, 1H), 3.81 (dd, J = 11.3, 3.9 Hz, 1H), 3.75 (s, 3H), 3.73–3.66 (m, 2H), 3.52–3.40 (m, 2H), 3.28–2.89 (m, 10H), 2.72–2.62 (m, 2H), 2.44–2.35 (m, 1H), 2.21–1.73 (m, 16H), 1.70–1.68 (m, 3H), 1.57–1.43 (m, 6H), 1.33 (s, 9H), 1.29–1.22 (m, 1H), 1.15–1.02 (m, 6H); MS (ESI): m/z calcd. for $[C_{56}H_{86}N_7O_9]^+$ 1000.64, found 1000.50.

Methyl N^6*-(bicyclo[2.2.1]heptan-2-yl)*-N^2*-((*S*)-2-((*S*)-2-((*S*)-2-(4-(*tert*-butyl)benzamido)-3-phenylpropanamido)-2-cyclobutylacetamido)-5-morpholinopentanoyl)*-N^6*-methyl*-L*-lysyl*-L*-serinate (UNC7263, 38)*.



Prepared according to general procedure B, E, and F using a 2-chlorotrityl Resin (200 mg) to yield the TFA salt of the desired product as a white solid (39.2 mg, 21.8%). ^{1}H-NMR (MeOD-d_4, 400 MHz) δ 7.65 (d, J = 8.5 Hz, 2H), 7.52–7.45 (m, 2H), 7.35–7.19 (m, 5H), 4.81 (dd, J = 9.4, 5.1 Hz, 1H), 4.53–4.50 (m, 1H), 4.45–4.40 (m, 2H), 4.27–4.23 (m, 1H), 4.01–3.98 (m, 2H), 3.92 (dd, J = 11.1, 5.0 Hz, 1H), 3.81 (dd, J = 11.3, 3.8 Hz, 1H), 3.74 (s, 3H), 3.73–3.67 (m, 2H), 3.48–3.43 (m, 3H), 3.28–2.96 (m, 8H), 2.85 (s, 3H), 2.76–2.63 (m, 2H), 2.35 (brs, 1H), 2.18–2.06 (m, 2H), 2.05–1.62 (m, 15H), 1.55–1.40 (m, 6H), 1.33 (s, 9H), 1.19–1.15 (m, 1H); MS (ESI): m/z calcd. for $[C_{53}H_{80}N_7O_9]^+$ 958.59, found 958.50.

QUANTIFICATION AND STATISTICAL ANALYSIS

The method of determining error bars is indicated in the corresponding figure legend with the replicate number also indicated. Statistical tests for cChIP-seq and RNA-seq data is outlined in the STAR Methods section under the relevant analysis. Data met the assumptions for all tests used.

Supplementary Material

Refer to Web version on PubMed Central for supplementary material.

ACKNOWLEDGMENTS

This work was supported by the National Cancer Institute (NCI) (R01CA218392) to S.V.F. and O.B.; the National Institute of General Medical Sciences (R01GM100919 to S.V.F., R35GM128705 to C.A.M., and R01GM132299 to D.B.K.); the New Frontiers Group of the Austrian Academy of Sciences (NFG-05), the Norris Comprehensive Cancer Center of USC and its NCI Award (P30CA014089) to O.B., S.M.M., and H.-J.L.; the National Institute on Drug Abuse (R61DA047023) to L.I.J.; the NCI (P30CA014089), Gloria Borges Wunder-Glo Foundation, Gene Gregg Pancreas Cancer Research Fund to H.-J.L.; the Holden Comprehensive Cancer Center at the University of Iowa and its NCI Award (P30CA086862) to C.A.M.; the NCI (R01CA172896) to Y.-M.K.; the UNC Lineberger Comprehensive Cancer Center and its NCI Award (P30CA016086) to S.V.F.; and by the UNC Eshelman Institute for Innovation (RX0351210 and RX03712105) to D.B.K. Probing of arrayed methyl-binding domains was made possible via the UT MDACC Protein Array & Analysis Core (PAAC) CPRIT (RP180804) to M.T.B. We acknowledge the UNC Longleaf supercomputer cluster and their staff for support. The authors thank Joshua Kritzer (Tufts) for sharing the CAPA HeLa cell line, Emily Dykhuizen (Purdue University) for sharing CBX8 expression constructs and Isabelle Engelberg, Catherine Foley, and Frances Potjewyd for reviewing primary data supporting this manuscript. The graphical abstract was created with Biorender.com.

REFERENCES

Anders S, Pyl PT, and Huber W (2015). HTSeq—a Python framework to work with high-throughput sequencing data. Bioinformatics 31, 166–169. [PubMed: 25260700]

Aranda S, Mas G, and Di Croce L (2015). Regulation of gene transcription by Polycomb proteins. Sci. Adv 1, e1500737. [PubMed: 26665172]

Beguelin W, Teater M, Gearhart MD, Calvo Fernandez MT, Goldstein RL, Cardenas MG, Hatzi K, Rosen M, Shen H, Corcoran CM, et al. (2016). EZH2 and BCL6 cooperate to assemble CBX8-BCOR complex to repress bivalent promoters, mediate germinal center formation and lymphomagenesis. Cancer Cell 30, 197–213. [PubMed: 27505670]

Blaskovich MA (2016). Unusual amino acids in medicinal chemistry. J. Med. Chem 59, 10807–10836. [PubMed: 27589349]

Cao R, Tsukada Y, and Zhang Y (2005). Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. Mol. Cell 20, 845–854. [PubMed: 16359901]

Cao R, Wang L, Wang H, Xia L, Erdjument-Bromage H, Tempst P, Jones RS, and Zhang Y (2002). Role of histone H3 lysine 27 methylation in Polycomb-group silencing. Science 298, 1039–1043. [PubMed: 12351676]

Chan HL, and Morey L (2019). Emerging roles for polycomb-group proteins in stem cells and cancer. Trends Biochem. Sci 44, 688–700. [PubMed: 31085088]

Chung CY, Sun Z, Mullokandov G, Bosch A, Qadeer ZA, Cihan E, Rapp Z, Parsons R, Aguirre-Ghiso JA, Farias EF, et al. (2016). Cbx8 Acts non-canonically with Wdr5 to promote mammary tumorigenesis. Cell Rep 16, 472–486. [PubMed: 27346354]

Connelly KE, Weaver TM, Alpsoy A, Gu BX, Musselman CA, and Dykhuizen EC (2018). Engagement of DNA and H3K27me3 by the CBX8 chromodomain drives chromatin association. Nucl. Acids Res gky1290.

Connelly KE, Weaver TM, Alpsoy A, Gu BX, Musselman CA, and Dykhuizen EC (2019). Engagement of DNA and H3K27me3 by the CBX8 chromodomain drives chromatin association. Nucl. Acids Res 47, 2289–2305. [PubMed: 30597065]

Creppe C, Palau A, Malinverni R, Valero V, and Buschbeck M (2014). A Cbx8-containing polycomb complex facilitates the transition to gene activation during ES cell differentiation. Plos Genet. 10, e1004851. [PubMed: 25500566]

Czermin B, Melfi R, McCabe D, Seitz V, Imhof A, and Pirrotta V (2002). Drosophila enhancer of Zeste/ESC complexes have a histone H3 methyltransferase activity that marks chromosomal Polycomb sites. Cell 111, 185–196. [PubMed: 12408863]

de Napoles M, Mermoud JE, Wakao R, Tang YA, Endoh M, Appanah R, Nesterova TB, Silva J, Otte AP, Vidal M, et al. (2004). Polycomb group proteins Ring1A/B link ubiquitylation of histone H2A to heritable gene silencing and X inactivation. Dev. Cell 7, 663–676. [PubMed: 15525528]

Delaglio F, Grzesiek S, Vuister GW, Zhu G, Pfeifer J, and Bax A (1995). NMRPipe: a multidimensional spectral processing system based on UNIX pipes. J. Biomol. NMR 6, 277–293. [PubMed: 8520220]

Dietrich N, Lerdrup M, Landt E, Agrawal-Singh S, Bak M, Tommerup N, Rappsilber J, Södersten E, and Hansen K (2012). REST-mediated recruitment of polycomb repressor complexes in mammalian cells. PLoS Genet. 8, e1002494. [PubMed: 22396653]

Dobin A, Davis CA, Schlesinger F, Drenkow J, Zaleski C, Jha S, Batut P, Chaisson M, and Gingeras TR (2013). STAR: ultrafast universal RNA-seq aligner. Bioinformatics 29, 15–21. [PubMed: 23104886]

Dobrinić P, Szczurek AT, and Klose RJ (2020). PRC1 drives Polycomb-mediated gene repression by controlling transcription initiation and burst frequency. bioRxiv. 10.1101/2020.10.09.333294.

