Design and synthesis of inhibitors targeting methyllysine reader proteins belonging to the polycomb paralog family

by

Natalia Milosevich
B.Sc., Dalhousie University, 2013,

A Dissertation Submitted in Partial Fulfillment
of the Requirements for the Degree of

DOCTOR OF PHILOSOPHY

in the Department of Chemistry

University of Victoria

Supervisory Committee

Design and synthesis of inhibitors targeting methyllysine reader proteins belonging to the polycomb paralog family

by

Natalia Milosevich
B.Sc., Dalhousie University, 2013

Supervisory Committee

Dr. Fraser Hof, Department of Chemistry
Supervisor

Dr. Jeremy Wulff, Department of Chemistry
Departmental Member

Dr. Leigh Anne Swayne, Division of Medical Sciences
Outside Member

Abstract

Supervisory Committee
Dr. Fraser Hof, Department of Chemistry
Supervisor
Dr. Jeremy Wulff, Department of Chemistry
Co-Supervisor or Departmental Member
Dr. Leigh Anne Swayne, Division of Medical Sciences
Outside Member

Methyl reader proteins recognize and bind to post-translationally methylated residues and have functional roles in diverse cellular processes including gene regulation, development and oncogenesis. The CBX polycomb paralog family of methyllysine readers recognize trimethyllysine lysine residues on histone tail 3 and repress transcription by compacting chromatin. The polycomb paralogs form multi-protein complexes that silence the expression of tumour suppressor genes, and play important roles in regulating cell cycle and differentiation. Each paralog is structurally similar, yet has distinct functions, of which many are unknown.

My work has focused on the design and synthesis of CBX inhibitors and on the development of new methodologies for the discovery of inhibitors targeting methyllysine readers. In this work, I report on a series of potent peptidic inhibitors that selectively target the CBX polycombs, as well as the first selective inhibitor for the family member CBX6, and dual-active inhibitors that target CBX6/CBX8. The results demonstrate the potential to achieve selectivity through interactions outside of the methyllysine binding domain. Structural determinants in the binding pocket of each protein that differ within the family and give rise to selectivity were discovered. I will also report on a series of peptidomimetic CBX inhibitors that are active in cells. Cellular active inhibitors are critical for understanding the biological role of each CBX protein and their potential as therapeutic targets.

New high-throughput approaches are needed to efficiently target methyllysine readers by chemical inhibition. I describe in this work a strategy for creating massive libraries of phage-displayed peptidic inhibitors containing methyllysine mimics. Synthetic optimization on cysteine-containing peptide phage constructs allowed for the successful installation of Kme3 mimics. This is the first report of a post-translational methylated peptide phage library. The methodology I developed can be used in a

synthetic chemistry-driven adaptation of traditional phage display for the screening of millions of peptide-based compounds. Strategies that allow for diversity and high throughput screening will aid in future efforts in targeting the highly similar CBX proteins.

Table of Contents

List of Tables

List of Figures

Abbreviations

Abu	2-aminobutyric acid
Ac	acetyl
Ac_2O	acetic anhydride
AlphaScreen	amplified luminescent proximity homogeneous assay screen
Asp	aspartic acid
Asn	asparagine
aDMA	asymmetric dimethyl arginine
BCL9	b-cell CLL/lymphoma 9 protein
BET	bromo- and extra-terminal domain
BIA	Iodoacetyl-PEG2-Biotin
BRD	bromodomain
BrNMe3	(2-bromoethyl)trimethylammonium bromide
CBX	chromobox
CBX1	chromobox 1
CBX2	chromobox 2
CBX3	chromobox 3
CBX4	chromobox 4
CBX5	chromobox 5
CBX6	chromobox 6
CBX7	chromobox 7
CBX8	chromobox 8
CDY	chromodomain Y
CDYL2	chromodomain Y-like 2
CHD	chromodomain helicase DNA binding protein
CHD4	chromodomain helicase DNA binding protein 4
ChIP	chromatin immunoprecipitation
Chromodomain	chromatin organization modifier
Cys	cysteine
DCM	dichloromethane
DIPEA	*N,N*-diisopropylethylamine
DMF	*N,N*-diisopropylethylamine
DMSO	dimethylsulfoxide
DNA	deoxyribonucleic acid
DTT	dithiothreitol
EED	embryonic ectoderm development
ESI	electrospray ionization
EZH2	enhancer of zeste 2 polycomb repressive complex 2 subunit
FACS	fluorescence-activated cell sorting
FDA	food and drug administration
FITC	fluorescein isothiocyanate
Fmoc	fluorenylmethyloxycarbonyl
FP	fluorescence polarization
Glu	glutamic acid
G9a	euchromatic histone-lysine N-methyltransferase 2

HCC	hepatocellular carcinoma
HEPES	2-[4-(2-hydroxyethyl)piperzin-1-yl]ethanesulfonic acid
H2A	histone 2A
H2B	histone 2B
H3	histone 3
H4	histone 4
H3K4me3	trimethylated lysine 4 on histone 3
H3K9me3	trimethylated lysine 9 on histone 3
H3K27me3	trimethylated lysine 27 on histone 3
H3K36me3	trimethylated lysine 36 on histone 3
H3K9me3	trimethylated lysine 9 on histone 3
H4K20me3	trimethylated lysine 20 on histone 4
H3S10	serine 10 on histone 3
HBTU	*N,N,N',N'*- tetramethyl-O-(1H-benzotriazol-1-yl)uronium hexafluorophosphate
HP1	heterochromatin protein
HPLC	high-performance liquid chromatography
IC_{50}	inhibitory concentration that reduces effect by 50 %
ING	inhibitor of growth factor
ITC	isothermal titration calorimetry
iTCEP	immobilized tris(2-carboxyethyl)phosphine
K	lysine
Kme	monomethylated lysine
Kme2	dimethyl lysine
Kme3	trimethyl lysine
K_d	dissociation constant
LCMS	liquid chromatography mass spectrometry
Leu	leucine
LR-ESI-MS	low resolution electrospray ionization mass spectrometry
L3MBTL1	lethal (3) malignant brain tumor-like protein 1
L3MBTL3	lethal (3) malignant brain tumor-like protein 3
MeCN	acetonitrile
MD	molecular dynamics
MLL5	mixed lineage leukemia 5
MMA	monomethyl arginine
MORF4L1	mortality factor 4-like protein
MTT assay	3-(4,5-Dimethylthiazol-2-yl)-2,5-diphenyltetrazolium bromide assay
NSD3	nuclear receptor binding SET domain protein 3
N.T	not tested
NUP98	nucleoporin 98
OBOC	one-bead-one-compound
Pc	polycomb complexes
PDB	protein data bank
PHD	plant homeodomain
Phe	phenylalanine

PHF20L1	PHD finger protein 20-like protein 1
PRC	polycomb repressive complex
PRC1	polycomb repressive complex 1
PRC2	polycomb repressive complex 2
PTM	post-translational modification
PWWP	proline tryptophan tryptophan proline
Pygo2	pygopus homolog 2
NMR	nuclear magnetic resonance
R	arginine
RING	really interesting new gene
RNA	ribonucleic acid
r.t	room temperature
SAM	s-adenosylmethionine
SAR	structure activity relationship
sDMA	symmetric dimethyl arginine
Ser	Serine
SET	Su(var)3-9 and enhancer of zeste proteins
SETDB1	histone-lysine-N-methyltransferase SETDB1 protein
SPR	surface plasmon resonance
SUMO	small ubiquitin-like modifier
TCEP	tris(2-carboxyethyl)phosphine
TFA	trifluoroacetic acid
TRIS	tris(hydroxymethyl)aminomethane
Trp	tryptophan
TTD	tandem tudor domain
Tyr	tyrosine
UHRF1	ubiquitin-like containing PHD and RING finger domains 1
WD40	tryptophan-aspartic acid repeat domain
WDR5	WD repeat-containing protein
53BP1	p53-binding protein 1

Acknowledgments

First, I would like to thank those who encouraged me (or at least taught me enough) to start this endeavour. Thank you to Dr. Alison Thompson and Dr. Estelle Marchal. Estelle taught me so many foundational synthetic skills and her patience was astounding. Alison instilled in me a passion for chemistry and I am grateful for all the support and encouragement she has provided me.

My biggest thank you goes to my PhD supervisor, Dr. Fraser Hof. I don't even have the words. Fraser's creativity, passion and humour are contagious. I am grateful to Fraser for the immense amount of knowledge that he has taught me, both scientifically and professionally. Fraser's endless wisdom, patience and support (even if the wisdom/support was sometimes saying "sort it out Milo") has been critical to my growth as scientist, writer, and communicator. Thank you Fraser for believing in me, pushing me to do better, and for cultivating a positive and fun lab environment that made this experience a challenging and enjoyable one. I cannot express how grateful I am for all of the amazing opportunities you have given me. I would also like to thank Rebecca Hof, for her continued scientific support, biochemistry advice (and life advice) and for all of the protein grows, driving lessons, lab parties and laughs.

Thank you to my girl gang (Aiko, Alex, Alli, Cait, Codie, Emily, Meagan), for lifting me up – it's an honour to be surrounded by such strong women. An extra big thank you to Meagan for putting up with me in the final months, weeks and days of writing this. Thank you to the Dream Team (Alok Shaurya, Jun Chen, Ronan Hanley) – without your friendship and scientific advice I truly would not have gotten through this degree. The three of you are among the smartest people I know – and inspire me everyday. Thank you to Charles – for always having my back in times of need. Nathan, thank you for helping me make the leap across the country to begin with.

Thank you to my friends, co-workers and colleagues and to all the current and past members of the Hof lab. A special thank you to Michael Gignac for the immense amount of work that helped make this thesis possible (you truly are a protein wizard). Thank you to James McFarlane for all his computational work that has helped me understand binding and selectivity of my ligands – and for all the nerdy chats over the years about

CBX proteins. Thank you to all the powerhouse undergraduate researchers whom I have had the pleasure of working with (special thanks to Janessa Li, Shanti Horvath, Katrina Vizely, Chelsea Wilson and Tyler Brown for their significant contributions). Janessa tried so many different ways to get the Chapter 2 compounds into cells, Shanti for setting up our CBX SPR assay, Chelsea Wilson made for making many of the peptides in Chapter 4, and Tyler for being a jack-of-all-trades helping to finish up all of my loose ends. Thank you to Katrina, for joining me years ago on the phage-rage train. Your efforts, ideas, and hard work were critical to the success of Chapter 5. Thank you to Chakri Simhadri for helping me get started in my first year here. Thank you to the Wulff group for synthesis advice and for always having the chemicals I needed.

I would like to thank my committee (Jeremy Wulff, Leigh Anne Swayne) and members of the Department of Chemistry who have helped me throughout my years here. A special thank you to Jeremy for all of his excellent advice in relation to my project, and for teaching me so much about organic chemistry. Thank you to Chris Barr (NMR support) and to Ori Granot (MS support), for all the hours (so many hours) of assistance and for all you have taught me. Thank you Andrew Macdonald and Sean Adams, for fixing all of our broken things, and for being a such a pleasure to work with.

Thank you to the West-Coast Ride-to-Live Vancouver Island and Prostate Cancer Foundation of BC for their support of my research. I am honoured to have been acknowledged and very grateful for the support that has enabled so much of my research.

Lastly, I would like to thank my family and chosen family. Thank you to my godparents Duncan Wilson and Rowly Johnson for the all of the love, treats, and unfaltering belief in me (and the rest of the Wilson gang - Katya and Greg!). Thank you to the Godoroja and Pearse family (Thank you Olga, Tania, Tony and Serg for being my adopted west coast family). Thank you to my stepfather, Ian, and my stepmother Georgia, for always supporting my interest in science. Thank you to my father, Winston, for reminding me to go outdoors and get fresh air. Last but not least, thank you to my mother, Olga. Thank you for your tireless efforts in trying to make me believe that I could achieve my goals. Thank you for teaching me the importance of independence and hard work. I have always been inspired by your incessant need to ask a million questions, and by your dedication to serve your community.

Dedication

This is for you Bubi,

You always told me that boredom is unacceptable, because there is always something to learn. Thank you for imparting upon me insatiable curiosity.

Chapter 1. Introduction

Sections of this chapter were previously published. Three publications have been adapted to form this chapter. The publications and my contributions are described below.

Natalia Milosevich and Fraser Hof*
Chemical inhibitors of epigenetic methyllysine reader proteins. **2016**. *Biochemistry*. 55, 1570.
Reproduced in part with permission. Copyright 2016 American Chemical Society.

N.M. conceived of and wrote the manuscript. F.H. defined the scope and provided input for revisions.

Natalia Milosevich, Zoey Warmerdam and Fraser Hof*
Structural aspects of small molecule inhibition of methyllysine reader proteins. **2016**. *Future Medicinal Chemistry*. 8, 1681.
Adapted with permission of Future Science Group.

N.M. conceived of and wrote the manuscript. F.H. defined the scope and provided input for revisions.

Natalia Milosevich* and Fraser Hof
Inhibitors of methyllysine reader proteins: foundations of next generation epigenetic therapies. In *Epigenetic Inhibitors: Targeting Diseases Through the Development of Epigenomic Medicine.* Campeau, E. Wiley & Sons. New Jersey. 2019
Reproduced with permission of Wiley & Sons 2019.

N.M. conceived of and wrote the manuscript, and served as corresponding author. F.H. provided input for revisions.

1.1. Prologue

Epigenetics is a term that describes heritable modifications that regulate gene expression without altering the DNA sequence. Epigenetic processes are essential to both normal and diseased organism pathways. From fundamental development events, to progression of cancer, epigenetic mechanisms control the organization and readout of our genetic code. Environmental factors such as stress, viruses, chemicals, nutrition and radioactivity are all suspected in causing epigenetic changes that can result in the development of a wide variety of diseases.

Modifications to chromatin complexes that package our DNA are one form of epigenetic control. Methylation of lysine is one of the many possible covalent and reversible modifications that control gene transcription. Molecular recognition of methylated lysines by proteins called methyllysine readers is an area of epigenetic research that has only recently began to be understood. The full story of methylation signaling and its effects on cellular processes is yet to be unraveled.

Methyllysine reader proteins recognize and bind to post-translationally methylated lysine residues. They read and execute the commands issued by protein methyltransferases and play functional roles in diverse cellular processes including gene regulation, development, and oncogenesis. Efforts to inhibit these proteins are relatively new. Only a small number of methyl reader proteins belonging to the chromodomain, malignant brain tumor domain, WD40, plant homeodomain finger and tudor domain families, have been targeted by chemical inhibitors. Chemical probes for methyllysine reader proteins are critical for enhancing our understanding of these proteins and their ability to regulate cellular events.

This thesis will summarize my work on creating inhibitors for methyllysine reader proteins. The main focus of my research has been on developing inhibitors for polycomb paralogs—an important family of methyllysine reader proteins. I will also discuss my efforts to develop new methodologies to better target methyllysine reader proteins. This chapter will review methyllysine reader proteins, including their biological role and the development of inhibitors. The structural determinants for inhibition of polycomb paralog proteins will be discussed, along with an analysis of the molecular interactions that control potency and selectivity for inhibitors within the family.

1.2. Epigenetics and mechanisms of epigenetic control

Epigenetic mechanisms play important roles in almost all cellular processes. Two main forms of epigenetic control are DNA methylation and modifications to the histone proteins that package DNA.[1] Epigenetic mechanisms related to DNA methylation have been reviewed elsewhere,[2] and DNA methylation inhibitors have been used clinically since 2004.[3] This thesis will focus mainly on histone modifications, which are relatively less well understood. Chemical modifications to histone proteins act in combination with other protein complexes to control chromatin structure and function.[4]

Chromatin is comprised of DNA wound around histone proteins that are packaged into structural units called nucleosomes. Each nucleosome consists of two copies each of four core histone proteins (H2A, H2B, H3 and H4) and on average 146 base pairs of DNA (Figure 1.1A). The spacing and compaction of nucleosomes is highly regulated and the density dictates whether the chromatin is transcriptionally active or inactive.[5] Heterochromatin refers to transcriptionally inactive chromatin that is densely packed, whereas euchromatin refers to chromatin that is loosely packed and transcriptionally active. The landscape of chromatin changes due to activity of histone-modifying writer, reader and eraser proteins.[6] These proteins act to add, recognize or remove covalent chemical modifications and the resulting chemical changes to the histone proteins influence the state and accessibility of chromatin.

1.3. Post-translational modifications and lysine methylation

Post-translational modifications (PTMs) are chemical signals that add increased diversity and control over cellular signaling processes.[7-8] The term PTM refers to covalent modifications of proteins that occur following translation of that protein in a cell. The roles of PTMs in epigenetic pathways were first discovered among the large number of PTMs that occur on unstructured protruding N-terminal tails of histone proteins, where some of the most prominent examples are acetylation, phosphorylation, ubiquitination, and methylation.[7, 9]

PTMs on histone proteins are diverse in terms of location, size and chemical structure. PTMs can occur on the solvent exposed face of nucleosomes, lateral surface, histone-histone interface and most commonly on the unstructured N-terminal tails of the histone proteins.[4] Histone tail modifications alter chromatin structure by changing electrostatic interactions and/or by recruitment of non-histone proteins.[6] The primary sites of covalent modifications are lysine, arginine, serine, tyrosine, threonine and glutamate histone residues.[10] Lysine and arginine residues are abundant on solvent-exposed areas of histone proteins (Figure 1.1B),[11] and lysine residues in particular are highly post-translationally modified.[12] PTM size varies greatly. Smaller PTMs such as methylation, acetylation, propionylation, butyrylation, crotonylation and phosphorylation are all a mass change of less than 100 Da whereas the largest modifications are the attachment of 8-15k Da ubiquitin and SUMO (small ubiquitin-like modifier) proteins. The chemical make-up of histone PTMs including size, charge, hydrophobicity and polarity, all effect the mobility, compaction and overall structure of chromatin.

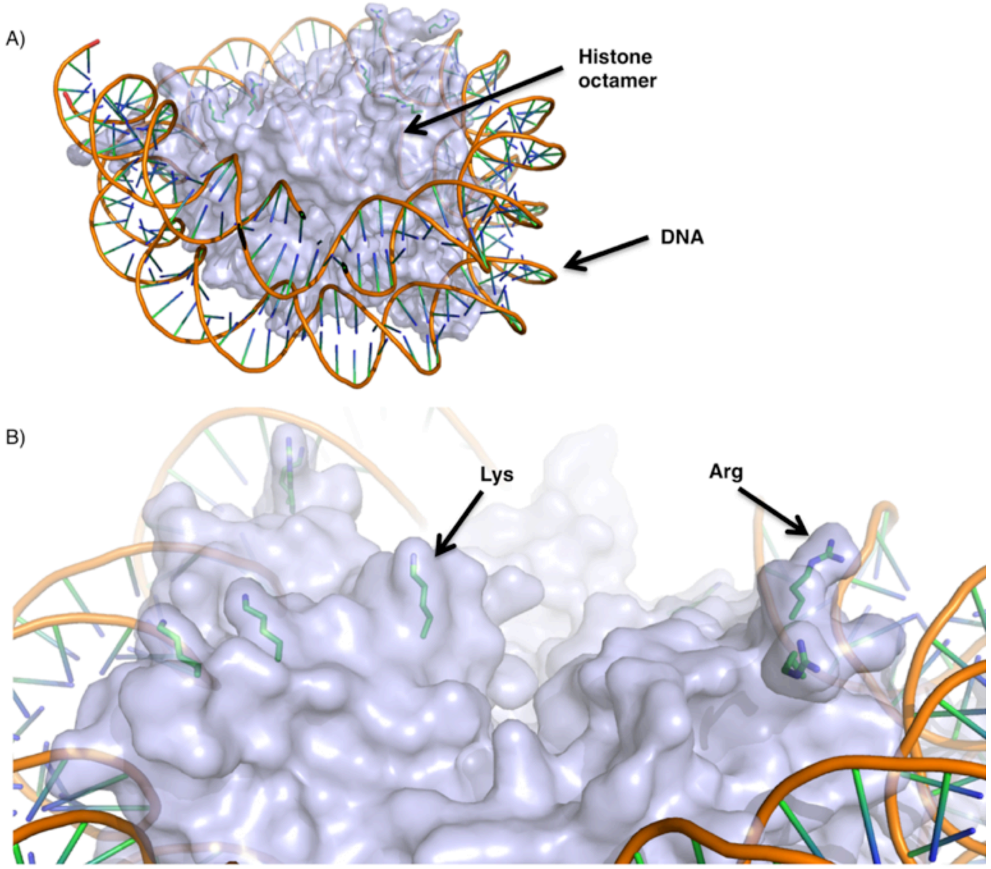


Figure 1.1. Crystal structure of nucleosome and solvent exposed lysine and arginine residues (pdb: 3C1C). A) Crystal structure of nucleosome highlighting the DNA and histone octamer proteins, B) Close-up of solvent exposed lysine and arginine residues that may be subjected to post-translational modifications.

Post-translational methylation is a term that most often describes amino acid side-chain methylations. “Methylation” in this context actually describes a family of related PTMs, including mainly mono, di and trimethyllysine (Kme, Kme2, Kme3), monomethylarginine (MMA), and asymmetric and symmetric isomers of dimethylarginine (aDMA and sDMA). Methylation of lysine residues on histone proteins controls gene expression by changing chromatin architecture, controlling genome stability, and altering the cell cycle.[13] Although the canonical ‘histone code’ methylation sites were the first discovered and are the most heavily studied,[9, 13-15] it is increasingly clear that methylation occurs on and regulates many non-histone proteins that participate in epigenetic and gene regulation pathways.[16] Non-histone protein lysine methylation affects protein stability, degradation, activation, and recruitment of binding partners, and is important in many disease pathways including cancer progression.[16-17] The number of known sites of lysine methylation continues to grow with the development of new technologies for mapping and understanding the protein methylome.[18]

Protein methylation is a dynamic and reversible modification. Lysine methylation is added by ‘writer’ methyltransferases,[19] removed by ‘eraser’ demethylases,[20] and recognized by methyllysine ‘reader’ proteins (Figure 1.2B). There are three families of ‘writer’ methyltransferases that add methyl groups from *S*-adenosylmethionine to their targets. Members of two of these families are known to methylate lysines—SET-domain-containing proteins and DOT1-like proteins.[21-23] Lysine demethylases were first discovered in 2004 and many others have since been discovered.[20, 24-25] Inhibitors of the enzymatic action of methyltransferases and demethylases are proceeding through pre-clinical and clinical trials for many indications, but are outside of the scope of this thesis.[26-28]

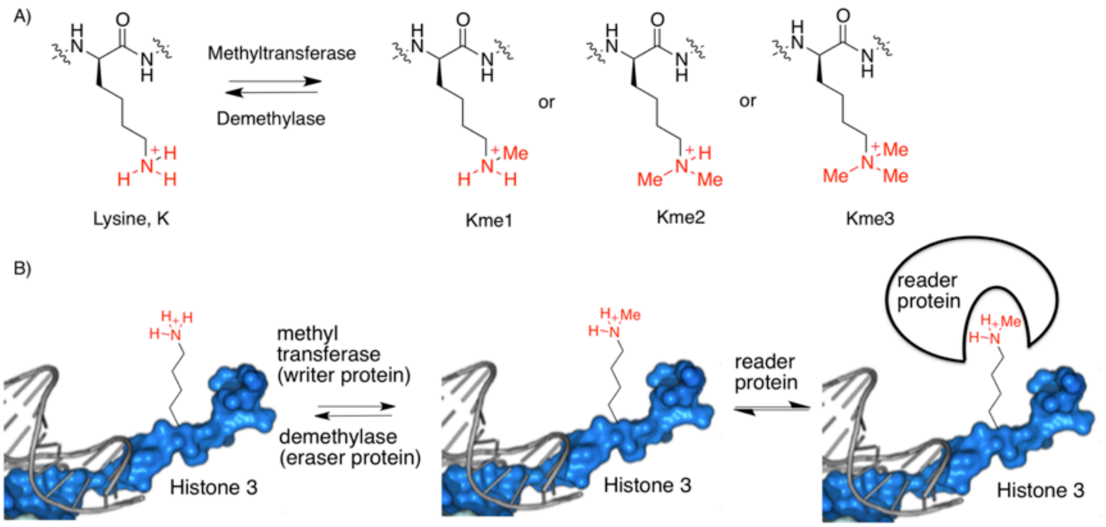


Figure 1.2. Depiction of histone tails showing the location and possible degrees of lysine methylation. A) Methylation states of lysine, B) illustration of interaction between methylated histone tails and writer/reader/eraser proteins.

Lysine methylation on histone tails can result in either gene activation or gene silencing of the associated chromatin, depending on the site of methylation. Methylation marks on Histone 3 Lysine 9 (H3K9), H3K27 and H4K20 are generally known to silences genes, whereas H3K4, H3K36 and H3K79 methylation marks result in active transcriptional states.[15] Different degrees of methylation at an individual site can also have distinct effects: for example, in chromatin H3K4me1 is located near enhancers and flanks promoters, H3K4me2 is often near transcription factor binding sites, and H3K4me3 is most often at promoters where it is associated with transcriptional initiation.[29] Understanding the mechanisms of gene regulation is further complicated by the interplay of interactions between methylated lysine marks on histones and other histone post-translational modifications and modifications of DNA. The bivalent methylation marks of H3K4me3 and H3K27me3 together control regulatory genes in embryonic stem cells maintaining pluripotency,[30] whereas the bivalent H3K4me3 and H3K9me3 marks control adipogenic regulatory genes in mesenchymal stem cells.[31] Other PTMs that are near methylation marks can alter a reader protein's ability to bind to that methylation mark, such as when the phosphorylation of H3S10 prevents recognition of H3K9me2 by the methyl reader heterochromatin protein 1 (HP1).[32] Research continues to illuminate the many interacting PTMs that work in concert to control gene transcription.

1.4. Methyllysine reader proteins

Methyllysine and methylarginine 'readers' are proteins with domains that recognize and bind to methylation marks.[33] They are responsible for conveying the methylation signal downstream, and they do so either by having their own catalytic functions or by recruiting other proteins to sites of action through the formation of multi-protein complexes.[34] The major families of methyl reader domains are the Plant homeodomain (PHD) fingers, WD40 repeat domains, Chromatin Organization Modifier domains (Chromodomains), tudor domains, agenet domains, Proline-Tryptophan-Tryptophan-Proline (PWWP) domains, and Malignant Brain Tumor (MBT) domains.[35] The latter five domains in this list are also referred to as the Royal Family domains. Reader proteins of methyllysine and methylarginine marks sometimes occur within the same families.

Molecular recognition by methyllysine reader proteins is carried out by a 'aromatic cage' motif that is made up of multiple aromatic amino acid side chains (Figure 1.3).[34]. Binding of the methylammonium group of the methyl lysine residue to the protein is achieved predominantly through multiple cation-pi interactions with the electron-rich aromatic residues of the protein's aromatic cage (Figure 1.3).[36-38] Hydrophobic interactions and desolvation of the quaternary amine and aromatic cage also contribute to the overall binding event.[37]

Most methyllysine readers have evolved to achieve recognition of a specific methylation state of lysine at one or more specific sites in the proteome. Methylated lysines Kme1, Kme2, and Kme3 vary in their size, distribution of positive charge, hydrophobicity, and ability to donate hydrogen bonds. These differences allow for specific methyllysine reader proteins to differentially recognize methylated lysine. For example, Kme1 and Kme2, which retain NH hydrogen bond donors, are often complemented by H-bond acceptor residues integrated into the aromatic cage pocket to provide methylation state specificity (Figure 1.3A, D, F).[34, 39-40]

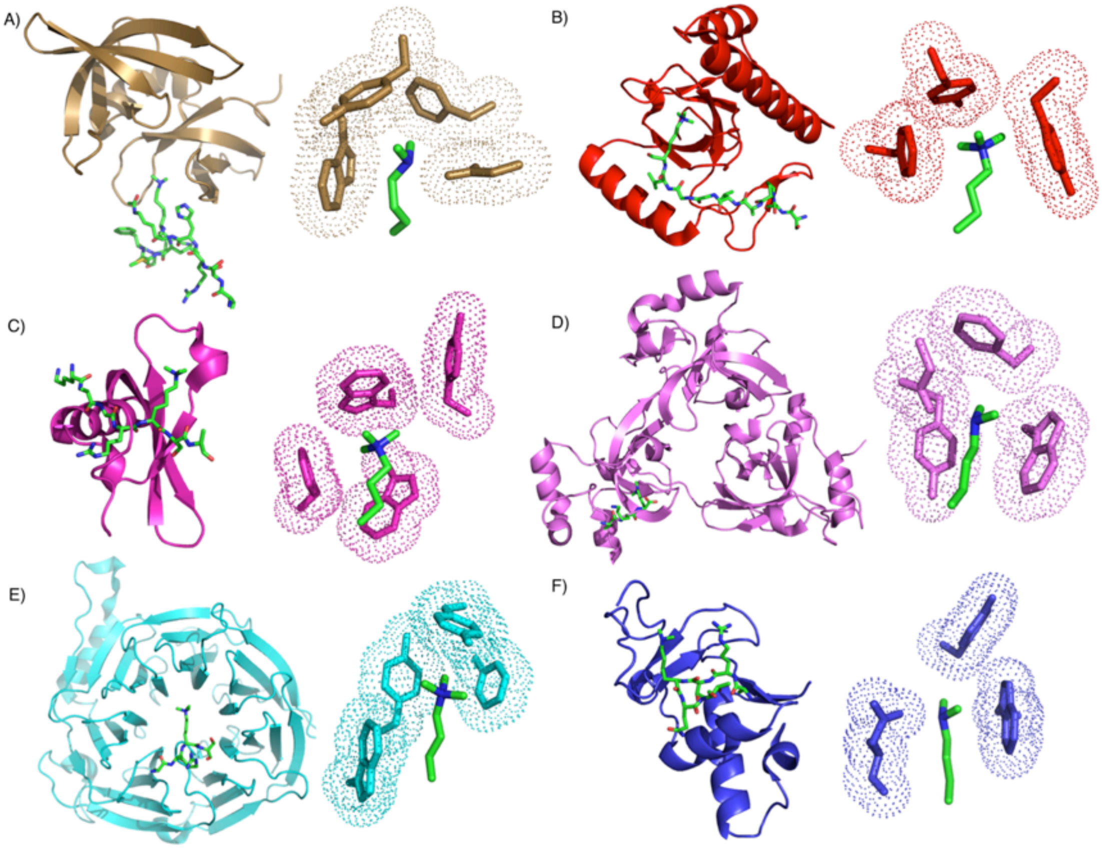


Figure 1.3. Aromatic cage motifs for different families of methyllysine reader domains. A) 53BP1 tandem tudor domain in complex with p53K381acK382me2 (pdb: 4X34), B) BRPF1 PWWP domain in complex with H3K36me3 (pdb: 2X4W), C) CBX7 chromodomain in complex with H3K27me3 (pdb: 4X3K), D) L3MBTL1 MBT domain in complex with H1.5K27me2 (pdb: 2RHI), E) EED WD40 repeat domain in complex with H3K27me3 (pdb: 3JZG) F) Pygo2 PHD finger in complex with the B9L HD1 domain and H3K4me2 (pdb: 4UP0).

In addition to their methyllysine recognizing motif, which dictates selectivity for different degrees of methylation, most methyllysine reader proteins have evolved to recognize motifs beyond the methylated lysine itself. Histone 3, for example, can be methylated at many positions, predominately lysine 4 (H3K4), H3K9, H3K27, H3K36, and H3K79. The position and degree of methylation is important for signaling, and reader proteins execute their signals faithfully by recognizing and binding to residues in the sequences surrounding each methylated lysine.[34, 41-44]

Methyllysine readers can act in combination with other epigenetic modules to achieve readout of multiple PTMs in concert.[45] The constant changing of chromatin structure is produced from the many PTM signaling pathways working together. Many methyllysine reader proteins form multi-protein complexes that contain multiple subunits of other PTM writers, erasers and readers. Specificity for particular PTMs can be increased by PTM binding modules recognizing multiple marks together.[46-47]

Mutations or changes in expression of lysine methylation, or in the writers, readers and erasers of methyllysines are observed in the development and progression of many disease states. This chapter provides further discussion of the cellular roles and disease implications for methyllysine reader proteins, with a focus on the polycomb paralog CBX proteins.

1.5. Biological and structural review of methyllysine reader proteins

Methyllysine reader proteins vary in their biological roles and many are implicated in a broad range of disease states. In general, these proteins play important biological roles in development, differentiation, and regulating cellular processes such as cell cycle regulation, DNA damage and stress responses.[48-49] Evidence suggests that histone methylation has a role in cancer and specific methylation marks are linked to more aggressive cancer and poor survival outcomes.[13, 50-52] Changes in expression of many methyl reader proteins are seen in a broad range of cancers.[13, 53-58]

Recognition of methylated lysine is categorized into two main types; surface groove and cavity insertion.[34] Domains that recognize lower methylated states contain a residue capable of accepting a hydrogen bond from the side chain of Kme1 or Kme2.[34, 39] In addition, these domains often have deep and narrow binding pockets and fall under the ‘cavity insertion’ category of binding. Domains that recognize Kme3 are categorized as ‘surface groove binders’ and have shallower binding pockets that contain more hydrophobic aromatic residues.[41, 59-63] This section will describe the structural aspects of methyllysine recognition and the biological role of the reader for each methyllysine reader domains that has been targeted by chemical inhibition.

1.5.1 MBT domains

Currently there are nine proteins that are known to contain MBT domains and each contain between 2–4 MBT domains.[64] MBT domains recognize mono- and dimethyl lysine on H3 and H4 tails which leads to transcriptional repression due to the compaction of chromatin.[56] Despite multiple repeating subunits, each of the MBT domain containing proteins has one MBT repeat that serves as the primary methyllysine binding motif.[56]

The aromatic cage of MBT domains consists of three aromatic residues (Trp, Tyr and Phe) and hydrophilic residues (Asp and Asn) capable of hydrogen bonding. The methyllysine binding pocket is too narrow to bind Kme3 and a conserved aspartate residue makes a key hydrogen bond with Kme1 and Kme2 residues. The hydrophobic aromatic cage favours interactions with Kme1 and Kme2 over unmethylated lysine.[65] MBT domains have a high-predicted druggability[66] and this can in part be explained by their cavity insertion mode of binding to methylated lysines.

The two MBT domains that have been targeted by chemical inhibition are lethal(3) malignant brain tumour-like protein (L3MBTL) 1 and 3 (Table 1.1, Figure 1.4). L3MBTL1/3 have nearly identical aromatic cages but differ in the residues and domain structure outside of the aromatic cage. Both proteins contain aromatic cages that are defined by the aromatic residues Phe, Trp and Tyr and the hydrogen-bonding residue Asp (Figure 1.4A).

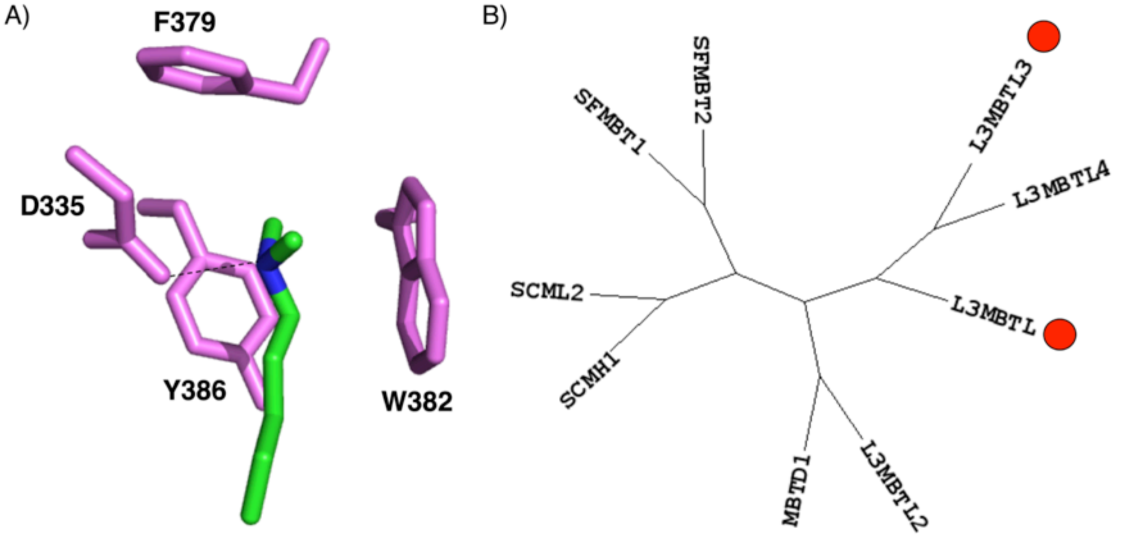


Figure 1.4. Aromatic cage of MBT domain and phylogenetic tree of MBT family. A) Aromatic cage of L3MBTL1 in complex with H4K20me2 (pdb: 2RJE), B) Phylogenetic tree of MBT family obtained and adapted from the Structural Genomic Consortium Chromohub.[64] MBT domains targeted by chemical inhibition are labeled with red circles.

MBT domains such as L3MBTL1/3 bind methylated histone peptides in a non-sequence specific manner. Promiscuity in recognition of Kme1 and Kme2 is because the molecular recognition of the methylation marks is largely confined to the narrow methyllysine binding aromatic cage.[67] However, *in vivo* studies have demonstrated that some MBT domains do bind in a sequence specific manner.[67-68]

L3MBTL1 contains 3 MBT domains in a propeller conformation and binds H4K20me1 and H1bK26me1/2. Compaction of chromatin by L3MBTL1 blocks transcription machinery and this is dependent on the aforementioned methylation marks.[69] Similarly, L3MBTL1 recognition of p53K382me1 compacts chromatin and represses p53 activity.[70]

Much of the research on the biological role of MBT domains has focused on the L3MBT family (L3MBTL1/2/3/4) because of their role in oncogenesis.[56] MBT domains are named after the tumour suppressor *lethal(3) malignant brain tumour (l(3)mbt)* genes that were discovered in *Drosophila*.[71] Loss-of-functions studies show that depletion of L3MBTL1 in human cells leads to replicative stress, DNA damage and activation of DNA damage response pathways.[72] Deletion or reduction in L3MBTL1 expression is proposed to impact hematopoietic malignancies by reducing genomic stability.[72] Loss of L3MBTL1 causes erythroid differentiation in human hematopoietic stem/progenitor cells and this contributes to the development of the chronic myeloproliferative disorder Polycythemia Versa.[73] Knockdown studies with L3MBTL1 have demonstrated its role in regulating neuronal excitability and synaptic efficacy.[74] Samples from patients with medulloblastomas found homozygous deletion of SCML2, L3MBTL2 and L3MBTL3.[75] There has been less research into the role of L3MBTL3 in the development of disease, however, it has been linked to hematopoiesis alongside L3MBTL1.[73, 76]

1.5.2 Tudor Domains

Tudor domains recognize methylated lysine and arginine residues and are critical for controlling a variety of RNA pathways as well as cellular responses to DNA damage.[77-80] Tudor domains that recognize methylated lysine residues contain tandem tudor domains (TTDs) and primarily read H3K4me3, H4K20me2/3, H3K9me3 and H3K36me3 marks.[81] Tudor domains engage histone residues surrounding the methylation mark allowing for selective recognition.[81] Binding of tudor domains to histone tails is in part regulated by the neighbouring reader modules in the tudor-domain containing protein complex.[46]

The aromatic cage of tudor domains capable of methyllysine recognition differ depending on their substrate. The majority bind to higher methylation states of lysine and contain 3-4 aromatic residues.[81] The tudor domains of 53BP1, spindlin1 and UHRF1 have been targeted by chemical inhibition (Table 1.1, Figure 1.5).

The tudor domain of p-53 binding protein 1 (53BP1) contains 4 aromatic cage residues (W1495, Y1502, F1519 and W1523) and an aspartic acid residue D1521 (Figure 1.5A). Selectivity for Kme1/2 over Kme3 is due to the hydrogen bonding ability of D1521. 53BP1 forms interactions with the residues surrounding its histone substrate H4K20me2, including H4K16 and H4H18 and this increases the specificity and affinity.[40]

The second tudor domain of Spindlin1 that binds to H3K4me3 contains 4 aromatic cage residues (F141, W151, Y170, and Y177) (Figure 1.5B). Spindlin1 binds with high affinity to H3K4me3 and this is in part due to the several hydrogen bonds formed between the neighbouring arginine residues H3R2 and H3R8 and negatively charged residues in the protein.[82]

The Ubiquitin-like with PHD and RING finger domains 1 (UHRF1) is a multi-domain protein consisting of a ubiquitin-like domain, tandem tudor domain (TTD), plant homeodomain (PHD), SET and RING associated domain (SRA) and RING domain. The TTD of UHRF1 recognizes the repressive H3K9me2/3 mark and binds in a cooperative fashion with the PHD domain that binds the N-terminal tail of unmodified H3.[83-85] The aromatic cage of UHRF1-TTD consists of F152, Y188 and Y191 (Figure 1.5C). The TTD-PHD module is highly selective for H3K9me3 and both modules including an intermodule linker are essential to the high affinity interaction.[85]

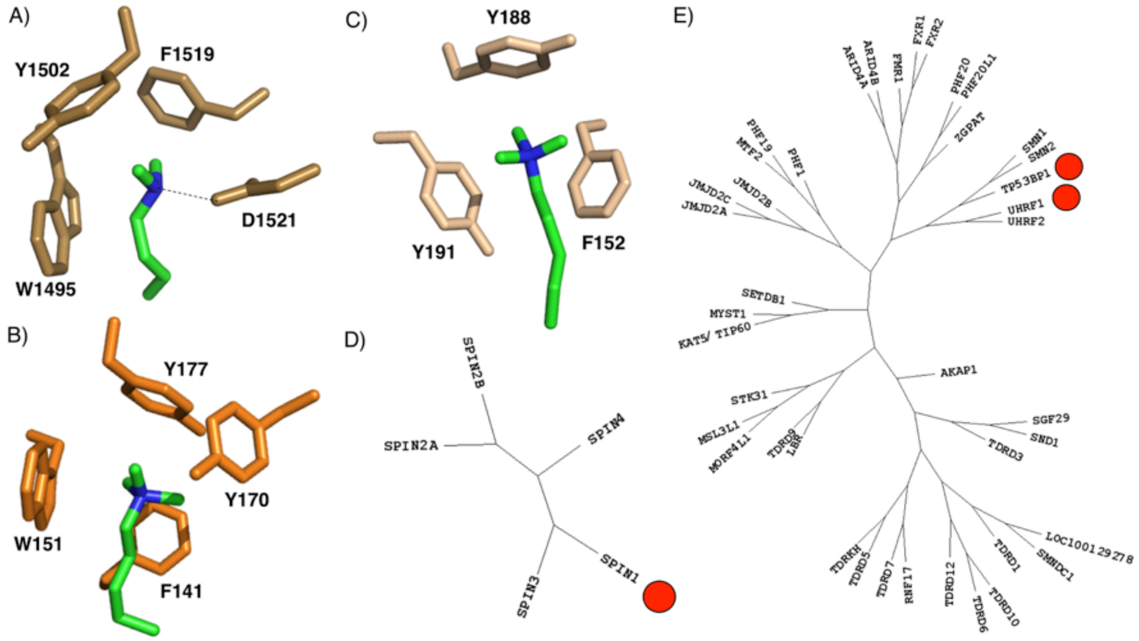


Figure 1.5. Tudor and SPIN domain aromatic cages and phylogenetic tree. A) Aromatic cage of 53BP1 in complex with p53K382me2 (pdb: 4X34), B) Aromatic cage of Spindlin1 in complex with H3K4me3 (pdb: 4H75), C) Aromatic cage of UHRF1 in complex with H3K9me3 (pdb: 4GY5), D) Phylogenetic tree of Spindlin family,[64] E) Phylogenetic tree of tudor family obtained and adapted from the Structural Genomic Consortium Chromohub.[64] Domains targeted by chemical inhibition are labeled with red circles.

The tudor domain containing methyllysine binding protein 53BP1 is a key player in DNA damage repair pathways. Methyllysine recognition by the tudor domain of mammalian p53 binding protein (53BP1) recruits DNA damage repair proteins to DNA double-strand break (DSB) sites and modulates the function of the tumour suppressor protein p53.[81] Ultimately this recruitment results in DNA damage repair and cell cycle arrest.[86-87] Recognition of H4K20me1/2 by 53BP1 has been proposed to promote the non-homologous recombination (NHEJ) pathway in response to DSB.[88]

53BP1 also has functions related to the cancer-associated protein BRCA1. The DNA damage repair gene BRCA1 is critical in maintaining and repairing DNA.[89] A major risk factor for breast and ovarian cancer is the presence of mutated BRCA1 genes.[90-91] Loss of 53BP1 promotes survival of BRCA1-deficient tumour cells after DNA damage is associated with BRCA-mutated breast cancers.[92] Patient samples of colorectal cancer showed decreased levels of 53BP1 was associated with poor survival and increased proliferation.[93] 53BP1 is also a substrate of ataxia telangiectasia mutated

(ATM) kinase.[94] 53BP1 knockout mice were shown to be immune deficient, growth retarded, and prone to cancer.[94] The role of 53BP1 as a tumour suppressor is evident in the many examples of decreased expression in cancerous cells.