Duchardt F, Fotin-Mleczek M, Schwarz H, Fischer R, and Brock R (2007). A comprehensive model for the cellular uptake of cationic cell-penetrating peptides. Traffic 8, 848–866. [PubMed: 17587406]

Espejo A, Cote J, Bednarek A, Richard S, and Bedford MT (2002). A protein-domain microarray identifies novel protein-protein interactions. Biochem. J 367, 697–702. [PubMed: 12137563]

Essmann U, Perera L, Berkowitz ML, Darden T, Lee H, and Pedersen LG (1995). A smooth particle mesh Ewald method. J. Chem. Phys 103, 8577–8593.

Foley CA, Potjewyd F, Lamb KN, James LI, and Frye SV (2020). Assessing the cell permeability of bivalent chemical degraders using the chloroalkane penetration assay. ACS Chem. Biol 15, 290–295. [PubMed: 31846298]

Fursova NA, Blackledge NP, Nakayama M, Ito S, Koseki Y, Farcas AM, King HW, Koseki H, and Klose RJ (2019). Synergy between variant PRC1 complexes defines polycomb-mediated gene repression. Mol. Cell 74, 1020–1036.e1028. [PubMed: 31029541]

Gil J, Bernard D, Martínez D, and Beach D (2004). Polycomb CBX7 has a unifying role in cellular lifespan. Nat. Cell Biol 6, 67–72. [PubMed: 14647293]

He Y, Selvaraju S, Curtin ML, Jakob CG, Zhu H, Comess KM, Shaw B, The J, Lima-Fernandes E, Szewczyk MM, et al. (2017). The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. Nat. Chem. Biol 13, 389–395. [PubMed: 28135237]

Herce HD, Garcia AE, Litt J, Kane RS, Martin P, Enrique N, Rebolledo A, and Milesi V (2009). Arginine-rich peptides destabilize the plasma membrane, consistent with a pore formation translocation mechanism of cell-penetrating peptides. Biophys. J 97, 1917–1925. [PubMed: 19804722]

Hess B, Bekker H, Berendsen HJC, and Fraaije JGEM (1997). LINCS: a linear constraint solver for molecular simulations. J. Comput. Chem 18, 1463–1472.

Hsu JH-R, Rasmusson T, Robinson J, Pachl F, Read J, Kawatkar S, O' Donovan DH, Bagal S, Code E, Rawlins P, et al. (2020). EED-targeted PROTACs degrade EED, EZH2, and SUZ12 in the PRC2 complex. Cell Chem. Biol 27, 41–46.e17. [PubMed: 31786184]

Knutson SK, Kawano S, Minoshima Y, Warholic NM, Huang KC, Xiao Y, Kadowaki T, Uesugi M, Kuznetsov G, Kumar N, et al. (2014). Selective inhibition of EZH2 by EPZ-6438 leads to potent antitumor activity in EZH2-mutant non-Hodgkin lymphoma. Mol. Cancer Ther 13, 842–854. [PubMed: 24563539]

Knutson SK, Warholic NM, Wigle TJ, Klaus CR, Allain CJ, Raimondi A, Porter Scott M, Chesworth R, Moyer MP, Copeland RA, et al. (2013). Durable tumor regression in genetically altered

malignant rhabdoid tumors by inhibition of methyltransferase EZH2. Proc. Natl. Acad. Sci. U S A 110, 7922–7927. [PubMed: 23620515]

Koppens M, and van Lohuizen M (2016). Context-dependent actions of Polycomb repressors in cancer. Oncogene 35, 1341–1352. [PubMed: 26050622]

Lamb KN, Bsteh D, Dishman SN, Moussa HF, Fan H, Stuckey JI, Norris JL, Cholensky SH, Li D, Wang J, et al. (2019). Discovery and characterization of a cellular potent positive allosteric modulator of the polycomb repressive complex 1 chromodomain, CBX7. Cell Chem. Biol 26, 1365–1379.e1322. [PubMed: 31422906]

Langmead B, and Salzberg SL (2012). Fast gapped-read alignment with Bowtie 2. Nat. Methods 9, 357–359. [PubMed: 22388286]

Laugesen A, and Helin K (2014). Chromatin repressive complexes in stem cells, development, and cancer. Cell Stem Cell 14, 735–751. [PubMed: 24905164]

Li H, Handsaker B, Wysoker A, Fennell T, Ruan J, Homer N, Marth G, Abecasis G, Durbin R, and Genome Project Data Processing, S. (2009). The sequence alignment/map format and SAMtools. Bioinformatics 25, 2078–2079. [PubMed: 19505943]

Love MI, Huber W, and Anders S (2014). Moderated estimation of fold change and dispersion for RNA-seq data with DESeq2. Genome Biol. 15, 550. [PubMed: 25516281]

Ma A, Stratikopoulos E, Park KS, Wei J, Martin TC, Yang X, Schwarz M, Leshchenko V, Rialdi A, Dale B, et al. (2020). Discovery of a first-in-class EZH2 selective degrader. Nat. Chem. Biol 16, 214–222. [PubMed: 31819273]

Mackay F, and Schneider P (2009). Cracking the BAFF code. Nat. Rev. Immunol 9, 491–502. [PubMed: 19521398]

Mark P, and Nilsson L (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. J. Phys. Chem. A 105, 9954–9960.

McGinty RK, Henrici RC, and Tan S (2014). Crystal structure of the PRC1 ubiquitylation module bound to the nucleosome. Nature 514, 591–596. [PubMed: 25355358]

McLean CY, Bristor D, Hiller M, Clarke SL, Schaar BT, Lowe CB, Wenger AM, and Bejerano G (2010). GREAT improves functional interpretation of cis-regulatory regions. Nat. Biotechnol 28, 495–501. [PubMed: 20436461]

Milosevich N, Gignac MC, McFarlane J, Simhadri C, Horvath S, Daze KD, Croft CS, Dheri A, Quon TTH, Douglas SF, et al. (2016). Selective inhibition of CBX6: a methyllysine reader protein in the polycomb family. ACS Med. Chem. Lett 7, 139–144.

Min J, Zhang Y, and Xu RM (2003). Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. Genes Dev. 17, 1823–1828. [PubMed: 12897052]

Moazed D (2011). Mechanisms for the inheritance of chromatin states. Cell 146, 510–518. [PubMed: 21854979]

Morey L, Aloia L, Cozzuto L, Benitah Salvador A., and Di Croce L (2013). RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. Cell Rep. 3, 60–69. [PubMed: 23273917]

Moussa HF, Bsteh D, Yelagandula R, Pribitzer C, Stecher K, Bartalska K, Michetti L, Wang J, Zepeda-Martinez JA, Elling U, et al. (2019). Canonical PRC1 controls sequence-independent propagation of Polycomb-mediated gene silencing. Nat. Commun 10, 1931. [PubMed: 31036804]

Müller J, Hart CM, Francis NJ, Vargas ML, Sengupta A, Wild B, Miller EL, O'Connor MB, Kingston RE, and Simon JA (2002). Histone methyltransferase activity of a *Drosophila* Polycomb group repressor complex. Cell 111, 197–208. [PubMed: 12408864]

Müller S, Ackloo S, Arrowsmith CH, Bauser M, Baryza JL, Blagg J, Böttcher J, Bountra C, Brown PJ, Bunnage ME, et al. (2018). Donated chemical probes for open science. eLife 7, e34311. [PubMed: 29676732]

Nosé S, and Klein ML (1983). A study of solid and liquid carbon tetrafluoride using the constant pressure molecular dynamics technique. J. Chem. Phys 78, 6928–6939.