Spindlin1 is important in spindle organization, chromosomal stability[95] and development of cancer.[96-98] Spindlin1 is a coactivator of transcription of rRNA genes, WNT-target genes and genes regulated by the MAZ transcription factor.[98-99] In terms of its oncogenic role, spindlin1 was shown to promote cancer cell proliferation through activation of the WNT/TCF-4 signaling pathway.[97] Increased expression levels of spindlin1 were found in clinical tumour samples, including ovarian, non-small cell lung, hepatic carcinomas and liposarcoma.[97-98]

UHRF1 is a key regulator of gene expression as it controls crosstalk between DNA methylation and histone modifications. UHRF1 is essential in the maintenance and propagation of heritable methylation of DNA. The SET and RING associated (SRA) domain of UHRF1 is responsible for recognition of hemi-methylated DNA and recruits DNMT1 to chromatin for maintenance of DNA methylation.

UHRF1 is overexpressed in many types of cancer and inhibitors apoptosis by silencing of tumour suppressor genes.[100-101] UHRF1 plays a key role in the hypermethylation of tumour suppressor genes. Expression of well-characterized tumour suppressor genes such as RB1, $p16^{INK4A}$, $p14^{arf}$ and BRCA1 are all regulated by UHRF1 binding to the methylated promoter regions of each gene.[101-104] UHRF1 is overexpressed in many human cancers,[105] and high circulatory levels and easy detection have identified it as a independent diagnostic marker.[106-108]

1.5.3 PHD Fingers

PHD fingers are a large family of proteins with diverse biological roles and targets. PHD fingers have been linked to a wide array of human diseases including neurological disorders, immunodeficiency and cancer.[109-110] PHD fingers recognize both methylated and unmethylated lysine residues, methylated arginine and acetyllysine residues on histones and other proteins.[111] PHD fingers are capable of recognizing multiple histone PTMs through distinct recognition sites on a single domain or recognize multiple PTMs through tandem domains.[111]

PHD fingers have diverse methyl-reader aromatic cages that can contain a wide variety of residues in the methyllysine binding pocket. PHD domains containing methyllysine reader motifs are best known for the recognition of methylated H3K4 and H3K9me3.[111] PHD fingers differ greatly within the family in terms of sequence and structure but do share a conserved cysteine rich zinc-binding motif.[111]

The third PHD finger of JARID1A is a methyllysine reading domain that binds to H3K4me3 promoting active transcription.[53] Similar to JARID1A, the PHD finger of Pygo2 recognizes H3K4me2/3 and leads to an active transcriptional state.[112-113] Pygo2 recruits other histone-modifying complexes to chromatin facilitating further deposition of active transcription histone marks.[114-118]

Pygo2 contains an aromatic cage made up of Y328, D339 and W353 and is capable of hydrogen bonding to the NH+ of H3K4me2 (Figure 1.6A). The aromatic cage of JARID1A is defined by W1625 and W1635 (Figure 1.6B). Selectivity for H3K4me3 recognition by both Pygo2 and JARID1A is in part due to the Thr residue at the (−1) position away from the Kme3. Steric hindrance in the pocket prevents binding to other histone methyllysine marks that contain larger residues such as H3K9me3 and H3K27me3.[110] Both PHD domains make extensive contacts with the H3 tail surrounding K4me3 and bind with high sequence specificity and selectivity.

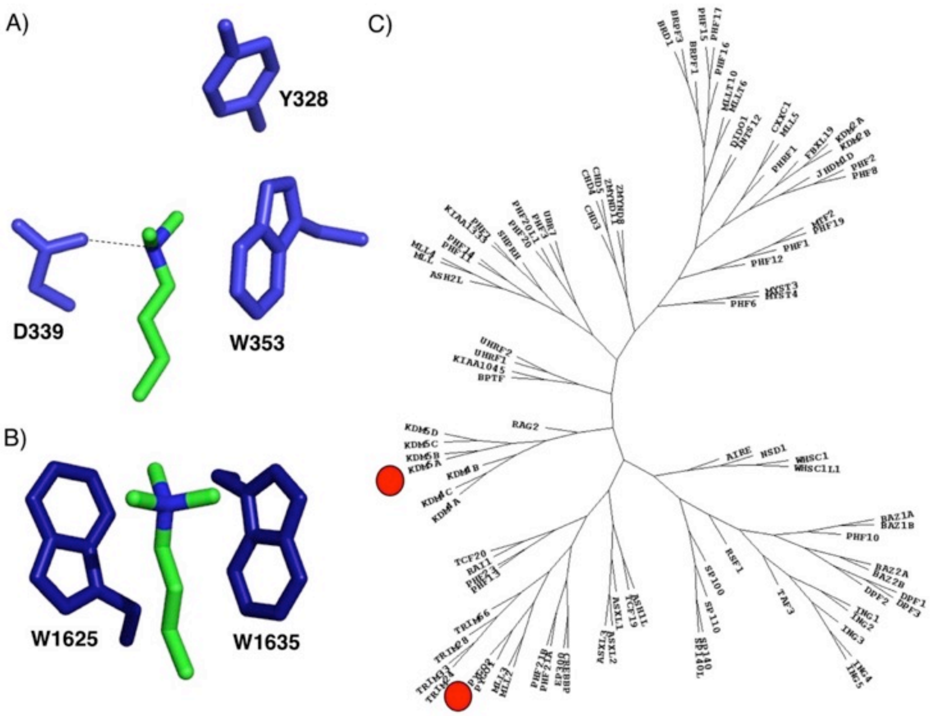


Figure 1.6. Aromatic cages and phylogenetic tree of PHD fingers. A) Aromatic cage of Pygo2 PHD finger in complex with H3K4me2 peptide (pdb: 4UP0), B) Aromatic cage of JARID1A-PHD3 complexed with H3K4me3 (pdb: 3GL6), C) Phylogenetic tree of PHD family obtained and adapted from the Structural Genomic Consortium Chromohub.[64] Domains targeted by chemical inhibition are labeled with red circles.

The pygo2 PHD finger interacts with both H3K4me2/3 and the co-factor BCL9 simultaneously and controls transcription as part of the Wnt-signalling pathway,[119-120] and mediates cross-talk between the Wnt and Notch signaling pathways in mammary stem cells.[121] Pygo2 is a player in the β-catenin signaling pathway and is overexpressed in malignant breast tumors as well as in epithelial ovarian cancer.[122-124] Pygo-BCL9 complex interacts with oncogenic β-catenin and promotes binding to H3K4 resulting in a active transcriptional state.[112-113, 116]

The long non-coding RNA prostate cancer gene expression marker 1 (PCGEM1) recruits Pygo2 to the enhancer-promoter regions of the androgen receptor gene to promote gene transcription.[125] Pygo2 overexpression was found to promote tumour growth in metastatic prostate cancer.[126] Increased mRNA expression was identified in primary glioma tissues and knockdown of Pygo2 inhibited proliferation, migration and invasion.[127]

Although much is known about the lysine demethylase activity of JARID1A, little is known about the biological functions of its PHD Kme3 reader domain. Research on the JARID1 family has focused on targeting the demethylase domains for cancer therapy.[128] The PHD finger of JARID1A forms a fusion protein with NUP98 that has been shown to be a leukemia oncogene.[53]

The ability of JARID1A to recognize multiple histone PTMs and recognition of multiple non-histone partners by Pygo2, illustrate the complexity of PHD finger read-out. Little is known about the biological roles of the majority of PHD fingers and as our knowledge increases it is likely that more of these proteins will be identified as drivers of disease.

1.5.4 WD40 repeat-containing proteins

WD40 repeat proteins are a large family of proteins named after a conserved motif in the WDR domain made up of a serine-histidine and tryptophan-aspartate (WD). WD40 domains are found in many multi-protein complexes that control processes including chromatin structure, gene expression, DNA damage and repair, immune response, and cell growth and division.[129] The WD40 domains known to recognize methyllysine are WDR5 and EED, and the methyllysine reader domain of EED is to date the only WD40 protein targeted by chemical inhibition (Figure 1.7).

EED consists of WD40-repeats that form a seven-bladed β-propeller with a central binding pocket. EED recognizes H3K27me3, H3K9me3, H4K20me3 and H1K26me3.[130] The aromatic cage is made up of three aromatic residues (Phe97, Tyr148 and Tyr365) that interact with the methylammonium and a 4th aromatic residue (W364) that makes hydrophobic interactions with the aliphatic chain of the lysine (Figure 1.7A).[130] Further interactions between the histone tail peptide and EED Kme binding pocket are made with the (−2) and (+2) residues on either side of the Kme. The histone alanine residue at the (−2) position fits into a small hydrophobic pocket in EED. This may give rise to the selectivity of methyllysine binding of EED as it primarily binds methylated lysine on histone tails possessing an alanine in this position.[130] Co-crystal structures of H3K27me3 with EED show that the peptide is bound in a sandwich conformation making interactions with both EED and the stimulated response motif (SRM) of EZH2.

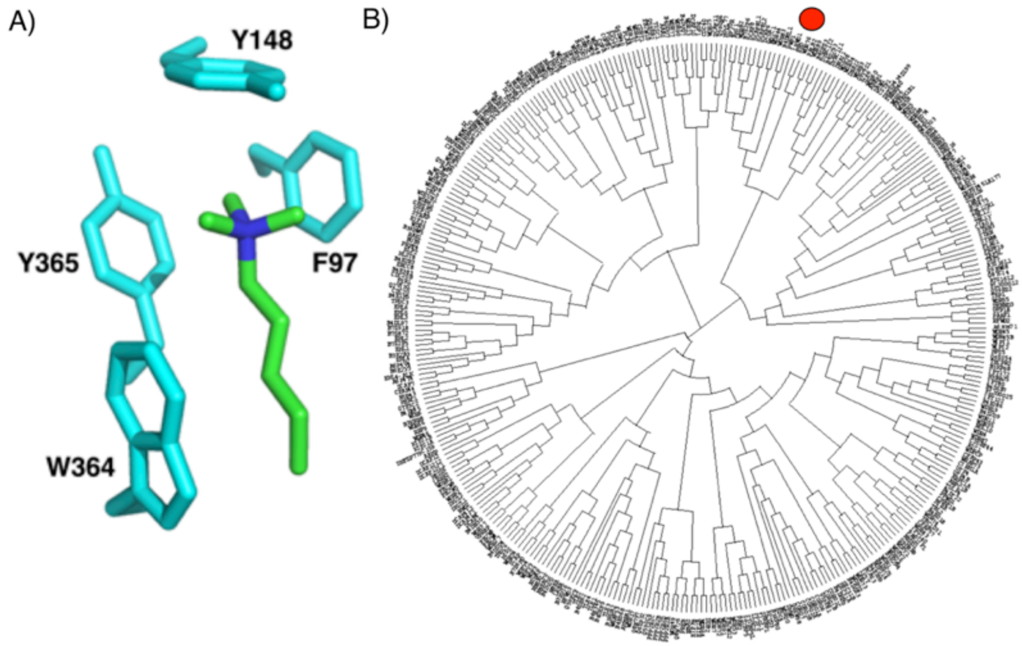


Figure 1.7. Aromatic cage of WD40 repeat EED and phylogenetic tree of WDR family. A) Aromatic cage of EED in complex with H3K27me3 (pdb: 3JZG), B) Phylogenetic tree of WDR family obtained and adapted from the Structural Genomic Consortium Chromohub.[64] Domains targeted by chemical inhibition are labeled with red circles.

The WD40 repeat embryonic ectoderm development (EED) protein is a core component of the multi-protein polycomb repressive complex 2 (PRC2). EED recognizes trimethyllysine residues that are known repressive chromatin marks and allosterically regulate the enzymatic activity of PRC2.[130-131] EZH2, a methyltransferase enzyme, is another member of the PRC2 complex that methylates lysine residues on histone proteins. EED allosterically regulates PRC2 activity and is essential to the formation of the protein complex.[131] EED regulation of PRC2 activity is dependent on which histone mark it binds.[132] The enzymatic activity of PRC2 is enhanced by EED binding with H3K27me3 and this propagates methylation of H3K27.[132] In contrast, recognition of H1K26me3 by EED inhibits methylation by PRC2. Trimethylation of K116 of JARID2 by PRC2 has recently been shown to contribute to methylation of H3K27 by binding to EED and allosterically regulating PRC2.[133]

WD40 domains are associated with a range of diseases including cancer, metabolic disorders, inflammation and viral infections.[134] Increased expression of PRC2 in various cancers has motivated significant research towards the development of therapeutic interventions.[135] The majority of anti-cancer therapies targeting PRC2 are inhibitors of EZH2.[136] The discovery that disruption of the EZH2-EED complex suppresses the activity of PRC2 has increased efforts in targeting the PRC2 complex by inhibition of EED.[137] The structural diversity and low sequence identity of the WD40 repeat domains provide more hope for selective inhibition compared to other families of methyllysine reader proteins.[129]

1.5.5 Chromodomains

Chromodomains (chromatin organization modifier domains) were named after their role in chromatin remodeling complexes and were among the first domains to be characterized as methyllysine readers.[138-139] Chromodomain proteins belonging to the Heterochromatin protein-1 and polycomb paralog families recognize higher methylation states (Kme2 and Kme3) on H3K9 and H3K27.[140] The HP1 chromodomain proteins (CBX1/3/5) regulate heterochromatic and euchromatic gene expression and contribute to centromere and telomere stability.[141] The polycomb paralog proteins (CBX2/4/6/7/8) all act as part of the multi-protein complex PRC1. PRC1 and PRC2 act together to control gene transcription by covalent modification of the tails of histone proteins.[142] Trimethylation of H3K27 is a known repressive transcriptional mark and is recognized by the CBX polycomb paralogs.

The polycomb paralog proteins contain a shallow binding pocket consisting of 3-4 aromatic cage residues. The Kme3 motif makes multiple cation-pi interactions with the hydrophobic pocket. The H3 peptide surrounding the Kme3 residue makes multiple key interactions with the surface groove of the protein and this allows for recognition of specific sequences surrounding the Kme3 residue.[41, 143] The proteins share a highly conserved hydrophobic pocket along their H3 binding groove that interacts with the H3A25 and a conserved hydrophobic clasp that folds around the H3 peptide.

All five human polycomb paralogs contain identical aromatic cages. The aromatic cage of CBX7 is primarily defined by three aromatic cage residues (F11, W32 and W35) and mutation of any of these residues greatly diminishes binding (Figure 1.8A).[143-144]. The aromatic cage of CBX6 is identical to that of CBX7, however, there are significant differences in the residues that line the H3 binding groove between the two proteins. In general, there are minimal differences in the H3 binding groove within the CBX family and more sequence diversity is seen beyond the groove and aromatic cage.

The aromatic cage of chromodomain Y-like (CDYL) proteins contain 2–3 aromatic residues that bind Kme3 with the strongest affinity. CDY proteins recognize H3K9me3 and H3K27me2/3 and repress gene transcription.[145-147] The residues that define the aromatic cage of CDYL2 are Y7, W29 and Y32 (Figure 1.8C). The CDY family of chromodomains have high homology to HP1 and polycomb paralog families.[148] All three families bind their substrates through a surface-groove mechanism of binding and contain shallow binding pockets with a high degree of hydrophobicity in their Kme binding domain. CBX and CDY proteins are the only methyllysine reader chromodomains that have been targeted by chemical inhibition (Figure 1.8D).[149-156]

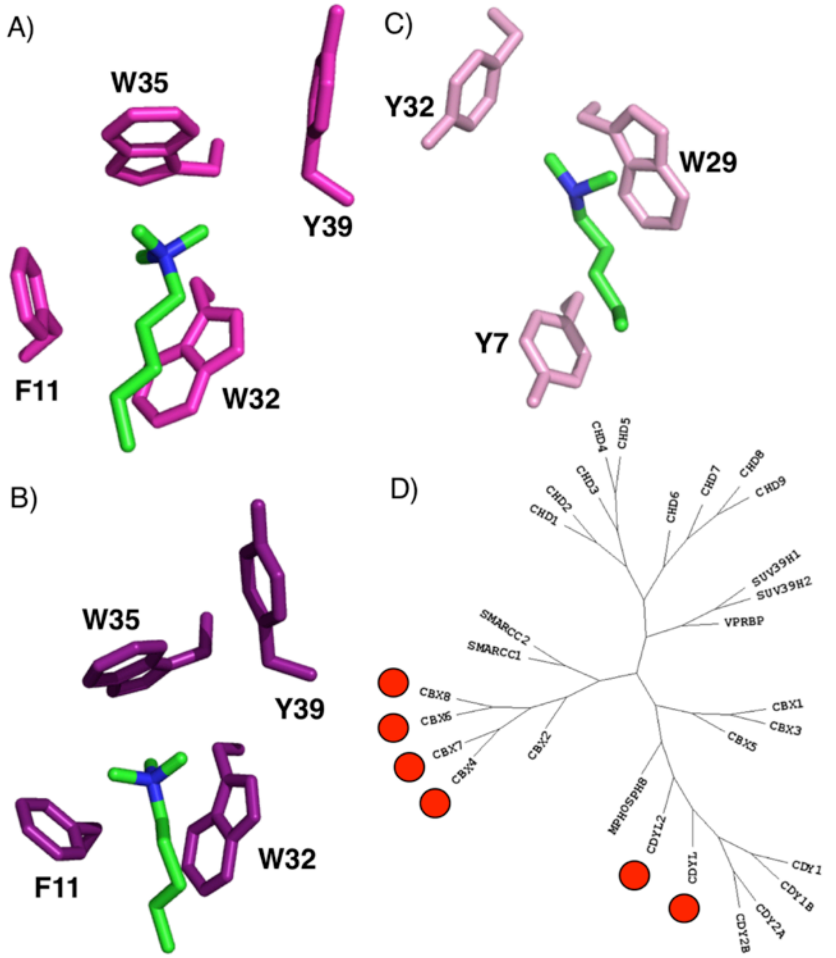


Figure 1.8. Aromatic cages of chromodomains and phylogenetic tree. A) Aromatic cage of CBX7 in complex with H3K27me3 peptide (pdb: 4X3K), B) Aromatic cage of CBX6 in complex with H3K27me3 peptide (pdb: 3I90), C) Aromatic cage of CDYL2 in complex with Kme2 (pdb: 5JJZ), D) Phylogenetic tree of chromodomain family obtained and adapted from the Structural Genomic Consortium Chromohub.[64] Domains targeted by chemical inhibition are labeled with red circles.

The five human polycomb paralogs are best known for their roles in regulating gene expression, stem cells differentiation and development.[157-158] These pathways can be hijacked by cancer cells to provide proliferative advantages.[58] In stem cells, several genes controlling cell fate are silenced by CBX7 and this allows for maintenance of pluripotency.[157] CBX7 regulates expression of the Ink4a/Ark locus and extends the cellular lifespan of both normal and cancerous cells.[159-160] This control is established following interactions of CBX7 with H3K27me3 and long noncoding RNA ANRIL.[144] CBX7 inactivates the well-known tumour suppressor gene *p16* and this is established to be a driving factor in several malignances.[143, 161-162] CBX6 regulates embryonic stem cell fate and interestingly this was shown to be independent of H3K27me3.[163]

Out of the five human polycomb paralogs, the majority of research has focused on CBX7. Biological data on the misregulation of CBX7 in cancer is vast. Most examples from human carcinomas show decreased levels of CBX7 expression to be correlated with poor patient outcomes.[164-165] Therapeutic interest in targeting CBX7 stems from research into aggressive stem-like cancers where the protein is overexpressed.[143, 159, 162, 166-168] Knockdown of CBX7 inhibited proliferation in various cancer cell lines such as ovarian, prostate, gastric, lymphoma and leukemia.[143, 159, 162, 166-168]

Within the polycomb paralog family, CBX6 is the least understood. CBX6 has been shown not to bind to H3K27me3 and H3K9me3 peptides, despite the evidence that these methylation sites are the targets of CBX proteins.[41, 169] There are very few studies reporting on the biological role of CBX6. Recently, CBX6 was shown to promote tumour growth in hepatocellular carcinoma,[170] and act as a tumour suppressor in breast cancer.[171] Increased research is needed to understand transcriptional regulation by CBX6 in normal and diseased tissues.

CDYL proteins are another group of poorly understand epigenetic modules with little known about their role in health or human disease. Recognition of methyllysine by CDYL proteins acts to propagate the methylation mark by interacting with and enhancing the enzymatic activity of EZH2-PRC2.[172] CDYL catalyzes downregulation of histone lysine crotonylation (Kcr) and acts to regulate gene transcription during spermatogenesis.[173] CDYL was reported to be upregulated in neurodevelopment Hirschsprung disease (HCSR),[174] and mice deficient in CDYL showed increased susceptibility to seizures.[175]

1.6. Polycomb paralog proteins and PRC1

The specific paralog utilized in the PRC1 complex is tissue and context-specific. The CBX used as part of PRC1 is in part controlled by the protein subunits in PRC1,[176] and by the varying binding affinities of individual CBX proteins.[60-61, 169] CBX proteins localize and bind to non-redundant sets of target genes.[177] The expression of CBX proteins is influenced by a variety of mechanisms. CBX proteins can repress transcription of other members of the family in autoregulatory feedback loops.[157, 168, 178-179] CBX7 expression levels are in part controlled by microRNAs in embryonic stem cells.[178, 180] Phosphorylation sites are known to exist on CBX2, CBX4 and CBX7, and can aid in the assembly of PRC1 and influence nucleosome binding.[181-182]

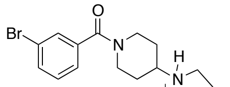
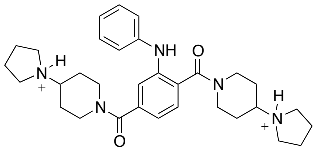
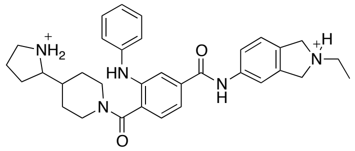
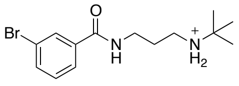
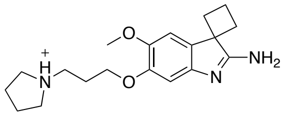
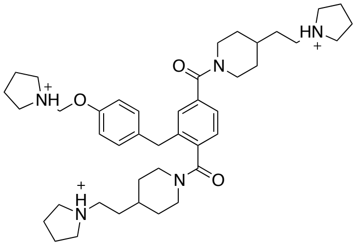
The main enzymatic functions of PRC1 and PRC2 are histone H2A ubiquitylation and H3 methylation. PRC2 consists of at least 4 subunits and one of the core units is the methyltransferase enzyme EZH2. The Set domain of EZH2 catalyzes the trimethylation of H3K27.[183] The CBX proteins act as one component of the PRC1 complex and recruit PRC1 to chromatin through H3K27me3 recognition.[41] Two core PRC1 components, RING1b and BMI-1, catalyze monoubiquitylation of H2AK119.[184-185] PRC1 represses transcription through several mechanisms, including H2A ubiquitylation, chromatin compaction and interactions with other transcription machinery.[186] Recruitment of the CBX proteins to histone tails and subsequent transcriptional repression is also controlled by interactions with long non-coding RNA.[144] Conventional beliefs that H3K27 methylation by PRC2 serves to recruit PRC1 are now being challenged by new research highlighting the complex interplay between the two complexes.[187] The actions of both polycomb complexes are key for the regulating stem cell differentiation and misregulation of these complexes can result in formation of cancer stem cells.

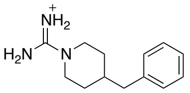
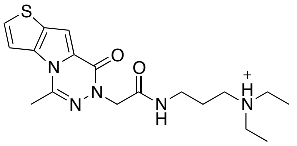
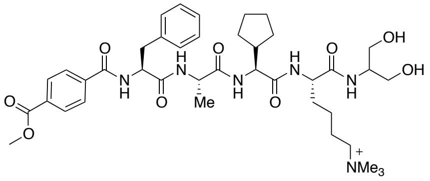
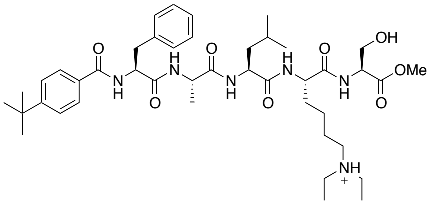
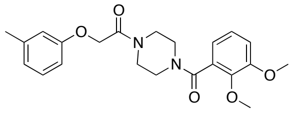
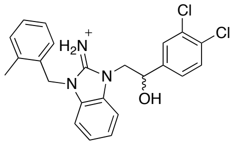
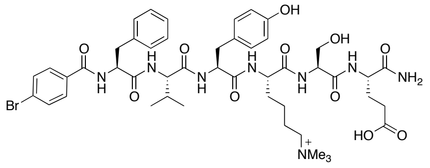
1.7. Inhibition of methyllysine reader proteins

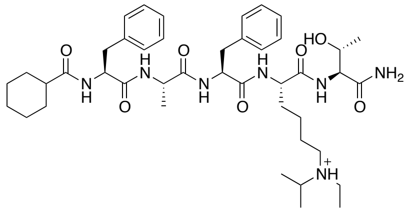
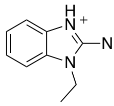
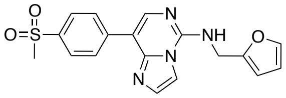
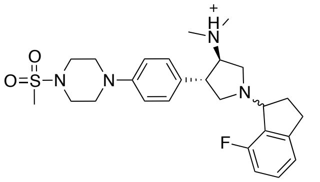
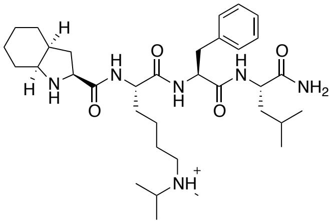
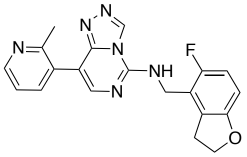
Epigenetic targets have been shown to be valid drug targets for diverse indications such as immune, pulmonary and neurological disorders as well as various types of cancer. Methyltransferases, deacetylases, acetyltransferases, and acetyllysine readers have been targeted for drug discovery with many inhibitors currently in clinical trials.[188-195] Inhibitors of methyl reader proteins are far behind in their development, but have increasingly been suggested as druggable targets.[66] Despite their potential for treating diseases, the creation of methyllysine reader protein antagonists presents a large challenge in molecular recognition. This challenge, along with the promise for reader protein inhibitors to drive new medicine, motivates this area of research.

Significant challenges are associated with inhibition of methyllysine reader proteins. Methyllysine reader proteins bind their native substrates with weak affinity and have shallow binding pockets. These qualities pose challenges to traditional drug discovery approaches such as small molecule high-throughput screening. Identifying inhibitors through *in silico* screens and docking using known crystal structures is hindered by conformational changes to the protein-binding pocket that occur during protein-ligand binding.[196] The term druggability is used describe the ability or predicted ability of a protein target to bind a drug.[197] Many methyllysine reader proteins are reported to have low druggability and are more challenging targets compared to, for example, bromodomains that are readers of lysine acetylation.[66] Despite these challenges, increased research in targeting these proteins has developed many chemical inhibitors (Table 1.1)

Table 1.1. Inhibitors of methyllysine reader proteins

Inhibitor	Structure	Potency	Target Protein	Reference
UNC926		IC_{50} 3.9 μM	L3MBTL1/ L3MBTL3 MBT domain	198
UNC1215		K_d 0.12 μM	L3MBTL3 MBT domain	199
56		K_d 0.35 μM	L3MBTL3 MBT domain	200
UNC2170		K_d 22 μM	53BP1 Tudor domain	201
A366		K_d 111 nM (Spindlin1) IC_{50} 3.3 nM (G9a)	Spindlin1 Tudor Domain G9a Histone Methyltransferase	202 203
EML631		K_d 3 μM	Spindlin 1 Tudor domain	204

BPC		K_d 50 μM	UHRF1 Tudor domain	205
NV01		K_d 5 μM	UHRF1 Tudor domain	206
Compound **64**		K_d 0.2 μM (CBX7) K_d 0.29 μM (CBX4) K_d 1.89 μM (CBX8)	CBX7/CBX4 Chromodomain	149
UNC3866		K_d 97 nM (CBX7) K_d 1.8 μM (CBX2) K_d 94 nM (CBX4) K_d 0.61 μM (CBX6) K_d 1.2 μM (CBX8)	CBX7/CBX4 Chromdomain	151
MS37452		K_d 29 μM (CBX7) K_d 95 μM (CBX4)	CBX7 Chromdomain	153
MS351		K_d 23.8 μM (CBX7)	CBX7 Chromdomain	154
Compound **5**		K_d 0.9 μM (CBX6) K_d 17.6 μM (CBX2) K_d 15.5 μM (CBX4) K_d 5.3 μM (CBX7)	CBX6 Chromodomain	155

		K_d 6.0 µM (CBX8)		
UNC4991		K_d 0.49 µM (CDYL) K_d 0.64 µM (CDYL2)	CDYL/CDYL2 Chromodomain	[156]
CF16		K_d 7.5 mM	Pygo2 PHD finger	[207]
EED226		K_d 82 nM	EED WD40 domain	[208]
A-395		K_d 1.5 mM	EED WD40 domain	[209]
UNC5114		K_d 0.68 µM	EED WD40 domain	[210]
MAK683		unknown	EED WD40 domain	(clinicaltrials.gov identifier NCTO 2900651).
BI-9321	unknown	K_d 166 mM	NSD3 PWWP domain	Un-published

1.8. Inhibitors of chromodomains

Drug discovery efforts towards chromodomain containing proteins have focused primarily on CBX7 because of its well-studied biological role, involvement in cancer proliferation, and predicted higher druggability compared to other polycomb paralogs.[66]

The first report of an inhibitor for any chromodomain containing protein was developed for CBX7.[149] Peptidic inhibitors were designed starting from a peptide sequence derived from the protein SETDB1-K1170Kme3, which had been shown to have higher binding affinity for CBX7 compared to the native substrate H3K27me3.[41] A truncated 5-mer peptide, Ac-FALKme3S-NH2, disrupted the CBX7-H3K27me3 complex in a competitive fluorescence polarization assay with an IC_{50} of 12 μM. Binding of Ac-FALKme3S-NH2 to CBX7 is driven by the Kme3 motif of the ligand forming multiple cation-pi contacts with the aromatic cage, backbone β-sheet hydrogen bonds and a hydrogen bond between the side-chain of the C-terminal serine residue with Glu37 of the protein (Figure 1.10A, C). Amino acids at each position of the peptide were substituted with a variety of natural and unnatural substitutions resulting in a series of peptidomimetic inhibitors with K_d values between 0.2-2.0 μM (Figure 1.9). Lead compounds were tested against CBX8, CBX4 and CBX1, a representative HP1 paralog. The alanine side chain of the ligand at the −2 position away from the Kme3 side chain sits in a small hydrophobic pocket, termed the (−2) pocket, and substitutions of this residue with anything larger than alanine resulted in diminished binding.

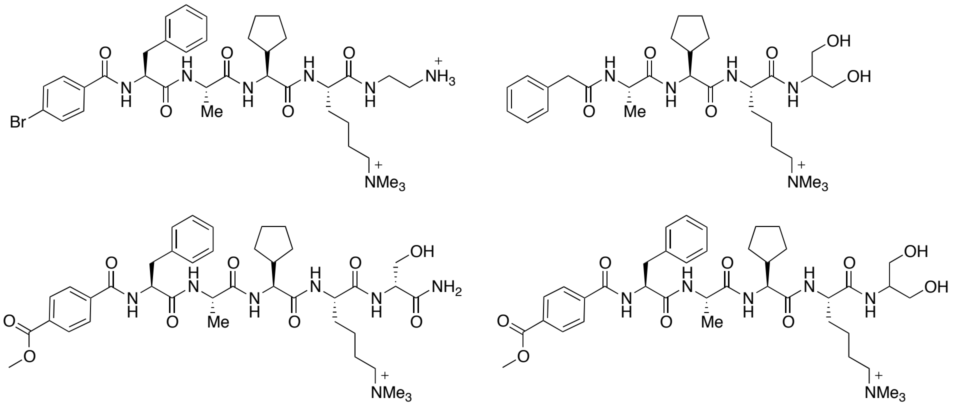


Figure 1.9. Lead peptidic inhibitors for CBX7/CBX4.[149]

A similarly structured peptidic probe was developed to target CBX7 with several chemical modifications giving improved cell permeability.[151] After initial failed small molecule screens, analogs of a peptidic ligand were synthesized with a various replacements for the quaternary Kme3 amine. Replacement with a diethyllysine residue showed nearly equipotent binding to CBX7. UNC3866 showed improvements in potency and drug-like qualities from the addition of an N-terminal hydrophobic group, leucine at the (−1) position, and a C-terminal methyl ester (Figure 1.10B, C). UNC3866 has a K_d of 97 nM for CBX7 and inhibits the CBX7-H3K27me3 interaction with an IC_{50} of 66 nM. UNC3866 was tested in a panel of 9 Kme readers belonging to the families of MBT, PHD and tudor, and was more than 100-fold selective for CBX7. Further selectivity studies were conducted with a biotinylated analog of UNC3866 in a protein microarray of 96 methyllysine readers and other chromatin-associated effector proteins. The biotinylated analog showed binding to CBX7, as well as the polycomb paralogs CBX2/4/6/8 and to members of the CDY family of chromodomains (CDY1, CDYL1b and CDYL23). UNC3866 binds CBX7 and CBX4 with equal affinity and is 18-, 6- and 12-fold selective for for CBX7 over CBX2, CBX6 and CBX8 respectively. UNC3866 is 65- and 9-fold selective for CBX7 over CDY1, and CDYL1/CDYL2.

Chemiprecipitation experiments in PC3 cell lysates with a biotinylated analog of UNC3866 pulled down CBX4/7/8. UNC3866 was found to have low cellular permeability by using a Caco-2 assay but did possess micromolar cellular potency and no observed toxicity. PC3 cells treated with 30 μM UNC3866 showed a significant decrease in cell proliferation with an EC_{50} of 7.6 μM.[151]

Additional details of the development and SAR of UNC3866 are described in a follow up paper.[152] Similar efforts to substitute the Kme3 motif and explore varying N-terminus substitutions in the peptide scaffold were carried out by the Hof group.[150] Both efforts explored multiple tertiary ammonium ion replacements for the Kme3 side chain, and both found the diethyllysine to be the optimal replacement.

The structural basis for the CBX7-selectivity of the diethyllysine containing UNC3866 was examined using molecular dynamics simulations.[211] UNC3866 binding to CBX2 and CBX7 was compared and UNC3866 was found to make more van der Waals contributions in the binding pocket of CBX7 than to CBX2. In addition, UNC3866 binding to CBX2 required unfavourable conformational changes to the diethyllysine side chain, providing another explanation for the weaker binding affinity.

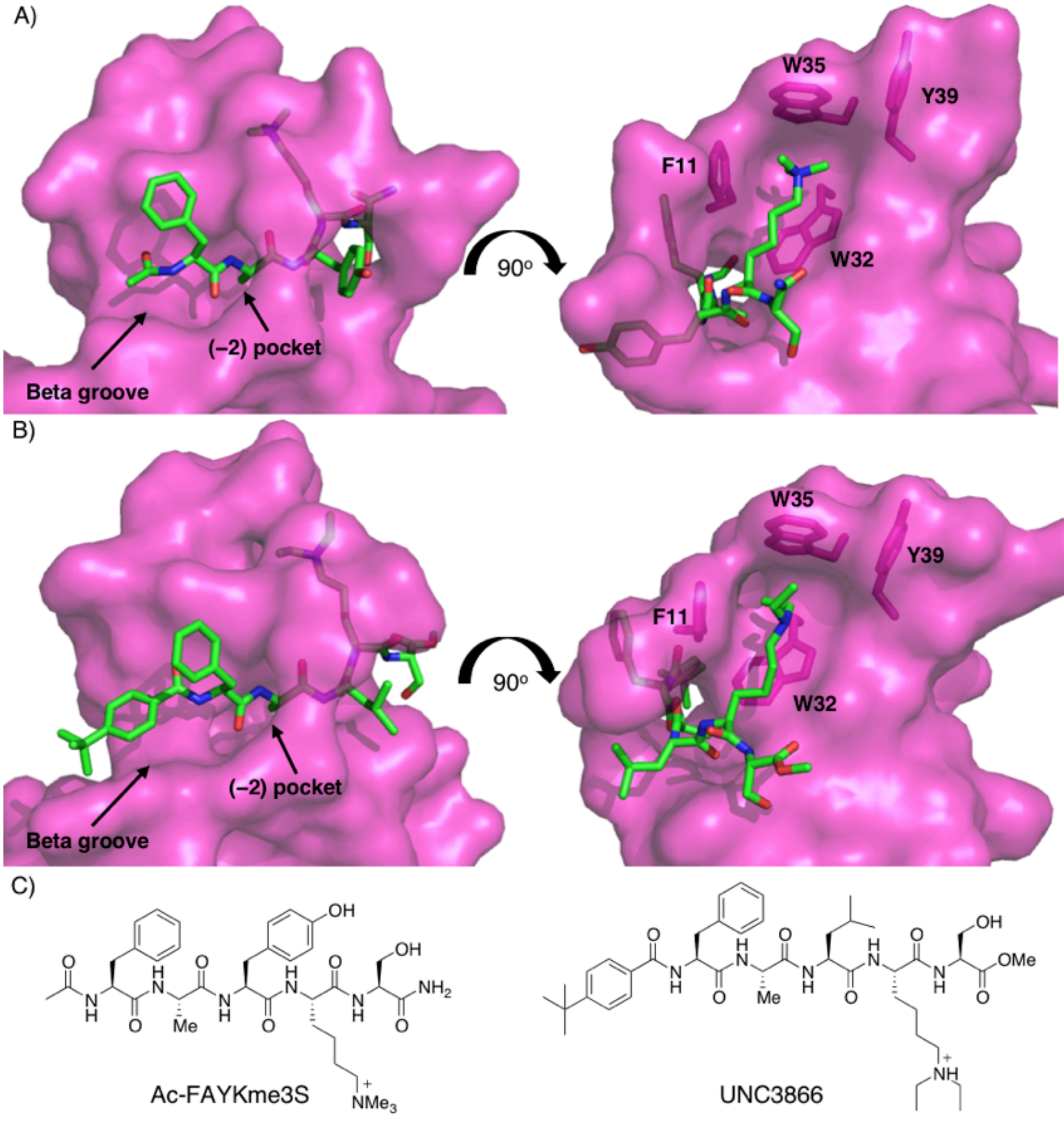


Figure 1.10. Peptidic inhibitors of CBX7. A) Co-crystal structure of CBX7 in complex with Ac-FAYKme3S (pdb:4MN3),[149] B) Co-crystal structure of CBX7 in complex with UNC3866 (pdb:5EPJ),[151] C) Chemical structures of peptidic CBX7 inhibitors.

The first non-peptidic small molecule inhibitor for CBX7, MS37542, was discovered from a high throughput screen of a library of drug-like small molecules (Figure 1.11).[153] The screen was carried out using a fluorescence anisotropy assay with FITC-labeled SETDB1-K1170me3 serving as the fluorescent probe. MS37542 has a reported K_d of 29 μM and disrupts CBX7-H3K27me3 binding with a K_i of 43 μM. MS37452 binds measurably to the other polycomb paralogs (CBX2/4/6/8) but not to the HP1 family members CBX1/3/5. MS37452 binds CBX4 with 3-fold weaker binding and greater than 10-fold weaker binding to CBX2/6/8 compared to CBX7. Chromatin immunoprecipitation (ChIP) experiments were used to study the effects on transcriptional repression from inhibiting CBX7 with MS37452. PC3 prostate cancer cells treated with high concentrations (250–500 μM) of MS37452 showed reduced CBX7 occupancy at the INK4A/ARF locus,[153] a known genetic target of CBX7 repression.[159] Further studies showed inhibition of CBX7 resulted in transcriptional derepression of the target genes p14/ARF and p16/INK4a.

A co-crystal structure of the CBX7-MS37452 complex showed the compound adopting two rotamer conformations. In both conformations, the dimethoxybenzene and piperazine motifs are bound in the same orientation. The dimethyoxybenzene moiety makes interactions with all three of the aromatic cage residues of CBX7 and the piperazine motif interacts with Phe11 and Trp32. The methylbenzene motif adopts a cis or trans conformation and in the latter extends to makes interactions with the peptide binding groove similar to the native H3K27me3 substrate. SAR studies with similar analogs in the library showed that the two methoxy groups are critical for binding, and that removing the methyl group and having no substitution on the methylbenzene ring reduced affinity. In total, the authors prepared 23 structural analogs of MS37452 without identifying any that were more potent than the original hit compound. The relatively weak binding of MS37452 is explained by the shallow binding pocket of CBX7 and the open-faced aromatic cage.

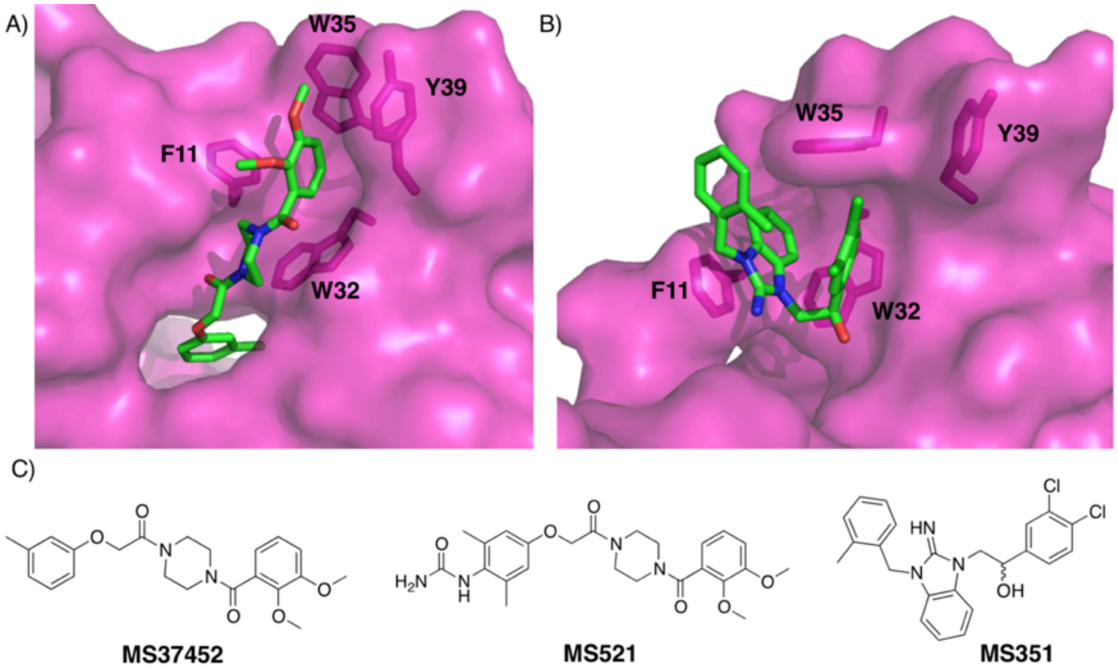
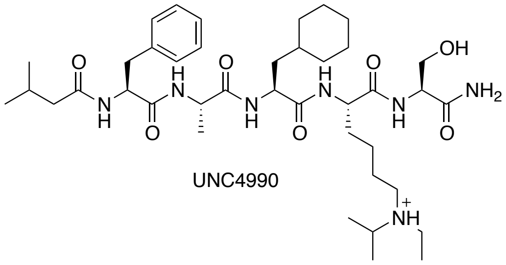


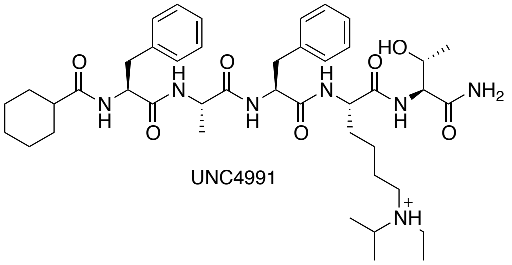
Figure 1.11. Small molecule inhibitors of CBX7. A) Co-crystal structure of CBX7 in complex with MS37542, (pdb:4X3T),[153] B) Co-crystal structure of CBX7 in complex with MS351 (pdb:5EJW),[154] C) Chemical structures of small molecule CBX7 inhibitors.

A follow up study reported on further SAR of MS37452 resulting in MS508 that binds with 7-fold improved affinity to CBX7 with a K_i of 4.8 μM (Figure 1.11C).[154] MS508 possesses a urea moiety that makes a hydrogen bond with His47 of CBX7, resulting in the improved potency. Despite the improved in vitro potency, MS508 did not show improved cellular efficacy. MS508 showed poor solubility and cell permeability and these problems also plagued other analogs from the series. An *in silico* screen in the same report identified a different iminobenzimidazole compound, MS351. MS351 formed a ternary complex with CBX7 and a hairpin RNA derived from ANRIL (which binds CBX7 at a different sub-site) with a K_d of 24 μM and showed 21-fold weaker affinity to free CBX7.[154] Fluorescence anisotropy assays showed enhanced RNA binding to CBX7 in the presence of MS351 in a dose-dependent manner. MS351 makes cation-pi contacts and pi-stacking interactions in the aromatic cage but does not make interactions with the hydrophobic (−2) pocket or extended beta groove. The hydroxyl group makes a unique hydrogen bond with a Thr41 next to the aromatic cage. MS351 induced transcriptional depression of p16/INK4a in PC3 cells similarly to the first reported inhibitor.

The off-target binding of UNC3866 to CDYL proteins was leveraged to find the first peptidic inhibitors of the CDYL chromodomain containing proteins.[156] The authors utilized a one-bead-one-compound (OBOC) approach with a cross screening strategy to achieve selectivity between the closely related CDYL and CBX protein families. A library based on the UNC3866 peptide was made screened against CDYL2 with counter-screening against CBX7, CBX8, CBX5 and another chromodomain containing protein, M-phase phosphoprotein 8 (MPP8). The authors used an on-bead binding assay with magnetic enrichment to identify several peptidic scaffolds that bind to CDYL2. Lead compounds UNC4990 (K_d = 0.67 μM and 0.57 μM for CDYL and CDYL2) and UNC4991 (K_d = 0.49 μM and 0.64 μM for CDYL and CDYL2) were found to be between 3- and 6-fold selective for the CDYL proteins over CBX7 (Figure 1.12). MD simulations comparing the binding of UNC4991 to CDYL2 and CBX7 suggested that the 5-fold selectivity was arising from the N-terminal cyclohexyl cap that did not fit into the smaller binding groove of CBX7 and in part from the isopropyl group on the lysine residue. A biotinylated version of UNC4991 pulled down both CDYL and CDYL2 from cellular lysates but not CBX7.



UNC4990



UNC4991

Figure 1.12. Lead peptidic inhibitors targeting CDYL and CDYL2.[156]

1.9. Conclusion and future directions of methyllysine reader inhibition

Chemical probe development for methyllysine reader proteins has only recently begun. The growing collection of chemical tools targeting methyllysine reader proteins will allow for understanding and investigation of the phenotypes generated by inhibition. Our poor understanding of epigenetic regulation combined with its inherent complexity, require more efforts on characterization and validation of methyllysine reader proteins as therapeutic targets. To better validate methyllysine readers as targets, we need to expand the collection of selective inhibitors. Out of the over 200 methyllysine reader proteins, only 13 of these proteins have published inhibitors. Many of the inhibitors are in early stages of development and do not meet the criteria of a chemical probe.[212]

It has proven extremely challenging to create chromodomain inhibitors that are selective for any member of the family besides CBX7. At the outset of this thesis work, all reports of chromodomain inhibitors bound to CBX7 most strongly. CBX7 is the most druggable member of the family,[66] and binds its native substrates with greater affinity compared to the other CBX proteins.[41] Collectively this information supports the idea that CBX7 is more amenable to screening and development of inhibitors. Overcoming the challenge of biased affinity of inhibitors for CBX7 is required to create selective inhibitors for the other CBX proteins. Chemical tools targeting CBX2/4/6/8 are essential for understanding the specific role of each protein in regulating gene transcription and in development of cancer and disease.

Progress is happening, but significant chemical and biochemical challenges exist for the creation of potent and selective inhibitors. Improved and focused screening efforts are needed to better target these challenging proteins. Better assay platforms that overcome the challenges associated with the weak binding affinity of these proteins to their native substrates, and that incorporate selectivity considerations are needed. Methyllysine reader proteins have very low *in vitro* affinities for native ligands that hamper high-throughput screening. There are relatively few pharmacologically compatible methyllysine isosteres known to engage aromatic cage pockets. This presents a challenge in design and mimicry of the molecular recognition elements required. Adding to the challenge of methyllysine isosteres and selective inhibition is the high degree of similarity of the aromatic cage binding pockets within each family of proteins. Beyond the aromatic cage, the similarity of surrounding structures within reader protein families is also very high. The ability to move forward and make selective inhibitors for these challenging proteins will rely on innovative chemical methods to search for Kme mimics and surrounding structures that provide selectivity advantages. Recently discovered allosteric modulators of methyllysine reader proteins[208-209] provide promise for successful inhibition that bypass the challenges often seen in targeting the aromatic cage.

Many of the structural and functional roles of methyllysine reader proteins have been discovered and this has motivated continued research in the field. However, there is a fundamental difference between background biology studies that rely on knockdown (which removes an entire multi-domain reader protein from the cell), and inhibition with a chemical agent. A protein whose reader domain is inhibited, but whose other domains remain functional, becomes disengaged from its methyllysine target but might be able to retain its other functions in the cell. Inhibiting a reader protein is better thought of as 'de-targeting' the protein's other functions, rather than to be thought of as a complete inhibition of the sort that would be achieved by direct inhibition of targets such as enzymes. Chemical probes that bind and inhibit just the Kme reader domain could provide a unique way to understand epigenetic regulation of methyllysine recognition. Such studies will be necessary in order to demonstrate the potential of these proteins as new therapeutic targets.

Future development must focus on preclinical in vivo studies. Validation of drug and target candidates is essential in achieving this goal. We need to better understand the biology and downstream effects of inhibiting specific methyllysine reader domains.

1.10. Summary of thesis

This chapter has reviewed the foundations of post-translational lysine methylation and the proteins responsible for recognition of the methylation mark. The structural and chemical basis for methyllysine recognition was discussed along with the biological roles of methyllysine reader proteins. Included above is a review of current inhibitors for chromodomain containing methyllysine readers and a perspective on future development in this field.

This thesis reports new inhibitors for CBX proteins, as well as the development of novel techniques for the generation and study of methyllysine reader antagonists. Chapter 2 of my thesis is on the structural determinants of recognition of the CBX proteins and my efforts to selectively target each family member. Through step-wise synthesis and screening, I created a small library of peptidomimetic compounds. This chapter examines general structure activity relationships (SAR) of the extended beta groove and my work towards creating chemical tools for biochemical and biophysical studies of the CBX proteins.

Chapter 3 reports on the development of selective inhibitors targeting CBX6. I will report on the design, synthesis and selectivity studies of the first non-CBX7 chromodomain antagonists and give an outline of ongoing and future work in CBX6 inhibition. Chapter 4 is a continuation on this work and will report on my second generation of cell-permeable CBX6/CBX8 inhibitors.

Chapters 2–4 discuss rational design and synthesis of peptidic inhibitors one at a time. The remainder of my thesis work focused on developing new methodologies to find potent and selective peptidomimetic inhibitors of CBX proteins as well as other methyllysine readers. Chapter 5 reports on chemical modifications to peptide libraries expressed on bacteriophages. This technique is a synthetic chemistry-driven adaptation of traditional phage display for the purpose of screening millions of peptide-based compounds simultaneously. Strategies that allow for diversity and high throughput screening will be key to successfully targeting the highly similar CBX proteins.

Chapter 6 concludes my thesis with a review of the main lessons learned throughout this research. A discussion of key questions answered, impact of research and future directions are included.

Chapter 2. Structural insights and selectivity studies with dye-labeled peptidic inhibitors of polycomb paralog proteins

Contributions

I conceived of and carried out the design, synthesis, purification, and characterization of all peptides with synthetic assistance from Caitlin Croft and Chelsea Wilson. I evaluated and interpreted all binding and computational data. James McFarlane conceived of and executed the computational studies, and wrote the corresponding paragraphs. Michael Gignac and Rebecca Hof carried out protein expression and purification. Michael Gignac, Tyler Brown and I carried out the FP assays. Compound **2.5** was previously made by Chakravarthi Simhadri and I resynthesized this compound for the purpose of determining binding affinity to other CBX proteins within the family. Protein microarray testing was done by a service at the MD Anderson Cancer Center, Dallas, TX. Dr. Lindsay Devorkin and Janessa Li carried out the MTT assays. Live cell imaging was carried out by the Michael Cox lab at the Vancouver Prostate Centre. I planned and wrote the manuscript with editing help from Dr. Hof.

2.1. Foreword

My work on developing inhibitors for CBX proteins began following the first publication of an inhibitor for any of the CBX proteins. The Hof group published on a series of peptidic inhibitors targeting CBX7 in 2014. Since this work, five other reports of inhibitors targeting CBX7 have been published. The unpublished work shown in this chapter started with two distinct goals: to create novel dye-labeled inhibitors that would enable a variety of new characterization studies, and to identify and understand the structural features required to tip inhibitors away from CBX7 inhibition and towards other members of the CBX family. This work was almost completely carried out before the more recent publications on CBX7 inhibition, and before my published work on CBX6 inhibition (Chapter 3).

2.2. Introduction

In Chapter 1, I described how each CBX protein has unique biological roles as well as distinct roles in different stages of cancer and disease.[55, 177, 213-214] To better understand the biology of these proteins and to test their potential as drug targets, new chemical tools are needed. Significant progress has been made in understanding the biological role of CBX7 but comparatively much less is understood about the other CBX proteins.

Typical approaches to generating small-molecule inhibitors have proven very challenging for CBX proteins. Early efforts in the Hof group at virtual and small molecule screening for CBX7 did not yield any promising inhibitors (unpublished results). Our most recent high-throughput screen, testing 350,000 compounds in collaboration with a major pharma company, has only yielded inhibitors with IC_{50} values ≥ 140 µM (unpublished results). The challenges associated with screening small molecules against the CBX proteins have been widely discussed among multiple groups at conferences, and have also been reported in the literature.[149, 151]

The Hof group used a peptide-driven approach to identify a series of sub-micromolar inhibitors targeting CBX4 and CBX7.[149-150] The peptidic inhibitors developed from this work are equipotent for CBX7 and CBX4, and have limited selectivity over other CBX proteins. First-in-class inhibitors often exhibit limited selectivity but are still important and useful molecules.[196, 198, 215] JQ1 was among the first inhibitors of the BET family of bromodomains that recognize acetylated lysine residues on N-terminal histone tails.[216-217] JQ1 has been used extensively to improve our understanding of bromodomain proteins and their role in disease.[218-219] JQ1 is not selective within the BET family, containing 4 different bromodomains, but has aided in the development of inhibitors with improved selectivity profiles.[216, 220] The challenges faced in selective BET bromodomain inhibition mirror those of CBX inhibition. The BET bromodomains and CBX chromodomains are similar in terms of their high degree of sequence homology within their respective families. The use and success of pan-specific probes for bromodomain proteins provide motivation for the development of CBX probes.

Peptide-based inhibitors are useful for modulation of protein-protein interfaces (PPIs).[221] The surfaces of PPIs are large and shallow and for these reasons it is difficult to find low molecular weight inhibitors.[222] Peptides are advantageous for inhibition of PPIs because of their size, similarity to native protein substrates, and because of the ability to add diverse functional groups. Peptides are particularly useful for promoting induced-fit binding because of the many interactions and hydrogen bonds that form between the peptide amide backbone and protein target. Challenges associated with peptide-based ligands include poor cell permeability and physiochemical properties. Despite these challenges, examples of peptide ligands used as drugs and as chemical tools are numerous.[223] As of 2018, there are 60 peptide drugs approved by the FDA and 150 currently in human clinical trials.[224] Strategies to aid in overcoming the drawbacks of peptidic ligands continue to increase in number.[221, 225]

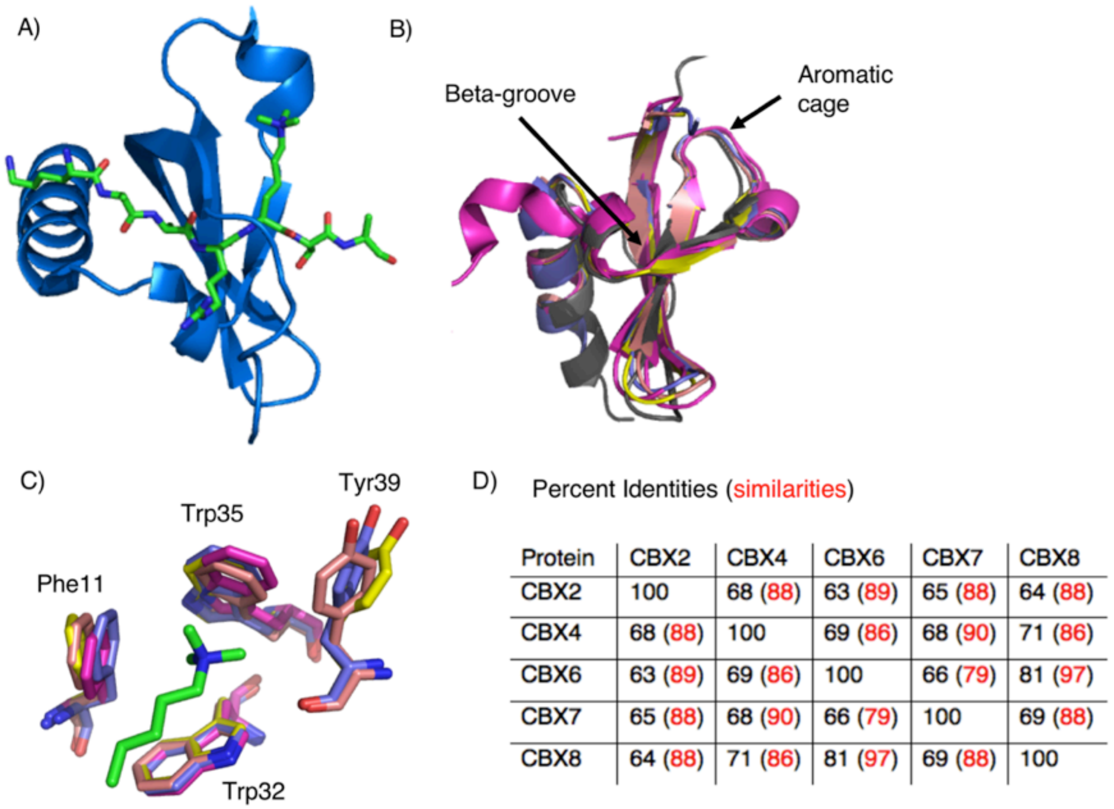
No selective inhibitors have been developed for the other CBX proteins (with the exception of my published CBX6 inhibitors—see Chapter 3). Relatively little is known about the many biological roles of CBX2/4/6/8, and nothing is known about the impacts of chemical inhibition of any of these proteins. The delay in progress is because of the many challenges in targeting the CBX proteins (reviewed in Chapter 1). CBX chromodomains bind to their native substrates with weak affinity,[41] undergo an induced fit mode of binding,[151] and are between 79-98% similar within the family (Figure 2.1, Table 2.1).

Each CBX protein binds with different affinities to the native H3 substrates despite the high degree of sequence homology (Figure 2.1, Table 2.1). To test the different affinities of each CBX protein for the H3 peptide, a histone peptide array screen was conducted using 13-mer H3 sequences.[41] The HP1 homologs (CBX1/3/5) show a clear preference for binding to H3K9me3. CBX4 and CBX7 bind both H3K9me3 and H3K27me3 with 2 and 3-fold weaker binding to H3K27me3. CBX6 and CBX8 show very weak binding to both methylation marks. CBX2 is unique within the family as it is the only member with a clear binding preference for H3K27me3 in biochemical binding assays.[41]

Table 2.1.Binding affinities of the HP1 homologs (CBX1/3/5) and polycomb paralogs (CBX2/4/6/7/8) to H3K9me3 and H3K27me3 peptides. Binding affinities determined by fluorescence polarization (FP) assays and reported as K_d values in μM.[41]

Protein	**H3K9me3**	**H3K27me3**
CBX1	5 ± 2	N/B
CBX3	15 ± 8	N/B
CBX5	30 ± 5	N/B
CBX2	>500	185 ± 20
CBX4	70 ± 7	205 ± 20
CBX6	>500	>500
CBX7	55 ± 5	110 ± 17
CBX8	>500	>500

Significant structural differences exist between the HP1 homologs (CBX1/3/5) and the polycomb paralogs (CBX2/4/6/7), but the differences within the polycomb paralog family are subtle. The polycomb paralog proteins bind Kme3 through cation-pi interactions in their conserved aromatic cage (Figure 2.1B, C). The aromatic cage is nearly identical within the family, but differences exist in the protein structure in the nearby beta-groove (Figure 2.1B). Discovery and exploitation of these structural differences outside of the methyllysine binding motif are necessary for the development of selective inhibitors.



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

Figure 2.1.Structural analysis of the highly similar CBX proteins. A) Cartoon depiction of CBX7 bound to native H3K27me3 (pdb: 4X3K). B) Overlays show the similarity of CBX proteins with aromatic cage and beta-groove indicated with arrows (magenta = CBX2 pdb code: 3H91, grey = CBX4 pdb code: 3I8Z, yellow = CBX6 pdb code: 3I90, purple = CBX7 pdb code: 4MN3, ligand in green (labeled using CBX7 numbering), C) Overlay of aromatic cage of CBX2/6/7/8. CBX4 not shown in the structural overlay of the aromatic cage, as the only available X-ray structure is not bound to histone tail ligand, D) Percent identities and similarity scores (in red) of CBX chromodomains, as determined by EMBOSS matcher software pairwise alignment.

Previous studies established the pentapeptide sequence Ac-FALKme3S and related analogs as potent inhibitors of CBX7 (Table 2.2).[149] A co-crystal structure of the Ac-FAYKme3S complex with CBX7 shows the peptide ligand forming several key binding interactions in the peptide-binding groove (Figure 2.2). The N-terminus of the ligand sits in the protein's beta-groove, and the Ala residue two positions towards the N-terminus from the key trimethyllysine, points into a small hydrophobic pocket called the (–2) pocket. The hydrophobic clasp of CBX7 is made up of Val10 and Leu49 and these side chains fold and clamp around the ligand (Figure 2.2A). The Kme3 group of the ligand forms several cation-pi interactions in the protein's aromatic cage, and the C-terminal Ser residue of the ligand peptide hydrogen bonds with the carboxylate side chain of Glu43 (Figure 2.2B).

Table 2.2. Binding affinities for Ac-FALKme3, Ac-FAYKme3S and p-Br-FALKme3S to CBX7 and CBX8.[149]

	IC_{50} values (μM) determined by CBX7-H3K27me3 disruption.	**K_d values (μM) determined by ITC**	
Ligand	**CBX7**	**CBX7**	**CBX8**
Ac-FALKme3S	11 ± 0.4	2 ± 0.2	14.2 ± 2
Ac-FAYKme3S	6.2 ± 0.4	1.7 ± 0.12	12.3 ± 1.4
p-Br-FALKme3S	4.6 ± 0.3	0.28 ± 0.05	4.6 ± 0.3

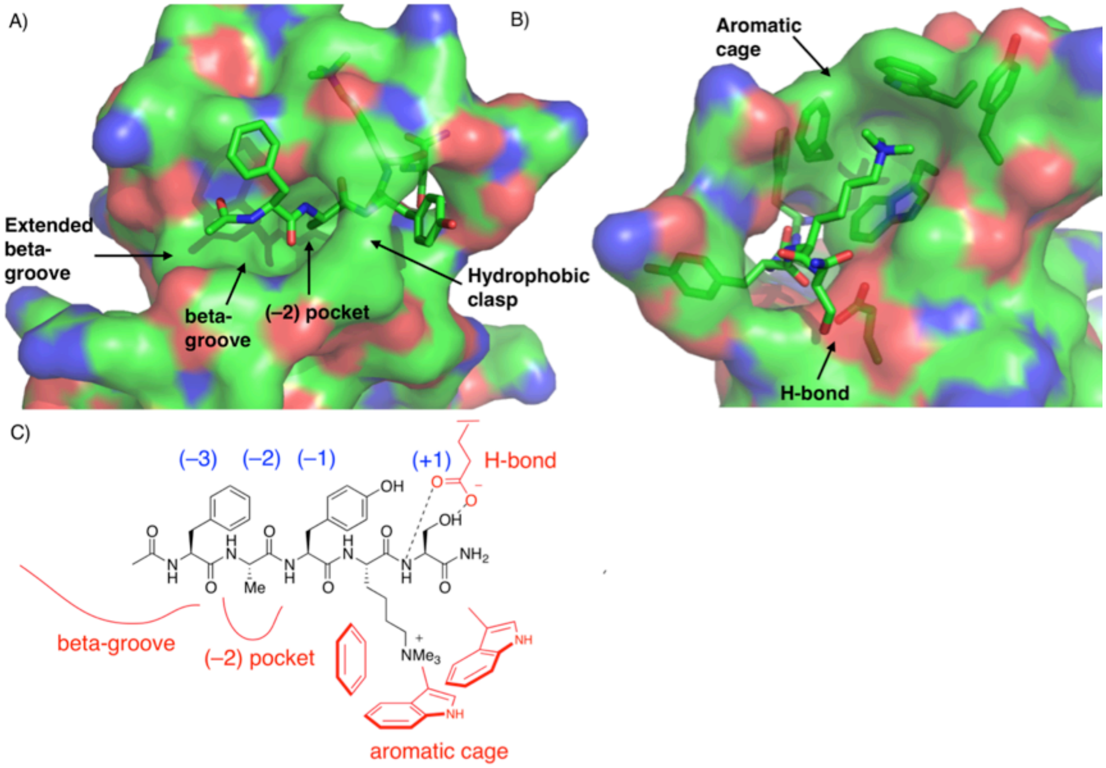


Figure 2.2. Structural analysis of key regions of the CBX proteins responsible for binding. A and B) Co-crystal structure of Ac-FAYKme3S in complex with CBX7 (pdb: 4MN3). Key structural features of the proteins binding pocket are labeled with black arrows, C) Structure of Ac-FAYKme3S and cartoon depiction of interactions with the protein. Ligand residues are numbered in relation to the Kme3 residue in blue and protein residues are labeled in red.

We chose to use direct titration fluorescence polarization (FP) assays to determine dissociation constants (K_d) for each compound against CBX1/2/4/6/7/8. FP measurements are used to study molecular interactions in solution such as receptor-ligand binding.[226] Fluorescence polarization is defined as the difference between the intensity of the emitted light parallel and perpendicular in respect to the plane polarized excitation light, normalized by the total fluorescence emission intensity (Polarization = $(F_{||} - F_{\perp})/(F_{||} + F_{\perp})$). The degree of polarization of a fluorophore is inversely related to the amount of tumbling of the molecule. The two main types of FP assays are direct (involving the binding of a dye-labeled ligand to a protein) and competitive (in which a dye-labeled probe ligand is displaced from a protein by a competing ligand). An illustration of the direct FP assay that I utilized to determine binding affinities of the dye-labeled peptidic ligands is shown in Figure 2.3. The CBX protein of interest is titrated into a solution containing the dye-labeled inhibitor and the change in the polarization signal is measured over the range of protein concentrations. Direct FP assays for determination of K_d values are advantageous because they are quick, reproducible and cost effective.

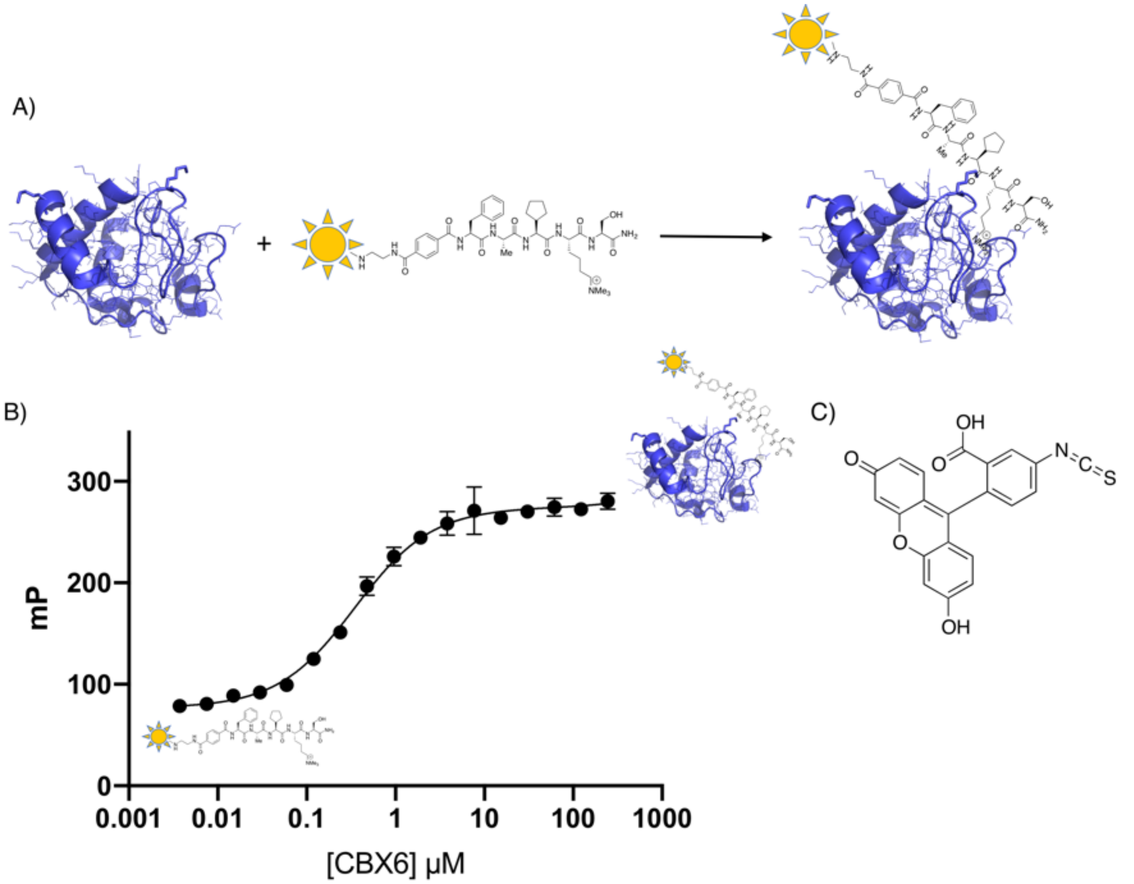


Figure 2.3. Illustration of direct fluorescence polarization (FP) assay. A) Cartoon depiction of protein binding to a dye-labeled peptide, B) fluorescence polarization graph highlighting the increase in millipolarization units (mP) from increased protein-probe complex formed upon titration of CBX6, C) chemical structure of fluorescein isothiocyanate (FITC) used for covalent labeling of peptides.

My work presented in this chapter is focused on the design and synthesis of peptidomimetic inhibitors for members of the CBX polycomb paralog family. The goal of this work was to study the structural determinants of recognition, so that we can selectively target each family member. A secondary goal was to create novel dye-labeled inhibitors as chemical tools that would allow for new biochemical and biophysical studies of the CBX proteins.

I created a small library of peptidomimetic compounds through iterative synthesis and tested each compound with a panel of CBX proteins. The library of peptides synthesized are labeled with the fluorescent dye fluorescein isothiocyanate (FITC). Dye-labeled inhibitors were used for multiple forms of testing, including direct fluorescence polarization assays (to determine affinity) and microarray testing (to determine selectivity). Each of these studies is unprecedented for the majority of methyllysine reader proteins in the CBX family.

2.3. Fluorescence polarization-driven attempts to understand polycomb paralog SAR

Peptides were synthesized using standard Fmoc solid-phase peptide synthesis protocols (Figure 2.4). Peptides contained either a beta-alanine residue at the N-terminus or a Lys(Mtt) residue at the C-terminus to allow dye labeling while still on resin. Fmoc-beta alanine was deprotected using standard protocols and then reacted with FITC to produce compounds **2.1** and **2.2**. Peptides containing a C-terminal Lys-(Mtt) residue were selectively deprotected under mildly acidic conditions, followed by a reaction with FITC to give compounds **2.3**, **2.4**, **2.8-2.11**.

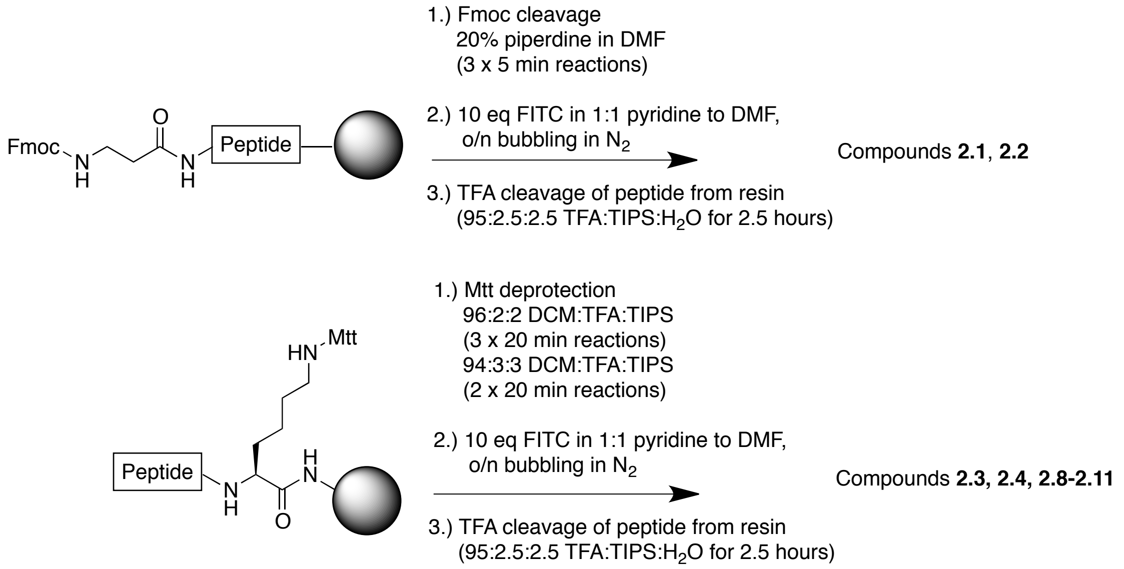


Figure 2.4. Synthetic approach for dye labeling peptides.

A dye-labeled analog of our previously identified CBX7 inhibitor FALKme3S was synthesized with a N-terminal beta-alanine residue (compound **2.1**) for covalent attachment to FITC. Compound **2.1** was screened against all CBX polycomb paralogs and the HP1 homolog CBX1. Compound **2.1** displayed potent binding to all CBX proteins tested, near equipotent binding to CBX2/4/6/7 and weaker binding to CBX1 and CBX8 (Figure 2.5A). The previously reported K_d values of Ac-FALKme3S with CBX7 and CBX8 are similar and within the same magnitude of those reported for the dye-labeled analog **2.1** (Table 2.2, Figure 2.5A).

The addition of a second phenylalanine to the N-terminus of **2.1** resulted in an increase in binding to CBX8 (2.6-fold) and decrease in binding to CBX7 (2.3-fold) (Figure 2.5B, **2.2**). Compound **2.2** showed a 19-fold decrease in binding to CBX1, 2 to 3-fold decrease in binding to CBX2/4 and no significant change in binding to CBX6.

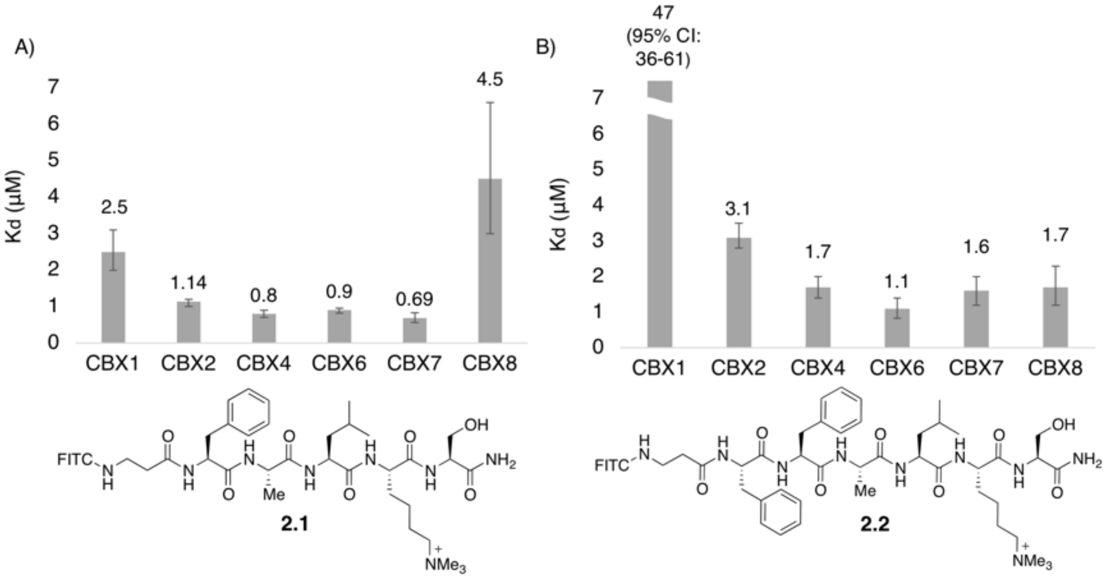


Figure 2.5. Peptidic inhibitors for CBX proteins and corresponding dissociation constants for CBX1/2/4/6/7/8. A) Binding affinities and chemical structure of compound **2.1**, B) binding affinities and chemical structure of compound **2.2**. Error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

One key structural difference between CBX7/CBX8 is the presence of an Arg or Glu residue in the protein where the C-terminus of the ligand interacts (Figure 2.6A, B). We predicted that an anionic substitution at the (+2) position would improve binding to CBX6/8 but not CBX7. To test this, two analogous compounds were synthesized containing either a neutral Leu residue at the (+2) position (compound **2.3**) or a anionic Glu residue (compound **2.4**) (Figure 2.6C). Compound **2.3** was a potent binder of CBX6 and CBX7 with K_d values of 78 and 11 nM. Addition of a Glu residue in the (+2) position (**2.4**) gave small increases in binding to CBX6 (1.6-fold) and CBX8 (1.4-fold) with no change in binding to CBX7 and a decrease in binding to CBX1 (2.7-fold) (Figure 2.6C, D).

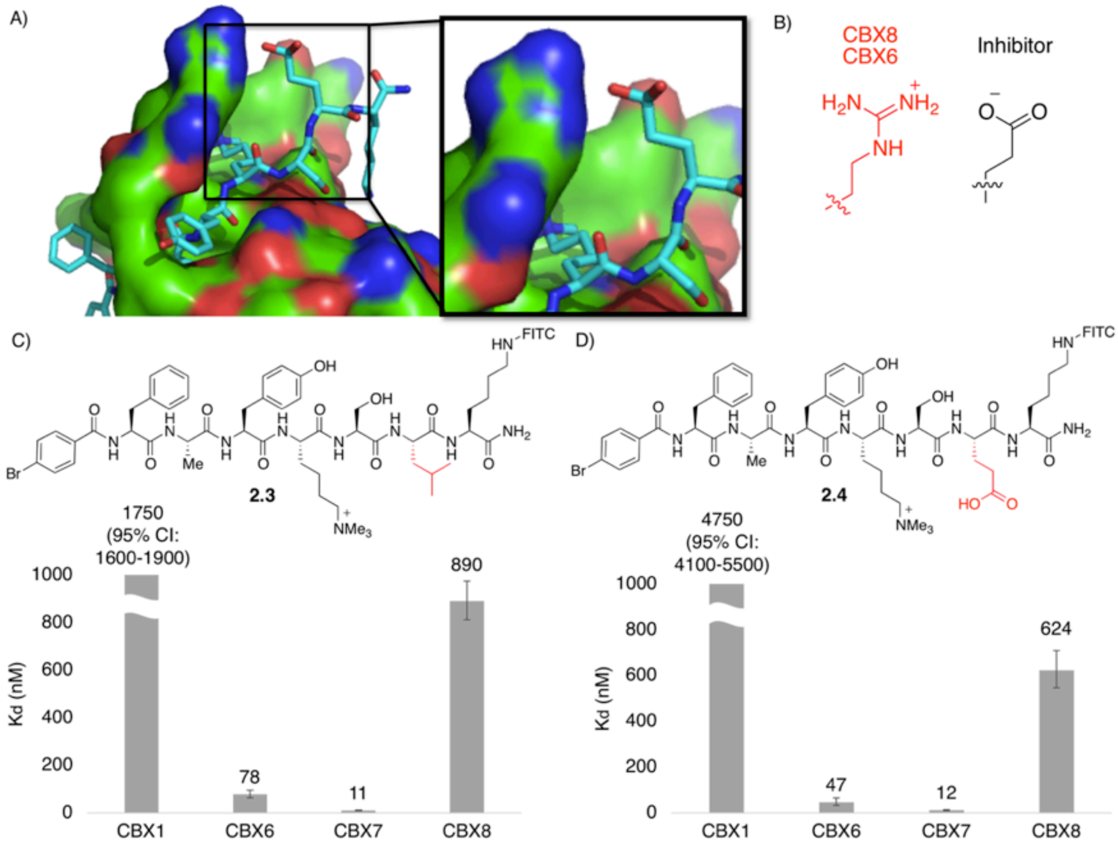


Figure 2.6. Interactions between anionic ligand substituent and cationic protein residue in CBX6/CBX8. A) Moloc model of compound **2.4** with CBX8 (pdb: 3I91) showing zoomed in interaction between Arg residue in protein and anionic Glu ligand substitution, B) Depiction of interaction between protein Arg and ligand Glu, C) Binding affinities of **2.3** to CBX1/6/7/8, D) Binding affinities of **2.4** to CBX1/6/7/8. K_d values are reported in nM and error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

To investigate further the role of the Glu residue and to rule out effects of the dye on binding trends, we synthesized analogues of compounds **2.3** and **2.4** lacking the C-terminal lysine and FITC label. A competitive FP assay was used to determine the IC_{50} values of the unlabeled compounds **2.5**, **2.6** and **2.7** (Figure 2.7). Compound **2.4** was used as the dye-labeled probe in the competitive FP assay because of its solubility and low K_d values for all CBX proteins. Disruption of the probe-protein complex by titration of the unlabeled inhibitor results in a decrease in polarization. This data can be plotted as a dose-response curve for the determination of IC_{50} values. Compound **2.5**, lacking a residue at the (+2) position, displayed binding to CBX7 with 5- and 30-fold selectivity over CBX6 and 8 respectively. The addition of a Leu residue at the (+2) position (**2.6**) did not significantly change binding to any of the CBX proteins tested. Compound **2.7** with (+2) Glu residue gave 1.7-fold and 1.4-fold improvements in binding to CBX6 and 8, respectively, and a 1.6-fold decrease in binding to CBX7. The trends in binding affinities observed with the dye-labeled compounds **2.3** and **2.4** are consistent with the trends seen for the unlabeled compounds **2.6** and **2.7.**

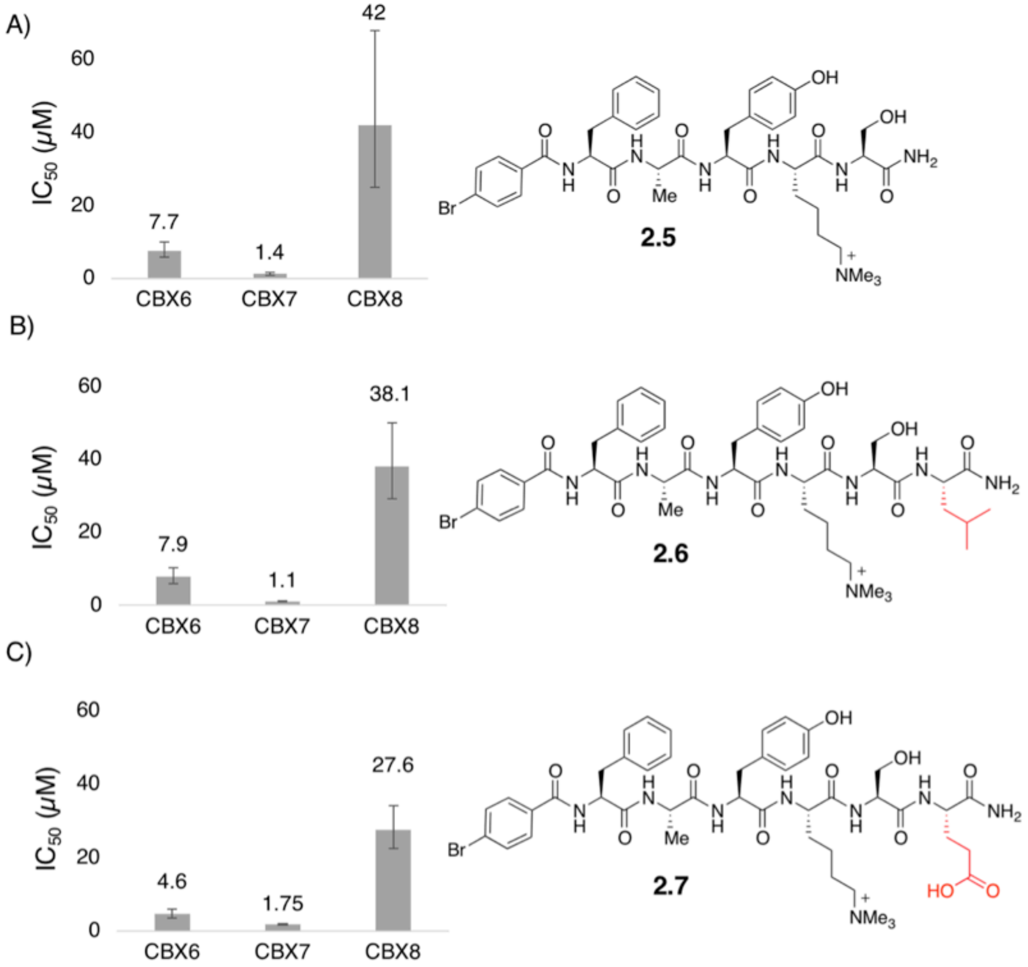


Figure 2.7. IC_{50} values of unlabeled ligands **2.5-2.7** for CBX6/7/8 determined by competitive FP. A) Binding affinities and chemical structure of compound **2.5**, B) Binding affinities and chemical structure of compound **2.6**, C) Binding affinities and chemical structure of compound **2.7**. Error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

We carried out computational ligand docking in order to understand better the effects of the proposed salt bridge interaction, and to provide starting structures for molecular dynamics (MD) simulations. We used the crystal structure of CBX8 complexed with H3K9me3 (pdb: 3I91), with H3K9me3 removed, for docking of unlabeled compounds **2.6** and **2.7** (Figure 2.8). The docking experiment showed a clear preference for a negatively charged residue in the glutamate position of ligand **2.7**. The salt bridge interaction was seen to occur mostly with Arg25, instead of the predicted Arg9 shown in Figure 2.6 (Figure 2.8B, C). A preference for a charged Glu residue in the ligand seems to be better explained by the significant patch of positive charge on the protein surface of both CBX6 and CBX8 surrounding the aromatic cage, for which the Arg9 residue is only partly responsible.[41] The docking process was not able to generate a properly bound configuration for the Leu-containing ligand (**2.6**). In the best docking pose generated for compound **2.6**, the ligand orientation in the binding site was completely reversed, bringing the meaning of this particular docked structure into question (Figure 2.8). In order to obtain useful data from MD simulations of **2.6**, the docked structure of **2.7** was used as a template for the starting structure of CBX complexes with **2.6**.