Pasini D, and Di Croce L (2016). Emerging roles for polycomb proteins in cancer. Curr. Opin. Genet. Dev 36, 50–58. [PubMed: 27151431]

Peraro L, Deprey KL, Moser MK, Zou Z, Ball HL, Levine B, and Kritzer JA (2018). Cell penetration profiling using the chloroalkane penetration assay. J. Am. Chem. Soc 140, 11360–11369. [PubMed: 30118219]

Peraro L, Zou Z, Makwana KM, Cummings AE, Ball HL, Yu H, Lin YS, Levine B, and Kritzer JA (2017). Diversity-oriented stapling yields intrinsically cell-penetrant inducers of autophagy. J. Am. Chem. Soc 139, 7792–7802. [PubMed: 28414223]

Potjewyd F, Turner AW, Beri J, Rectenwald JM, Norris-Drouin JL, Cholensky SH, Margolis DM, Pearce KH, Herring LE, and James LI (2020). Degradation of polycomb repressive complex 2 with an EED-targeted bivalent chemical degrader. Cell Chem. Biol 27, 47–56.e15. [PubMed: 31831267]

Qi W, Zhao K, Gu J, Huang Y, Wang Y, Zhang H, Zhang M, Zhang J, Yu Z, Li L, et al. (2017). An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. Nat. Chem. Biol 13, 381–388. [PubMed: 28135235]

Qvit N, Rubin SJS, Urban TJ, Mochly-Rosen D, and Gross ER (2017). Peptidomimetic therapeutics: scientific approaches and opportunities. Drug Discov. Today 22, 454–462. [PubMed: 27856346]

Rada-Iglesias A, Bajpai R, Swigut T, Brugmann SA, Flynn RA, and Wysocka J (2011). A unique chromatin signature uncovers early developmental enhancers in humans. Nature 470, 279–283. [PubMed: 21160473]

Ramirez F, Ryan DP, Gruning B, Bhardwaj V, Kilpert F, Richter AS, Heyne S, Dundar F, and Manke T (2016). deepTools2: a next generation web server for deep-sequencing data analysis. Nucl. Acids Res 44, W160–W165. [PubMed: 27079975]

Rectenwald JM, Hardy PB, Norris-Drouin JL, Cholensky SH, James LI, Frye SV, and Pearce KH (2019). A general TR-FRET assay platform for high-throughput screening and characterizing inhibitors of methyl-lysine reader proteins. SLAS Discov. 24, 693–700. [PubMed: 31017815]

Reinberg D, and Vales LD (2018). Chromatin domains rich in inheritance. Science 361, 33–34. [PubMed: 29976815]

Ren C, Morohashi K, Plotnikov Alexander N., Jakoncic J, Smith Steven G., Li J, Zeng L, Rodriguez Y, Stojanoff V, Walsh M, et al. (2015). Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. Chem. Biol 22, 161–168. [PubMed: 25660273]

Ren C, Smith SG, Yap KL, Li S, Li J, Mezei M, Vincek AS, Aguilo F, Walsh MJ, and Zhou M-M (2016). Structure-guided discovery of selective antagonists for the chromodomain of polycomb repressive protein CBX7. ACS Med. Chem. Lett 7, 601–605. 10.1021/acsmedchemlett.6b0004. [PubMed: 27326334]

Schuettengruber B, Bourbon HM, Di Croce L, and Cavalli G (2017). Genome regulation by polycomb and trithorax: 70 years and counting. Cell 171, 34–57. [PubMed: 28938122]

Scott CL, Gil J, Hernando E, Teruya-Feldstein J, Narita M, Martínez D, Visakorpi T, Mu D, Cordon-Cardo C, Peters G, et al. (2007). Role of the chromobox protein CBX7 in lymphomagenesis. Proc. Natl. Acad. Sci. U S A 104, 5389–5394. [PubMed: 17374722]

Shimokawa M, Ohta Y, Nishikori S, Matano M, Takano A, Fujii M, Date S, Sugimoto S, Kanai T, and Sato T (2017). Visualization and targeting of LGR5+ human colon cancer stem cells. Nature 545, 187–192. [PubMed: 28355176]

Simhadri C, Daze KD, Douglas SF, Quon TT, Dev A, Gignac MC, Peng F, Heller M, Boulanger MJ, Wulff JE, et al. (2014). Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). J. Med. Chem 57, 2874–2883. [PubMed: 24625057]

Simon JA, and Kingston RE (2013). Occupying chromatin: polycomb mechanisms for getting to genomic targets, stopping transcriptional traffic, and staying put. Mol. Cell 49, 808–824. [PubMed: 23473600]

Stuckey JI (2016). Design and Characterization of Chemical Probes Targeting the Chromodomains of Polycomb Repressive Complex I (University of North Carolina at Chapel Hill).

Stuckey JI, Dickson BM, Cheng N, Liu Y, Norris JL, Cholensky SH, Tempel W, Qin S, Huber KG, Sagum C, et al. (2016a). A cellular chemical probe targeting the chromodomains of polycomb repressive complex 1. Nat. Chem. Biol 12, 180–187. [PubMed: 26807715]

Stuckey JI, Dickson BM, Cheng N, Liu Y, Norris JL, Cholensky SH, Tempel W, Qin S, Huber KG, Sagum C, et al. (2019). Addendum: a cellular chemical probe targeting the chromodomains of polycomb repressive complex 1. Nat. Chem. Biol 15, 846.

Stuckey JI, Simpson C, Norris-Drouin JL, Cholensky SH, Lee J, Pasca R, Cheng N, Dickson BM, Pearce KH, Frye SV, et al. (2016b). Structure-activity relationships and kinetic studies of peptidic antagonists of CBX chromodomains. J. Med. Chem 59, 8913–8923. [PubMed: 27571219]

Tan J, Jones M, Koseki H, Nakayama M, Muntean AG, Maillard I, and Hess JL (2011). CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. Cancer Cell 20, 563–575. [PubMed: 22094252]

Tunnemann G, Martin RM, Haupt S, Patsch C, Edenhofer F, and Cardoso MC (2006). Cargo-dependent mode of uptake and bioavailability of TAT-containing proteins and peptides in living cells. FASEB J. 20, 1775–1784. [PubMed: 16940149]

van Zundert GCP, Rodrigues J, Trellet M, Schmitz C, Kastritis PL, Karaca E, Melquiond ASJ, van Dijk M, de Vries SJ, and Bonvin A (2016). The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes. J. Mol. Biol 428, 720–725. [PubMed: 26410586]

Vanommeslaeghe K, Hatcher E, Acharya C, Kundu S, Zhong S, Shim J, Darian E, Guvench O, Lopes P, Vorobyov I, et al. (2010). CHARMM general force field: a force field for drug-like molecules compatible with the CHARMM all-atom additive biological force fields. J. Comput. Chem 31, 671–690. [PubMed: 19575467]

Vranken WF, Boucher W, Stevens TJ, Fogh RH, Pajon A, Llinas M, Ulrich EL, Markley JL, Ionides J, and Laue ED (2005). The CCPN data model for NMR spectroscopy: development of a software pipeline. Proteins Struct. Funct. Bioinform 59, 687–696.

Wang H, Wang L, Erdjument-Bromage H, Vidal M, Tempst P, Jones RS, and Zhang Y (2004a). Role of histone H2A ubiquitination in Polycomb silencing. Nature 431, 873–878. [PubMed: 15386022]

Wang L, Brown JL, Cao R, Zhang Y, Kassis JA, and Jones RS (2004b). Hierarchical recruitment of polycomb group silencing complexes. Mol. Cell 14, 637–646. [PubMed: 15175158]

Wang S, Denton KE, Hobbs KF, Weaver T, McFarlane JMB, Connelly KE, Gignac MC, Milosevich N, Hof F, Paci I, et al. (2020). Optimization of ligands using focused DNA-encoded libraries to develop a selective, cell-permeable CBX8 chromodomain inhibitor. ACS Chem. Biol 15, 112–131. [PubMed: 31755685]

Wassenaar TA, van Dijk M, Loureiro-Ferreira N, van der Schot G, de Vries SJ, Schmitz C, van der Zwan J, Boelens R, Giachetti A, Ferella L, et al. (2012). WeNMR: structural biology on the grid. J. Grid Comput 10, 743–767.