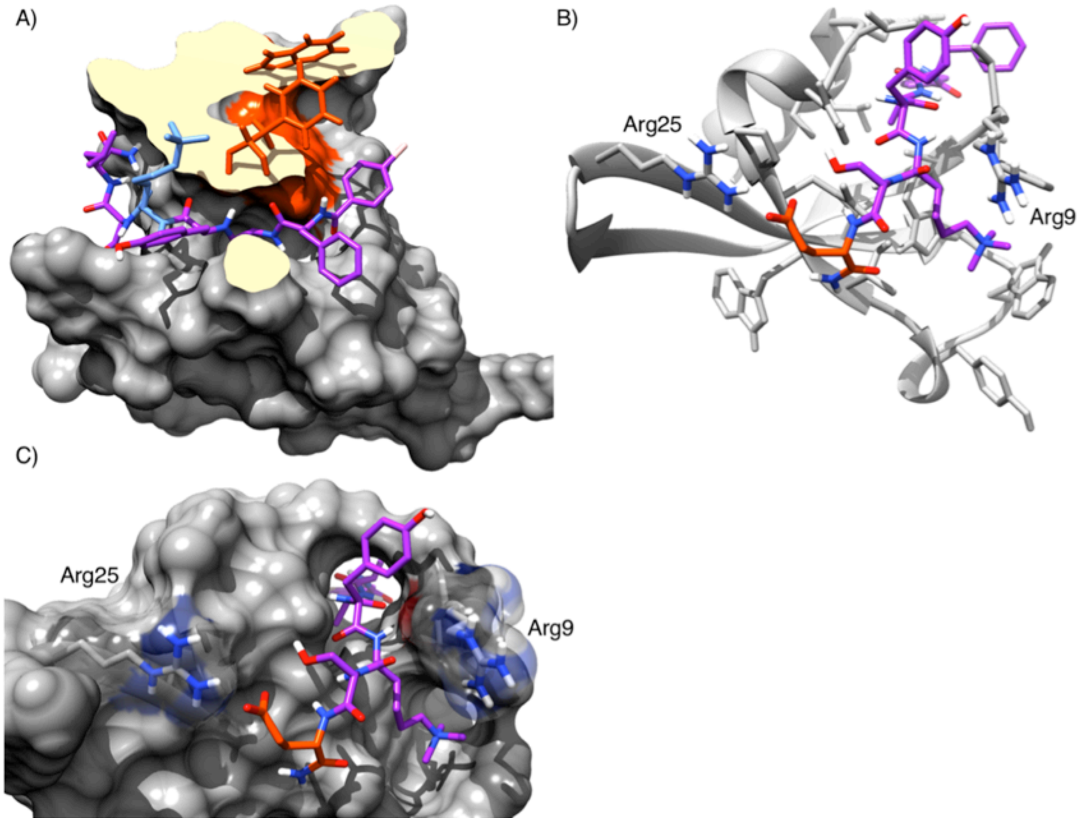


Figure 2.8.Best poses for compounds **2.6** and **2.7** with CBX8 (pdb:3I91) generated using Autodock Vina. A) Compound **2.6** (–7.8 Docking Vina docking score, Kme3 residue coloured blue and aromatic cage residues coloured orange) was unable to be docked. The structure of **2.7** was used for docking of **2.6**. B and C) compound **2.7** (–8.3 Docking Vina docking score) with glutamate residue participating in a salt bridge interaction with Arg25 coloured orange.

Both compounds in MD simulations with CBX8 showed stability in respect to two of the interaction sites, the (–2) pocket and the aromatic cage. The only difference observed in a key interaction site was a slight and possibly insignificant increase in Ser-Glu43 hydrogen bonding (Figure 2.2, Figure 2.9). From the combined molecular dynamics trajectories, compound **2.6** shows a tendency to retain the serine hydrogen bond for longer periods of time resulting in a more stable configuration below the aromatic cage. Monitoring of the interactions with compound **2.6** showed the ligand's Leu residue interacting with both of the two Arg residues with no preference in either position (Figure 2.9C). With respect to the salt bridge with Arg (compound **2.7**), there is formation on and off of this interaction shown by the flat 5 angstrom distances formed with Arg9 (Figure 2.9D). Our hypothesis for the Glu residue of compound **2.7** interacting with Arg9 seems correct, however this interaction comes at a cost. We observe a compromise in binding whereby the serine hydrogen bond anchor is destabilized by the formation of this salt bridge interaction.

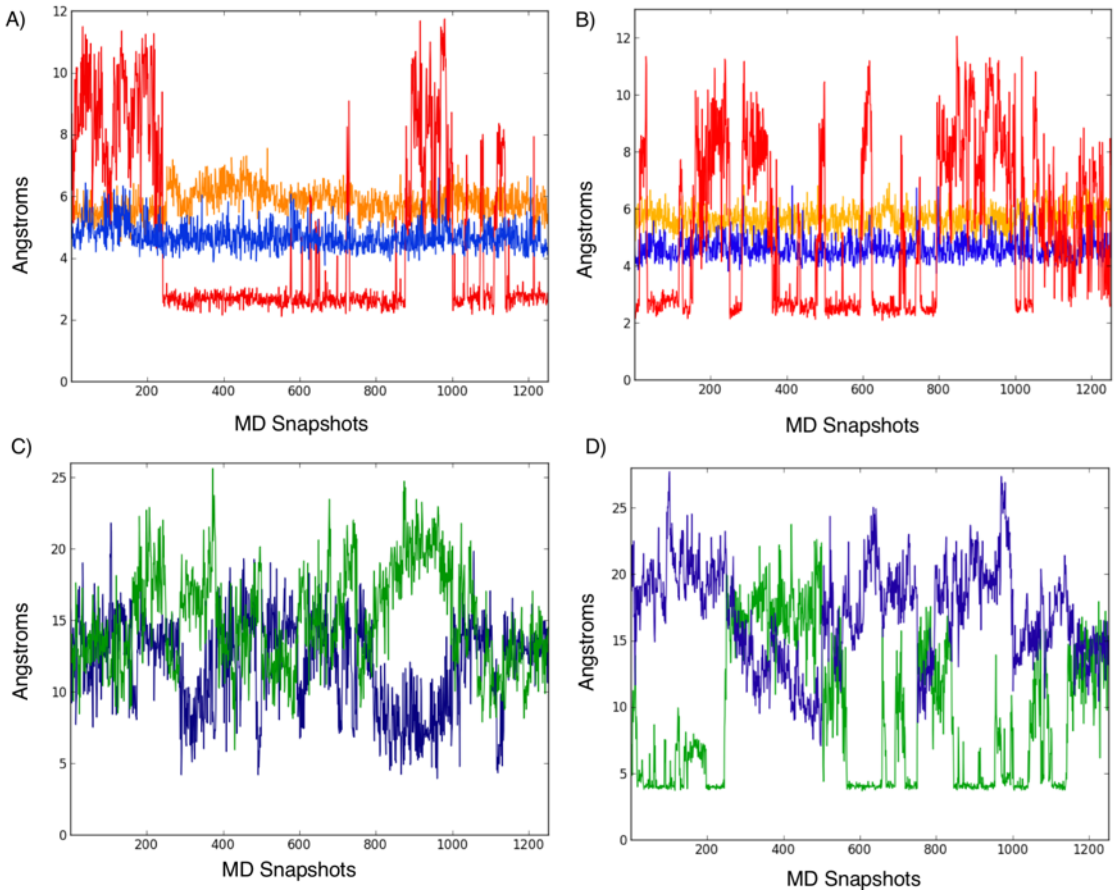


Figure 2.9. Key residue distances calculated from MD simulations of ligands **2.6** and **2.7** with CBX8. A) MD simulation for compound **2.6** and B) MD simulation for compound **2.7**. A and B describe anchor residue distances. Hydrophobic clasp (ligand-Ala – Val10 distance in orange), aromatic cage (ligand-Kme3 – Trp32 distance in blue), anchor hydrogen bond below aromatic cage (ligand-Ser – Glu43 distance in red). C) MD simulation for compound **2.6** and D) MD simulation for **2.7**. Graphs C and D describe salt bridge residue distances. Arginine clasp (ligand Glu or ligand-Leu – Arg9 distance in green and ligand-Glu or ligand-Leu – Arg25 distance in blue). The distance between ligand Glu or ligand-Leu – Arg25 (blue) was the initial docking preferred arginine. 1250 snapshots were out over 50 ns of simulation time.

Returning to synthesis and testing of new ligands, we sought to explore substitutions of the (–2) residue within the scaffold of **2.4**. An additional difference between CBX7 and CBX8 is the hydrophobic (–2) pocket. The (–2) pocket in CBX7 is partially defined by a Val residue and is replaced with an Ala residue in CBX8. We sought to exploit this difference by adding a larger alkyl substituent at the (–2) position. Within the scaffold of compound **2.4**, the methyl substituent at the (–2) position was replaced with an ethyl group to produce compound **2.8** (Figure 2.10A). Significantly weaker binding was observed for **2.8** with CBX1 (31-fold), between a 2- to 5-fold decrease in binding for CBX2/4/6/7 and no change in binding to CBX8.

We next replaced the Phe residue with a cyclopentyl moiety at the (–1) position to give compound **2.9** (Figure 2.10B). Addition of the cyclopentyl group in this position was previously shown to improve binding of peptidic inhibitors to CBX7.[149] This swap produced the most potent CBX8 inhibitor made to date with a K_d value of 120 nM. In addition to improving potency for CBX8, the cyclopentyl side chain in the (–1) position increased binding to all CBX proteins tested. Compound **2.9** is 26-fold selective for CBX8 over CBX1, 3-fold selective over CBX2/6, and 5-fold selective over CBX4. The difference in affinity of **2.9** for CBX7 and CBX8 is not significant. Due to its broad-spectrum ability to bind all polycomb CBX chromodomains, we predict **2.9** to be a useful dye-labeled probe for future FP screening efforts.

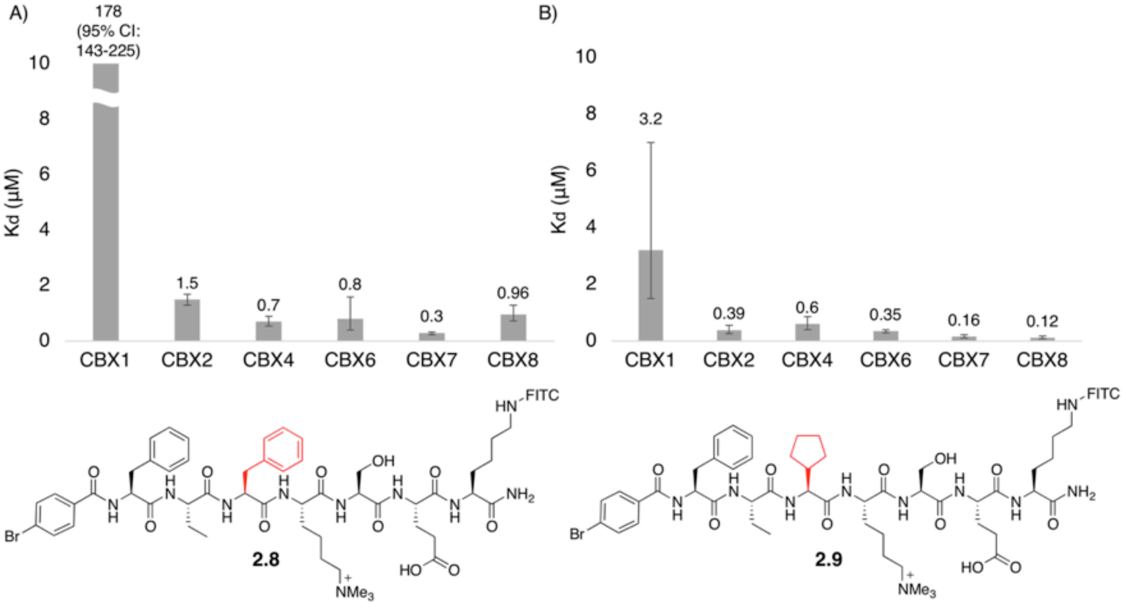


Figure 2.10. Peptidic inhibitors **2.8** and **2.9** and corresponding dissociation constants for CBX1/2/4/6/7/8. A) Binding affinities and chemical structure of compound **2.8**, B) binding affinities and chemical structure of compound **2.9**. Error bars are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

We hoped to improve the selectivity of **2.9** by adding a second Phe residue as the N-cap substituent. To this end, we synthesized compound **2.10** containing an acetylated N-capping Phe residue in combination with the (–2) ethyl group and (–1) cyclopentyl side chain (Figure 2.11A). To our knowledge, **2.10** exhibits the greatest affinity for CBX8 compared to CBX7. **2.10** is most potent for CBX6, and is 4 and 2-fold selective for CBX6 and 8 over CBX7. In respect to CBX8, **2.10** is between 14- and 135-fold selective over CBX1/2/4.

Replacing the (–2) and (–1) substituents of **2.10** to Ala-Leu residues known to be favoured by CBX1/2/4/7 did not significantly change binding to these proteins, but did decrease binding to CBX6 and 8 by a factor of 2.7 and 3.5 (Figure 2.11B). Compounds **2.10** and **2.11** are selective for CBX6/7/8 over CBX1/2/4.

Selective inhibition of CBX7 over CBX4 has not yet been reported. CBX4 is the most similar CBX protein to CBX7, with a similarity score of 90% (Figure 2.1), and both bind the native histone substrate with similar affinity (Table 2.1). Interestingly, both **2.10** and **2.11** show significantly weaker binding to CBX4 compared to CBX7. Future efforts on selective inhibition of CBX7 may benefit from extended engagement of the peptide-binding groove.

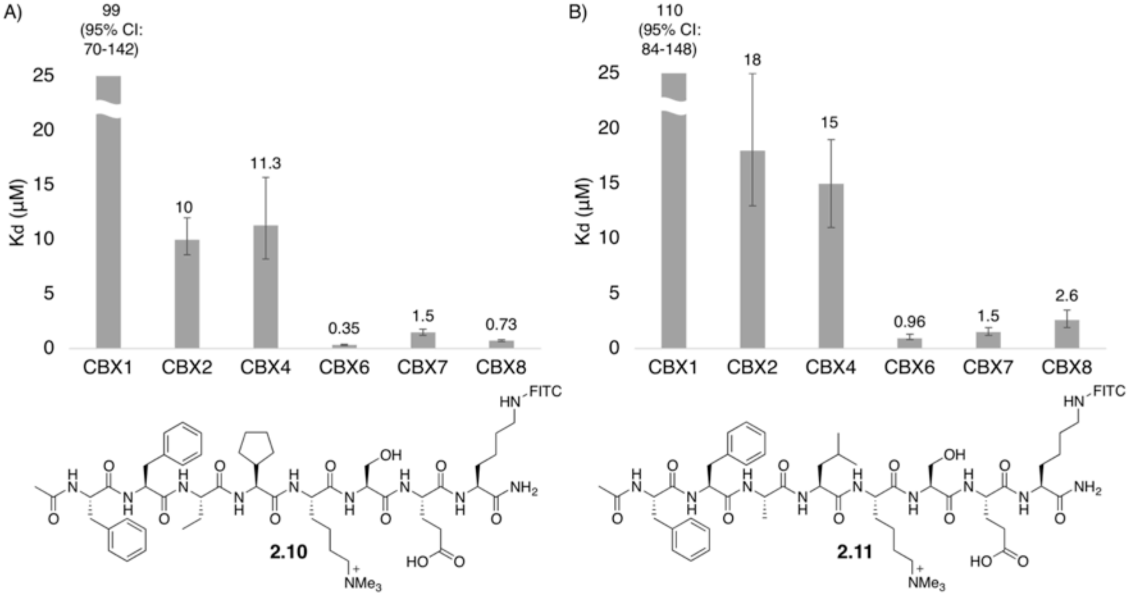


Figure 2.11. Peptidic inhibitors **2.10** and **2.11** and corresponding dissociation constants for CBX1/2/4/6/7/8. A) Binding affinities and chemical structure of compound **2.10**, B) binding affinities and chemical structure of compound **2.11**. Error bars are reported as

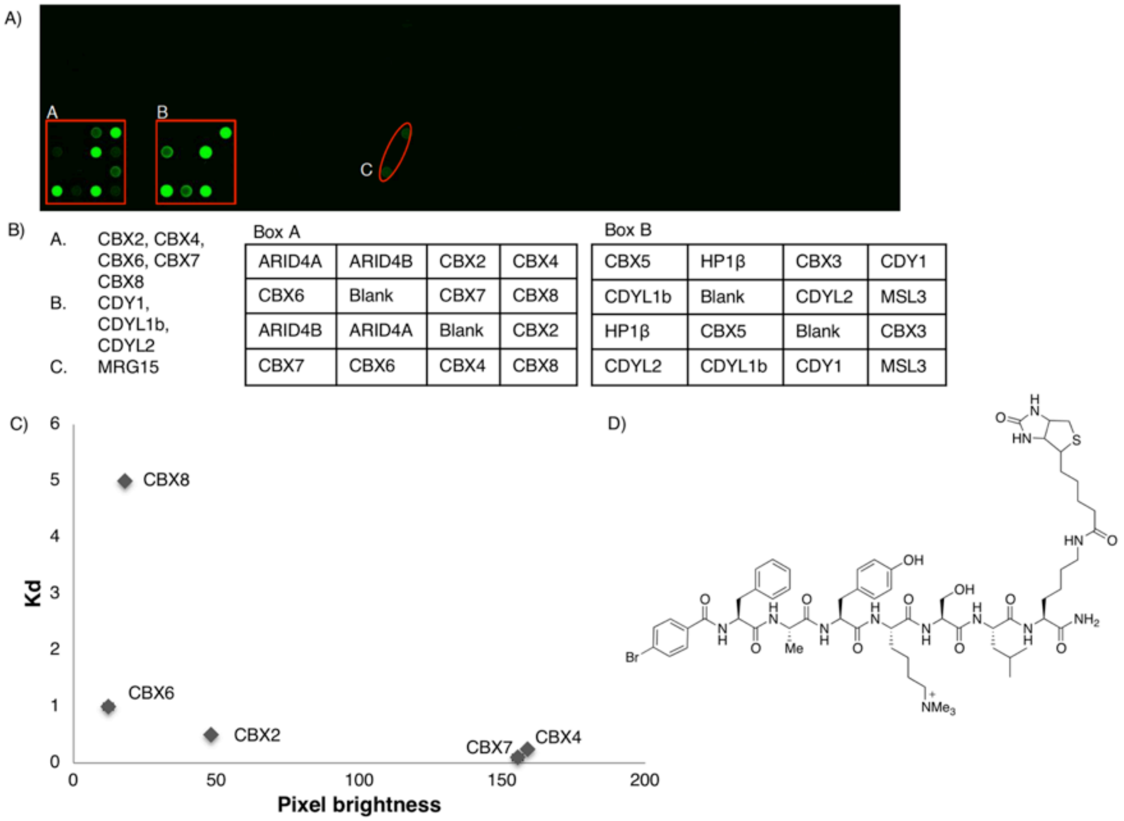
asymmetric 95% confidence intervals from experiments done in duplicate for CBX1/2/4/6/7 and triplicate for CBX8.

2.4. Selectivity studies using a methyl reader protein microarray

Tagged inhibitors allow for visualization of binding in protein microarrays and this informs on the inhibitor's selectivity against a set of protein targets. Functional protein microarrays are made by attaching proteins to a surface and are useful for characterization of protein-protein or protein-ligand interactions.[227] After production and purification, the proteins of interest are reacted with and immobilized onto a solid support.

Using a microarray to test our fluorescently tagged peptides provides both validation of our in-house FP results and more diverse knowledge on selectivity beyond the CBX family of proteins. Our collaborator, Dr. M. Bedford, from the MD Anderson Cancer Center, created a protein microarray made up of 98 different recombinant human methyl reader proteins (including all CBX proteins) arrayed in duplicate. Initial testing of tetramethylrhodamine isothiocyanate (TRITC) and FITC dye-labeled inhibitors produced the expected binding trends but with a high degree of background fluorescence. To prevent this, a biotinylated analog of **2.3** was synthesized (compound **2.12**, Figure 2.12C). The microarray was incubated with the probe, and the binding of the probe to microarrayed reader proteins was imaged using a streptavidin fluorescent reagent.

The inhibitor tested in the microarray showed excellent selectivity for CBX proteins over almost all of the other microarrayed methyl readers. Some off-target binding was observed to the chromodomain Y like (CDYL) proteins and weak off-target binding to the chromodomain-containing mortality factor 4-like protein (MORF4L1 or MRG15). Compound **2.12** showed the highest selectivity for CBX4 and 7, with weaker binding observed to CBX2/6/8 (Figure 2.12).



A.	CBX2, CBX4, CBX6, CBX7 CBX8
B.	CDY1, CDYL1b, CDYL2
C.	MRG15

Box A

ARID4A	ARID4B	CBX2	CBX4
CBX6	Blank	CBX7	CBX8
ARID4B	ARID4A	Blank	CBX2
CBX7	CBX6	CBX4	CBX8

Box B

CBX5	HP1β	CBX3	CDY1
CDYL1b	Blank	CDYL2	MSL3
HP1β	CBX5	Blank	CBX3
CDYL2	CDYL1b	CDY1	MSL3

Figure 2.12. Protein microarray made up of 98 human methyl reader proteins shows peptidic inhibitors are selective against a broad set of methyl reader targets. Proteins were coated onto the wells in each plate with each square representing a family of proteins. A) Protein microarray with compound **2.12**. B) Legend of proteins in box A and B. Box A contains polycomb paralogs, box B contains HP1 homologs and CDYL proteins, and box C highlights weak off target binding to MRG domains. The full map of the protein microarray is available in Figure 2.38. C) K_d values of compound **2.3** (dye labeled analog of **2.12**) with CBX proteins plotted against brightness of each spot from protein microarray (quantified using ImageJ), D) Chemical structure of biotin-tagged compound **2.12**. *Microarray data from Dr. Mark Bedford at the MD Anderson Cancer Center*.

To ensure this data was in agreement with K_d values we collected from our FP assays, ImageJ software was used to quantify the brightness of each microarray spot and this was plotted against the K_d values for analogous dye-labeled compound of **2.12** (Figure 2.12C). In general, it is clear that proteins with weak K_d values also have weak pixel intensity showing that the microarray data is in qualitative agreement with our solution-phase results. CBX6 shows significantly less pixel brightness than would be expected from its K_d for this probe ligand, suggesting that microarrayed CBX6 does not accurately represent solution activity.

2.5. Cell-based data

We sought to determine the ability of the dye-labeled peptidic agents to be used in cell-based studies. Our efforts to study the inhibitors in cells included cell imaging, fluorescence-activated cell sorting (FACS), and MTT-based viability cell studies.

Cell-based studies showed that the compounds could enter cells, but did not generate a biological effect. For these studies we used TOV21G cells, an ovarian clear cell carcinoma cell line in which CBX signaling has been implicated as a driver of growth.[166] Live cell imaging and immunofluorescence confocal imaging done by our collaborators using TOV21G and PC3 cells treated with compounds **2.2** and **2.4** did not show compounds entering the cytoplasm or nucleus (data not shown). FACS experiments with TOV21G cells treated with compound **2.2** showed uptake of the inhibitor, supported by the formation of a population of fluorescent cells in each case (Figure 2.13A). Sorting the cells according to their cell cycle stage showed that the presence of inhibitor led to no observable change in cell cycle (Table 2.3). This is also consistent with compounds being taken up in endosomes, but not being able to escape to cytoplasm or nucleus in order to have a biological effect. We also carried out MTT assays, used to measure metabolic activity and cell viability, with TOV21G cells treated with inhibitor **2.4**, and unlabeled analogs **2.6** and **2.7**. A slight decrease in cell viability was seen for the cells treated with **2.4**, however we did not observe a dose-response for the unlabelled analogs **2.6** and **2.7** (Figure 2.14). These experiments were carried out as technical triplicates and further experiments were not conducted due to the lack of promising evidence supporting the inhibitors to be cell permeable.

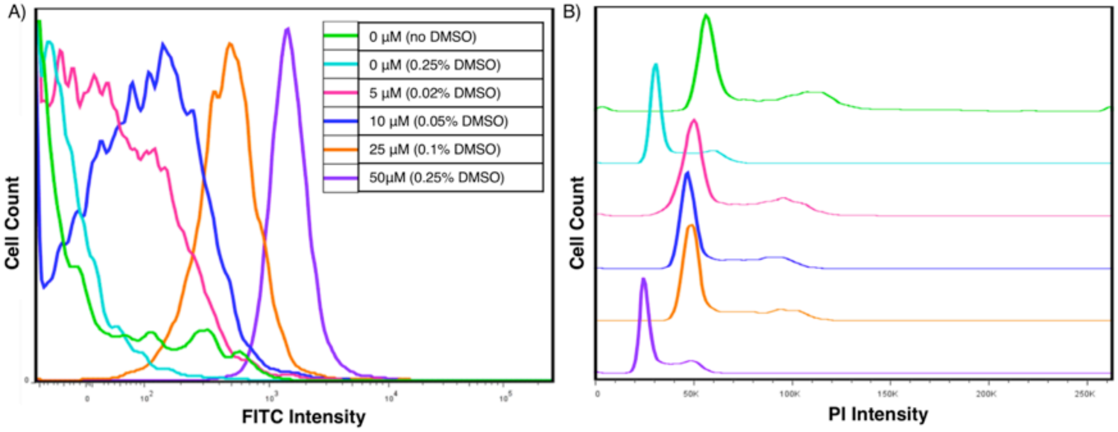


Figure 2.13. Uptake efficiency and PI based cell cycle analysis for TOV21G cells treated with **2.2** show uptake of inhibitor with no change to cell cycle. A) Uptake efficiency based on FITC intensity (BluFL1). Samples of equivalent DMSO (light blue trace compared to purple trace) show increased fluorescent signal arising from 50 µM treatment of **2.2**. B) Cell cycle distribution based on PI intensity (BluFL2).). Samples of equivalent DMSO (light blue trace compared to purple trace) show no change in cell cycle. Cells seeded at 2.5×10^5 cells/well in a 6-well plate, grown overnight at 37°C, 5% CO_2, treated for 21 h with 0 µM, 5 µM, 10 µM, 25 µM or 50 µM of **2.2**. After trypsinization, cells were stained with Fixable Viability Dye eFlour 780, followed by ethanol fixation and PI staining. Samples were run on Becton Dickinson FACS Calibur^{TM} and analyzed using FlowJo 7.6.5 software.

Table 2.3. Changes in percent distribution of cells in G1, S and G2 stage of cell cycle at different concentrations of treatment (measured from Figure 2.13).

Concentration of 2.2	**G1 (%)**	**S (%)**	**G2 (%)**
0 µM (no DMSO)	49	14	21
0 µM (0.25% DMSO)	66	14	12
5 µM (0.02% DMSO)	60	13	16
10 µM (0.05% DMSO)	65	14	12
25 µM (0.1% DMSO)	66	15	11
50 µM (0.25% DMSO)	68	12	12

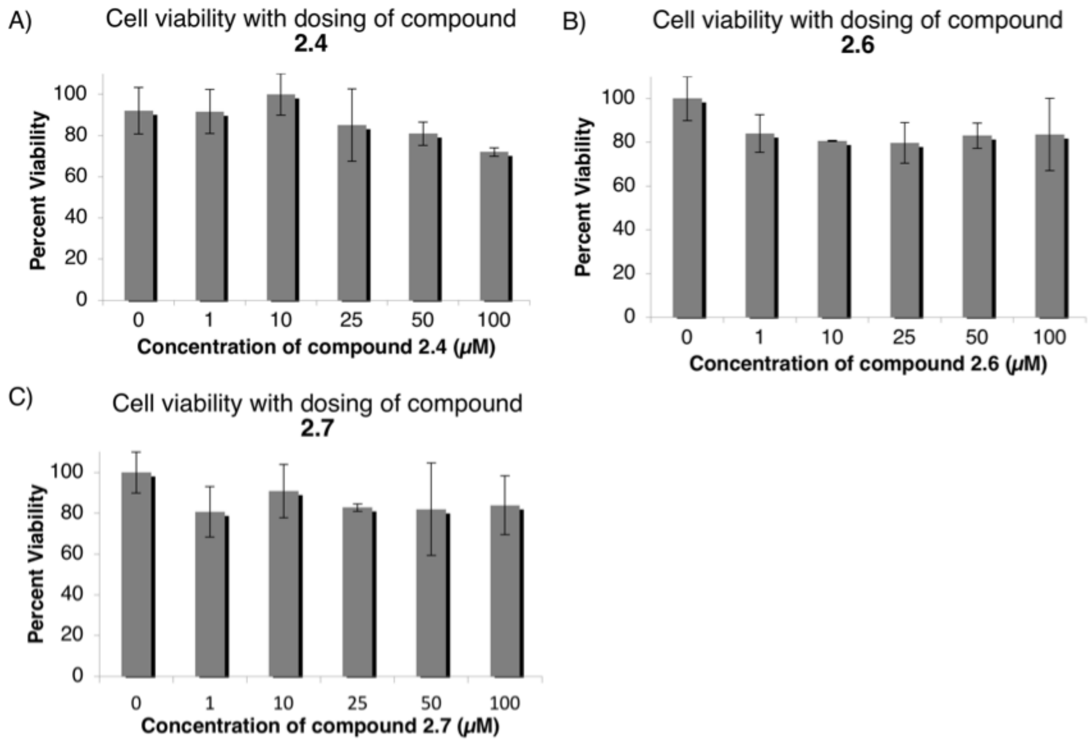


Figure 2.14. Cell viability determined using an MTT assay carried out with TOV21G cells treated with compounds **2.4**, **2.6** and **2.7**. A) Percent viability of TOV21G cells treated with **2.4**, B) Percent viability of TOV21G cells treated with **2.6**, C) Percent viability of TOV21G cells treated with **2.7**. Experiment carried out in triplicate with error bars representing one SD.

2.6. Discussion

Potent inhibitors were developed that are either pan-specific or partially selective within the polycomb paralog family. None of the inhibitors developed were selective for a single CBX protein. We have identified several compounds with sub-micromolar affinity including two inhibitors that bind CBX6/7 with IC_{50} values < 100 nM (**2.3**, **2.4**). Although the desired selectivity was not achieved, these inhibitors have potential as pan-specific chemical tools.

Several compounds displayed partial selectivity, with preferential binding to CBX6/7/8 over CBX1/2/4. We report inhibitors with greater affinity for CBX6/8 over CBX7 and this has not previously been reported. This includes the most potent CBX8 inhibitor to date (**2.9**, 120 nM) and an inhibitor that is 2-fold selective for CBX8 over CBX7 (**2.10**). Our results cumulated in the first peptide inhibitor that is 10-fold selective for CBX7 over the highly similar CBX4 (compound **2.10**).

We have identified areas of the peptide-binding groove that are different within the family of CBX proteins and form distinct interactions with the peptide ligands reported. Substitutions at the (–1) position effect the hydrophobic clasp of the CBX proteins and alter binding affinities differentially within the family. Addition of a cyclopentyl moiety in the (–1) position increases binding to all CBX members with the greatest increase observed for binding to CBX8. The addition of two Phe residues at the N-terminus of the ligand diminishes binding to CBX1/2/4 to give inhibitors that are selective for CBX6/7/8 (**2.9**, **2.10**). We predict that future efforts targeting the extended beta-groove, (–1) and (–2) position of the protein will aid in the discovery of selective inhibitors.

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

The inhibitors developed are highly selective for the CBX polycomb proteins over other methyllysine reader proteins. The protein microarray studies with **2.12** show the inhibitors to be highly selective for the CBX polycomb paralogs over many other methyllysine readers. Potent off-target binding was observed with the CDY proteins. Future optimization of the inhibitors is needed to differentiate binding between the two highly similar families of chromodomains.

Our efforts to use the dye-labeled peptide in cell-based studies were unsuccessful due poor cell permeability and presumed endosomal entrapment. Poor solubility of the peptidic inhibitors was another challenge and limited our ability to test higher concentrations of the inhibitors. Future efforts to use the reported inhibitors in cell-based studies will require investigating alternative delivery strategies.

2.7. Conclusion

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

The SAR learned from this work provides new insights into the structure and molecular recognition properties of these proteins. In particular, we have furthered our knowledge of differences in the peptide-binding groove within the CBX family. Future progress is needed to develop selective cell-permeable chemical tools to study the role of CBX proteins in epigenetic regulation in novel and dynamic ways.

2.8. Experimental methods and supplementary data

2.8.1 General Synthesis

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

2.8.2 Side-chain functionalization with fluorescein isothiocyanate (FITC)

Side-chain functionalization with fluorescein isothiocyanate (FITC) was achieved by incorporation of either a Fmoc-beta-Alanine or a side-chain Mtt-protected lysine (Figure 2.4). For functionalization of Fmoc-beta-Alanine, Fmoc was removed following standard protocols after synthesis was completed on resin. A solution of 20% piperdine in DMF (10 mL) was added to the resin and bubbled with N_2 for 5 minutes (repeated 3×). The resin was washed 3× with DMF between each reaction. FITC (5 equiv) was added to a 10 mL solution of 1:1 DMF and pyridine. This solution was added to the resin and bubbled with N_2 at ambient temperature overnight.

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

Following the functionalization with FITC, all peptide/resin solutions were filtered and the resin washed with DMF (3 × 5 mL) and CH_2Cl_2 (3 × 5 mL) and air-dried. The product was cleaved from resin with 10 mL of 95:2.5:2.5 TFA/H_2O/triisopropylsilane for 2.5 h. The solution was then concentrated in vacuo and added to cold diethyl ether to yield a crude yellow precipitate that was collected by centrifugation. Purifications by preparative HPLC were carried out using a Phenomenex Luna, 5 µm, C-18 column, 250 x 21.20 mm column. Peptides were characterized using LC-MS and ESI-MS and purity was determined to be >95% by analytical LC-MS. Retention times reported arise from analytical traces done on a Thermo Scientific C-18 column, 5 µm, 4.6 mm x 250 mm, flow rate of 1.5 mL/min, gradient running from 90:10 water (0.1 % TFA) and MeCN (0.1 % TFA) to 10:90 water (0.1 % TFA) and MeCN (0.1 % TFA) over 30-38 min.

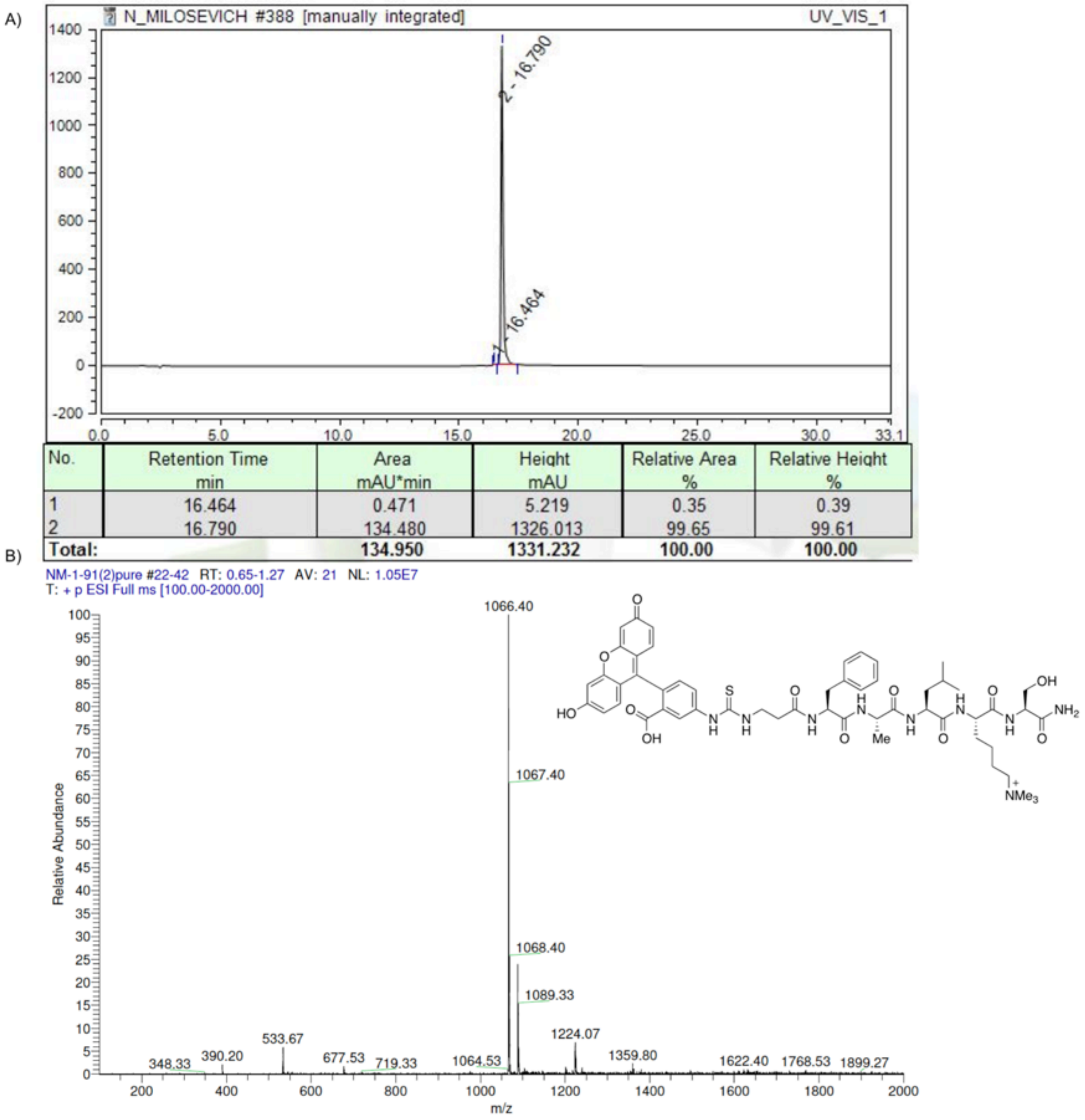
2.8.3 Side-chain functionalization with biotin

For side-chain functionalization with biotin, a side-chain (Mtt)-protected lysine was used. After synthesis on resin, the Mtt protecting groups were cleaved as described above. (+)-Biotin *N*-hydroxysuccinimide ester (5 equiv) was added to a 10 mL solution of DMF with DIPEA (10 equiv). This solution was then added to the resin and bubbled with N_2 at ambient temperature overnight. Cleavage and purification is the same as described above.

2.8.4 *N*-terminal acetylation of peptides

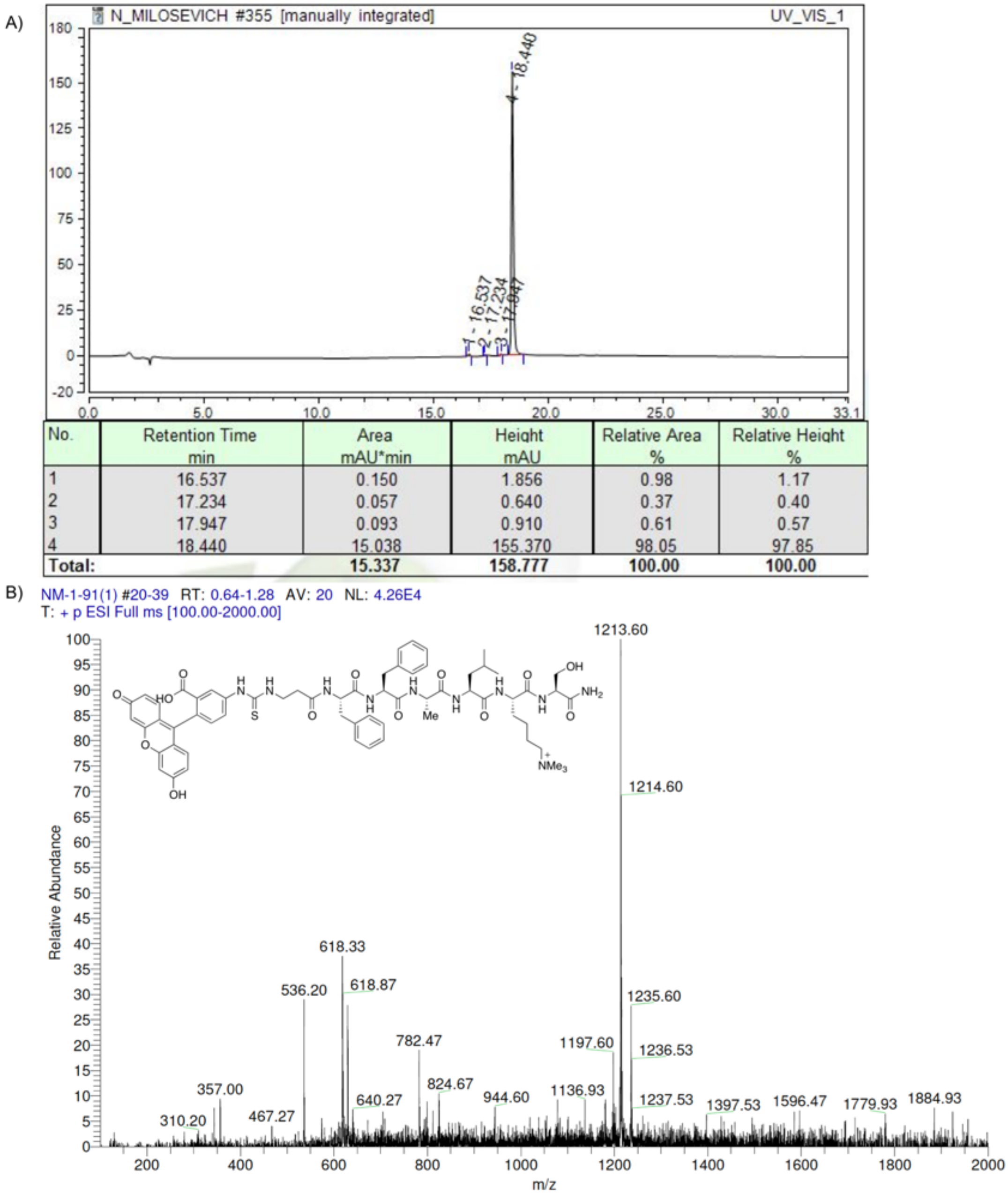
Following deprotection of *N*-terminal Fmoc (as described above), a mixture of pyridine/Ac_2O/CH_2Cl_2 in a ratio of 30:20:50 was added to the peptide on resin. The reaction was bubbled with N_2 for two hours at ambient temperature. Cleavage and purification is the same as described above.

2.8.5 Compound characterization data



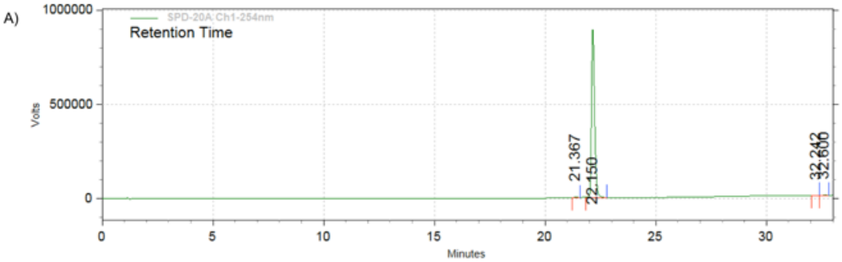
No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	16.464	0.471	5.219	0.35	0.39
2	16.790	134.480	1326.013	99.65	99.61
Total:		**134.950**	**1331.232**	**100.00**	**100.00**

Figure 2.15. Characterization data for compound **2.1**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{54}H_{68}N_9O_{12}S^+$: 1066.47; found: 1066.40.



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

Figure 2.16. Characterization data for compound **2.2**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{63}H_{77}N_{10}O_{13}S^+$: 1213.54; found: 1213.60.



SPD-20A
Ch1-254nm
Results

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

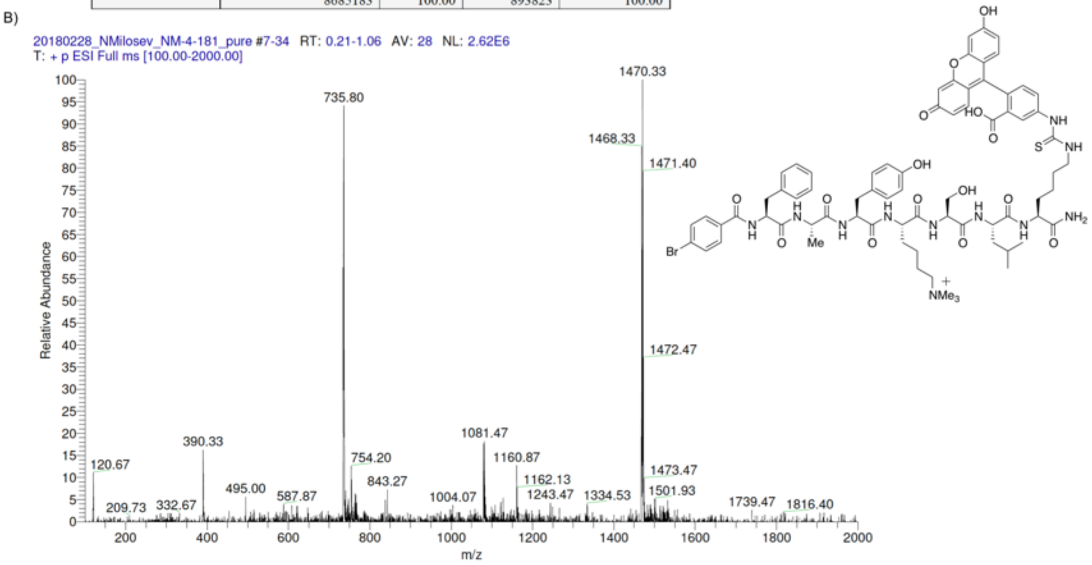
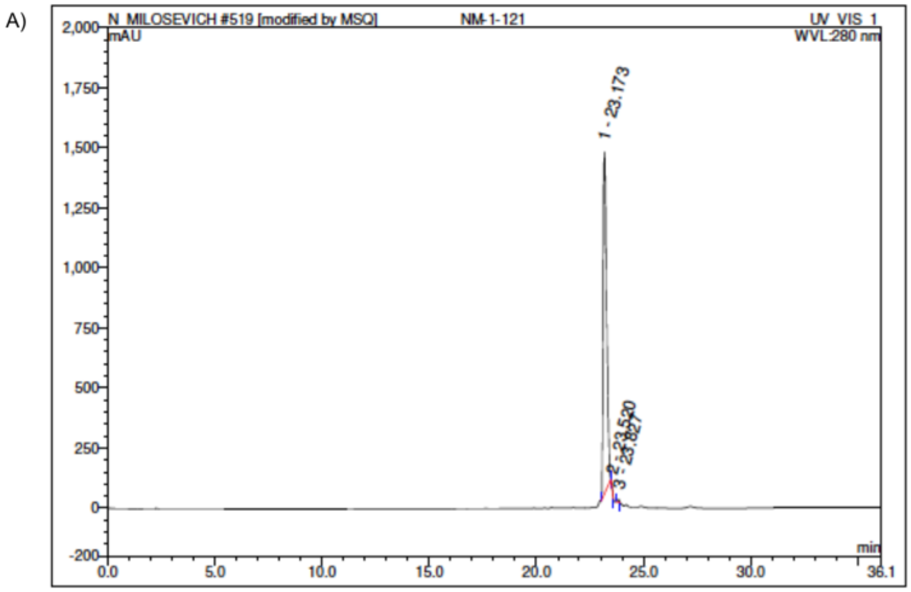


Figure 2.17. Characterization data for compound **2.3**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{73}H_{89}BrN_{11}O_{15}S^{+}$: 1470.54; found: 1470.33.



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

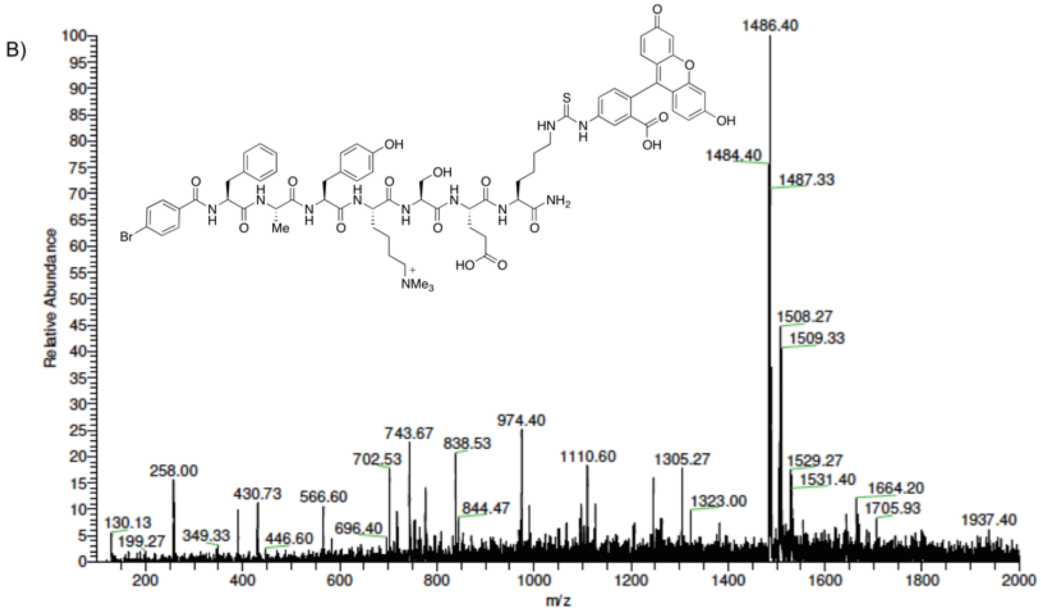
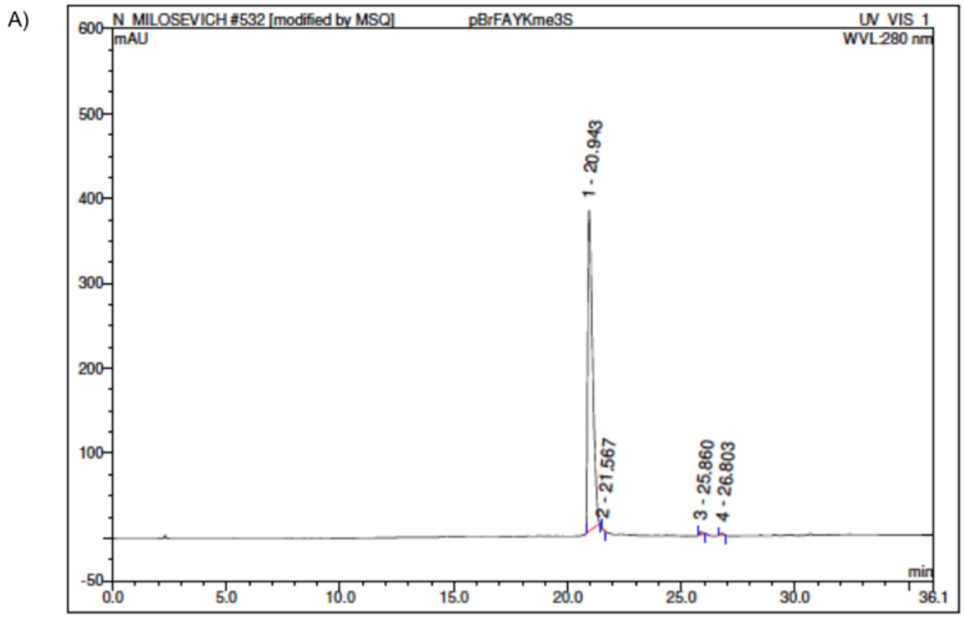


Figure 2.18. Characterization data for compound **2.4**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{72}H_{85}BrN_{11}O_{17}S^{+}$: 1486.50; found: 1486.40.



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

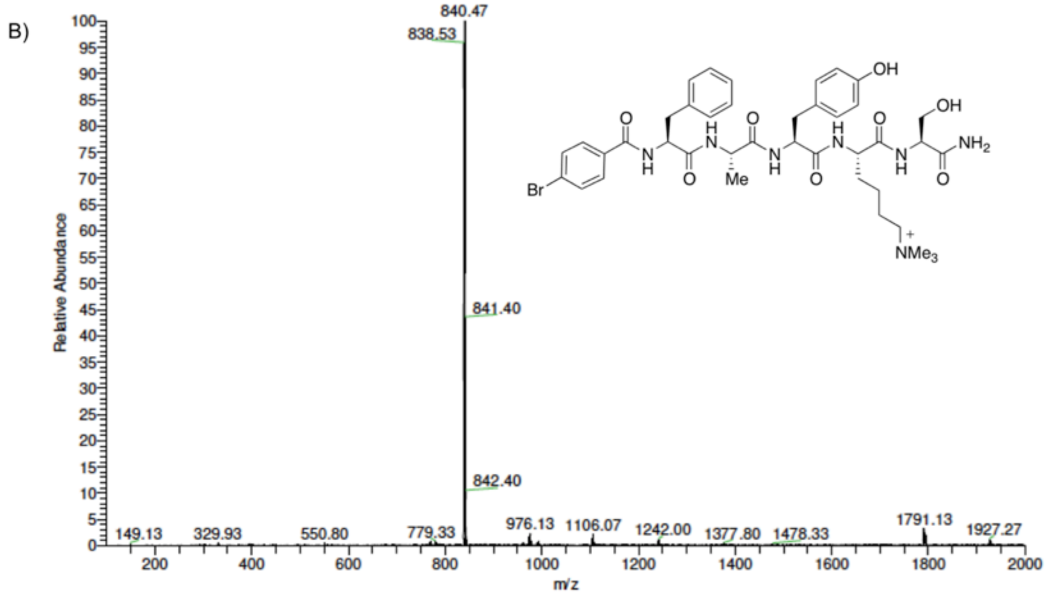
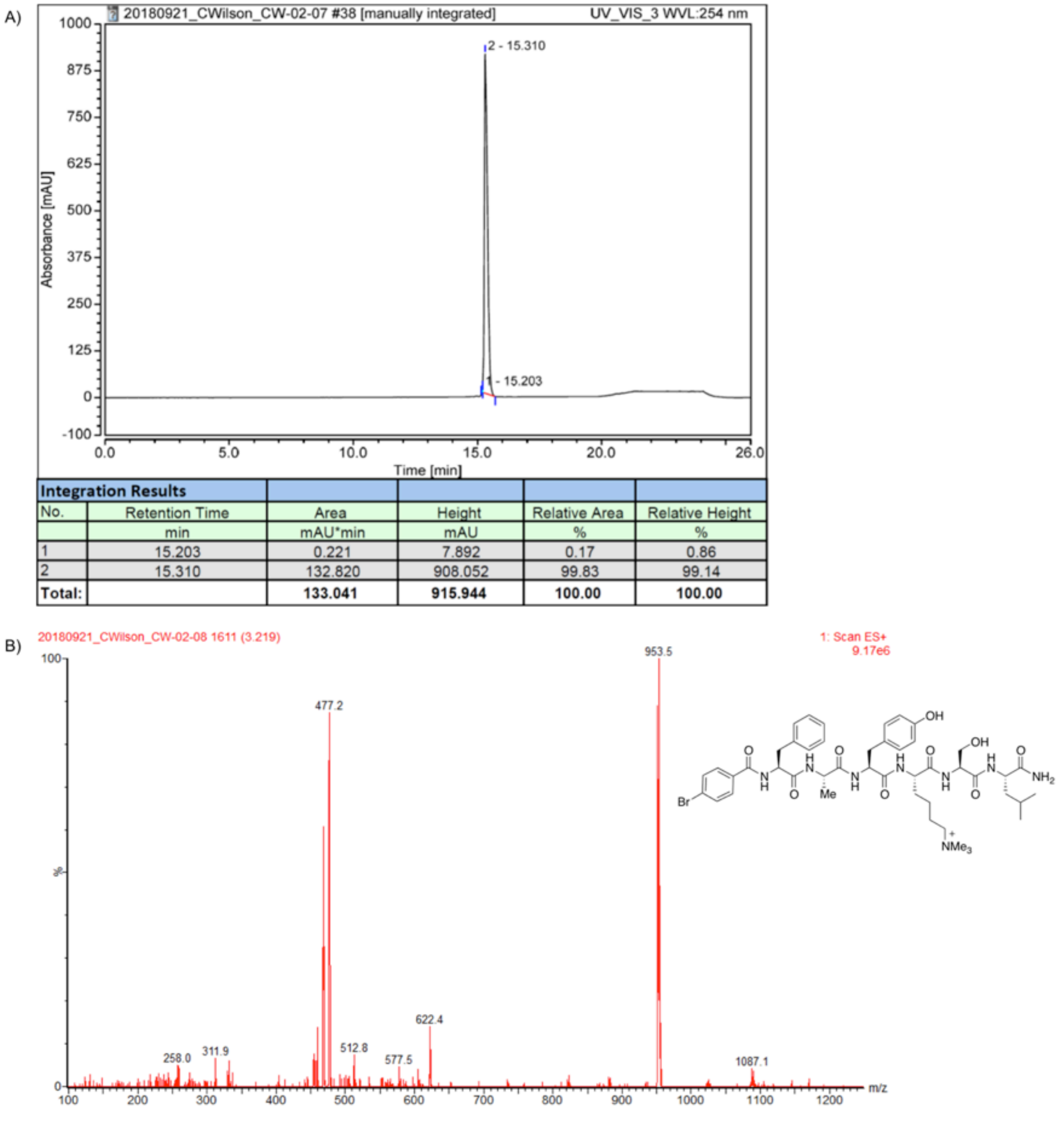


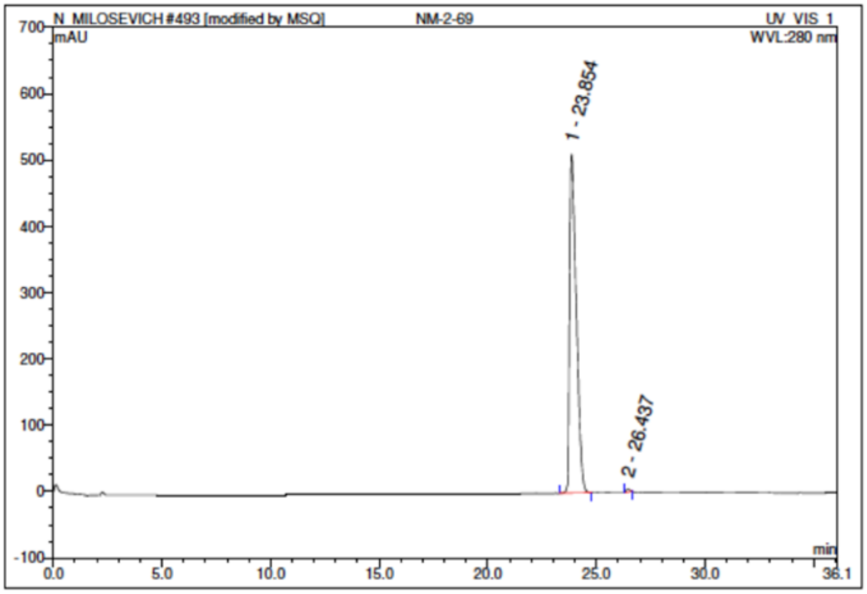
Figure 2.19. Characterization data for compound **2.5**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{40}H_{53}BrN_7O^+$: 838.31; found: 838.53.



Integration Results					
No.	Retention Time	Area	Height	Relative Area	Relative Height
	min	mAU*min	mAU	%	%
1	15.203	0.221	7.892	0.17	0.86
2	15.310	132.820	908.052	99.83	99.14
Total:		**133.041**	**915.944**	**100.00**	**100.00**

Figure 2.20. Characterization data for compound **2.6**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{46}H_{64}BrN_8O_9^+$: 953.40; found: 953.50.

A)



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

B)

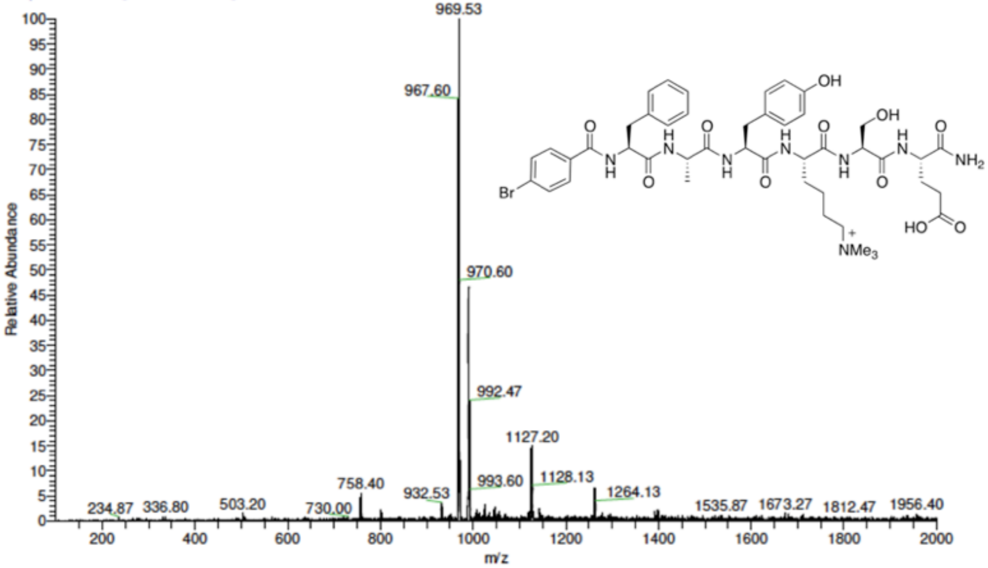
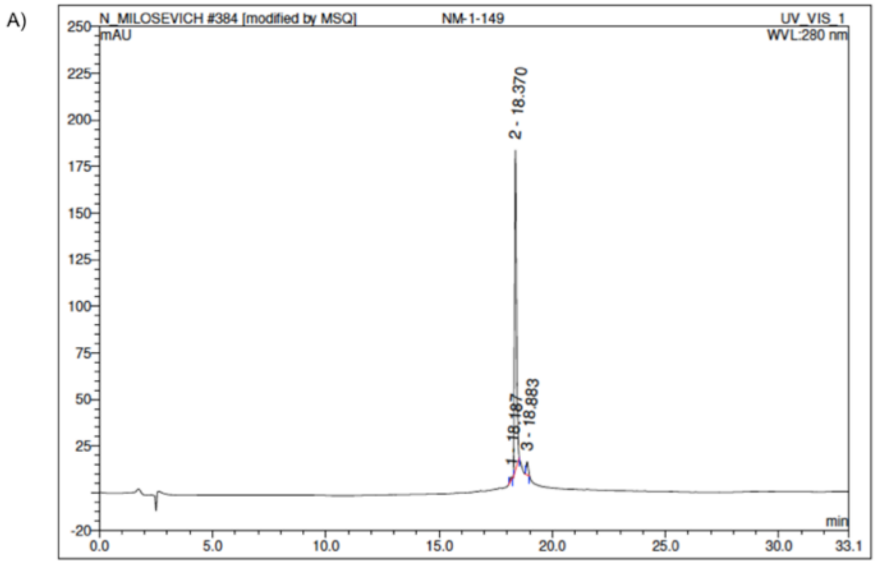


Figure 2.21. Characterization data for compound **2.7**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{45}H_{60}BrN_8O_{11}^+$: 967.36; found: 967.60.



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

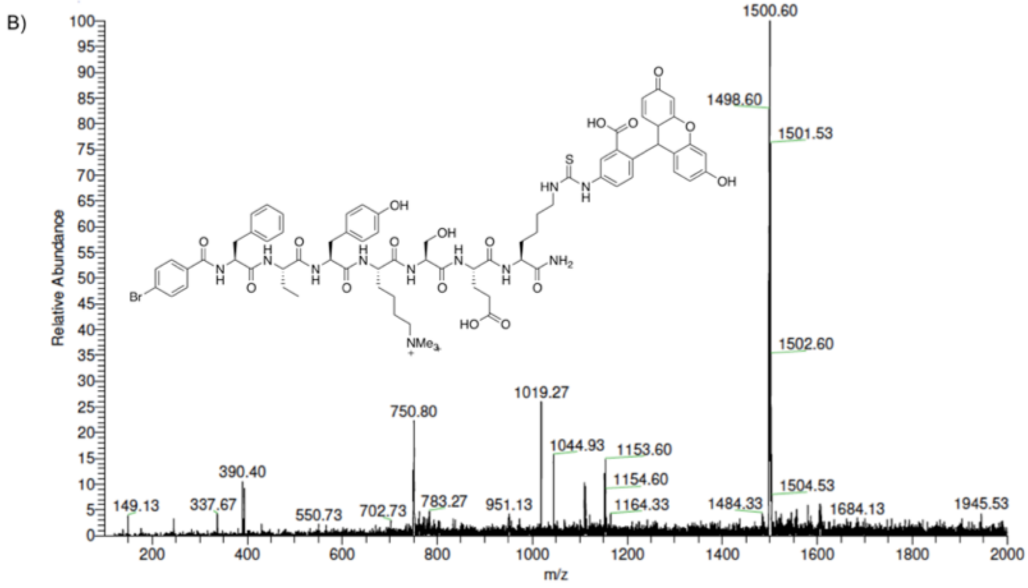
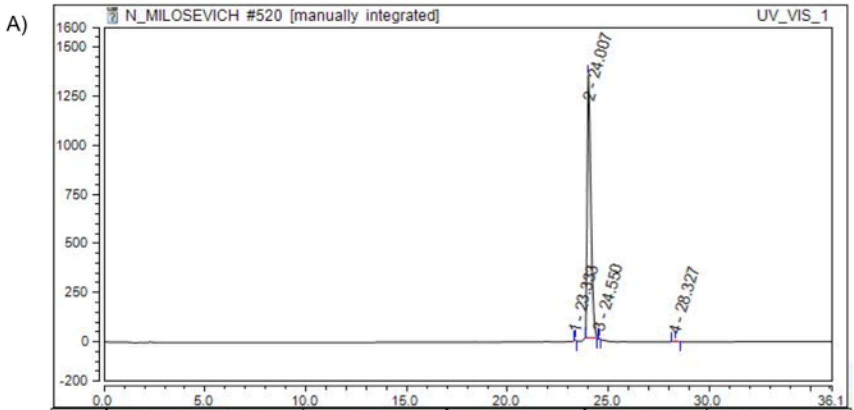


Figure 2.22. Characterization data for compound **2.8**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. For $C_{73}H_{87}BrN_{11}O_{17}S^{+}$: 1500.52; found: 1500.60.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	23.333	0.448	5.786	0.16	0.43
2	24.007	274.794	1335.288	99.45	98.96
3	24.550	0.436	5.077	0.16	0.38
4	28.327	0.637	3.197	0.23	0.24
Total:		**276.315**	**1349.347**	**100.00**	**100.00**

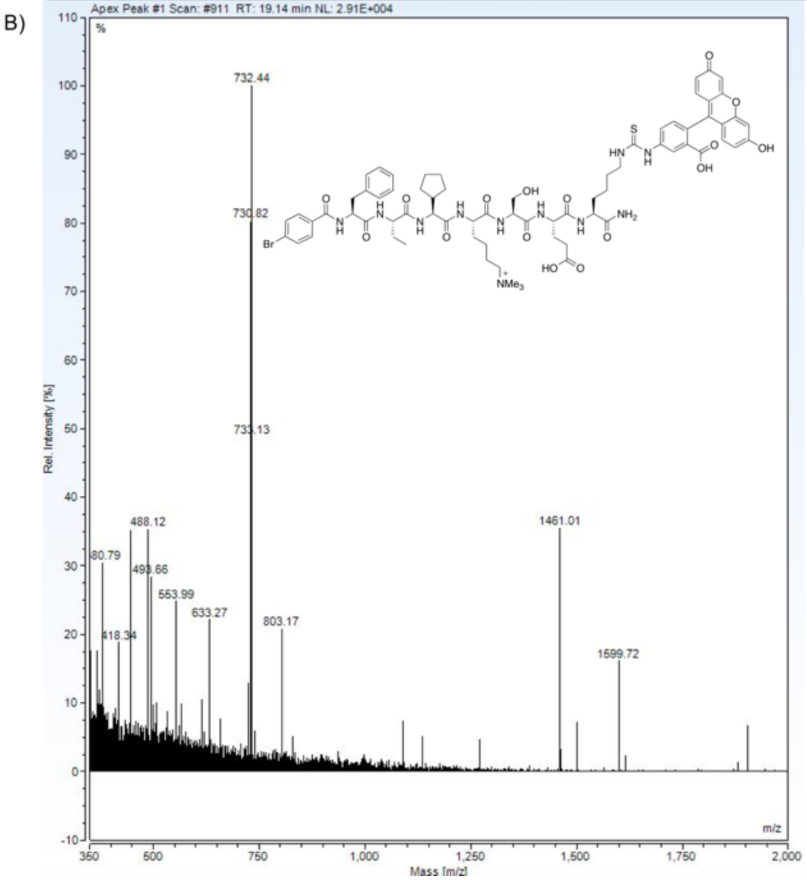
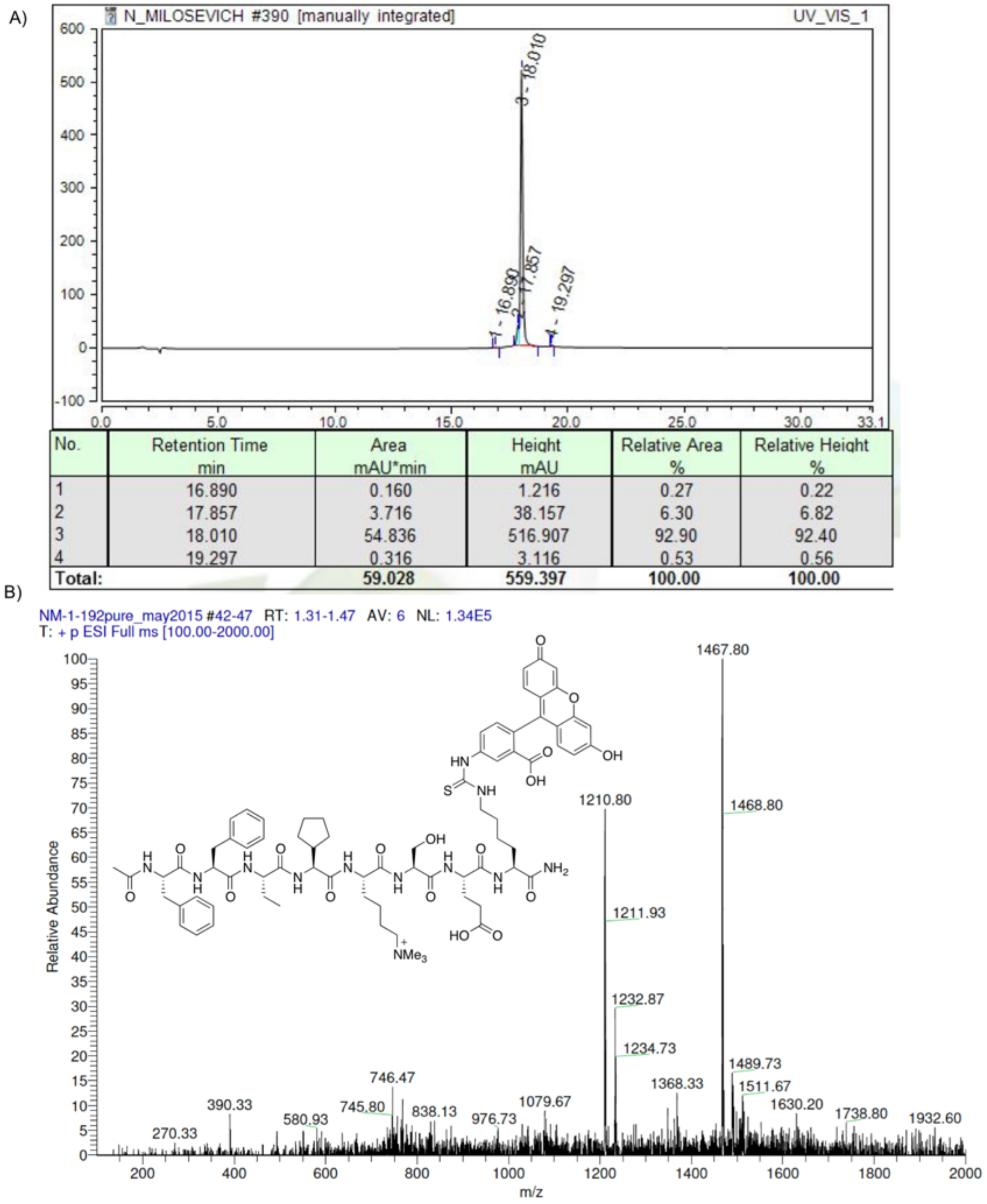
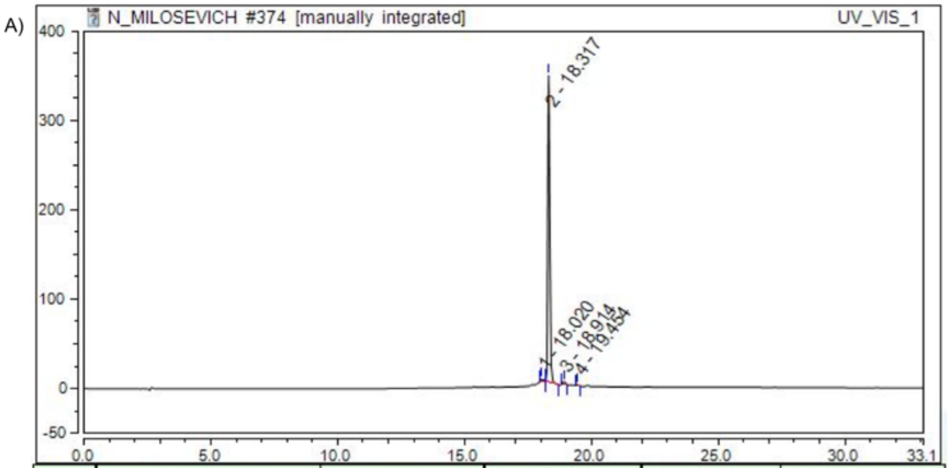


Figure 2.23. Characterization data for compound **2.9**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. For $C_{71}H_{89}BrN_{11}O_{16}S^{+}$: 732.27, 1464.54; found: 732.44, 1461.01.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	16.890	0.160	1.216	0.27	0.22
2	17.857	3.716	38.157	6.30	6.82
3	18.010	54.836	516.907	92.90	92.40
4	19.297	0.316	3.116	0.53	0.56
Total:		**59.028**	**559.397**	**100.00**	**100.00**

Figure 2.24. Characterization data for compound **2.10**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{75}H_{95}N_{12}O_{17}S^+$: 1467.67; found: 1467.80.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	18.020	0.434	3.518	1.41	1.00
2	18.317	29.706	342.705	96.77	97.64
3	18.914	0.469	3.629	1.53	1.03
4	19.454	0.090	1.142	0.29	0.33
Total:		**30.699**	**350.993**	**100.00**	**100.00**

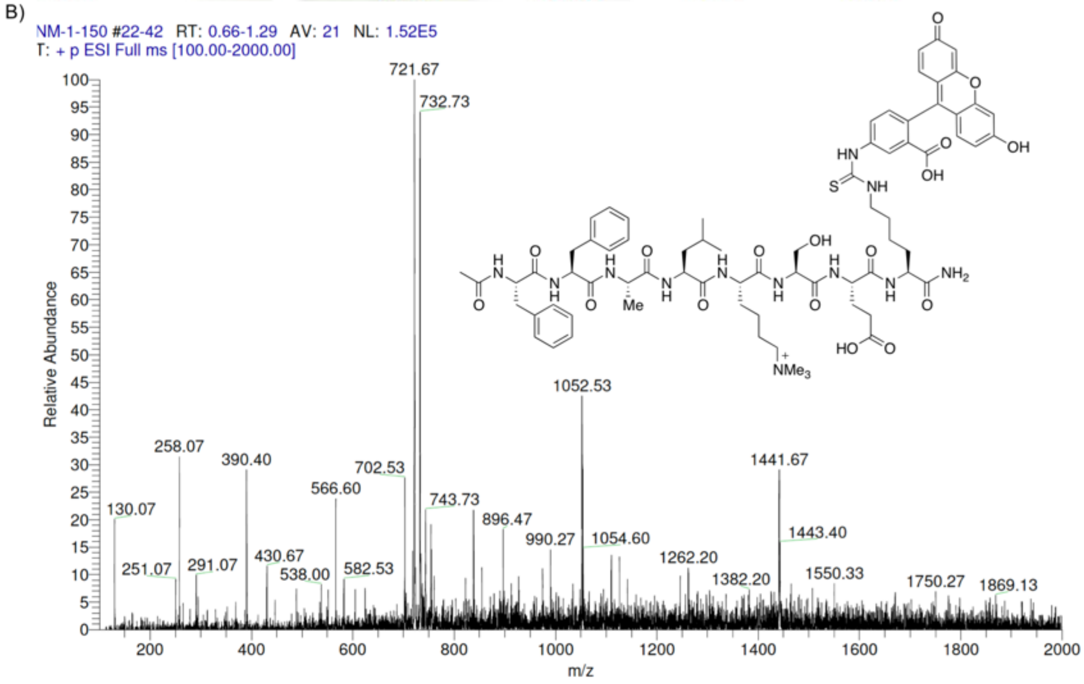
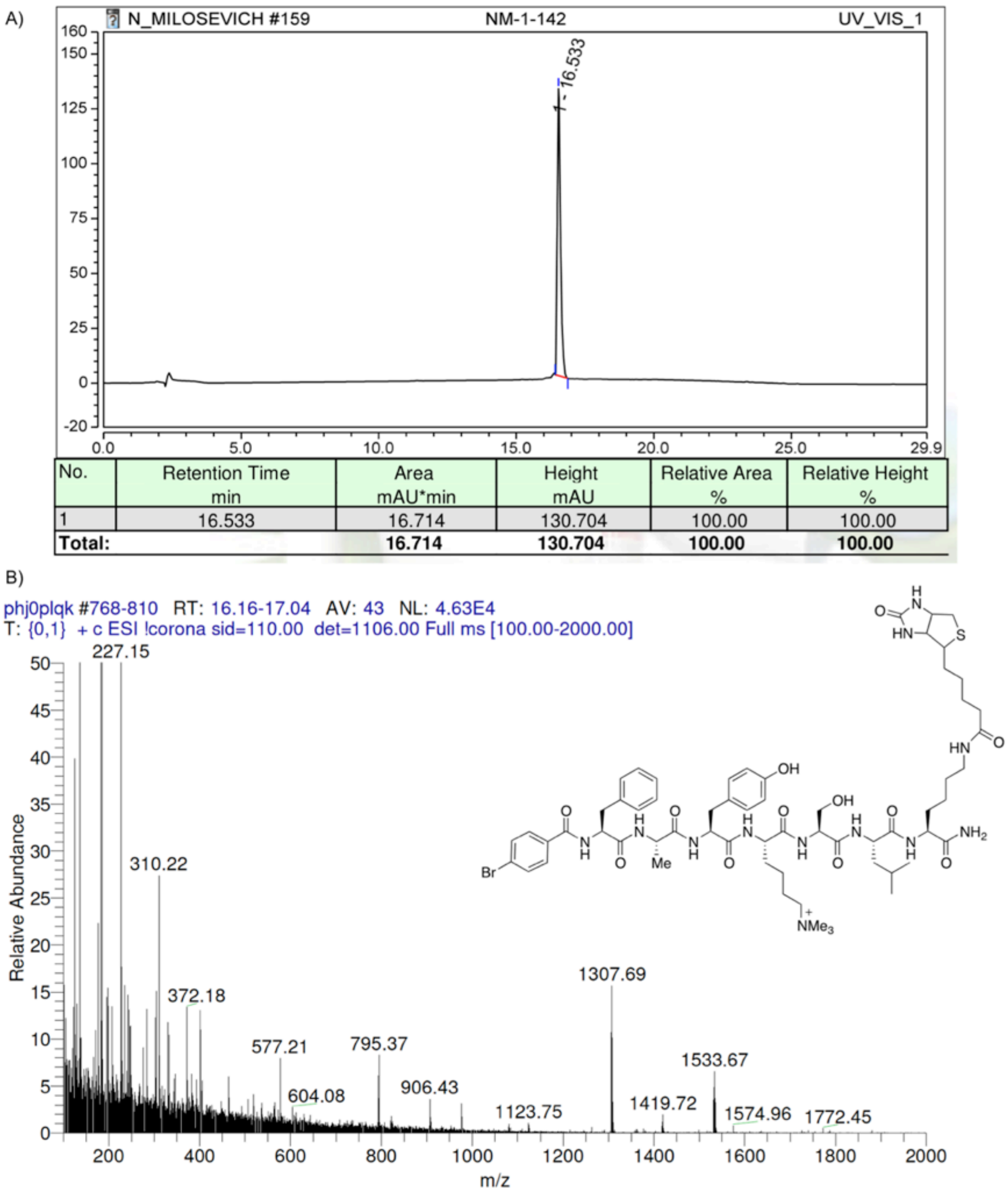


Figure 2.25. Characterization data for compound **2.11**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{73}H_{95}N_{12}O_{17}S^{+}$: 721.35, 1441.65; found: 721.67, 1441.67.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	16.533	16.714	130.704	100.00	100.00
Total:		**16.714**	**130.704**	**100.00**	**100.00**

Figure 2.26. Characterization data for compound **2.12**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{62}H_{90}BrN_{12}O_{12}S^{+}$: 1307.57; found:1307.69.

2.8.6 Protein expression and purification

Protein expression

Addgene plasmids 25245 (CBX1), 25158 (CBX2), 25237 (CBX4), 25296 (CBX6), 25241 (CBX7) and a CBX8 plasmid donated directly by C. Arrowsmith (Structural

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

Protein purification

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

2.8.7 Fluorescence polarization methods

Direct FP method

Direct FP was done by titration of CBX proteins into FITC-labelled probes. Binding assays were performed in black 384 well plates with optical bottoms at a total volume of 50 µL per well. All solutions were dissolved in 'full' FP buffer made from the addition of DTT, PMSF and benzamidine (1 mM final concentration) into 'minimal' FP buffer (20 mM Tris HCl, 250 mM NaCl, 0.01% Tween-20). A solution containing 100 nM of the labelled probe and a varying concentration of the CBX protein in question, was titrated into wells containing a solution with a constant concentration (100 nM) of FITC-labelled probe. 100 μL of the CBX-containing solution was added to well 19. Wells 2-18 contained 50 μL of the FITC-probe only solution. 50 μL of the well 19 solution was added to well 18 and mixed 4 times. These dilutions continued until well 3, whereupon 50 μL was discarded after mixing, leaving the total volume of all wells at 50 μL. Well 1 contained only full FP buffer and well 2 contained only the FITC-probe in full FP buffer. The maximum concentrations used for each CBX protein varied (0.001 μM – 600 μM) based on availability and solubility of the protein at high concentration, as well as the expected strength of the FITC-probe being tested. Assays were performed in triplicate or duplicate depending on CBX protein availability (noted in raw data). After removing any bubbles with a needle, plates were incubated in the dark at RT for 15 minutes and then read with a SpectraMax M5 plate reader (Molecular Devices). Raw data for perpendicular (P) and parallel (S) light was collected with SoftMax Pro software using a fluorescence polarization endpoint read. 100 reads were made and averaged per well with 5 second automix before reading and medium PMT sensitivity. Excitation was 450 nm for FITC and emission was 530 nm (with a 515 nm cut-off).

Analysis of direct FP data

The background of the blank buffer was subtracted from the parallel and perpendicular intensities of emission. Average values for parallel and perpendicular emission intensities (minus the average values for buffer control wells) were determined for sets of replicates using Microsoft Excel and millipolarization units (mP) were calculated using the formula ((Parallel –Perpendicular)/(Parallel + Perpendicular))*1000. Data were then plotted using Graphpad Prism 8 and fitted using a 'one-site' total binding model. The mathematical model for the fit was described by the equation $y = (\mathrm{Bmax} \times x)/(K_d+x) + \mathrm{NS} \times x + \mathrm{background}$ (Bmax = maximum specific binding, NS = slope of nonspecific binding and background = amount of nonspecific binding with no added labeled ligand). For certain assays, some of the data points (0 to 5 points) representing the highest CBX concentrations had to be excluded due to abnormally high mP values, likely arising from protein aggregation.

Competitive FP method

Competitive FP analysis of **2.5-2.7** binding to CBX6/7/8 was used to determine IC_{50} values as with compound **2.4** as the competitive binding probe (Figure 2.31-Figure 2.33). Compound **2.4** and peptide inhibitors were dissolved in DMSO and diluted into distilled H_2O. The assay was carried out in NUNC black 96-well plates and 'full' FP buffer was used as described above (20 mM Tris-HCl, pH = 8.0, 250 mM NaCl, 1 mM DTT, 1 mM benzamidine, 1 mM PMSF, 0.01% Tween). Two solutions were used, one containing protein, FITC probe **2.4**, and peptide inhibitor and another with just the protein and FITC probe **2.4**. Probe **2.4** was at a constant concentration of 100 nM in both solutions. Each CBX protein was at a constant concentration dependent on the K_d of the protein to probe **2.4**. CBX1 was used at 10 μM, CBX2 at 0.7 μM, CBX4 at 0.7 μM, CBX6 at 0.7 μM, CBX7 at 0.7 μM and CBX8 at 4 μM. Inhibitor concentrations varied between 0-500 μM. All wells in the plate were made to final volume of 100 μL. 200 μL of the inhibitor-containing solution was added to well 12. Wells 2-11 contained 100 μL of the CBX and **2.4** only solution. 100 μL of the well 12 solution was added to well 11 and mixed. These dilutions continued to well 3, whereupon 100 μL was discarded after mixing, leaving the total volume of all wells at 100 μL. Well 1 contained only full FP buffer and well 2 contained only CBX protein and **2.4** in full FP buffer. Following serial dilution of the inhibitor across the plate, plates were incubated for 15 minutes in the dark and then were read with the SpectraMax M5 plate reader as described above.

Analysis of competitive FP data

The background of the blank buffer was subtracted from the parallel and perpendicular intensities of emission. Values were graphed using Graphpad Prism 8 and fitted using a sigmoidal curve function from which IC_{50} values were extrapolated. The equation used is described as follows: $y = \text{bottom} + (\text{top}–\text{bottom}) / (1 + 10\text{^}(X–\text{Log}IC_{50}))$. Errors are reported as 95% confidence intervals. Experiments were done in duplicate or triplicate and this is indicated in the raw data shown in Figure 2.31-Figure 2.33.

2.8.8 Fluorescence polarization data

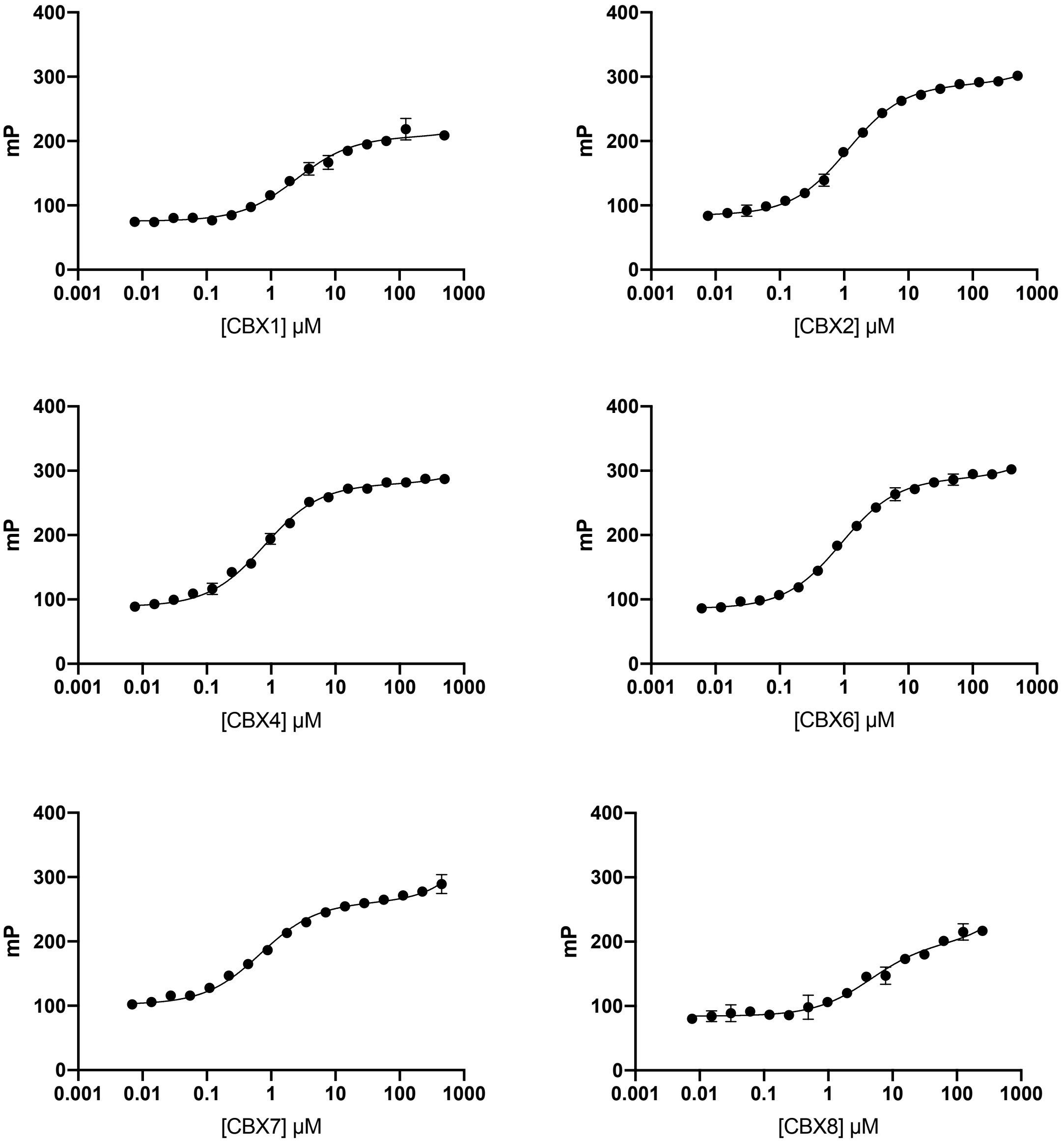


Figure 2.27. Direct fluorescence polarization data of compound **2.1** with CBX1/2/4/6/7/8. CBX1 (K_d = 2.5 μM, 95% CI 2.0 μM to 3.1 μM, R^2 = 0.983), CBX2 (K_d = 1.14 μM, 95% CI 1.0 μM to 1.2 μM, R^2 = 0.997), CBX4 (K_d = 0.79 μM, 95% CI 0.70 μM to 0.90 μM, R^2 = 0.994), CBX6 (K_d = 0.90 μM, 95% CI 0.82 μM to 0.96 μM, R^2 = 0.998), CBX7 (K_d = 0.69 μM, 95% CI 0.57 μM to 0.83 μM, R^2 = 0.993), CBX8 (K_d = 4.5 μM, 95% CI 3.0 μM to 6.8 μM, R^2 = 0.967). Error reported as asymmetrical 95% confidence intervals. Titrations for CBX1/2/4/8 were performed in triplicate and CBX6/7 titrations were performed in duplicate.