Weaver T, Morrison E, and Musselman C (2018). Reading more than histones: the prevalence of nucleic acid binding among reader domains. Molecules 23, 2614.

Yap KL, Li S, Munoz-Cabello AM, Raguz S, Zeng L, Mujtaba S, Gil J, Walsh MJ, and Zhou MM (2010). Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. Mol. Cell 38, 662–674. [PubMed: 20541999]

Zepeda-Martinez JA, Pribitzer C, Wang J, Bsteh D, Golumbeanu S, Zhao Q, Burkard TR, Reichholf B, Rhie SK, Jude J, et al. (2020). Parallel PRC2/cPRC1 and vPRC1 pathways silence lineage-specific genes and maintain self-renewal in mouse embryonic stem cells. Sci. Adv 6, eaax5692. [PubMed: 32270030]

Zhang Y, Kang M, Zhang B, Meng F, Song J, Kaneko H, Shimamoto F, and Tang B (2019). m6A modification-mediated CBX8 induction regulates stemness and chemosensitivity of colon cancer via upregulation of LGR5. Mol. Cancer 18, 185. [PubMed: 31849331]

Zhang Y, Liu T, Meyer CA, Eeckhoute J, Johnson DS, Bernstein BE, Nusbaum C, Myers RM, Brown M, Li W, et al. (2008). Model-based analysis of ChIP-seq (MACS). Genome Biol. 9, R137. [PubMed: 18798982]

Zhu A, Ibrahim JG, and Love MI (2019). Heavy-tailed prior distributions for sequence count data: removing the noise and preserving large differences. Bioinformatics 35, 2084–2092. [PubMed: 30395178]

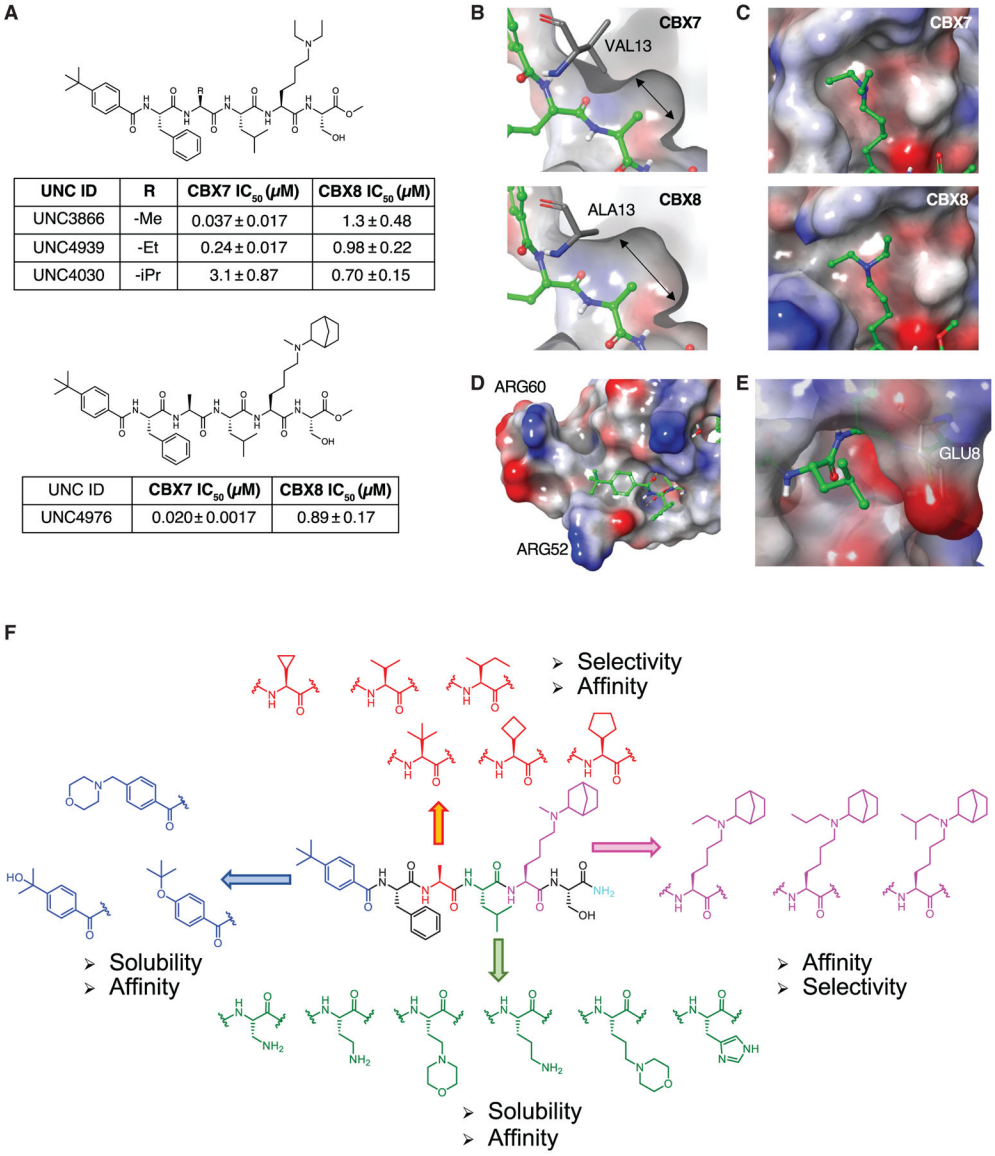
Zoete V, Cuendet MA, Grosdidier A, and Michielin O (2011). SwissParam: a fast force field generation tool for small organic molecules. J. Comput. Chem 32, 2359–2368. [PubMed: 21541964]

Highlights

- Structure-based design of selective, high-affinity chemical probe against CBX8
- Cellular reporter identifies UNC7040, a PAM of CBX8
- PAM activity blocks H3K27me3 binding and enhances CBX8 affinity for nucleic acids
- UNC7040 evicts PRC1 from H3K27me3 targets triggering DLBCL cell differentiation

SIGNIFICANCE

Multivalency is a central theme in chromatin regulatory processes wherein multiple low-affinity interactions can result in sufficient specific binding to selectively control biological processes. An aspect of Polycomb CBX chromodomains that has been underappreciated during inhibitor development is the ability of these methyl-lysine reader domains to bind to nucleic acids in addition to the methylated histone substrate peptide in a multivalent fashion. UNC7040 simultaneously modulates each of these binding phenomena: directly competing with H3K27me3 binding, while acting as a positive allosteric modulator to enhance nucleic acid affinity. This results in superior cellular activity relative to a ligand with identical binding potency but diminished allosteric efficacy, UNC7263, by increasing equilibration of CBX8-containing PRC1 away from H3K27me3 sites. The positive allosteric activity of UNC7040 results in sub-micromolar efficacy in a CBX8-specific cellular reporter assay, strong displacement from H3K27me3 sites in chromatin, and antiproliferative activity in diffuse large B cell lymphoma and colorectal cancer cell lines. A TR-FRET-based ternary complex assay (UNC7040/CBX8-CD/DNA), NMR, and molecular dynamics support the proposed mechanism of action. In addition, the selectivity of UNC7040 is thoroughly profiled via TR-FRET competition assays versus multiple chromodomains, protein microarrays displaying 274 purified methyllysine readers, and ChIP experiments versus CBX2, 6, and 7. Discovery and characterization of UNC7040 not only reveals the most cellularly potent CBX8-specific chemical probe to date, but also corroborates a mechanism of Polycomb regulation by non-specific CBX nucleotide binding activity.