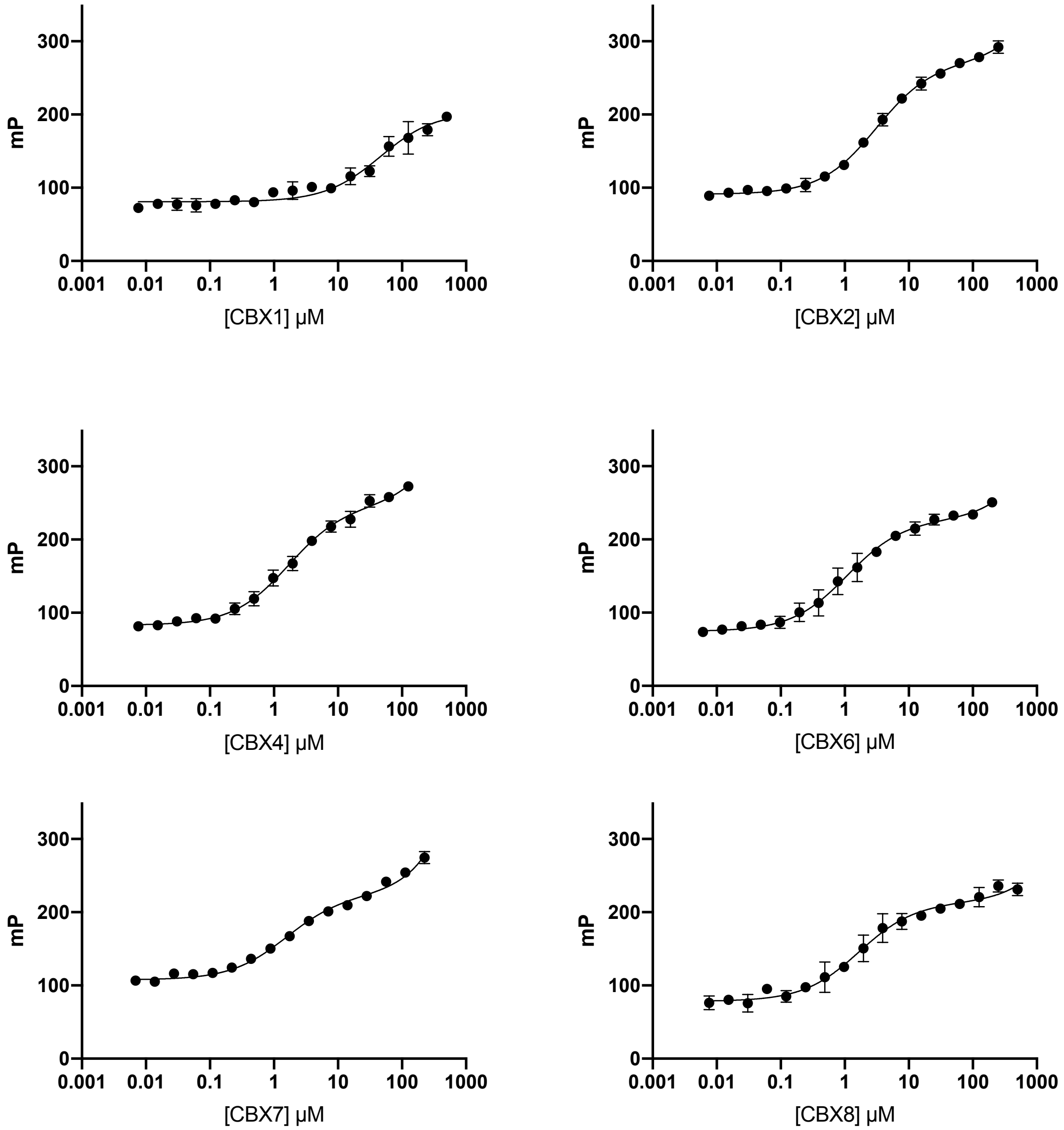


Figure 2.28. Direct fluorescence polarization data of compound **2.2** with CBX1/2/4/6/7/8. CBX1 (K_d = 47 μM, 95% CI 36 μM to 61 μM, R^2 = 0.944), CBX2 (K_d = 3.1 μM, 95% CI 2.8 μM to 3.5 μM, R^2 = 0.995), CBX4 (K_d = 1.7 μM, 95% CI 1.4 μM to 2.0 μM, R^2 = 0.991), CBX6 (K_d = 1.1 μM, 95% CI 0.83 μM to 1.4 μM, R^2 = 0.987), CBX7 (K_d = 1.6 μM, 95% CI 1.2 μM to 2.0 μM, R^2 = 0.992), CBX8 (K_d = 1.7 μM, 95% CI 1.2 μM to 2.3 μM, R^2 = 0.977). Error reported as asymmetrical 95% confidence intervals. CBX1/2/4/7 were performed in triplicate, and CBX6/8 titrations were performed in duplicate.

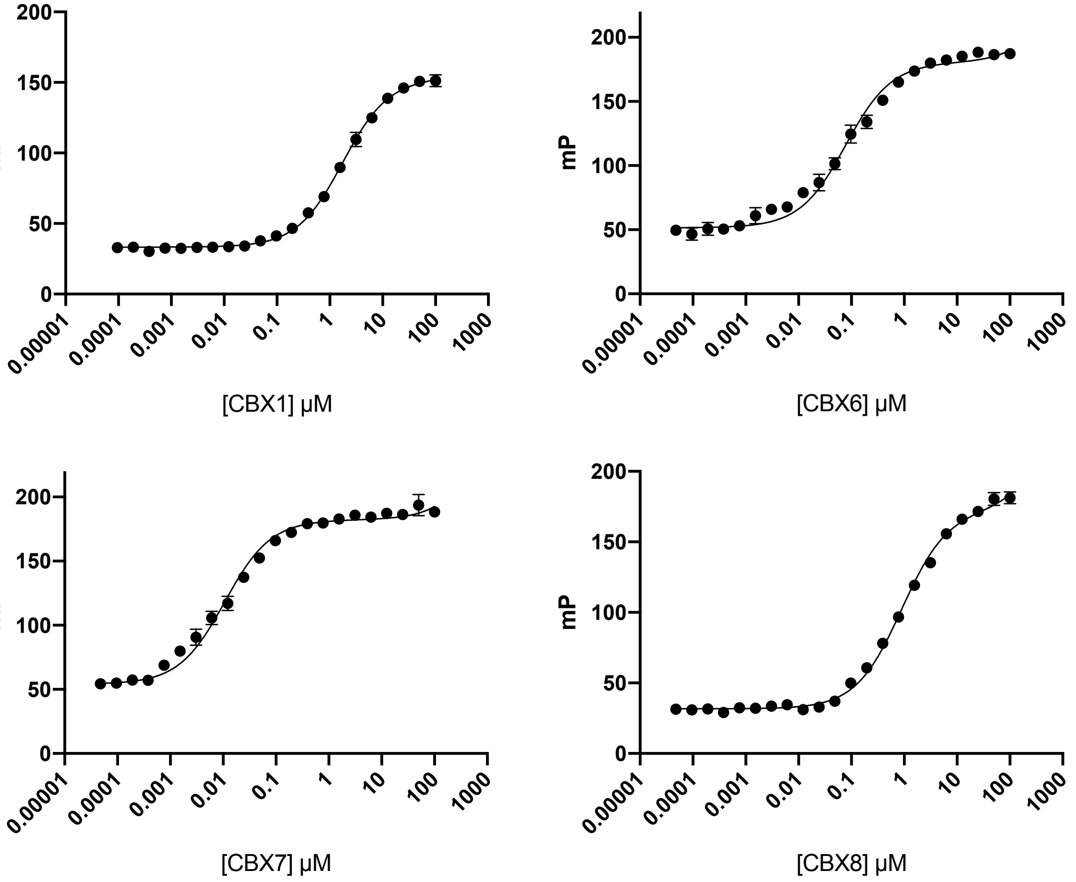


Figure 2.29. Direct fluorescence polarization data of compound **2.3** with CBX1/6/7/8 performed in triplicate. CBX1 (K_d = 1.75 µM, 95% CI 1.62 µM to 1.89 µM, R^2 = 0.998), CBX6 (K_d = 0.078 µM, 95% CI 0.063 µM to 0.097 µM, R^2 = 0.984), CBX7 (K_d = 0.011 µM, 95% CI 0.0090 µM to 0.013 µM, R^2 = 0.987), CBX8 (K_d = 0.89 µM, 95% CI 0.81 µM to 0.98 µM, R^2 = 0.997). Error reported as asymmetrical 95% confidence intervals.

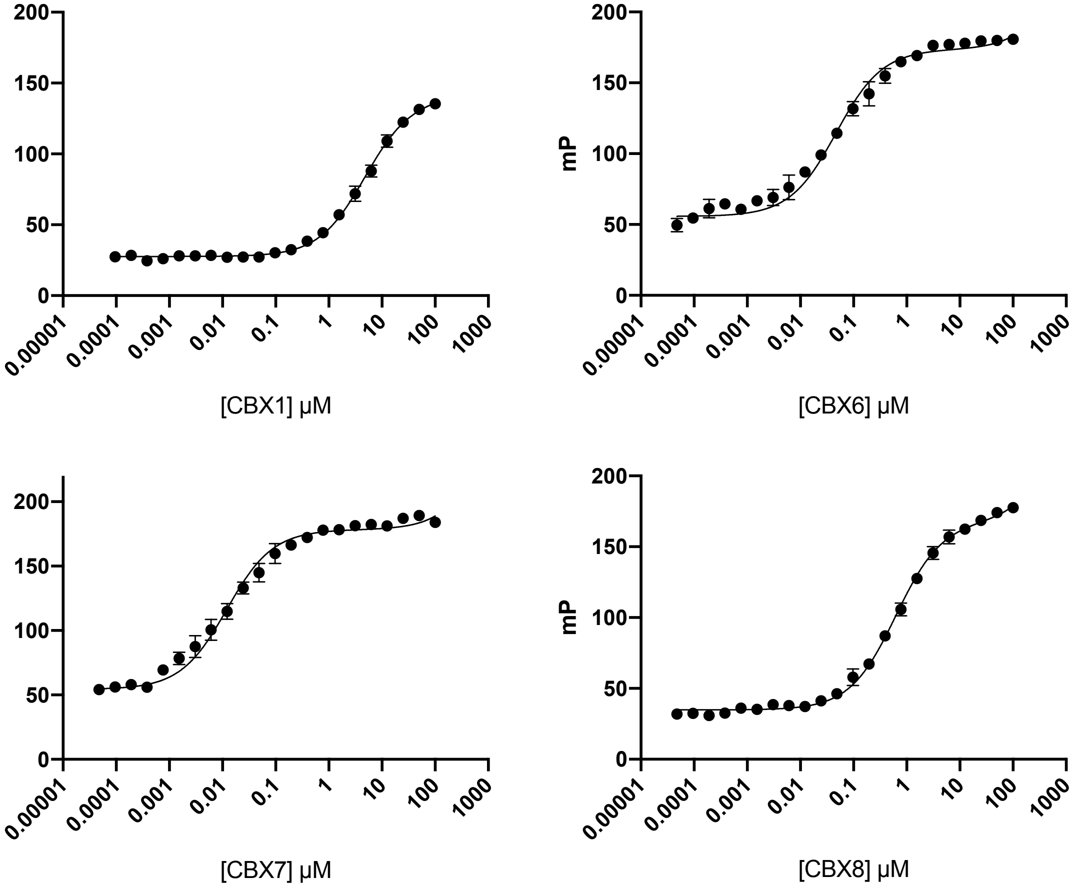


Figure 2.30. Direct fluorescence polarization data of compound **2.4** with CBX1/6/7/8 performed in triplicate. CBX1 (K_d = 4.8 µM, 95% CI 4.1 µM to 5.5 µM, R^2 = 0.944), CBX6 (K_d = 0.047 µM, 95% CI 0.035 µM to 0.062 µM, R^2 = 0.972), CBX7 (K_d = 0.012 µM, 95% CI 0.0097 µM to 0.015 µM, R^2 = 0.982), CBX8 (K_d = 0.62 µM, 95% CI 0.57 µM to 0.68 µM, R^2 = 0.998). Error reported as asymmetrical 95% confidence intervals.

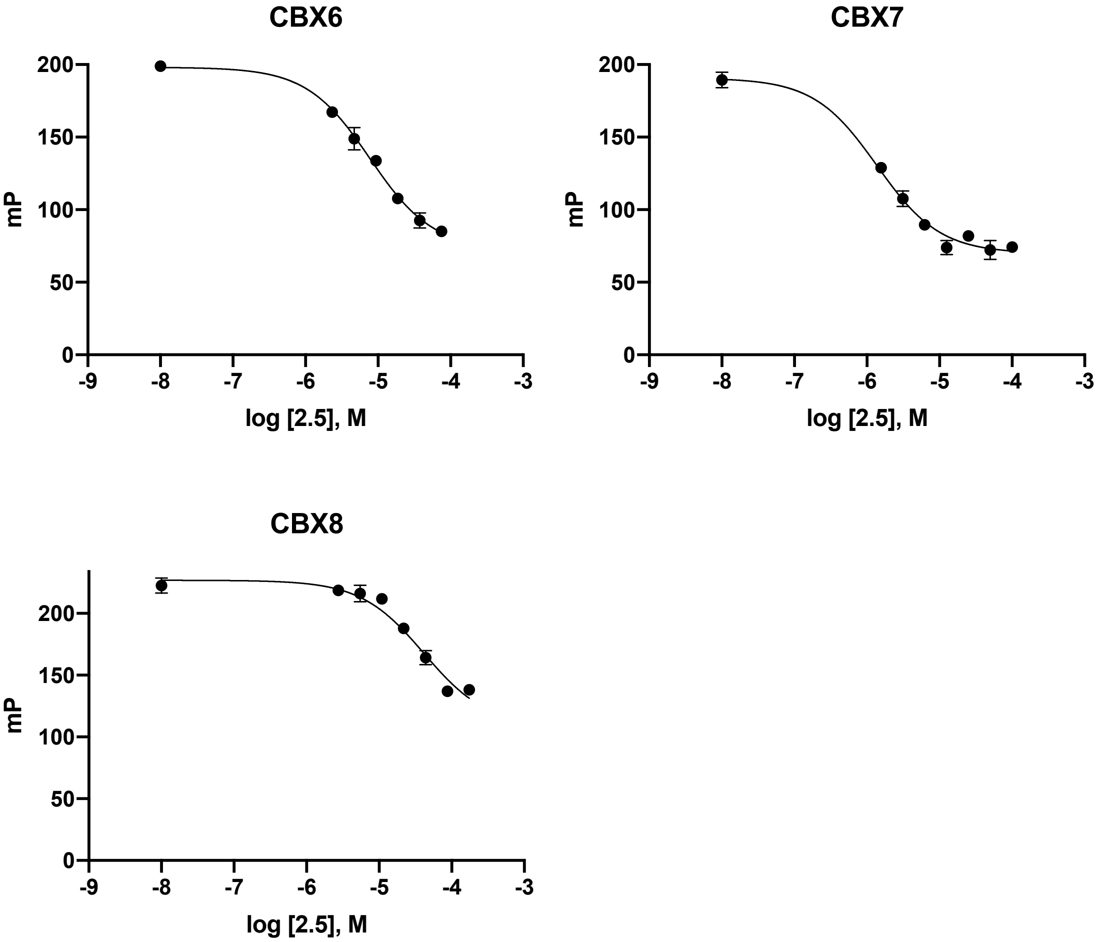


Figure 2.31. Competitive fluorescence polarization data of compound **2.5** with CBX6/7/8 performed in duplicate. CBX6 (logIC$_{50}$ = –5.113 M, 95% CI –5.222 M to –5.000 M, R^2 = 0.990), CBX7 (logIC$_{50}$ = –5.862 M, 95% CI –5.983 M to –5.746 M, R^2 = 0.983), CBX8 (logIC$_{50}$ = –4.385 M, 95% CI –4.593 M to –4.162 M, R^2 = 0.967). Error reported as asymmetrical 95% confidence intervals.

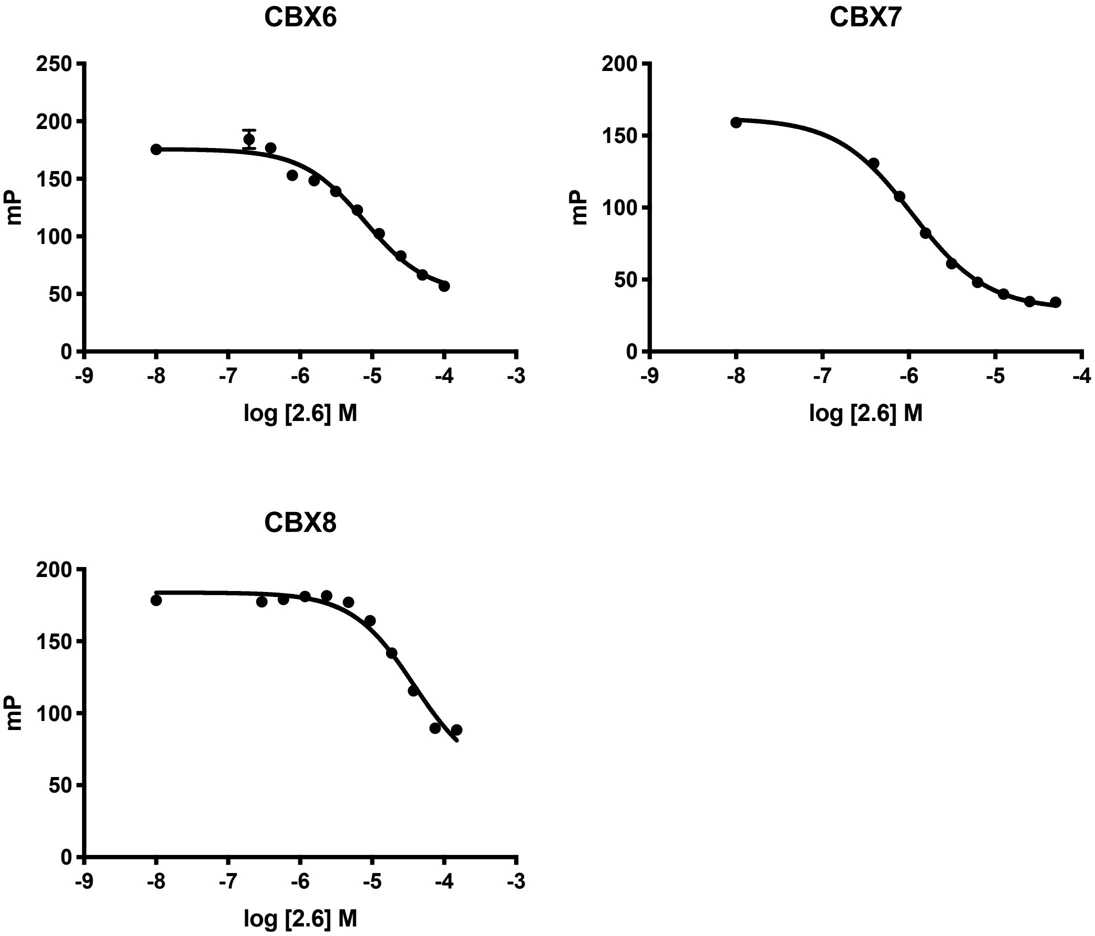


Figure 2.32. Competitive fluorescence polarization data of compound **2.6** with CBX6/7/8 performed in triplicate. CBX6 (logIC$_{50}$ = –5.102 M, 95% CI –5.219 M to –4.985 M, R^2 = 0.980), CBX7 (logIC$_{50}$ = –5.962 M, 95% CI –5.996 M to –5.982 M, R^2 = 0.997), CBX8 (logIC$_{50}$ = –4.419 M, 95% CI –4.533 M to –4.300 M, R^2 = 0.975). Error reported as asymmetrical 95% confidence intervals.

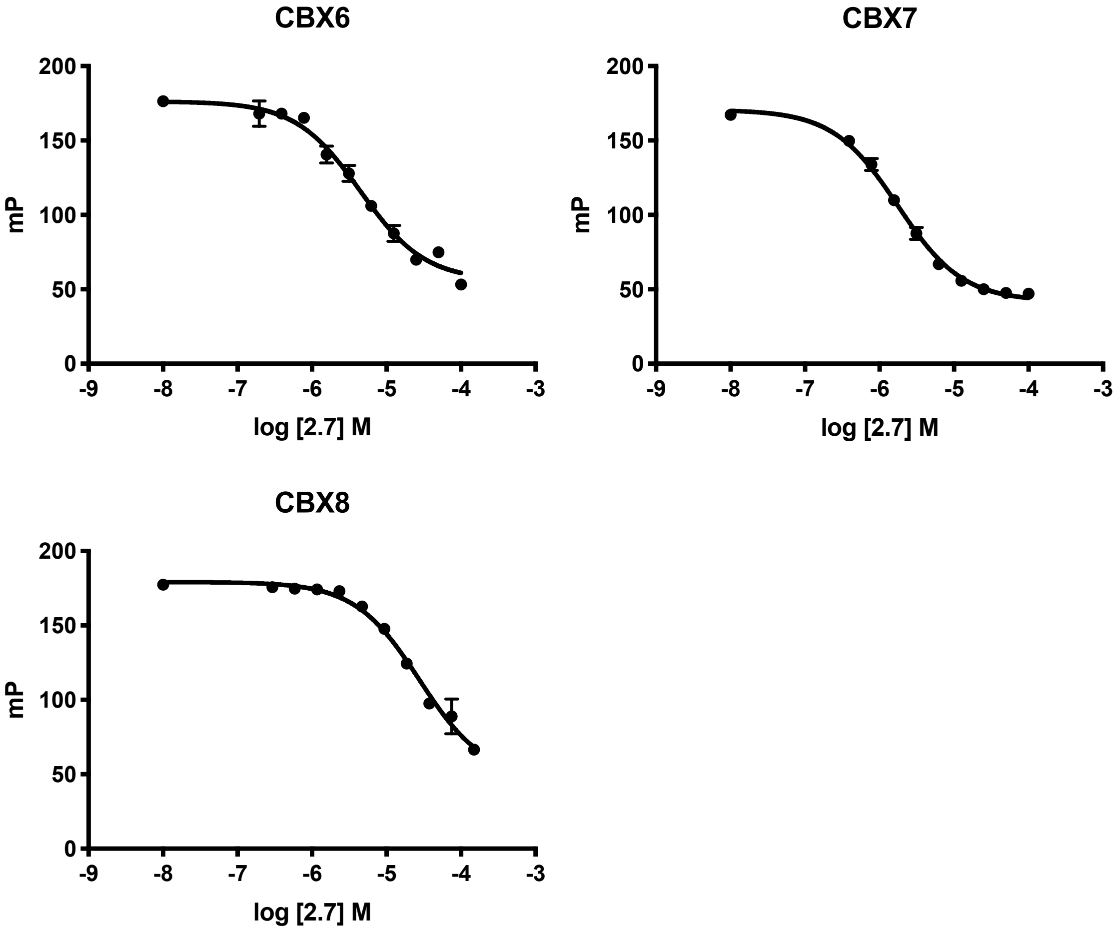


Figure 2.33. Competitive fluorescence polarization data of compound **2.7** with CBX6/7/8 performed in triplicate. CBX6 ($\log IC_{50}$ = –5.338 M, 95% CI –5.447 M to –5.227 M, R^2 = 0.979), CBX7 ($\log IC_{50}$ = –5.756 M, 95% CI –5.800 M to –5.711 M, R^2 = 0.995), CBX8 ($\log IC_{50}$ = –4.559 M, 95% CI –4.648 M to –4.467 M, R^2 = 0.986). Error reported as asymmetrical 95% confidence intervals.

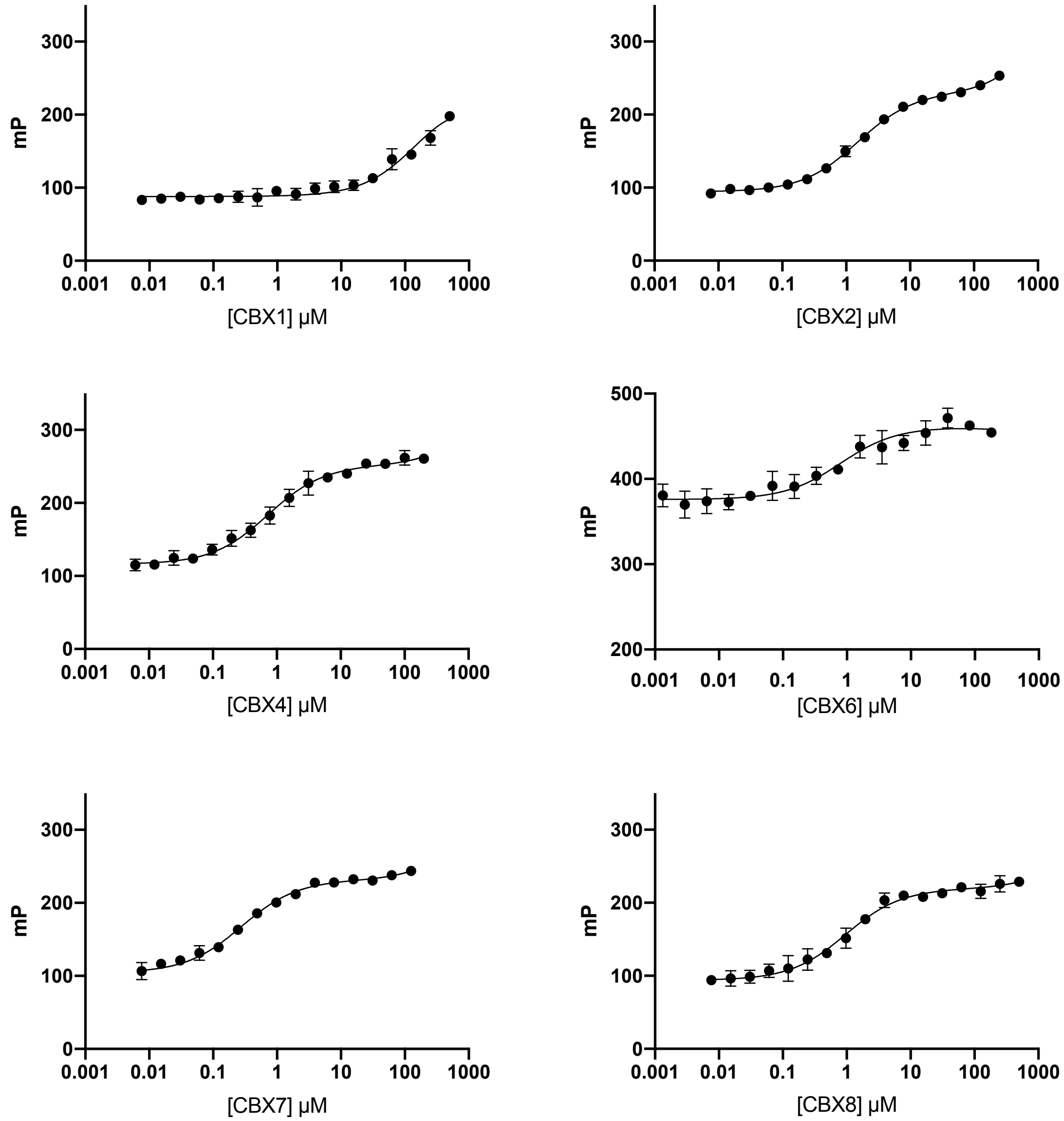


Figure 2.34. Direct fluorescence polarization data of compound **2.8** with CBX1/2/4/6/7/8. CBX1 (K_d = 178 μM, 95% CI 143 μM to 225 μM, R^2 = 0.941), CBX2 (K_d = 1.5 μM, 95% CI 1.3 μM to 1.7 μM, R^2 = 0.996), CBX4 (K_d = 0.73 μM, 95% CI 0.57 μM to 0.93 μM, R^2 = 0.981), CBX6 (K_d = 0.80 μM, 95% CI 0.4 μM to 1.6 μM, R^2 = 0.896), CBX7 (K_d = 0.30 μM, 95% CI 0.23 μM to 0.33 μM, R^2 = 0.988), CBX8 (K_d = 0.96 μM, 95% CI 0.73 μM to 1.30 μM, R^2 = 0.980). Error reported as asymmetrical 95% confidence intervals. CBX1/2/4/6/7 titrations were performed in triplicate and CBX8 titrations were performed in duplicate.

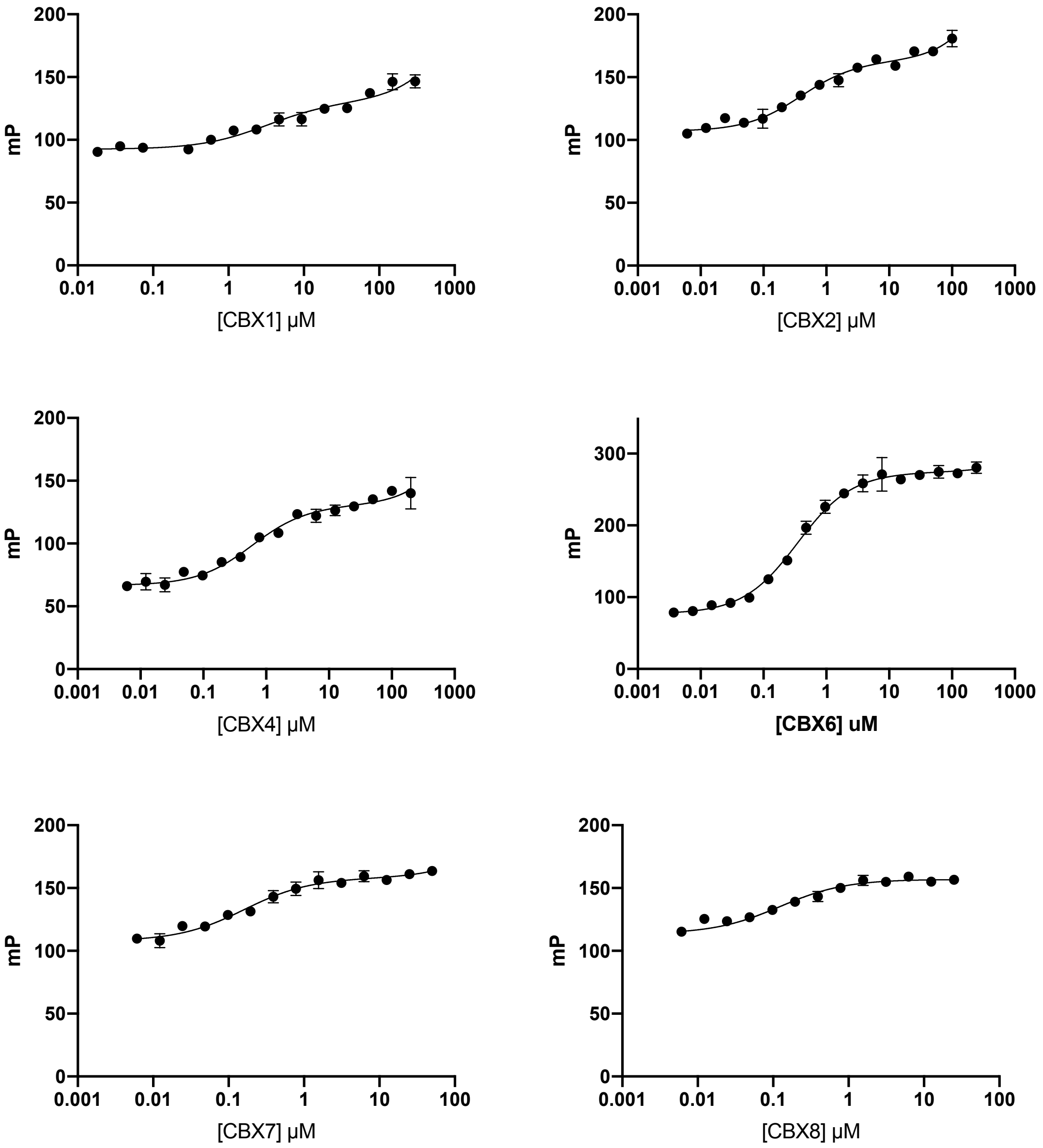


Figure 2.35. Direct fluorescence polarization data of compound **2.9** with CBX1/2/4/6/7/8. CBX1 (K_d = 3.2 µM, 95% CI 1.5 µM to 7.0 µM, R^2 = 0.956), CBX2 (K_d = 0.39 µM, 95% CI 0.27 µM to 0.56 µM, R^2 = 0.971), CBX4 (K_d = 0.60 µM, 95% CI 0.41 µM to 0.87 µM, R^2 = 0.974), CBX6 (K_d = 0.35 µM, 95% CI 0.299 µM to 0.401 µM, R^2 = 0.993), CBX7 (K_d = 0.16 µM, 95% CI 0.11 µM to 0.23 µM, R^2 = 0.965), CBX8 (K_d = 0.12 µM, 95% CI 0.08 µM to 0.19 µM, R^2 = 0.960). Error reported as asymmetrical 95% confidence intervals. CBX2/7/8 titrations were performed in triplicate. CBX1/4/6 titrations were performed in duplicate.

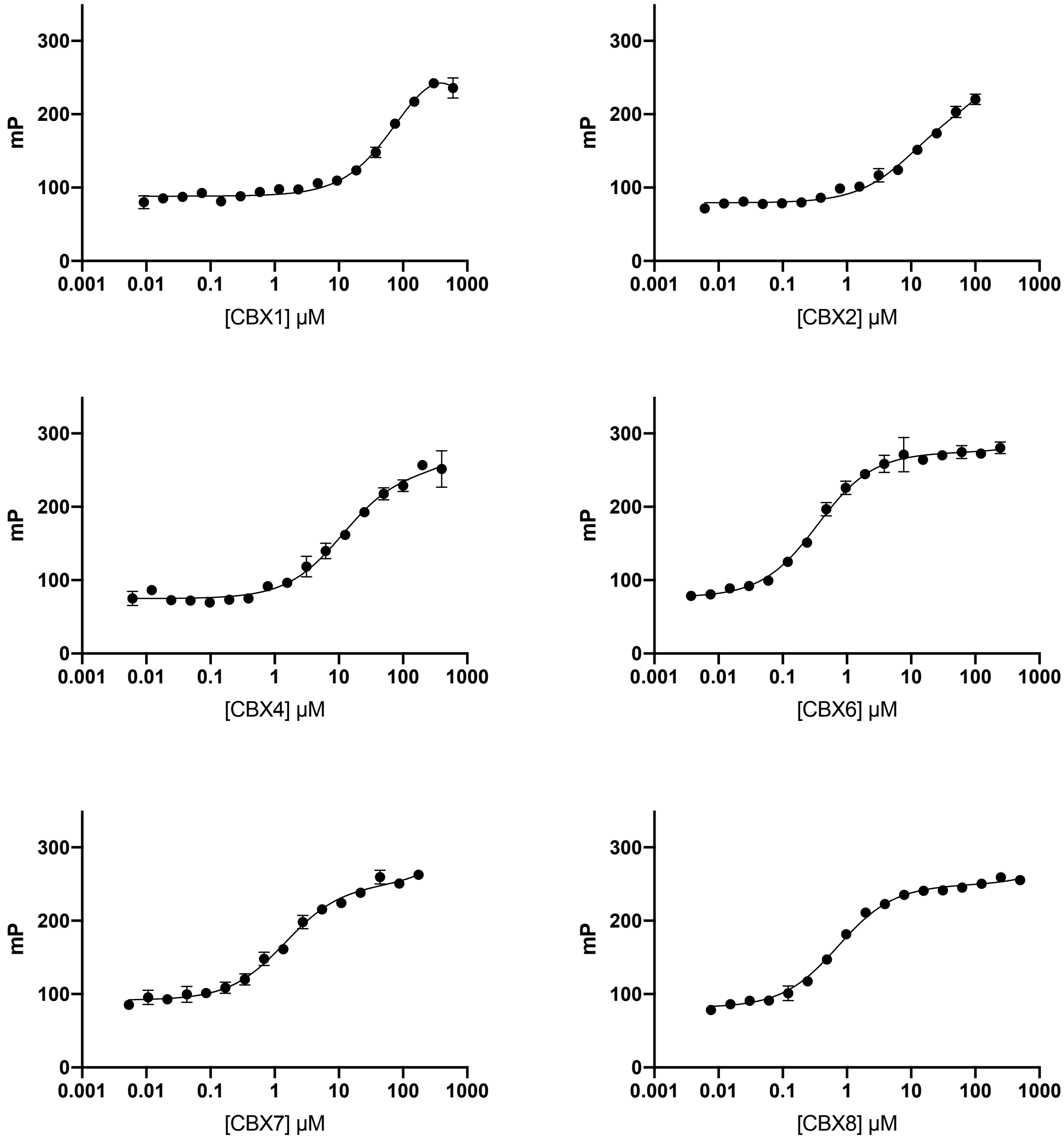


Figure 2.36. Direct fluorescence polarization data of compound **2.10** with CBX1/2/4/6/7/8. CBX1 (K_d = 99 μM, 95% CI 70 μM to 142 μM, R^2 = 0.989), CBX2 (K_d = 10 μM, 95% CI 8.6 μM to 12 μM, R^2 = 0.984), CBX4 (K_d = 11.3 μM, 95% CI 8.2 μM to 15.7 μM, R^2 = 0.985), CBX6 (K_d = 0.35 μM, 95% CI 0.30 μM to 0.41 μM, R^2 = 0.993), CBX7 (K_d = 1.5 μM, 95% CI 1.21 μM to 1.78 μM, R^2 = 0.989), CBX8 (K_d = 0.73 μM, 95% CI 0.63 μM to 0.84 μM, R^2 = 0.991). Error reported as asymmetrical 95% confidence intervals. All titrations were performed in duplicate.

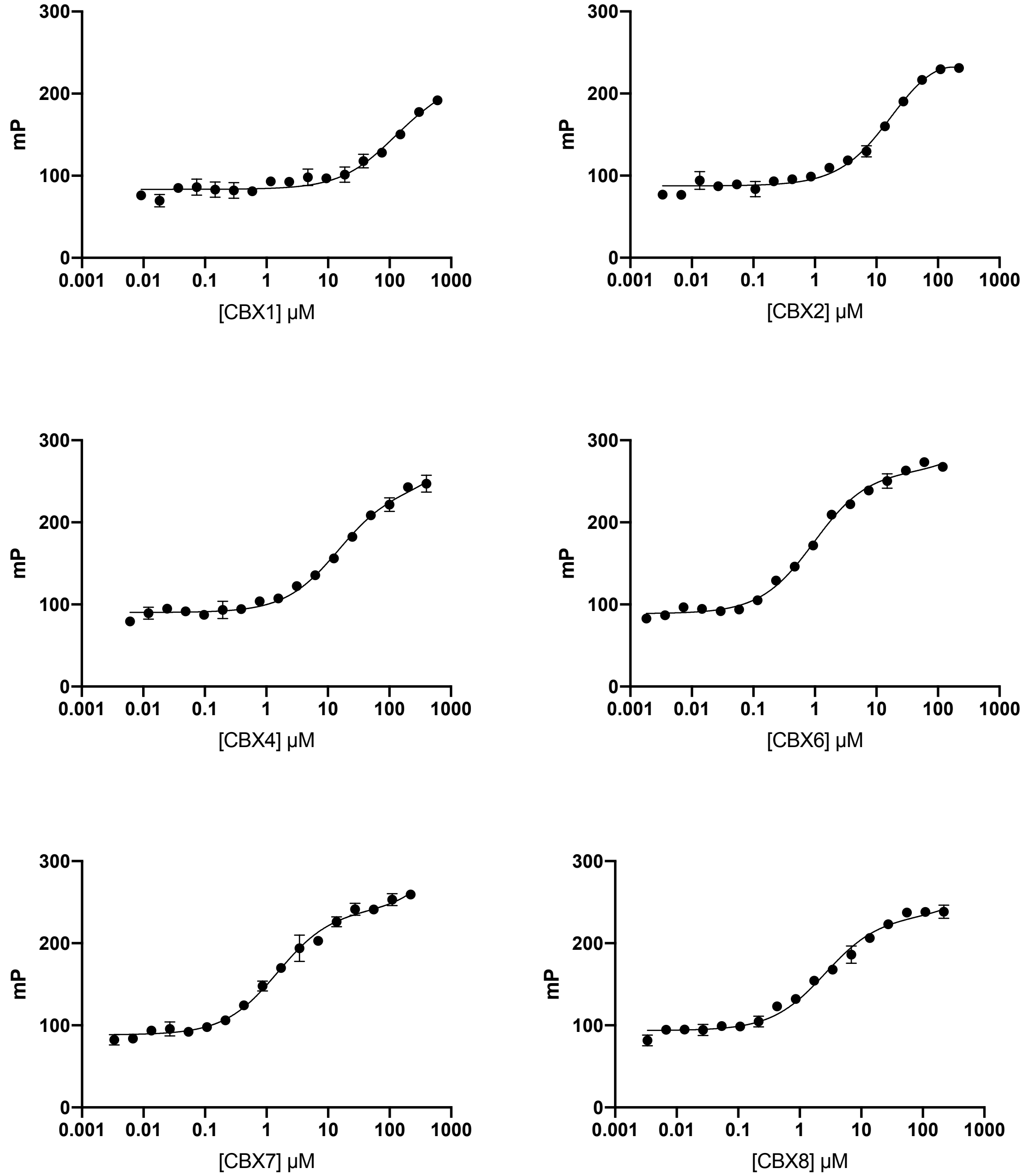


Figure 2.37. Direct fluorescence polarization data of compound **2.11** with CBX1/2/4/6/7/8. CBX1 (K_d = 110 μM, 95% CI 84 μM to 148 μM, R^2 = 0.958), CBX2 (K_d = 18 μM, 95% CI 13 μM to 25 μM, R^2 = 0.998), CBX4 (K_d = 15 μM, 95% CI 11 μM to 19 μM, R^2 = 0.991), CBX6 (K_d = 0.96 μM, 95% CI 0.82 μM to 1.3 μM, R^2 = 0.995), CBX7 (K_d = 1.5 μM, 95% CI 1.2 μM to 1.9 μM, R^2 = 0.992), CBX8 (K_d = 2.6 μM, 95% CI 1.9 μM to 3.5 μM, R^2 = 0.987). Error reported as asymmetrical 95% confidence intervals. All titrations were performed in duplicate.

2.8.9 Docking and MD methods

3i91 crystal structure of CBX8 complexed with H3K9me3 was used as the structure for docking. Ligand was removed, structure was cleaned, and 500 poses were generated using Autodock Vina for compounds **2.7** and **2.6**. UCSF Chimera was used to generate the initial PDBQT files for both host and inhibitors.

The docking for compound **2.7** was successful in generating the classic poses previously seen on the CBX proteins with the ligand Kme3 and (–2) Ala in the correct positions. However, **2.6** was unable to be docked in this fashion. For molecular dynamics of compound **2.6**, a lower energy pose resembling that of the docked configuration of **2.7** was used for comparative purposes. The docking poses for **2.7** consistently showed a preference for a salt bridge formation between ARG25 and the ligand glutamate.