UNC ID	R	CBX7 IC_{50} (μM)	CBX8 IC_{50} (μM)
UNC3866	-Me	0.037 ± 0.017	1.3 ± 0.48
UNC4939	-Et	0.24 ± 0.017	0.98 ± 0.22
UNC4030	-iPr	3.1 ± 0.87	0.70 ± 0.15

UNC ID	CBX7 IC_{50} (μM)	CBX8 IC_{50} (μM)
UNC4976	0.020 ± 0.0017	0.89 ± 0.17

Figure 1. Structure-based design of selective CBX8 compounds

(A) Reported CBX7 compounds and their *in vitro* potency in an AlphaScreen competition assay.

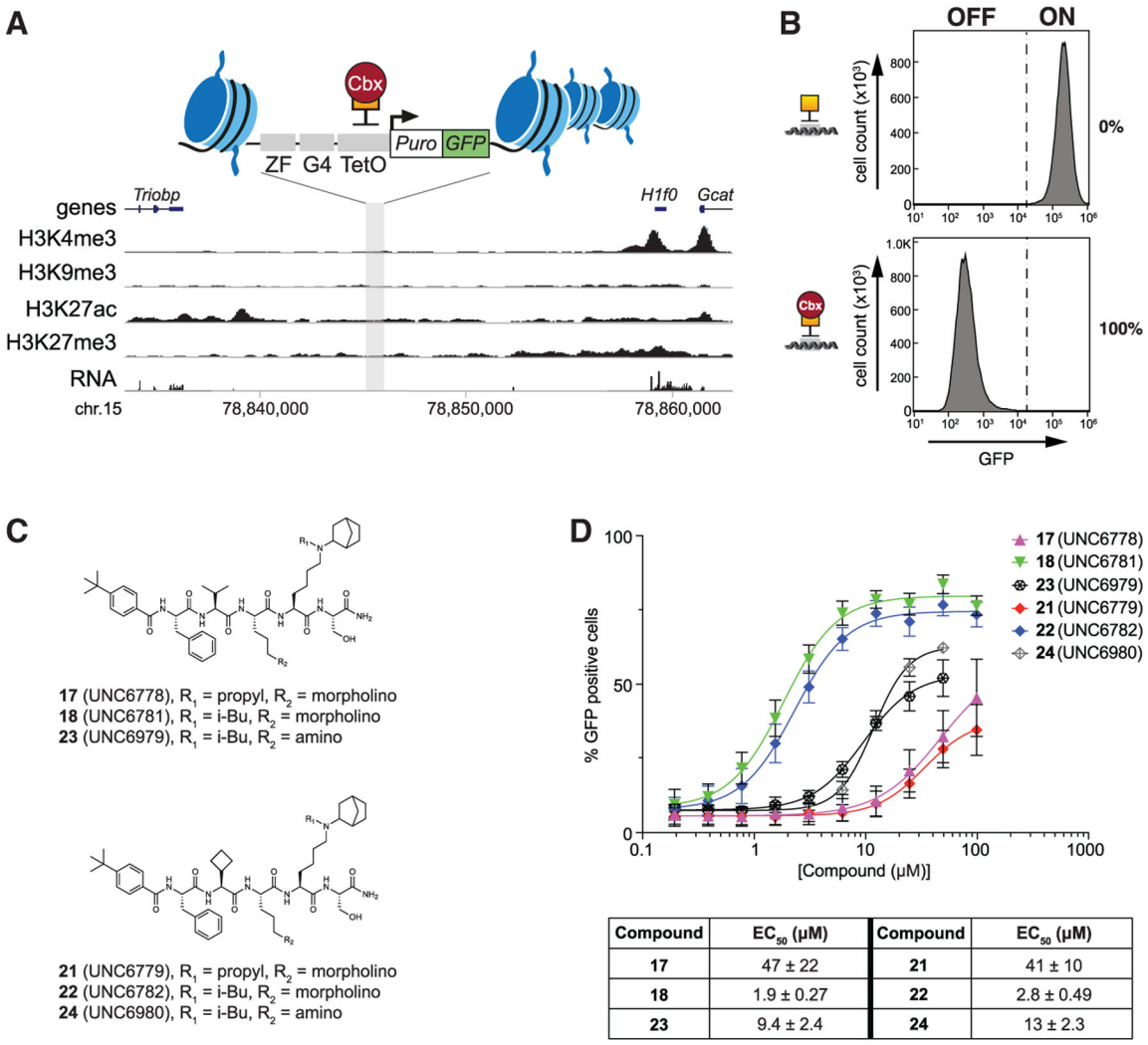
(B) Comparison of the alanine binding pocket size of CBX7 (top) and CBX8 (bottom).

(C) Comparison of the aromatic cage region of CBX7 (top) and CBX8 (bottom).

(D) The binding mode of the N-terminal *tert*-butyl benzoyl cap of UNC3866 in CBX8.

(E) The binding mode of the UNC3866 leucine side chain in CBX8. UNC3866 is displayed in ball-and-stick with carbons colored green.

(F) SAR strategy for new compounds (red, Ala position; blue, N-cap position; green, Leu position; magenta, lysine mimetics position).



Compound	EC_{50} (μM)	Compound	EC_{50} (μM)
17	47 ± 22	**21**	41 ± 10
18	1.9 ± 0.27	**22**	2.8 ± 0.49
23	9.4 ± 2.4	**24**	13 ± 2.3

Figure 2. Compound screening with CBX8 reporter cells

(A) Schematic representation of the CBX8-specific GFP reporter assay in mESCs.

(B) Flow cytometry histogram shows percent repression of GFP expression in response to recruitment of TetR (top) or TetR-CBX8 (bottom) in the absence of CBX8 antagonists.

(C) Compounds with the best potency profile from the SAR studies.

(D) The CBX8 reporter assay results of the six optimized compounds. Data shown are mean ± SD, n = 9.

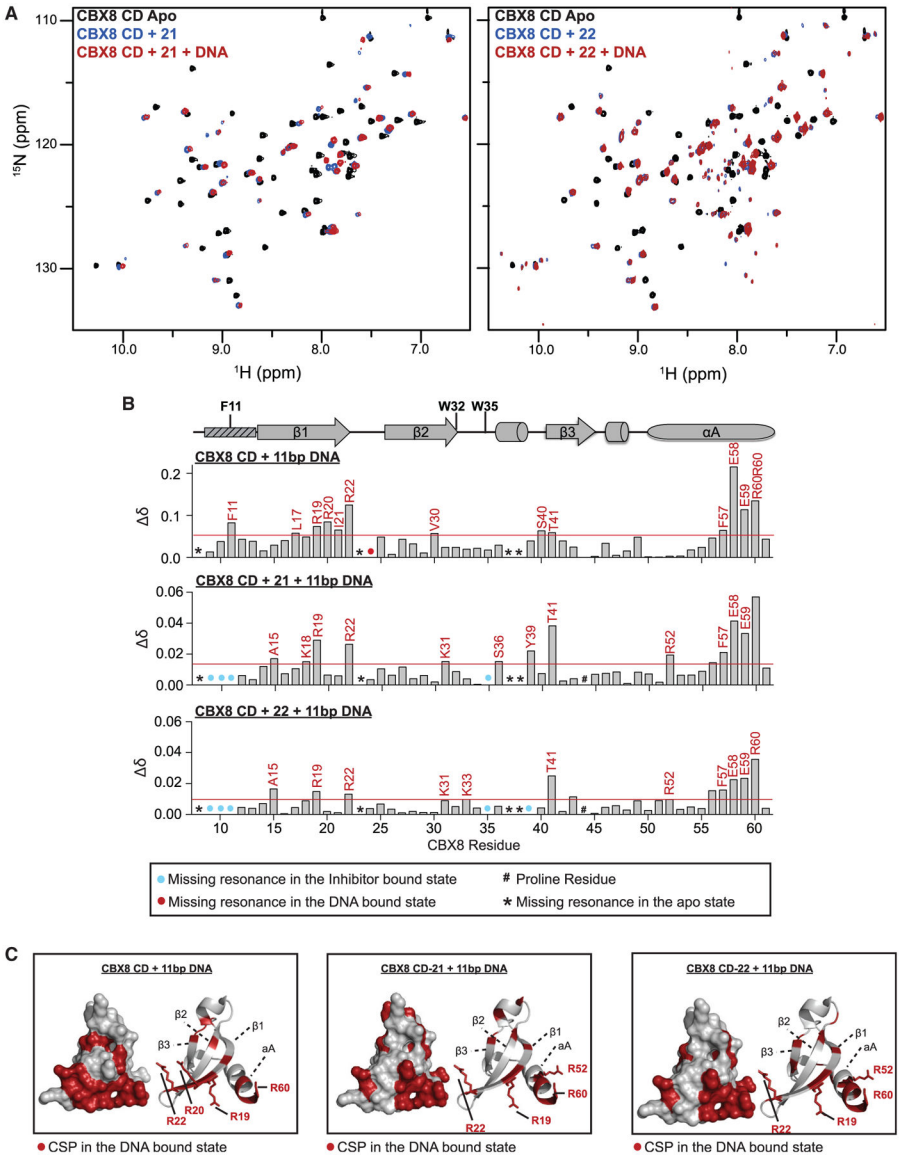


Figure 3. Compounds 21 and 22 bind the CBX8-CDs and form a DNA-ternary complex

(A) Full ^{1}H-^{15}N-HSQC overlays for ^{15}N-CD upon addition of compound **21** (left) or **22** (right) in the absence and the presence of 11bp DNA.

(B) Normalized CSP ($\Delta\delta$) between 11bp DNA-bound (top), 11bp DNA and **21**-bound (middle), and 11bp DNA and **22**-bound (bottom) spectra are plotted against CBX8 residue number.

(C) Residues with significant CSPs upon addition of 11bp DNA only, 11bp DNA with compound **21**, and 11bp DNA with compound **22** plotted onto a cartoon and surface representation of the CDs and colored red.

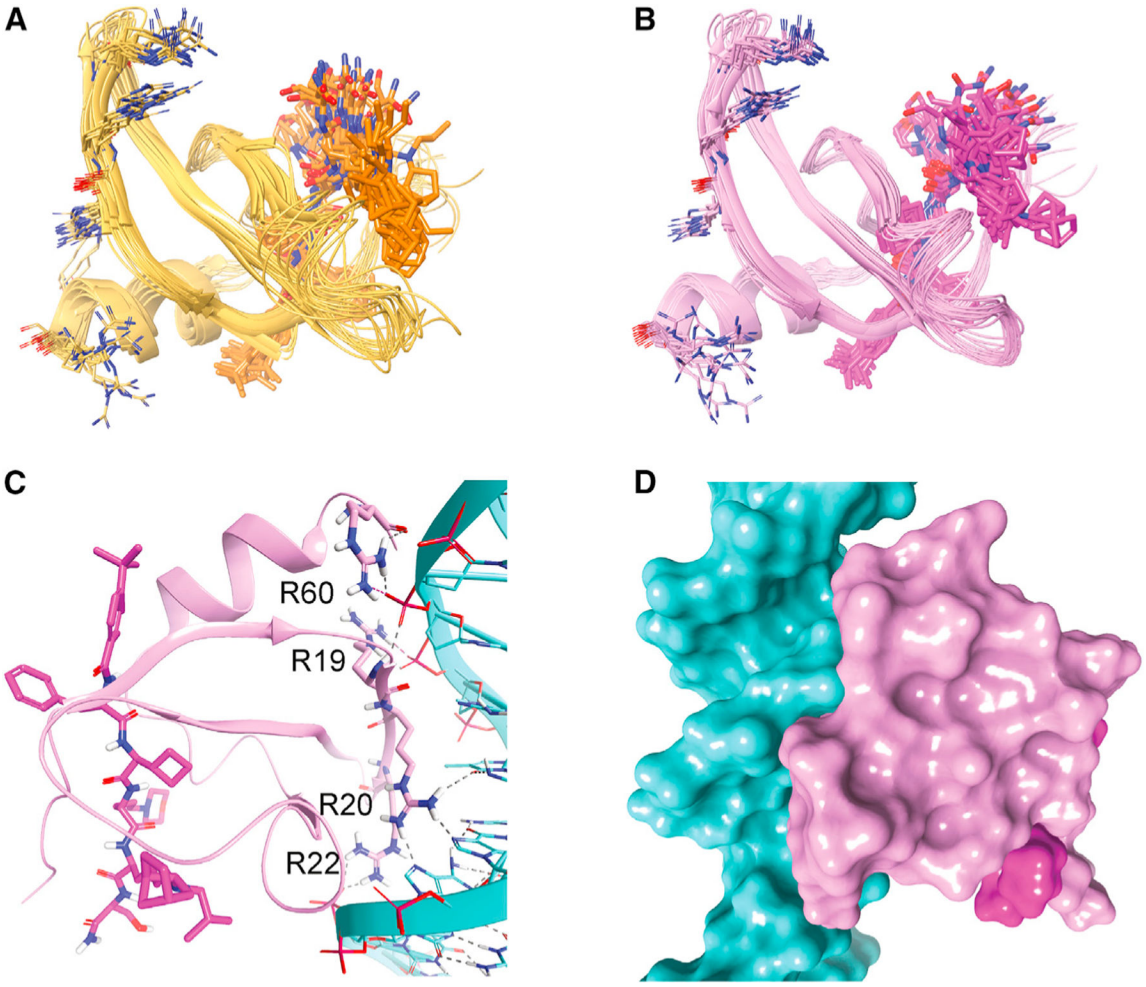
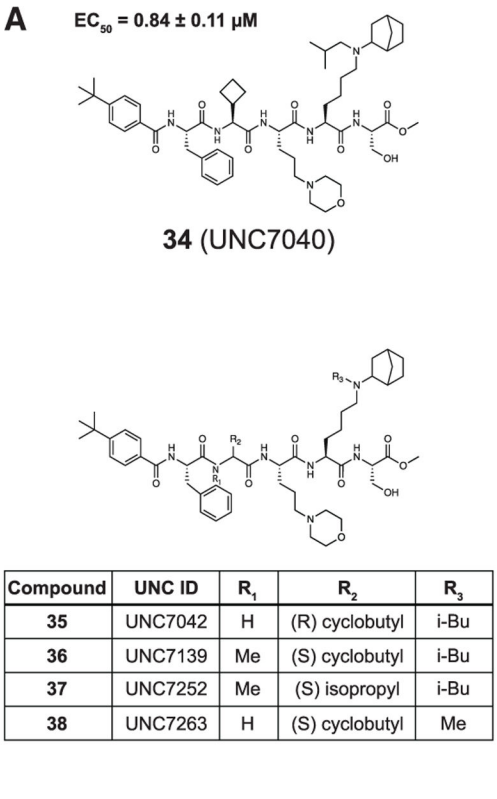


Figure 4. MD supports positive allosteric effect of 22

(A) MD simulation snapshots sampled from conformational ensembles of CBX8 in complex with **21** (yellow cartoon/orange sticks) and **22** (pink cartoon/magenta sticks).

(B and C) An example of a top-ranked docking pose showing a large contact surface area between the CBX8–22 complex (pink and magenta surfaces) and DNA double helix (cyan surface) (B) with the K18-G24 β strand binding deep into the major groove (C). Only **22**-bound CBX8 (pink cartoon/magenta sticks) has shown binding modes with all four essential DNA binding residues (R19, R20, R22, and R60) simultaneously interacting with DNA (cyan cartoon/sticks) (C).



Compound	UNC ID	R_1	R_2	R_3
35	UNC7042	H	(R) cyclobutyl	i-Bu
36	UNC7139	Me	(S) cyclobutyl	i-Bu
37	UNC7252	Me	(S) isopropyl	i-Bu
38	UNC7263	H	(S) cyclobutyl	Me

Figure 5. Cellular activity and toxicity of final CBX8 probe and negative controls

(A) Structures of the final probe and the negative controls.