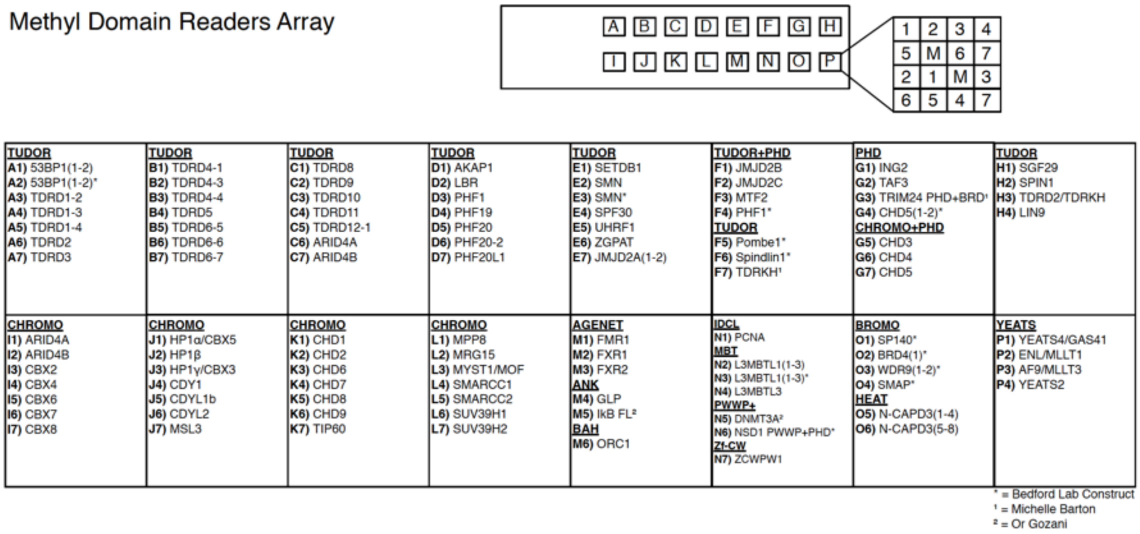
All molecular dynamics were done using the AMBER16 suite of programs. The generation of partial charges for non-standard residues in the inhibitors was done using a combination of Gaussian09 at the HF-6-31G* level of theory and the *residuegen* utility in AmberTools16. Simulation topologies and coordinates were set up in the xLEap graphical environment using the AMBER ff14SB force field.

For the molecular dynamics, the top five poses generated from **2.7** docking were treated with minimization, temperature ramping to 300K over 200 ps, and equilibration for 10 ns at 300K in explicit solvent (approximately 6000 TIP3PBOX waters). The simulations were then run for 50 ns for data collection for a combined 250 ns of simulation time. Replicates of the similar pose for **2.6** were also modeled in the same fashion. Key residue distances are reported in Figure 2.9.

Each 50 ns simulation required 60 hours on 28 cores (1 node with 2 Intel Xeon E5-2680 v4 processors). Initial docking for the poses was done on a single 2.8 GHz Intel Core i5, requiring 12 hours of processing time for the generation of 500 poses for each inhibitor.

2.8.10 Protein microarray

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



TUDOR **A1)** 53BP1(1-2) **A2)** 53BP1(1-2)* **A3)** TDRD1-2 **A4)** TDRD1-3 **A5)** TDRD1-4 **A6)** TDRD2 **A7)** TDRD3	**TUDOR** **B1)** TDRD4-1 **B2)** TDRD4-3 **B3)** TDRD4-4 **B4)** TDRD5 **B5)** TDRD6-5 **B6)** TDRD6-6 **B7)** TDRD6-7	**TUDOR** **C1)** TDRD8 **C2)** TDRD9 **C3)** TDRD10 **C4)** TDRD11 **C5)** TDRD12-1 **C6)** ARID4A **C7)** ARID4B	**TUDOR** **D1)** AKAP1 **D2)** LBR **D3)** PHF1 **D4)** PHF19 **D5)** PHF20 **D6)** PHF20-2 **D7)** PHF20L1	**TUDOR** **E1)** SETDB1 **E2)** SMN **E3)** SMN* **E4)** SPF30 **E5)** UHRF1 **E6)** ZGPAT **E7)** JMJD2A(1-2)	**TUDOR+PHD** **F1)** JMJD2B **F2)** JMJD2C **F3)** MTF2 **F4)** PHF1* **TUDOR** **F5)** Pombe1* **F6)** Spindlin1* **F7)** TDRKH[1]	**PHD** **G1)** ING2 **G2)** TAF3 **G3)** TRIM24 PHD+BRD[1] **G4)** CHD5(1-2)* **CHROMO+PHD** **G5)** CHD3 **G6)** CHD4 **G7)** CHD5	**TUDOR** **H1)** SGF29 **H2)** SPIN1 **H3)** TDRD2/TDRKH **H4)** LIN9
CHROMO **I1)** ARID4A **I2)** ARID4B **I3)** CBX2 **I4)** CBX4 **I5)** CBX6 **I6)** CBX7 **I7)** CBX8	**CHROMO** **J1)** HP1α/CBX5 **J2)** HP1β **J3)** HP1γ/CBX3 **J4)** CDY1 **J5)** CDYL1b **J6)** CDYL2 **J7)** MSL3	**CHROMO** **K1)** CHD1 **K2)** CHD2 **K3)** CHD6 **K4)** CHD7 **K5)** CHD8 **K6)** CHD9 **K7)** TIP60	**CHROMO** **L1)** MPP8 **L2)** MRG15 **L3)** MYST1/MOF **L4)** SMARCC1 **L5)** SMARCC2 **L6)** SUV39H1 **L7)** SUV39H2	**AGENET** **M1)** FMR1 **M2)** FXR1 **M3)** FXR2 **ANK** **M4)** GLP **M5)** IkB FL[2] **BAH** **M6)** ORC1	**IDCL** **N1)** PCNA **MBT** **N2)** L3MBTL1(1-3) **N3)** L3MBTL1(1-3)* **N4)** L3MBTL3 **PWWP+** **N5)** DNMT3A[2] **N6)** NSD1 PWWP+PHD* **Zf-CW** **N7)** ZCWPW1	**BROMO** **O1)** SP140* **O2)** BRD4(1)* **O3)** WDR9(1-2)* **O4)** SMAP* **HEAT** **O5)** N-CAPD3(1-4) **O6)** N-CAPD3(5-8)	**YEATS** **P1)** YEATS4/GAS41 **P2)** ENL/MLLT1 **P3)** AF9/MLLT3 **P4)** YEATS2

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

Figure 2.38. List of the methyl reader domains evaluated in the protein microarray seen in Figure 2.12. The letter associated with each domain specifies the section of the array containing the protein and the number designates the position of each replicate within that section.

2.8.11 Methods for cell-based studies

TOV21G cell line maintenance

TOV21G cell line (below passage 20) were obtained from Lindsay Devorkin (B.C. Cancer Agency Deeley Research Center, Victoria, B.C. Canada). Cell lines were cultured in TOV21G media (RPMI-1640 medium containing 25 mM HEPES and L-Glutamine (Thermo Scientific (Hyclone) SH30255.01), supplemented with 10% fetal bovine serum (FBS)(Life Technologies, 16000044), 1% L-Glutamine (Hyclone SH3003401) and 1% penicillin/streptomycin (Hyclone SV30010), at 37°C under 5% CO_2 atmosphere.

Treatment of TOV21G cells with inhibitors

Inhibitors were dissolved in DMSO, diluted with dsH_2O and further diluted in TOV21G media. The diluted inhibitor solutions were used to treat TOV21G cells (seeded at

2.5×10^5 cells/well in a 6-well plate, grown overnight at 37°C, 5% CO_2. Note that the samples were kept in the dark (covered in foil) as much as possible to reduce photobleaching of the FITC dye.

Treatment of TOV21G cells with inhibitors

Following a 21 hour incubation, the inhibitor solution was replaced with fresh media. Samples were imaged on EVOS® FL Cell imaging System (Life Technologies) exciting with the GFP laser (470 nm excitation, 525 nm emission) using either the 20× or 40× objective lens.

2.8.12 FACS analysis of inhibitor uptake, cell viability and cell cycle distribution

After cell imaging, cells were rinsed with warm 1× PBS and trypsinized for 5 min at 37°C, 5% CO_2. Trypsin was deactivated with TOV21G media and cells were collected by centrifuging at 1200 RPM for 5 min at 4°C. Cells were resuspended and washed once with 1× PBS, centrifuged using the same settings, decanted and resuspended in 500 µL ice cold 1× PBS. 0.5 µL of Fixable Viability Dye eFluor 780 (eBioscience, 65-0865-14) was added and the solution was incubated on ice for 30 min. After incubation, cells were collected by centrifuging at 1400 rpm for 5 min at 4°C and washed twice with ice cold 1× PBS. Cells were simultaneously fixed with 500 µL ice cold 70% ethanol and stained with propidium iodide (PI) (30 µL of 20 mg/mL RNaseA (Invitrogen, 60216-RN) and 26 µL PI (eBioscience, BMS5-FI/100)) for 30 min at room temperature. After centrifuging at 1400 rpm for 8 min at 4°C, cells were washed once with 1× PBS. Samples were resuspended with 0.3% FBS in 1× PBS and stored at 4°C until analysis. Samples were filtered (BD Falcon 352235) immediately before running on a Becton Dickinson FACSCalibur™ and analyzed using FlowJo 7.6.5 software.

MTT assay

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

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

Chapter 3. Selective CBX6 inhibition

Contributions

Sections of this chapter were previously published. The publication and my contributions are described below.

Milosevich, N., Gignac, M.C., Mcfarlane, J., Simhadri, C., Horvath, S., Daze, K.D., Croft, C.S., Dheri, A., Quon, T.T.H., Douglas, S.F., Wulff, J.E., Paci, I. Hof, F. **2016**. Selective Inhibition of CBX6: A Methyllysine Reader Protein in the Polycomb Family. *ACS Medicinal Chemistry Letters.* 7, 139.
Reproduced with permission. Copyright 2016 American Chemical Society.

I wrote and conceived of the manuscript with guidance from Dr. Wulff and Dr. Hof. I designed and synthesized the compounds with assistance from Caitlin Croft. James McFarlane wrote and conceived of the computational studies under supervision of Dr. Paci. Michael Gignac carried out protein expression and binding studies with assistance of Aman Dheri. I conceived of and designed SPR experiments, which were carried out with assistance of Shanti Horvath. Sarah Douglas and Taylor Quon carried out site-directed mutagenesis of CBX7.

3.1. Foreword

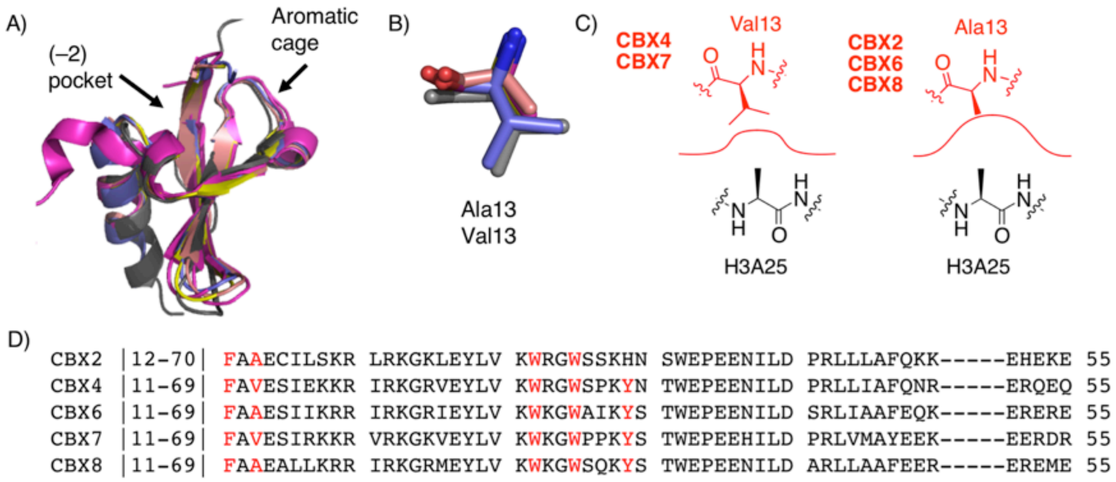
The work described in this Chapter developed from my initial work on creating peptidic inhibitors (Chapter 2). At the time of this work, I was focused on discovering chemical functionalities that would give rise to selective chemical antagonists for CBX2 and CBX8. My focus was on designing modifications to my ligands that would exploit subtle differences in the peptide-binding groove of the CBX proteins. Serendipitously, I discovered that one of my ligands was selective for CBX6 instead of my desired target, CBX8. This discovery prompted me to delve deeper into understanding the structural aspects that gave rise to this selectivity. At the time of publication, there was little known about the biology of CBX6 (in fact zero papers had been published on the subject). The main question driving the research described in the chapter 2 and 3 is ‘can we make selective chemical tools that can aid in understanding of the biology of the CBX proteins?’. The results in this chapter show that the answer is yes.

3.2. Introduction

My goal was to identify selective chemical or peptidic tools that could target other members within the polycomb CBX family of epigenetic modifiers. The sequence and structural similarities within their chromodomains are very high (reviewed in Chapter 2, and Figure 3.1). Their highly diverse in vivo functions are partially understood as arising from their divergent domain architecture outside of their methyl-reading chromodomains.[228] However, there has been no indication of any sites for chemical binding outside of the chromodomains themselves (with the exception of the RNA binding patch in CBX7, reviewed in Chapter 1), so any efforts to create selective ligands must rely on being able to discriminate among the five highly similar CBX chromodomains.

Chapter 2 proves that the selective inhibition of chromodomains is challenging. Modifications that tilt the selectivity in one direction or another were discovered, but no single, highly selective agent was identified after extensive variation of residues in the peptide-binding groove.

A small pocket binds the conserved histone Ala side chains that uniformly occur (–2) to the trimethyllysine sites H3K9me3 and H3K27me3—the two sites known to be targeted by CBX proteins. The floor of this (–2) pocket is partially defined by a Val residue in CBX4 and 7, but is replaced by an Ala residue in CBX2/6/8 (Figure 3.1B, C). Others have recently shown that a Val/Ala exchange at this position can make CBX7 display CBX2-like binding affinities and functions, suggesting the importance of these residues at the floor of the (–2) pocket in the CBX proteins' intrinsic biological functions.[229]



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

Figure 3.1. Similarity of CBX proteins. A) Overlays of CBX2, CBX6, CBX7, and CBX8 with (–2) and aromatic cage labeled. (magenta = CBX2 pdb code: 3H91, grey = CBX4 pdb code: 3I87, yellow = CBX6 pdb code: 3I90, purple = CBX7 pdb code: 4MN3, salmon = CBX8 pdb code: 3I91), B) Overlay of the key residues of the (–2) pocket, C) Depiction of pocket size in each CBX protein D) Sequence alignments that highlight the aromatic cage residues and residues lining the bottom of the (–2) pocket which is divergent within the family, as determined by ClustalW2 alignment. CBX4 not shown in structural overlay as the only available X-ray structure is not bound to histone tail or other ligand.

We decided to vary the ligand side chain that is directed into this (–2) pocket in an attempt to prepare potent and paralog-selective ligands. In this chapter I report on a small set of peptidomimetic agents in which a simple chemical modification switches the ligands from one with promiscuity across all polycomb paralogs to one that provides selective inhibition of CBX6. The structural basis for this selectivity, which involves occupancy of a small hydrophobic pocket adjacent to the aromatic cage, was confirmed through molecular dynamics simulations. Our results demonstrate the increases in affinity and selectivity generated by ligands that engage extended regions of chromodomain binding surfaces. What follows in this chapter is the work from a published paper on the first non-CBX7 chromodomain inhibitor.

The biological role of CBX6 has not been well studied and is poorly understood. Currently, there are three publications that report on the functions of CBX6 and its role in regulating pluripotency of stem cells, in the development and progression of hepatocellular carcinoma and in tumour suppression of breast cancer.[163, 170-171] Other reports of analyzing gene and protein databases have shown increased expression of CBX6 in hepatocellular carcinoma and malignant gliomas.[230-231] The complexity of CBX paralog utilization in the multi-protein PRC1 complex makes elucidating the individual roles of each CBX family member challenging. Chemical antagonists of CBX6 will provide a unique avenue to understand the protein interacting partners, phenotypes of inhibition, and biological roles.

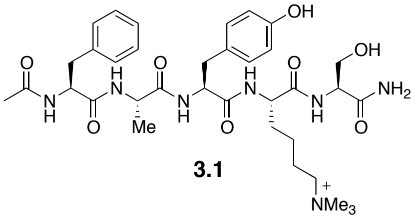
CBX6 is an essential regulator of embryonic stem cells (ESCs).[163] CBX6/7are both up-regulated in pluripotent ESCs.[157] Depletion of CBX6 in ESCs leads to rapid cell differentiation. Differentiation of ESCs was not observed following depletion of CBX7, highlighting the distinct roles of each polycomb paralog.[163] Cell-based studies with CBX6 mutants lacking the two methyllysine-binding tryptophan residues in the aromatic cage showed that the function of CBX6 in ESCs is strongly linked to the methyllysine reading function of the chromodomain.[163] Genome-wide localization studies in ESCs showed enrichment of CBX6 on genes involved in regulation of developmental processes and signaling pathways regulating pluripotency.[163] CBX6 participates as a component in both canonical PRC1 and non-canonical PRC1. Unlike CBX7 (and apparently all other polycomb paralogs), CBX6 does not bind to H3K27me3.[41, 169] Ablation of the H3K27 methyl-depositing PRC2 complex in ESCs did not change CBX6 recruitment to chromatin providing further evidence to support the non-redundant functions of CBX6 and CBX7.[163]

Overexpression of CBX6 promotes tumour growth and is linked to aggressive hepatocellular carcinoma (HCC).[170] The fifth-most common type of cancer in men is primary liver cancer, and HCC makes up 90% of primary liver cancers.[232-233] CBX6 mRNA levels were increased in 6 different HCC cell lines and in primary HCC tumour tissue samples.[170] In HCC patients, high CBX6 expression levels were correlated to poor patient outcomes after hepatectomy.[170] *In vitro* studies in multiple HCC cell lines showed increased CBX6 expression to promote cell growth whereas knockdown of CBX6 showed reductions in proliferation. *In vivo* studies also showed CBX6 overexpression to promote HCC tumour growth.[170] These findings suggest the potential for CBX6 to be a diagnostic biomarker of HCC and a therapeutic target.

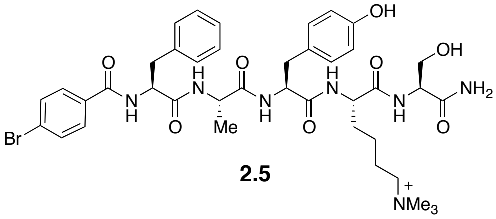
3.3. Results

The histone peptides that are the canonical binding partners for all of these chromodomains actually bind with relatively low in vitro affinities (55 – >500 μM).[41, 169, 234] In spite of the canon that defines H3K9me3 and/or H3K27me3 as the targets of CBX proteins, H3K9me3 and H3K27me3 peptides have been repeatedly shown not to bind measurably with multiple members of the CBX family.[41, 235] In order to assure strong baseline affinity for our ligands, we started with a peptidic sequence (**3.1**) that we previously identified as a moderate-strength CBX7 binder. Re-synthesis and testing of **3.1** in a FP assay with **2.4** as the competitive binding probe gave an IC_{50} value of 77 μM for CBX7. The chemically modified version (**2.5**, Chapter 2) has improved affinity (IC_{50} = 1.4 μM) arising from a *p*-bromobenzamide group at the N-terminus.[236]

Competitive fluorescence polarization (FP) studies are limited in their ability to measure K_d values for potent ligands, especially for comparison across different proteins.[237] In order to enable direct FP titrations of an entire panel of proteins into ligands, we modified **2.5** by adding a linking residue and fluorescent dye at the C-terminus to create compound **2.4** (Chapter 2). We had previously established a panel of proteins for use in fluorescence polarization assays of ligand affinities and selectivities (Chapter 2). In my work described in Chapter 2, the addition of an anionic Glu residue at the C-terminus of the ligand generally increased potency to all CBX proteins. It is from this lead scaffold **2.4** that further modifications were based. Compounds **2.8** and **3.2** involved modifications of the side chain directed into the (–2) pocket.



3.1



2.5

Figure 3.2. Peptide sequences Ac-FAYKme$_3$S-NH$_2$ (**3.1**) and pBr-FAYKme$_3$S-NH$_2$ (**2.5**) identified as CBX7 binders.

The (–2) substitutions in **2.4**, **2.8** and **3.2** have dramatic effects on potencies and selectivity for CBX-proteins. Figure 3.3 shows K$_d$ values arising from direct FP titrations of all six chromodomains into all three ligands (see also Chapter 2). Compound **2.4**, bearing the methyl substituent at the (−2) position was potent and promiscuous. Its K$_d$ values are, from strongest to weakest, CBX6/7 < CBX2 < CBX4/8. However, even the affinity for the weakest partner CBX4/8, at 0.65-0.85 μM, is >30-fold stronger than the affinities of any small molecule inhibitor for any chromodomain yet reported,[153-154] and near equal potency compared to the best peptidic inhibitors.[149-152, 156, 238] Inhibitor **2.4** is moderately selective for all polycomb paralogs (0.012 to 0.85 μM) over the HP1 paralog CBX1 (5 μM).

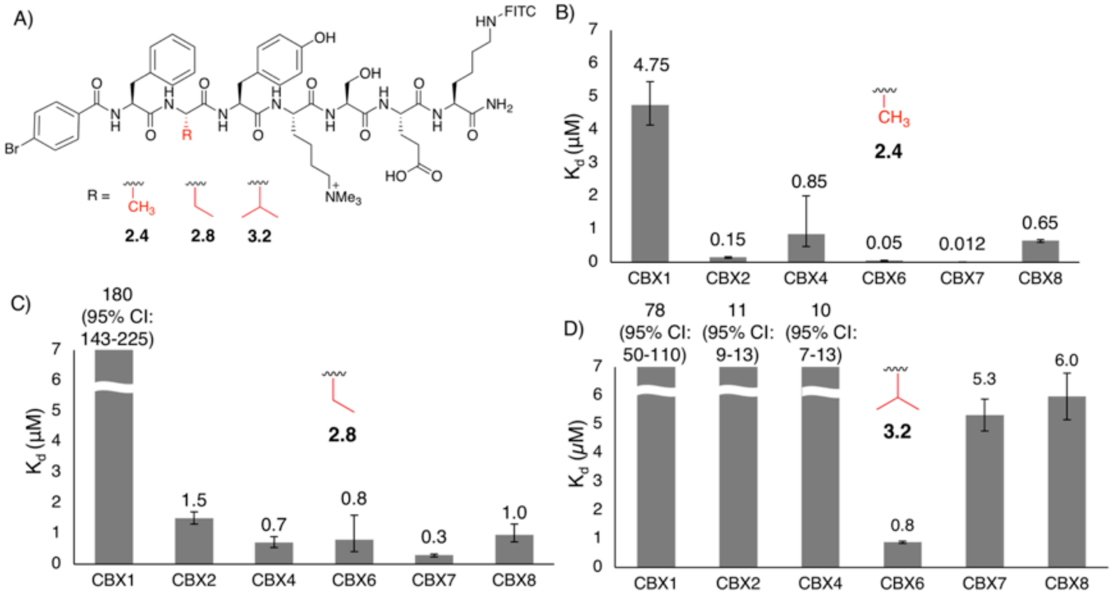


Figure 3.3. Series of CBX ligands with varying alkyl substitutions at the (–2) position and binding affinities determined by direct FP. A) Structure of ligands **2.4**, **2.8** and **3.2**, B) K_d values for **2.4**, C) K_d values for **2.8**, D) K_d values for **3.2**. Errors are reported as asymmetric 95% confidence intervals (or in parentheses for values that exceed axis limits) from experiments done in triplicate.

Addition of the ethyl substituent in **2.8** weakened binding to CBX1 (HP1β) by >30-fold, while also generating decreases in binding potency to CBX2/6/7 and no significant change in binding to CBX4/8. The isopropyl substitution in compound **3.2** decreased binding affinity of the peptide to all of CBX2/4/7/8, while not significantly changing binding to CBX6. Compound **3.2** is 98-, 14-, 13-, 7-, and 8-fold selective for CBX6 over CBX1/2/4/7/8, respectively. Analogues of compounds **2.4** and **3.2** lacking FITC labels were tested using a competitive FP assay (Figure 3.4, compounds **2.7** and **3.3**). The IC_{50} values determined showed 7-fold selectivity for CBX6 over CBX7 for unlabeled **3.2**, demonstrating that the FITC tag alone is not the source of CBX6 selectivity.

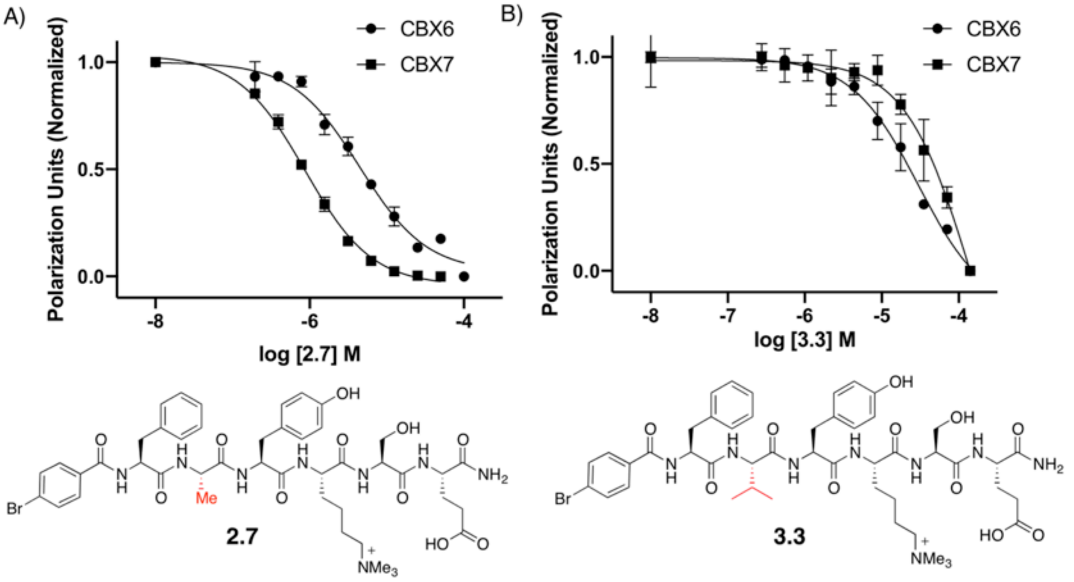


Figure 3.4. Overlays of normalized competitive fluorescence polarization titrations with **2.7** and **3.3** against CBX6 (circles) and CBX7 (squares). A) **2.7** studied with CBX6 (IC_{50} = 4.5 µM, 95% CI: 3.5-5.7 µM) and CBX7 (IC_{50} = 1.5 µM, 95% CI: 1.2-1.8 µM). B) **3.3** studied with CBX6 (IC_{50} = 20 µM, 95% CI: 18-29 µM) and CBX7 (IC_{50} = 150 µM, 95% CI: 85-300 µM). Errors are reported as asymmetric 95% confidence intervals from experiments done in triplicate.

We further confirmed the affinities and selectivities of **2.4** and **3.2** by preparing **2.4**-biotin (**3.4**) and **3.2**-biotin (**3.5**) to enable orthogonal characterization of the complexes by surface plasmon resonance (SPR) (Figure 3.5, Table 3.1). SPR is a established technique used to determine binding constants of biomolecular interactions.[239] Association and dissociation of a free molecule binding to an immobilized molecule are measured by the change in refractive index at the SPR sensor-chip.[239] We chose to design our SPR experiments using streptavidin-functionalized sensor-chips and immobilizing biotin-labeled ligands onto the chip. CBX6 and CBX7 were free in solution and flowed over the chip to allow for association with the immobilized peptide ligand. The results from these experiments echo the selectivity trends obtained by direct FP, while also generally agreeing with the absolute K_d values. This is the first reported study utilizing SPR to study binding affinities of a methyllysine reader protein.

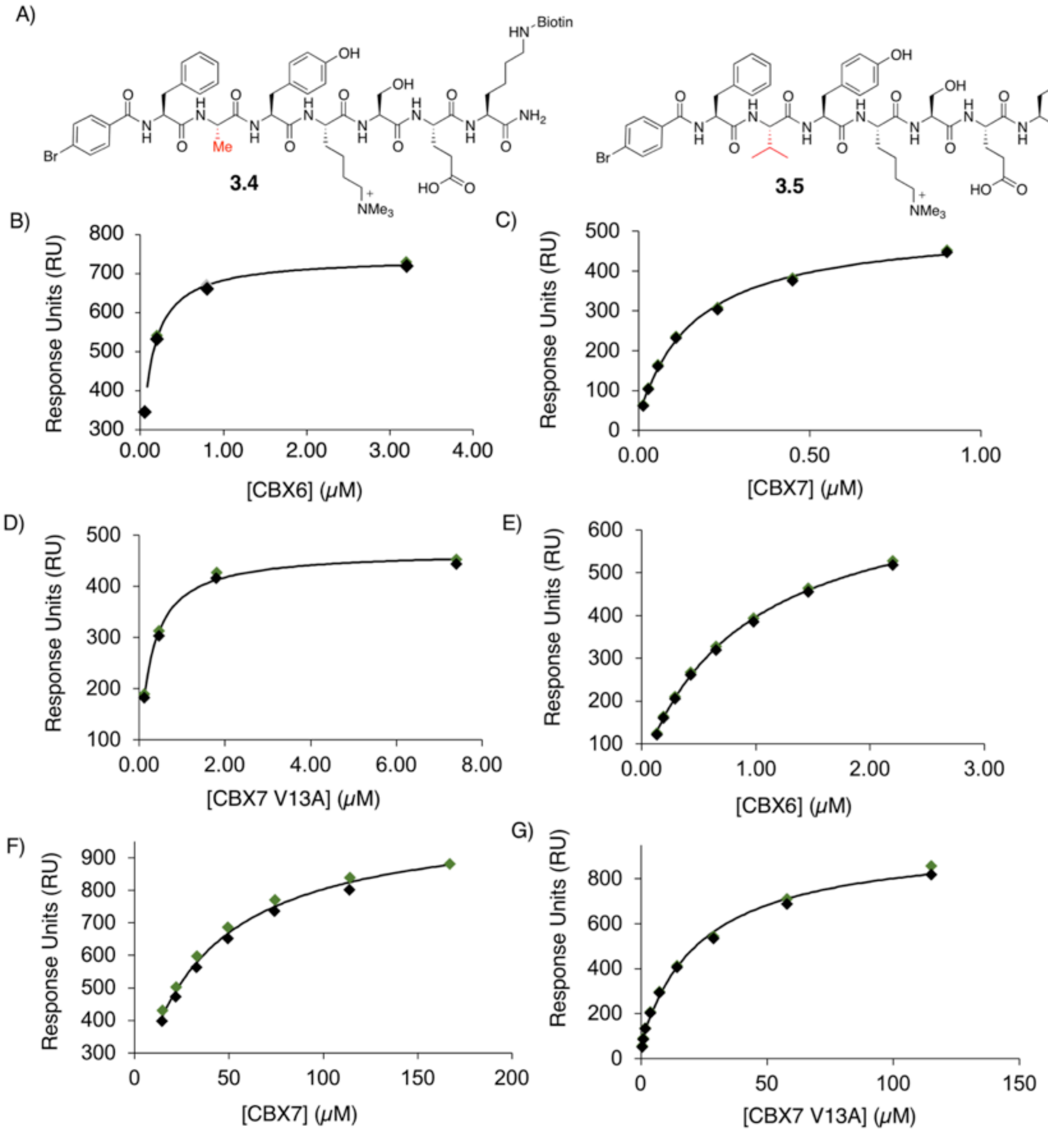


Figure 3.5. Fitted SPR binding curves for SA chip functionalized with **3.4** or **3.5** studied with CBX6, CBX7 and CBX7-V13A. A) Structures of **3.4** and **3.5**, B) **3.4** on SA chip with CBX6, C) **3.4** on SA chip with CBX7, D) **3.4** on SA chip with CBX7-V13A, E) **3.5** on SA chip with CBX6, F) **3.5** on SA chip with CBX7, G) **3.5** on SA chip with CBX7-V13A. See Table 3.1 for K_d values. All experiments were done in duplicate.

Table 3.1. Binding affinities of **3.4** and **3.5** determined by SPR

Compound	**Protein**	**K_d (µM)**
3.4	CBX6	0.11 ± 0.01
	CBX7	0.16 ± 0.01
	CBX7 V13A	0.29 ± 0.05
3.5	CBX6	0.91 ± 0.05
	CBX7	34 ± 8
	CBX7 V13A	24 ± 2

To probe the role of Val/Ala substitutions at the floor of the (−2) pocket (Figure 3.1B, C) in defining CBX protein selectivity, the CBX7-V13A mutant was prepared and tested by SPR (Figure 3.5D and G, Table 3.1). The weak binding to compound **3.5** by this CBX6-like mutation of CBX7 shows that the V13A substitution alone is not sufficient to drive CBX6-like selective binding of **3.5**. This is consistent with the observed low potency of **3.5** for binding CBX2 and CBX8, which also have the V13A substitution but differ in other residues.This observed selectivity for CBX6 only, and not CBX2 and CBX8, was not predicted by the simple model for Ala/Val swaps shown in Figure 3.1B and C.

We carried out energy minimizations (Moloc) and ligand docking (SeeSAR) for ligands **2.7** and **3.3** using the crystal structures for CBX6 (PDB: 3I90) and CBX7 (4MN3). Docking of compounds **2.7** and **3.3** was done by superposition of the ligand backbone coordinates over those of the crystal structure native peptide H3K27me3 in CBX6. Both methods showed that the large side chain on ligand **3.3** could be accommodated in identical modes and with identical energies in the (–2) pockets of CBX6 and CBX7, confounding the simple picture shown in Figure 3.1C and assumed to hold true for all CBX proteins. The overall picture is that the (−2) pocket is critical for the native binding preferences of CBX proteins, but that the way in which ligand **3.3** provides for CBX6 selectivity by occupying this pocket is too complicated to be understood using static X-ray structures.

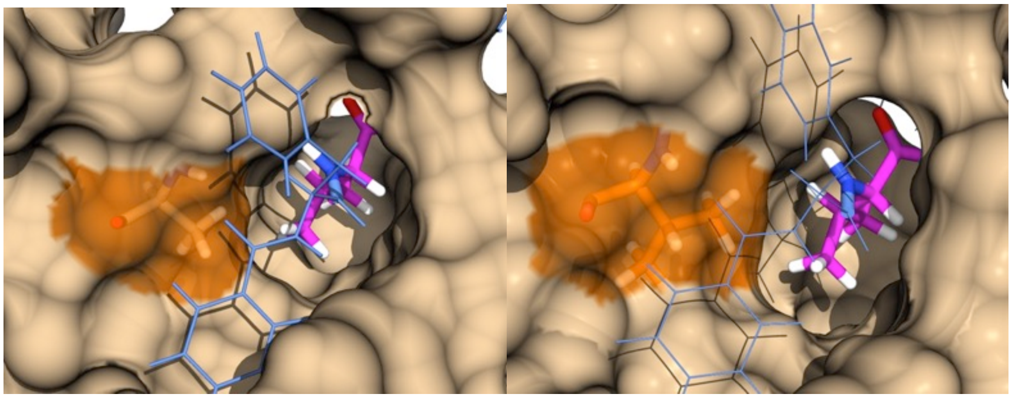
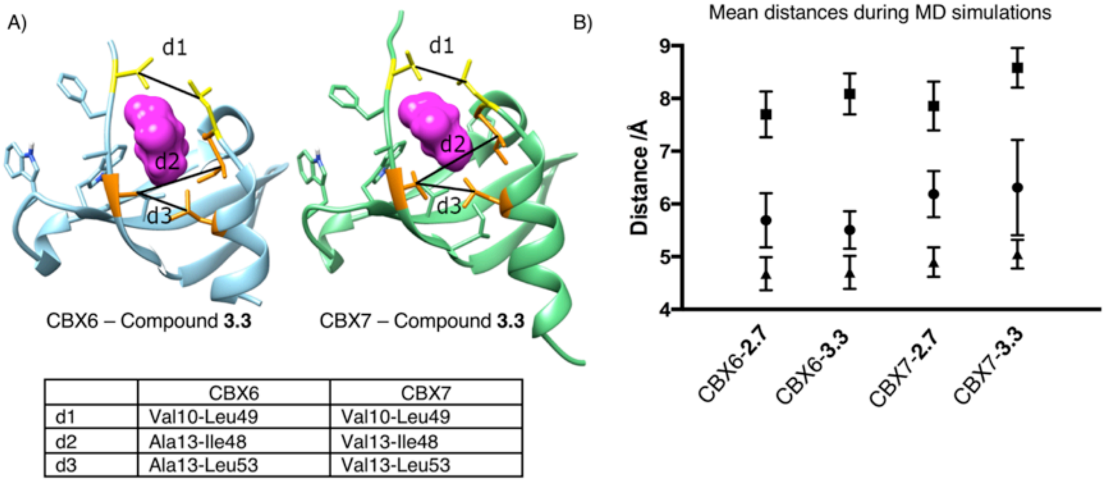


Figure 3.6. Minimized fit of **3.3** in CBX6 (Left) and CBX7 (Right). Orange highlighted regions indicate the alanine (CBX6) and valine (CBX7) residues on the outside of the hydrophobic clasp. While these residues appear to partially form the rim of the (−2) pocket, they have limited impact on the shape at the bottom of the pocket or how much steric interaction with the compound **3.3** valine (pink) occurs.

To gain a dynamic view of the complexes, we carried out molecular dynamics (MD) simulations of the complexes of CBX6 and CBX7 with each of **2.7** and **3.3**. Our goal was to better understand the important structural changes that explain the observed selectivity. The trajectories show that the complex of CBX7 with ligand **3.3** (mismatched) undergoes significant opening of the hydrophobic clasp that envelopes the ligand. In contrast, the complex of CBX6 with **3.3** (matched) remains completely wrapped around the ligand throughout the simulation, as illustrated by clasp distance d1 in Figure 3.7. The data in Figure 3.7 also include further key geometric parameters that quantify how CBX6 and CBX7 differ in their engagement of ligands. The ligand side chains always stay in the (–2) pockets, but the pocket shapes respond in different ways to different ligands. The overall 'external pocket width' (d3 in Figure 3.7) is the same in all complexes, showing that the mouth of the pocket is (surprisingly) the same regardless of the Val13Ala swap at one edge of this measured distance. The 'internal pocket width' (d2 in Figure 3.7) and hydrophobic clasp distance (d1) both show significant increases for CBX7-**3.3** that are not observed for CBX6-**3.3**, showing increased strains that account for the observed experimental data.



	CBX6	CBX7
d1	Val10-Leu49	Val10-Leu49
d2	Ala13-Ile48	Val13-Ile48
d3	Ala13-Leu53	Val13-Leu53

Figure 3.7. Molecular dynamic simulation results showing the change in distances within CBX6 and CBX7 when in complex with compound **2.7** and **3.3**. A) Representative snapshots from MD trajectories, showing only those parts of the ligands that occupy the (–2) pocket as magenta surfaces. The distances d1, d2, d3 define changes in pocket shape that can be compared between simulations. The distance d1 is a measure of the hydrophobic clasp that covers the ligand and is a major determinant of the 'closedness' of the binding site. Distance d2 measures the width at the deepest part of the (–2) pocket while d3 measures the width across the mouth of the (–2) pocket. B) Mean values for distances d1 (circle), d2 (square), and d3 (triangle) show the changes induced in and around the (–2) pocket when the bulkier compound **3.3** is bound to CBX6 and CBX7.

MD simulations show that inclusion of the bulky compound **3.3** under the CBX7 hydrophobic clasp induces significant changes in the Kme3 recognition site. The opening of the clasp leads to a shift in the aromatic cage phenylalanine (Phe11) residue, reducing its interaction strength with the Kme3. The distance between the two relevant residues was calculated along the trajectory, and is presented in Figure 3.8. These experiments highlight important structural information on how the position of the clasp is tied to the recognition of the Kme3 in the aromatic pocket.

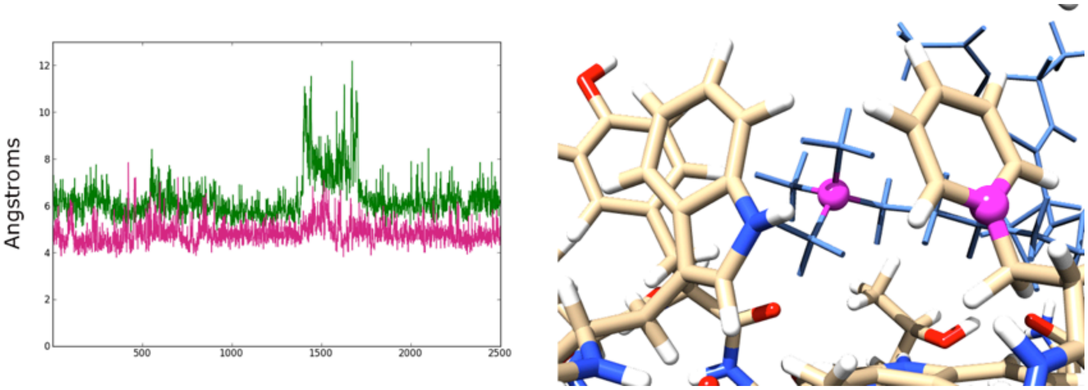


Figure 3.8. Hydrophobic clasp distance correlations with aromatic cage structure. To show the complex nature of the induced fit in the (−2) pocket region, a plot relating the hydrophobic clasp distance to the Kme3 interaction shows a correlation between these two binding regions. Left: distance plot over the course of 50 ns shows the clasp gap distance (green) compared to the distance between the Phe11 residue and Kme3 residues on CBX7 and compound **3.3**, respectively (pink). Right: aromatic cage interaction with the reference atoms coloured in pink.

3.4. Discussion

The aromatic cage of CBX proteins are known to be important binding sites, but our results show how ligand affinity also depends on occupying other binding sub-sites. The histone tail ligands in native CBX−histone complexes occupy the beta groove and interdigitate between the existing two protein strands in order to make a short three-strand beta-sheet motif. The previously reported small molecule ligands for CBX7 engage only the region around the aromatic cage and have modest potencies (23−67 μM).[153-154] The modified peptidic ligands we report here reach affinities as strong as 0.01 μM (for the complex CBX7-**2.4**) and are routinely sub-μM. We attribute this to their occupation of the beta groove, which includes the aforementioned (−2) pocket as well as a hydrophobic cleft that extends further from the Kme3 binding site. Comparison of **3.1** and **2.5** shows that occupying that cleft provides >40-fold enhanced potency.

The origins of selectivities for the small molecule vs. the peptidic ligands are more subtle. The small-molecule ligand of Zhou shows impressive 3- to 22-fold selectivities for CBX7 over CBX2/4/6/8 even though it only binds to the aromatic cage region.[153] The aromatic cages of all CBXs are highly similar to each other in structure, so we infer that the aromatic cage of CBX7 has better preorganization/reduced solvent accessibility relative to its family members, rather than a large difference in protein-ligand interactions in the bound state. Peptidic compound **2.4** also shows stronger binding to CBX7 than any other CBX protein, suggesting that it is benefitting from similar effects.

Large groups in the (–2) pocket are able to overcome the inherent bias toward CBX7 binding. The particular CBX6 selectivity is not simply explained by the Val/Ala difference at position 13 of the chromodomains, as would have been predicted both by simple modeling and by the Val/Ala mutation results reported for CBX2 and CBX7 functions.[240] Kaustov et al. showed using a peptide array that CBX8 (but not CBX7) could accommodate a valine side chain in the (–2) position of a histone tail sequence,[41] but their qualitative array-blotting result did not include CBX6 for comparison. Our solution phase data agree with this result to an extent, in that they show similar solution-phase affinities for valine-containing ligand **3.3** binding to CBX7 or CBX8 (ca. 5 μM each). But the affinities for CBX6 out-strip those affinities, providing an unanticipated degree of selectivity for this polycomb paralog.

These results also uncover the previously unknown and divergent binding preferences of CBX6's chromodomain. While it has been assumed to be a canonical polycomb reader of H3K9me3 (-ARKme3ST-) and/or H3K27me3 (-ARKme3SA-), the in vitro affinities of CBX6 for these marks are in fact unmeasurably weak in experiments done alongside other CBX chromodomains.[41, 235] Each of these histone marks has an alanine residue occupying the (–2) pocket, and according to our results would be poorly suited to bind CBX6. Our results, in which a ligand's isopropyl side chain (valine residue) occupies the (–2) pocket and provides high affinity binding, suggest indirectly that CBX6 might be a reader of a different, as-yet undetermined trimethyllysine site.

3.5. Conclusions

Chromodomain-containing proteins are increasingly suggested as targets for therapeutic intervention, and the functional biology of polycomb paralogs is an important frontier of epigenetics and stem cell biology. Studying the chemical biology and therapeutic potential of CBX proteins requires selective ligands that, until now, have not been available. Our results provide new peptidic probes for CBX6, which has been the least studied of the human polycombs. They also inform the design requirements of the next-generation of potent and selective small-molecule ligands for CBX proteins.