(B) CBX8 GFP reporter assay results.

(C) Compound cell viability results. n = 9.

A

Chromo-domain	TR-FRET IC50(μM)			SPR *Kd*(μM)
	34 (UNC7040)	**35** (UNC7042)	**38** (UNC7263)	**40** (UNC7045)
CBX2	4.4 ± 1.2	>10	1.4 ± 0.24	16
CBX4	4.0 ± 1.2	>10	4.4 ± 0.88	0.81
CBX6	3.2 ± 0.4	>100	3.2 ± 1.4	6.9
CBX7	14 ± 2.0	>100	17 ± 0.26	N.D.
CBX8	0.65 ± 0.03	>100	0.70 ± 0.09	0.16
CBX5	>100	>100	>100	N.D.
CDYL2	~ 100	>100	0.96 ± 0.14	54
MPP8	>100	>100	>10	N.D.

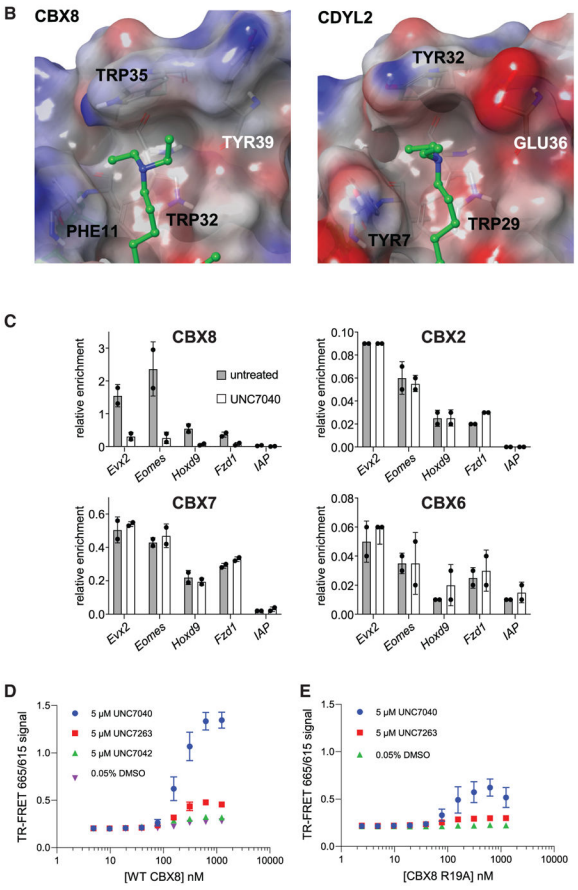


Figure 6. Analyses of probe specificity for CD proteins and effect on CBX8 DNA binding

(A) The binding profile of 34 (UNC7040), 35 (UNC7042), 38 (UNC7263), and 39 (UNC7045) for CBXs, CDYL2, and MPP8 chromodomains. For CBX2 and CBX4, the top 2 concentrations were discarded for technical issues in TR-FRET assay. Data shown are mean ± SD, n = 6.

(B) The side chain of Tyr39 in CBX8 faces out from the aromatic cage making it more expansive than that of CDYL2 (right). The side chain of Glu36 in CYDL2 facing into the aromatic cage makes it more compact (left).

(C) ChIP-qPCR analysis shows relative enrichments of CBX proteins at endogenous Polycomb target genes in CBX8-expressing mESCs. IAP serves as negative control locus. Data are mean ± SD (error bars) of two independent experiments.

(D) TR-FRET ternary complex assay results with WT CBX8, biotin-dsDNA, and compounds or DMSO. Data are mean ± SD (error bars) of two independent experiments, n = 6.

(E) TR-FRET ternary complex assay results with CBX8 R19A mutant, biotin-dsDNA, and compounds or DMSO. Data are mean ± SD (error bars) of three independent experiments, n = 9.

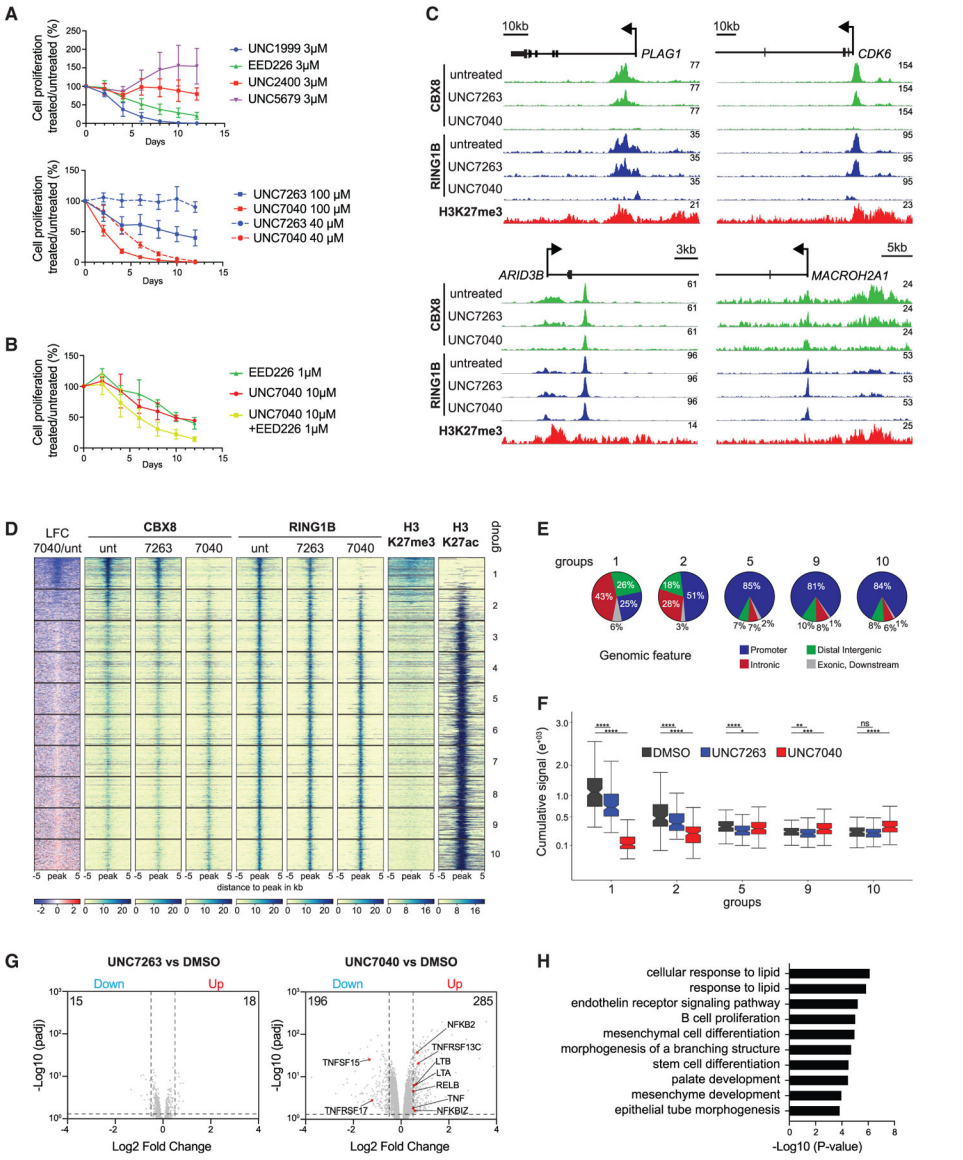


Figure 7. UNC7040 impairs DLBCL cell proliferation by blocking cPRC1 targeting via CBX8

(A) Effects of treatment with PRC2 (UNC1999, EED226) or CBX8 (UNC7040) antagonists on SUDHL4 cell proliferation relative to vehicle control. UNC2400, UNC5679, and UNC7263 served as negative controls for UNC1999, EED226, and UNC7040, respectively.

(B) Comparison of individual and combined treatment with UNC7040 and EED226 on SUDHL4 cell proliferation. Data in (A) and (B) represent an average of four to six replicates ± SD.

(C) Genomic screenshots of CBX8 and RING1B cChIP-seq signals at four selected PRC2 target genes in SUDHL4 cells treated for 6 h with DMSO, UNC7263, or UNC7040. Bottom track shows H3K27me3 signal in untreated cells.

(D) Heatmaps display cChIP-seq enrichment of CBX8 and RING1B in SUDHL4 cells treated for 6 h with DMSO, UNC7263 at 40 μM, or UNC7040 at 40 μM. Also shown are signals of H3K27me3 and H3K27ac in untreated cells. The signals from two independent

cChIP-seq experiments are centered around RING1B peaks ±5 kb and plotted as spike-in normalized mapped read counts. Genomic intervals were separated into 10 equal groups and sorted based on log2 fold change in CBX8 occupancy upon UNC7040 treatment relative to DMSO.

(E) Pie charts display relative distribution of genomic features in selected cChIP-seq groups based on (D).

(F) Boxplots compare cumulative CBX8 signal (±0.5 kb around RING1B peaks) for selected groups in untreated and treated SUDHL4 cells. Significance (p value) was calculated using Wilcoxon signed-rank test (*$p \leq 0.05$, **$p \leq 0.01$, ***$p \leq 0.001$, ****$p \leq 0.0001$; ns, not significant).

(G) Volcano plots show gene expression changes of UNC7263-treated (left) or UNC7040-treated SUDHL4 cells relative to DMSO. Numbers in represent repressed (blue) and upregulated (red) genes. Differential gene expression ($p_{adj} < 0.05$; LFC = <–0.5 and >0.5, respectively) was calculated from three independent replicates. Highlighted are genes associated with B cell maturation.

(H) For functional analysis of canonical CBX8-PRC1 targets, differential peaks resulting from UNC7040 treatment were analyzed by GREAT (v.4.0.4) using the nearest gene within 100 kb to generate enrichment of biological processes.

KEY RESOURCES TABLE

REAGENT or RESOURCE	SOURCE	IDENTIFIER
Antibodies		
IRDye 800CW Goat anti-Rabbit IgG (H+L)	LI-COR	Cat#: 926-32211; RRID: AB_621843
IRDye 680RD Goat anti-Mouse IgG (H+L)	LI-COR	Cat#: 926-68070; RRID: AB_10956588
Rabbit anti-H3K27me3	Diagenode	Cat#: C15410195; RRID: AB_2753161
Rabbit anti-H2AK119ub	Cell Signaling	Cat#: 8240; RRID: AB_10891618
Rabbit anti-CBX8	Bethyl Labs	Cat#: A300-882A; RRID: AB_2071525
Rabbit anti-CBX7	Abcam	Cat#: ab21873; RRID: AB_726005
Rabbit anti-Cbx7	Merck Millipore	Cat#: 07-981; RRID:AB_10807034
Rabbit anti-RING1B	Cell Signaling	Cat#: 5694; RRID: AB_10705604
Rabbit anti-SUZ12	Cell Signaling	Cat#: 3737; RRID: AB_2196850
Rabbit anti-PHC1	Cell Signaling	Cat# 13505, RRID:AB_2798239
Rabbit anti-RYBP	Cell Signaling	Cat# 41787; RRID:AB_2799210
Mouse anti-FLAG	Sigma-Aldrich	Cat# F1804, RRID:AB_262044
Rabbit anti-H3K27ac	Abcam	Cat#: ab4729, RRID:AB_2118291
Chemicals, peptides, and recombinant proteins		
Methanol-free Formaldehyde	Thermo Fisher Scientific	Cat# 10751395
UNC1999	Selleck Chemicals	Cat# S7165
EED226	Selleck Chemicals	Cat# S8496
Dimethyl Sulfoxide	Millipore Sigma	Cat# D8418
Critical commercial assays		
CellTiter-Glo® Luminescent Cell Viability Assay	Promega	Cat#G7573
NEXTflex ChIP-Seq Kit	Bioo Scientific	Cat#: NOVA-5143-01
KAPA Real-Time Library Amplification Kit	KAPA Biosystems	Cat#: 07959028001
High-Capacity cDNA Reverse Transcription Kit	Applied Biosystems	Cat#: 4368814
Monarch Total RNA Miniprep Kit	NEB	Cat#: T2010S
Deposited data		
RNA-seq	this manuscript	GEO accession number: GSE182592
ChIP-seq	this manuscript	GEO accession number: GSE182592
Experimental models: cell lines		
Mouse: CBX8 reporter ES Cells	this manuscript	N/A
Mouse: CBX7 reporter ES Cells	Lamb et al., 2019	N/A
Human: Cell Line HeLa-GFP-Mito	Peraro et al., 2018	N/A
Human: Cell Line SUDHL4	ATCC	Cat#: CRL-2957
Human: Cell Line LOVO	ATCC	Cat#: CCL-229
Human: Cell Line HCT116	ATCC	Cat#: CCL-247
Human: Cell Line Caco2	ATCC	Cat#: HTB-37
Human: Cell Line DLD-1	gift from Dr. Yun at Baylor College of Medicine	N/A
Experimental models: organisms/strains		
One Shot Stbl3 competent cells	Thermo Fisher Scientific	Cat# C737303

REAGENT or RESOURCE	SOURCE	IDENTIFIER
DH5a competent cells	Thermo Fisher Scientific	Cat# 18265017
Oligonucleotides		
Hoxd9 forward; GGATAATCGCCTAGGTGTGACTTAG	this manuscript	N/A
Hoxd9 reverse; CATCTCTTCTTGCCTCTCTGGG	this manuscript	N/A
Eomes forward; GGCGCAGGGAATCTTAACTG	this manuscript	N/A
Eomes reverse; AAGACCCAACATGAGCCTGA	this manuscript	N/A
Fzd1 forward; AGCCCGTAAACCTTGGTGGG	this manuscript	N/A
Fzd1 reverse; GGGAAAGTTCTCCTGCCCG	this manuscript	N/A
IAP forward; CTCCATGTGCTCTGCCTTCC	this manuscript	N/A
IAP reverse; CCCCGTCCCTTTTTTAGGAGA	this manuscript	N/A
Evx2 forward; CGCAGCCCATCATTAAGAC	this manuscript	N/A
Evx2 reverse; CGGACAAACTGGAGAACCTC	this manuscript	N/A
Recombinant DNA		
CP5-TetR-P2A-mCherry	Moussa et al., 2019	N/A
CP55-TetR-CBX8-P2A-mCherry	this manuscript	N/A
Software and algorithms		
GraphPad Prism Software	Prism	https://www.graphpad.com/
FlowJo Software	FlowJo, LLC	https://www.flowjo.com/
ProteOn Manager™ Software #1760200	Bio-Rad	https://www.bio-rad.com/
Maestro molecular modeling suite 2016-2	Schrödinger	https://www.schrodinger.com/maestro
PyMOL	Schrödinger	https://pymol.org/2/
GROMACS 2018.2	GROMACS project	www.gromacs.org
Discovery Studio 4.0	BIOVIA	www.3dsbiovia.com
Pipeline Pilot (Data Processing Software)	BIOVIA	www.3ds.com
Image Studio Software	LI-COR	https://www.licor.com/bio/image-studio/
Bowtie2	Langmead and Salzberg, 2012	http://bowtie-bio.sourceforge.net/bowtie2/index.shtml
SAMtools	Li et al., 2009	http://samtools.sourceforge.net/
MACS2	Zhang et al., 2008	https://github.com/taoliu/MACS
PICARD	“Picard Toolkit.” 2019. Broad Institute	https://broadinstitute.github.io/picard/
deepTools	Ramirez et al., 2016	https://github.com/deeptools/deepTools
STAR	Dobin et al., 2013	https://github.com/alexdobin/STAR
DESeq2	Love et al., 2014	http://www.bioconductor.org/
HTSeq	Anders et al., 2015	https://htseq.readthedocs.io/en/master/overview.html#citation
Other		
Attune NxT Analyzer	Thermo Fisher Scientific	
Odyssey® CLx Near-Infrared Imaging System	LICOR	