3.6. Experimental methods and data

3.6.1 Synthesis

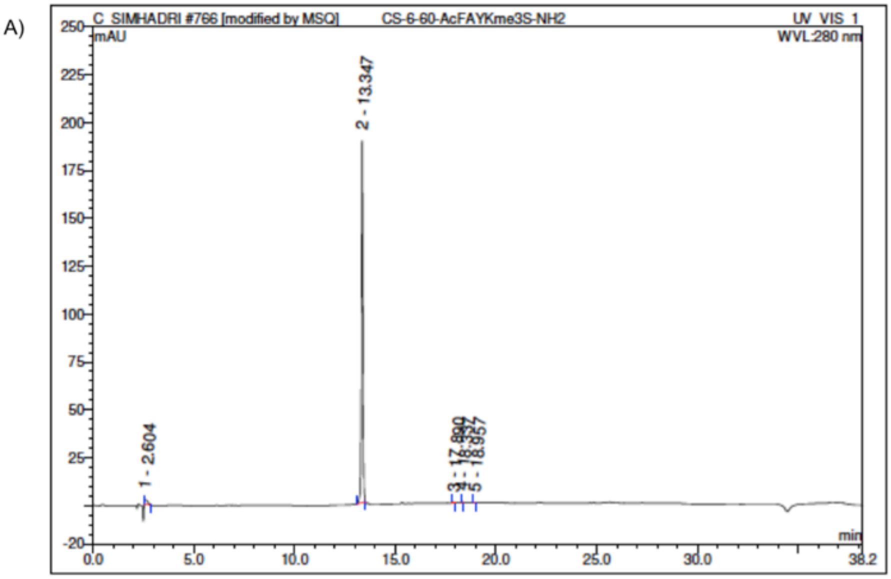
General synthesis, chemicals used, and Fmoc protocols for peptide synthesis are the same as described in Chapter 2, Experimental methods and supplementary data section 2.8.1.

Methods for lysine side-chain functionalization with fluorescein isothiocyanate (FITC) are described in Chapter 2, Experimental methods and supplementary data section 2.8.2.

Methods for lysine side-chain functionalization (+)-Biotin N-hydroxysuccinimide ester are described in Chapter 2, Experimental methods and supplementary data section 2.8.3.

Methods for N-terminal acetylation of peptides are described in Chapter 2, Experimental methods and supplementary data section 2.8.4.

3.6.2 Compound characterization data



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

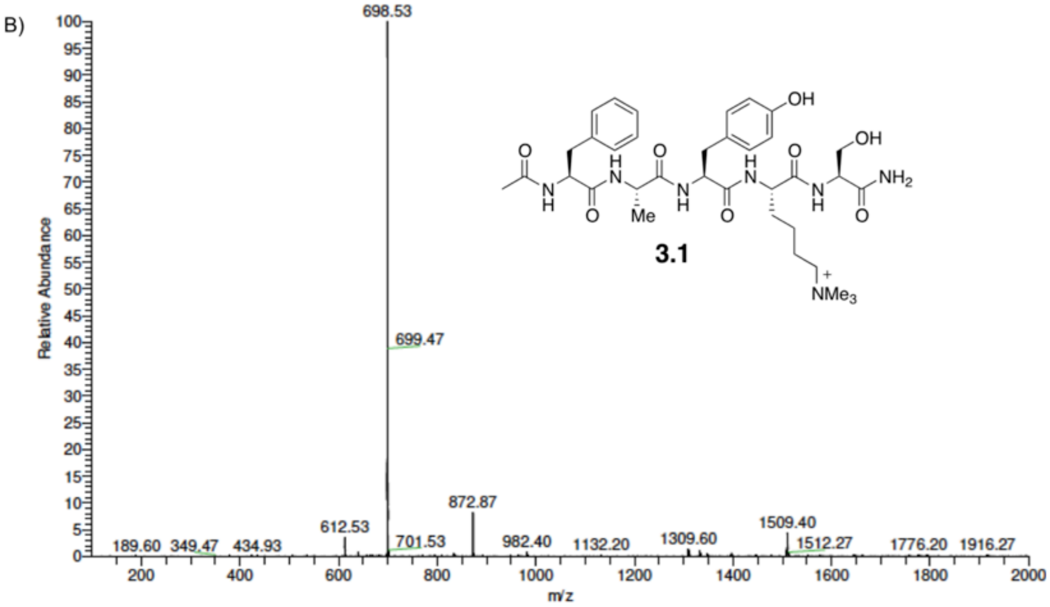
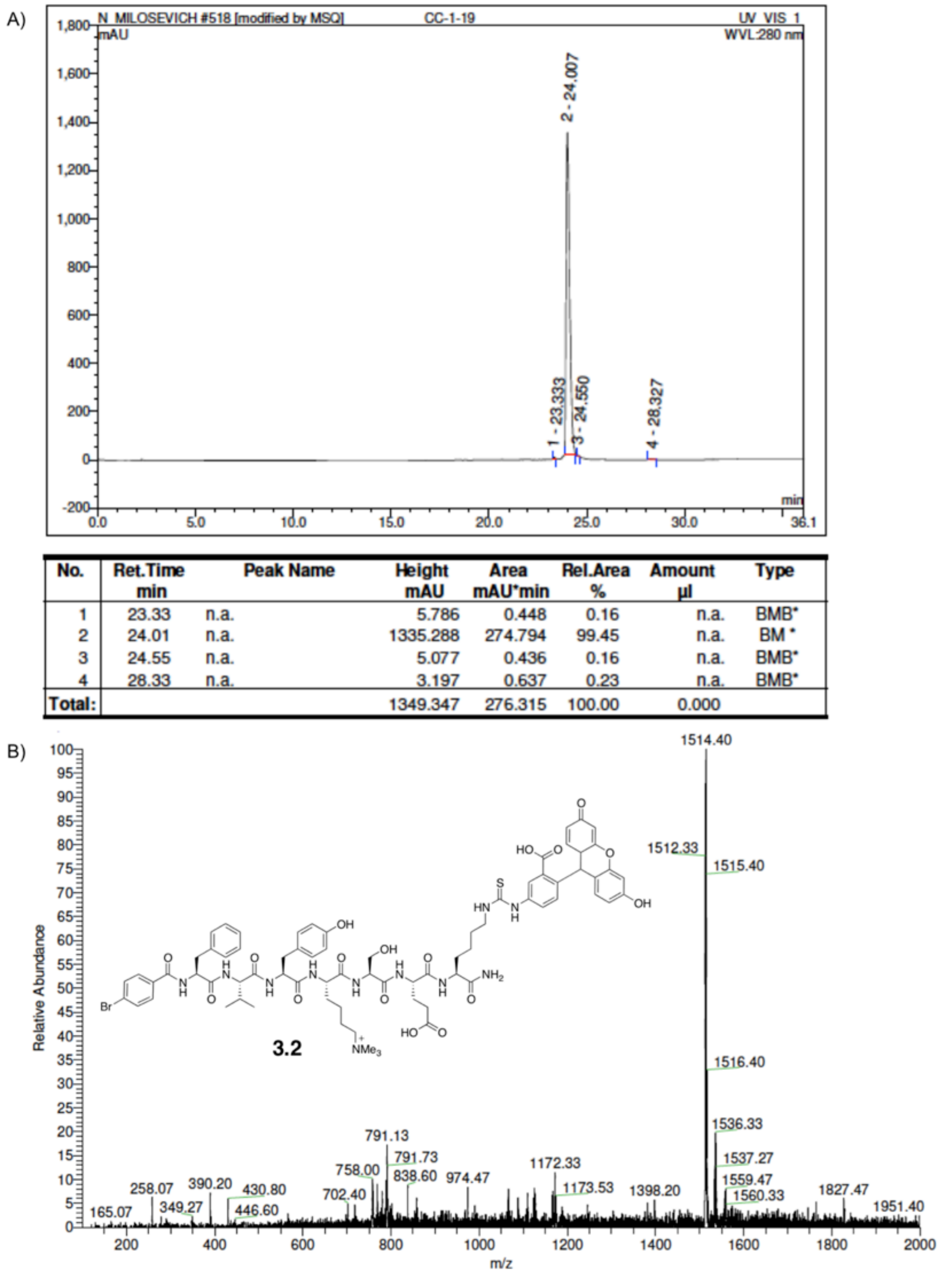
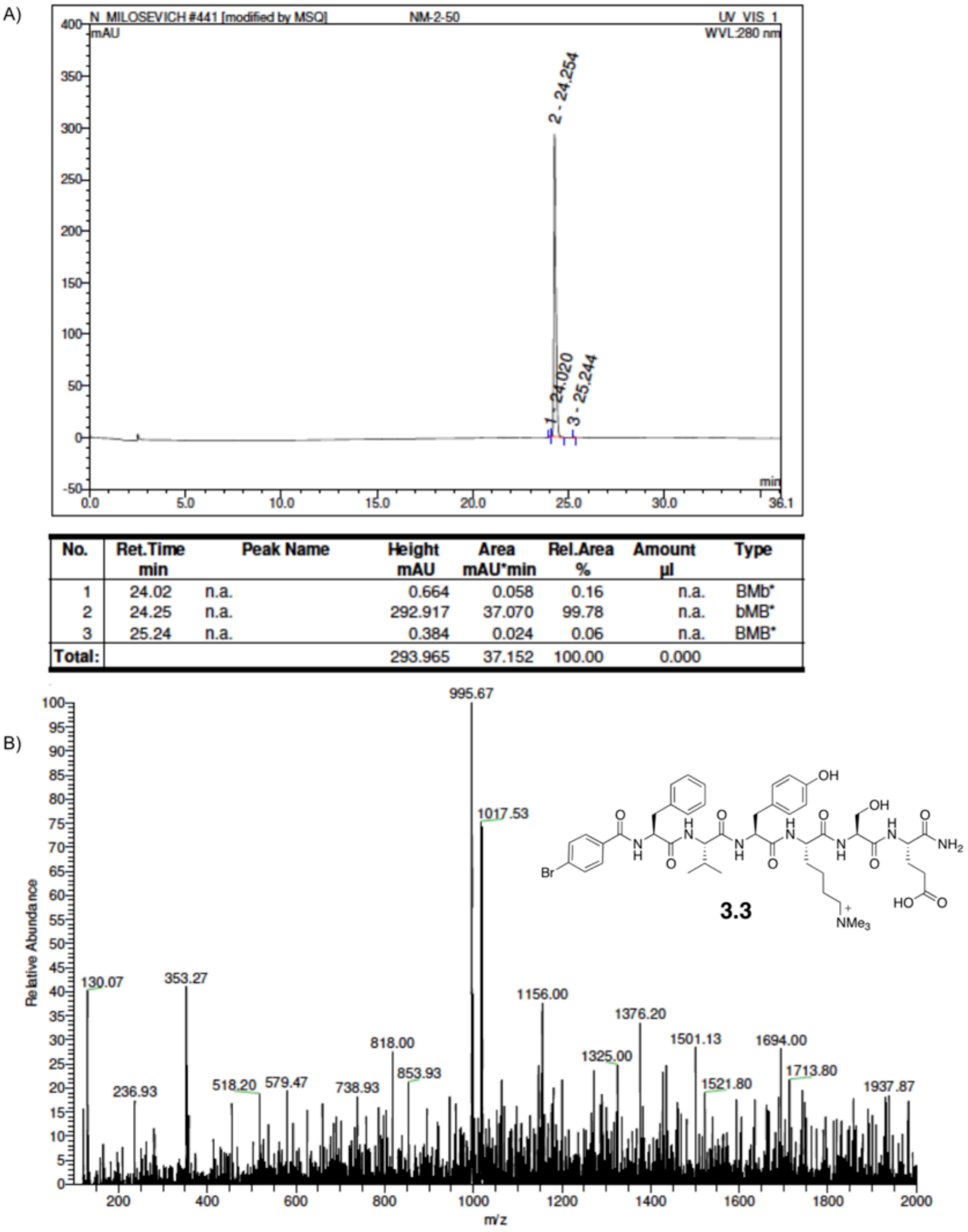


Figure 3.9. Characterization data for compound **3.1**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{35}H_{52}N_7O_8^+$: 698.39; found: 698.53.



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

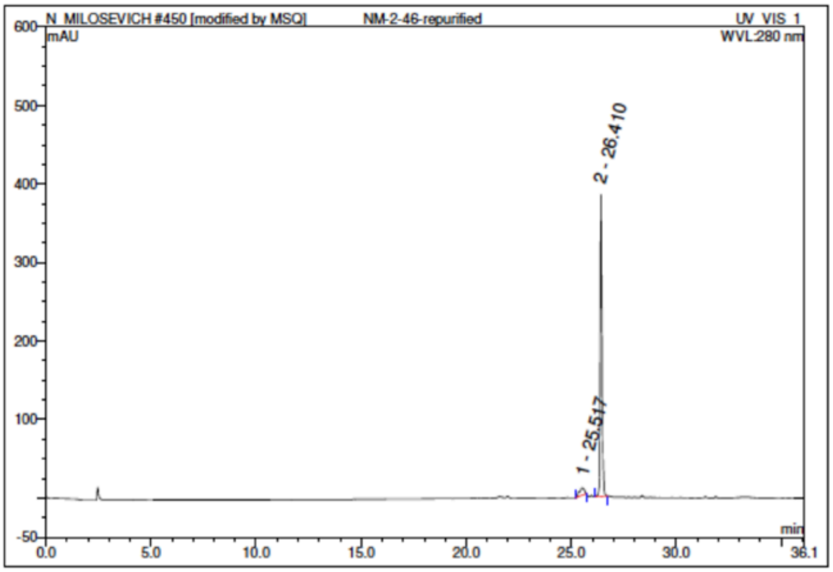
Figure 3.10. Characterization data for compound **3.2**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{74}H_{89}BrN_{11}O_{17}S^+$: 1514.53; found: 1514.40.



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

Figure 3.11. Characterization data for compound **3.3.** A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{47}H_{64}BrN_8O_{11}^+$: 995.39; found: 995.67.

A)



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

B)

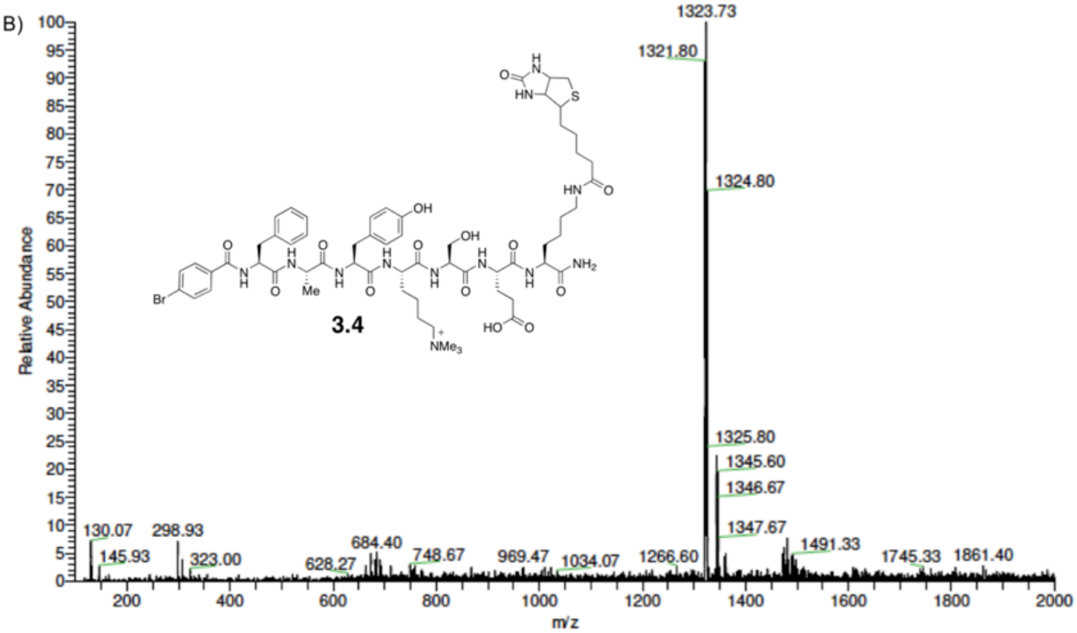
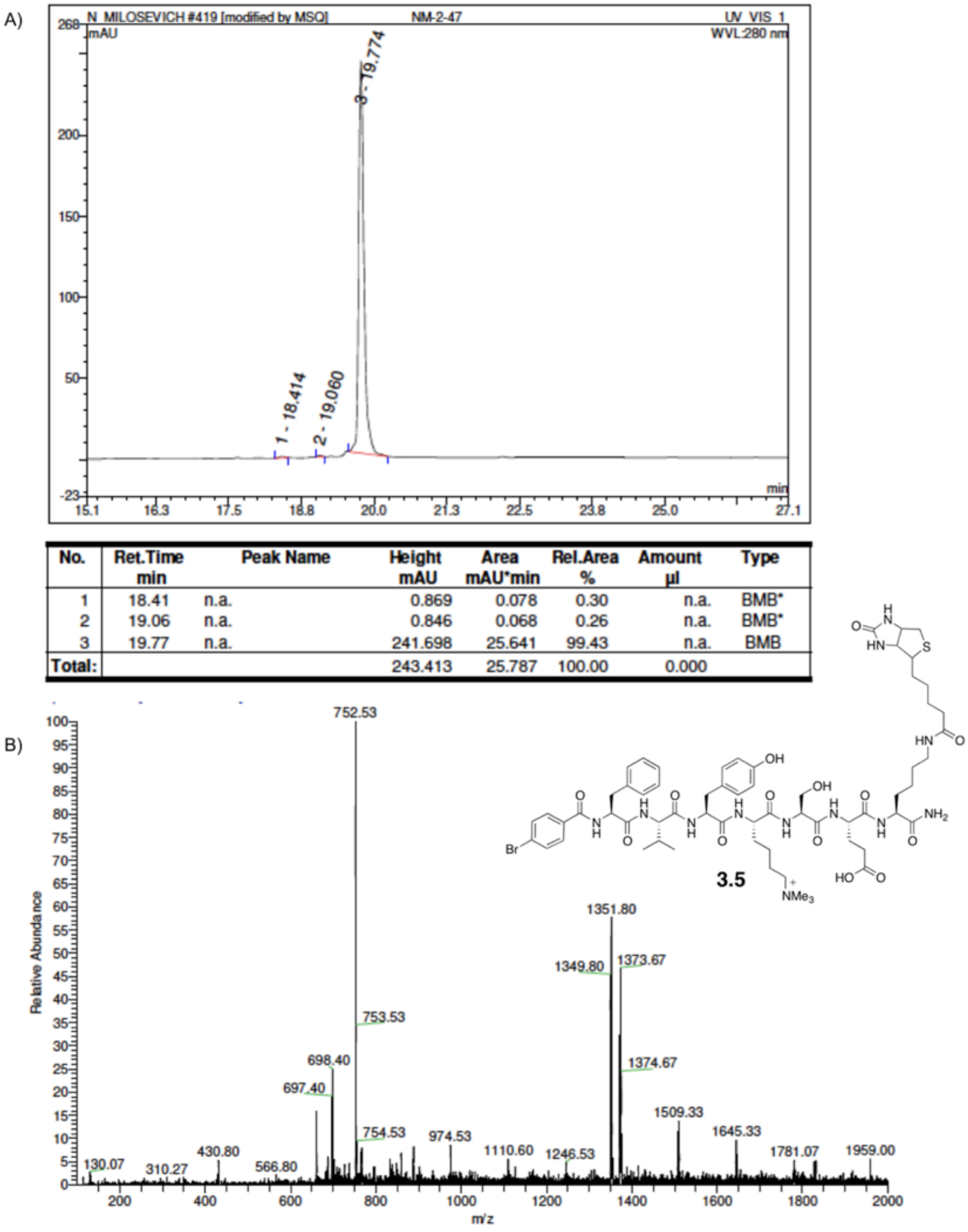


Figure 3.12. Characterization data for compound **3.4**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{61}H_{86}BrN_{12}O_{14}S^{+}$: 1323.53; found: 1323.73.



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

Figure 3.13. Characterization data for compound **3.5**. A) Analytical LC/MS trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{63}H_{90}BrN_{12}O_{14}S^{+}$: 1351.56; found: 1351.80.

3.6.3 Protein expression and purification

Methods used were the same as those described in Chapter 2, Experimental methods and data section 2.8.6.

3.6.4 Generation of CBX7 point mutant

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

3.6.5 Fluorescence polarization methods

Direct FP was done by titration of CBX proteins into FITC-labelled probes **2.4**, **2.8** and **3.1**. Methods used for direct FP assays and analysis of data is described in Chapter 2, Experimental methods and supplementary data section 2.8.7.

Competitive FP analysis of **2.5**, **2.7**, **3.1**, and **3.3** binding to CBX6/7 was used to determine IC_{50} values using the new FITC-peptide **2.4** as the competitive binding probe (Figure 3.4). The method used for the competitive FP assays and analysis of data is described in Chapter 2, Experimental methods and supplementary data section 2.8.7.

The experiments on **2.7** and **3.3** were added in order to confirm that the FITC tag on compound **3.3** is not somehow responsible for its observed CBX6 selectivity. The data, although qualitative, do confirm this by showing that **2.7** actually has a 3-fold preference for CBX7 that is reversed for **3.3**, which has 7-fold preference for CBX6 (Figure 3.4). In spite of these results agreeing with our conclusion that **3.2** is selective for CBX6 because of its isopropyl substituent, we emphasize that competitive FP in such systems provides IC_{50} values that depend strongly on the intrinsically different affinities between each protein and the FITC-peptide probe. Direct FP and/or SPR experiments (see below) don't rely on a third binding partner as a reporter of the interaction, and are therefore better measures of protein-ligand K_d values when comparing data from a panel of different proteins.

Raw data for compounds **2.4** and **2.8** are shown in Chapter 2.

3.6.6 Fluorescence polarization data

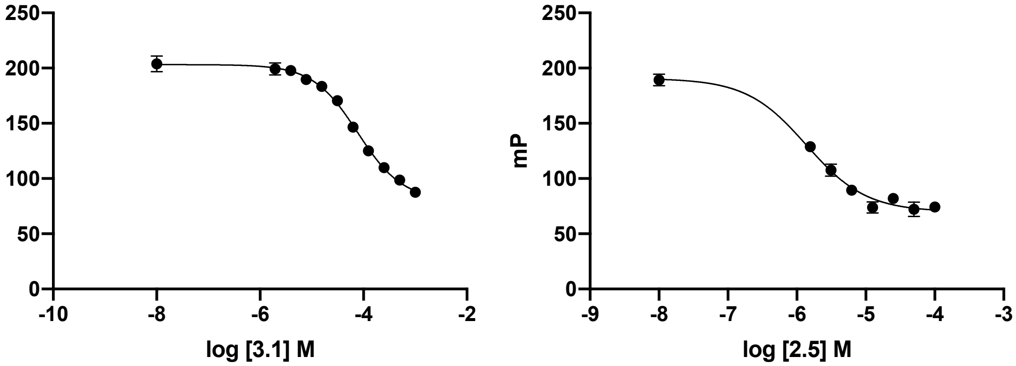


Figure 3.14. Competitive fluorescence polarization for compound **3.1** and **2.5** with CBX7 using compound **2.4** as the fluorescent probe. Left: **3.1** with CBX7 (IC_{50} = 77 μM, 95% CI 66 μM to 90 μM, R^2 = 0.995), Right: **2.5** with CBX7 IC_{50} = 1.4 μM, 95% CI 1.0 μM to 1.8 μM, R^2 = 0.983). Error reported as asymmetrical 95% confidence intervals from experiments done in triplicate

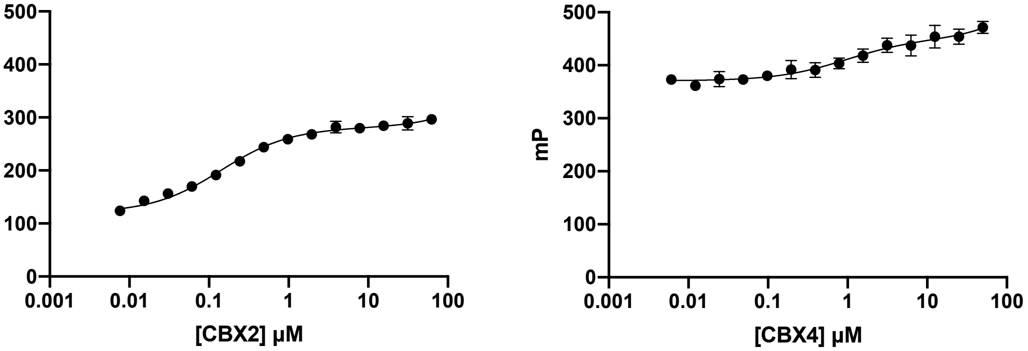


Figure 3.15. Direct fluorescence polarization data of compound **2.4** with CBX2 and CBX4. Left: CBX2 (K_d = 0.15 μM, 95% CI 0.11 μM to 0.17 μM, R^2 = 0.990), right: CBX4 (K_d = 0.85 μM, 95% CI 0.47 μM to 2 μM, R^2 = 0.900). Error reported as asymmetrical 95% confidence intervals from experiments done in triplicate. Direct FP curves and binding constants for **2.4** with CBX1/6/7/8 are reported in Chapter 2.

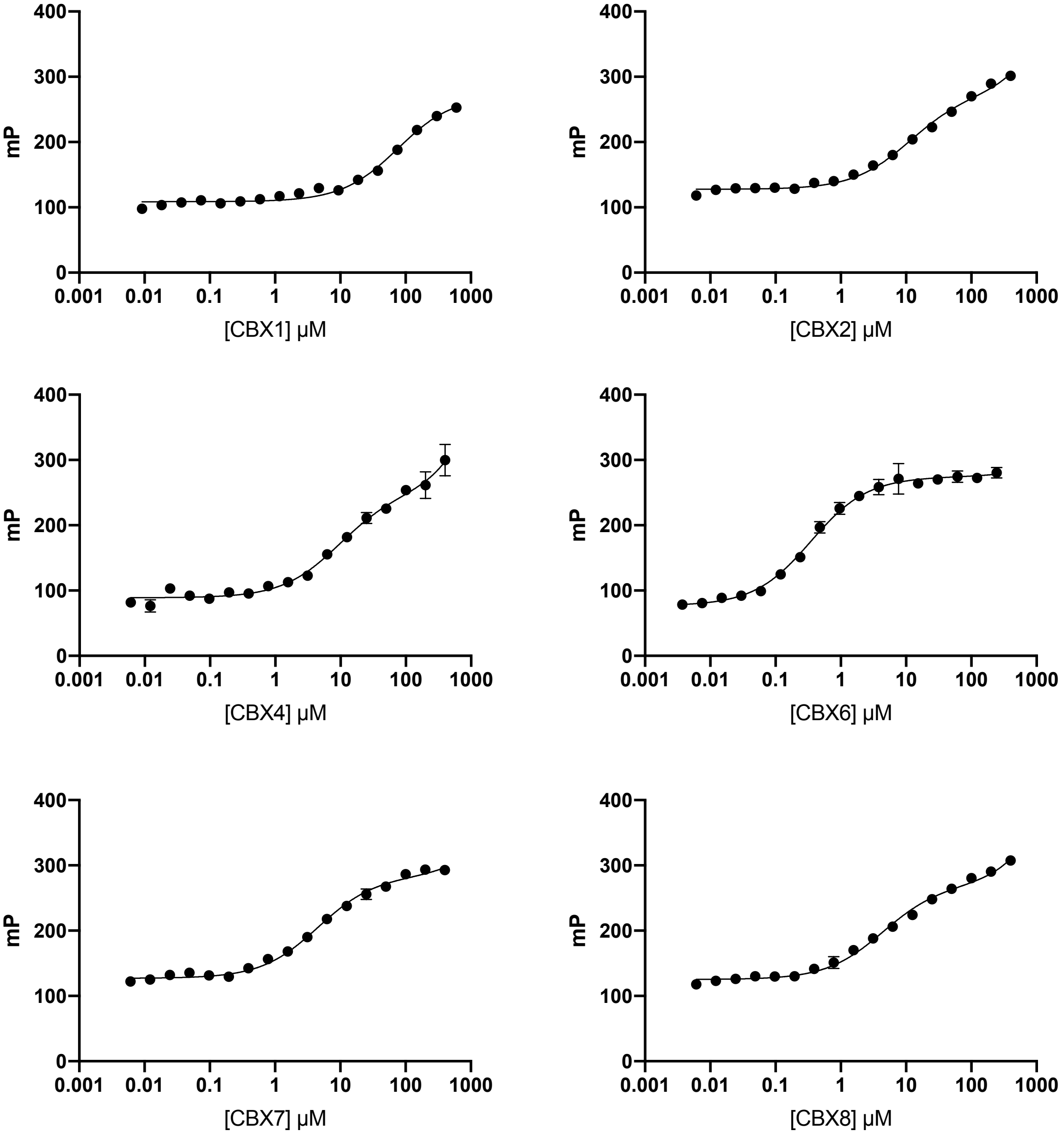


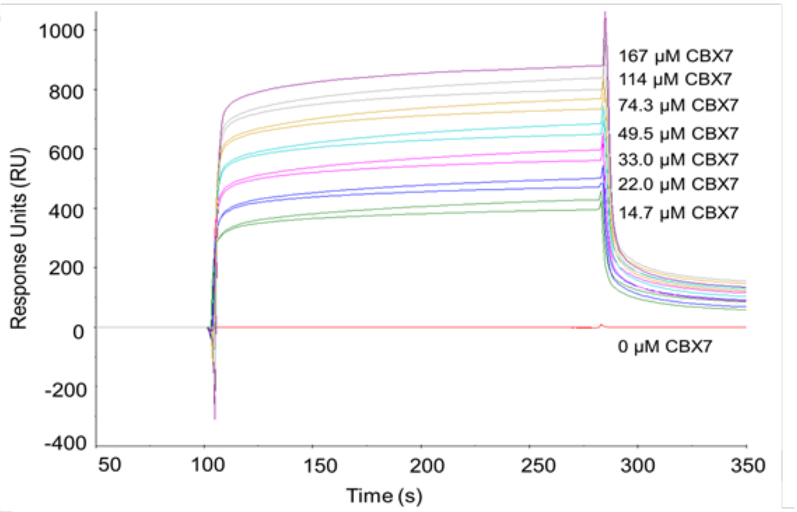
Figure 3.16. Direct fluorescence polarization data of compound **3.2** with CBX1/2/4/6/7/8 performed in triplicate. CBX1 (K_d = 79 µM, 95% CI 50 µM to 125 µM, R^2 = 0.986), CBX2 (K_d = 11 µM, 95% CI 9 µM to 13 µM, R^2 = 0.992), CBX4 (K_d = 9.5 µM, 95% CI 7 µM to 13 µM, R^2 = 0.986), CBX6 (K_d = 0.75 µM, 95% CI 0.65 µM to 0.92 µM, R^2 = 0.993), CBX7 (K_d = 4.8 µM, 95% CI 3.9 µM to 6 µM, R^2 = 0.992), CBX8 (K_d = 4.5 µM, 95% CI 3.7 µM to 5.2 µM, R^2 = 0.993). Error reported as asymmetrical 95% confidence intervals from experiments done in triplicate.

3.6.7 SPR methods

All surface plasmon resonance experiments were performed on a Biacore X100 (GE Healthcare). Research-grade streptavidin (SA) sensor chips were used for ligand immobilization. The biotinylated ligands (**3.4** and **3.5**) were prepared in HBS-EP+ buffer (10 mM HEPES, 150 mM NaCl, 3mM EDTA, 0.005% P20, pH 7.4) and 10% dimethylsulfoxide. Ligand immobilization in flow cell 2 was carried out by injection at a flow rate of 5 µl/min for 18 minutes for a density of 632 RU and 1313 RU for compound **3.4** and **3.5** respectively. Flow cell 1 was left blank as a reference. Stock protein solutions were stored in HBS-EP+ running buffer at –80°C and dilution series were prepared as required. Binding data was collected by injecting the analyte at a flow rate of 30 µL/min over two flow cells at a temperature of 25°C. Association and dissociation of the complex was 3 minutes and 6 minutes respectively for compound **3.4**, and 3 minutes and 4.5 minutes for compound **3.5**. The flow cell surfaces were regenerated with a 30 second injection of 10 mM glycine-HCl solution pH 2.0 (GE Healthcare). Duplicate injections for each sample, including the blank, were carried out for each protein-ligand interaction. Affinity curves, dissociation constants and sensograms were determined using Biacore X100 Evaluation Software.

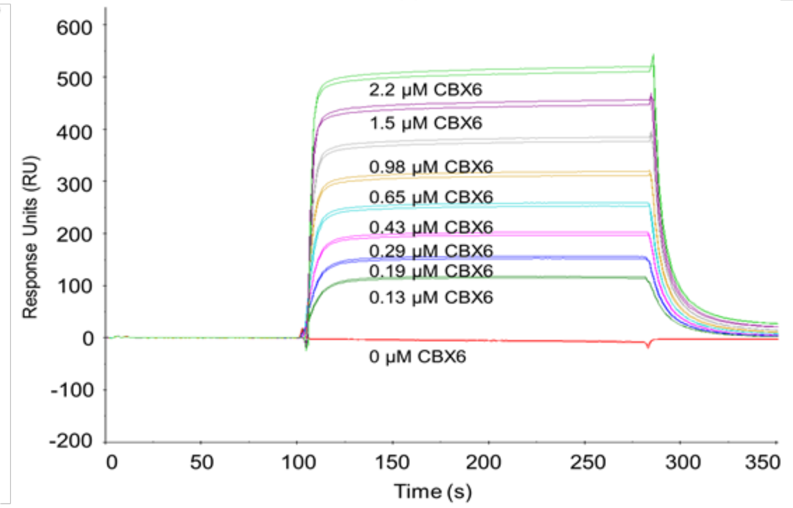
3.6.8 SPR data

A)



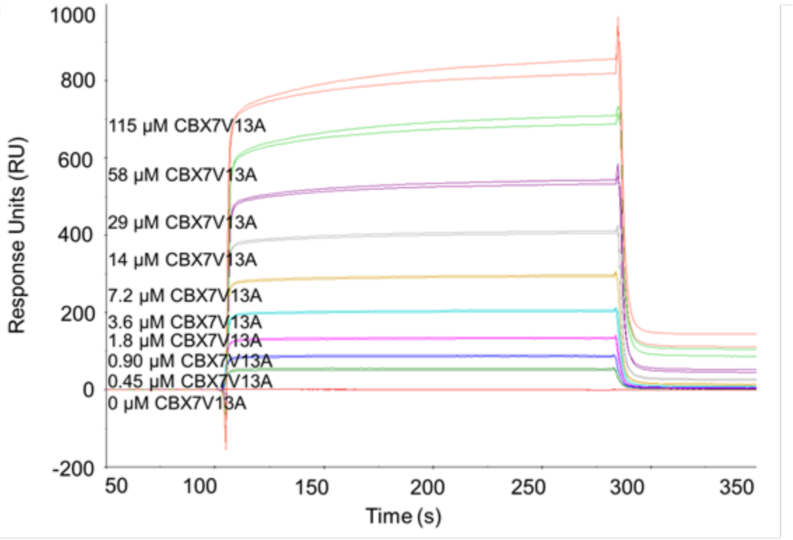
Response Units (RU)
1000
800
600
400
200
0
-200
-400
50
100
150
200
250
300
350
Time (s)
167 µM CBX7
114 µM CBX7
74.3 µM CBX7
49.5 µM CBX7
33.0 µM CBX7
22.0 µM CBX7
14.7 µM CBX7
0 µM CBX7

B)



Response Units (RU)
600
500
400
300
200
100
0
-100
-200
0
50
100
150
200
250
300
350
Time (s)
2.2 µM CBX6
1.5 µM CBX6
0.98 µM CBX6
0.65 µM CBX6
0.43 µM CBX6
0.29 µM CBX6
0.19 µM CBX6
0.13 µM CBX6
0 µM CBX6

C)



Response Units (RU)
1000
800
600
400
200
0
-200
50
100
150
200
250
300
350
Time (s)
115 µM CBX7V13A
58 µM CBX7V13A
29 µM CBX7V13A
14 µM CBX7V13A
7.2 µM CBX7V13A
3.6 µM CBX7V13A
1.8 µM CBX7V13A
0.90 µM CBX7V13A
0.45 µM CBX7V13A
0 µM CBX7V13A

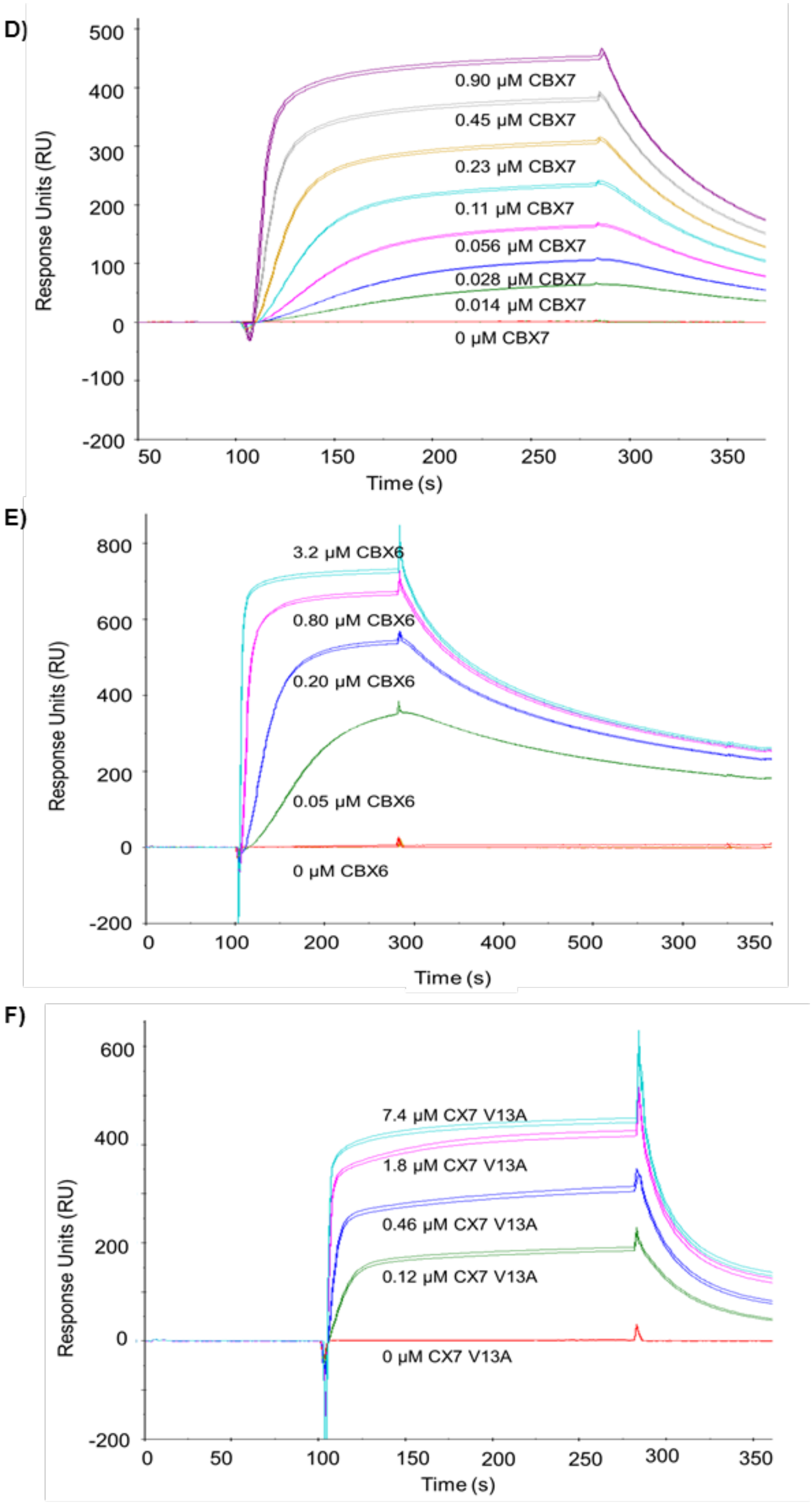


Figure 3.17. SPR sensorgrams for SA chip functionalized with **3.4** and **3.5** studied against CBX7, CBX6 and CBX7-V13A. All sensograms were adjusted to baseline = 0 RU. A) **3.5** on SA chip with CBX7, B) **3.5** on SA chip with CBX6, C) **3.5** on SA chip with CBX7-V13A, D) **3.4** on SA chip with CBX7, E) **3.4** on SA chip with CBX6, F) **3.4** on SA chip with CBX7-V13A.

3.6.9 MD methods and data

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

Docking of compounds **2.7** and **3.3** was done by superposition of the ligand backbone coordinates over those of the crystal structure native peptide H3K27 in CBX6 (See Figure 3.6). Successful equilibration of the complex was achieved in this scheme, and trajectories of CBX6 and 7, both with compounds **2.7** and **3.3**, were run for 50 ns each with a 2 fs time step.

Compound **3.3** valine was stable in the (−2) binding pocket of both CBX6 and CBX7, despite the bulkier valine residue at the edge of the pocket in the latter. The zoomed-in region presented in Figure 3.18 indicates the stability of **3.3** valine in the CBX7 binding pocket, at the 25 ns mark.

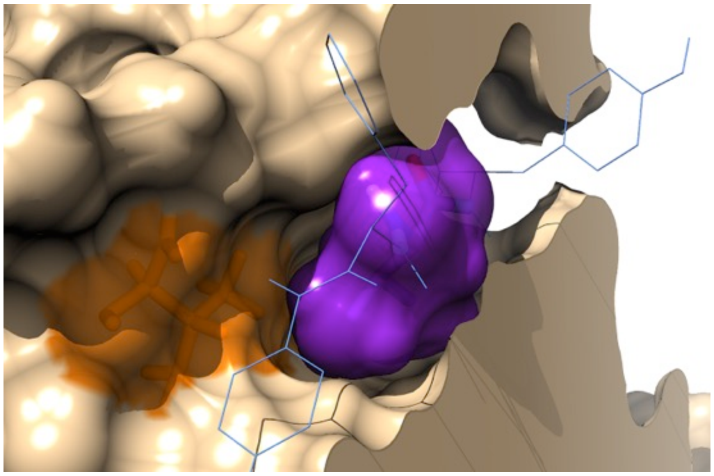


Figure 3.18. Clipped view of CBX7-Compound **3.3** interaction after equilibration. Following several nanoseconds of simulation, **3.3** valine still sits in the (−2) hydrophobic pocket. Clasp residues have been clipped out of the image and a surface area has been applied to the guest residue to show the fit in the bottom of the pocket. The same colour scheme as Figure 3.6 applies.

Chapter 4. Polycomb inhibitors active against both CBX6 and CBX8

Contributions

I conceived of and carried out the design, synthesis, purification and characterization of the peptides with synthetic assistance from Chelsea Wilson and Tyler Brown. I evaluated and interpreted all binding data. Tyler Brown, Kirsty Sinclair and I carried out the FP assays. Michael Gignac and Rebecca Hof carried out protein expression and purification. Modeling and MD simulations were conducted by James McFarlane. Design of several peptides was inspired by data from a DNA encoded library conceived of by Casey Krusemark at Purdue University. Cell-based studies were carried out by Emily Dykhuizen at Purdue University.

4.1. Foreword

Chapters 2 and 3 report on the synthesis and testing of dye-labeled inhibitors targeting the CBX polycomb paralog family. The compounds reported had limited use in cell-based experiments due to their poor cell permeability. The goals of this chapter were twofold: to create inhibitors with improved potency and selectivity towards CBX6 and CBX8, and to improve the cell permeability of these inhibitors.

4.2. Introduction

Our group reported the first inhibitors of any CBX protein or chromodomain, first targeting CBX7,[149-150] and more recently CBX6 (Chapter 3).[155] Subsequently, a cell permeable peptidic inhibitor UNC3866 was built on a similar scaffold.[151] In order to improve cell permeability, the quaternary amine trimethyllysine residue (Kme3) was replaced in UNC3866 with a tertiary amine diethyllysine (Ket2), providing a compound with similar potency and modest cell permeability. This result arises because quaternary ammonium ions like Kme3 side chains possess a permanent positive charge and because of this have limited ability to permeate cell membranes.[241-242] Comparatively, tertiary amines like the side chain of Ket2 have improved cell permeability due to their ability to be partially deprotonated at biologically relevant pH.

Our inhibitor of CBX6 (**3.2**) displayed selectivity over the other polycomb paralogs and we hypothesized that the valine residue in the (–2) position of the ligand was giving rise to the selectivity.[155] Despite the use of this compound in biochemical assays, the high molecular weight, the inclusion of a quaternary trimethyllysine residue, and the presence of a dye label limited its use in cellular studies due to poor cell permeability.

Developing potent and cellular active chromodomain inhibitors is challenging and there are limited examples that have achieved this goal. Of the 9 reported chromodomain inhibitors (targeting CBX6, CBX7 and CDYL1/2),[149-156, 238] there are 2 examples of cellular active compounds. The small molecule inhibitor MS37452 (K_d 30 μM) was shown to be active in cells but required compound treatment of >250 μM to observe decreased amounts of CBX7 on a known target locus.[153] The aforementioned UNC3866, with approximately 5% cell permeability, is the only example of a cellular active peptidic chromodomain inhibitor.[151] The other peptidic inhibitors of CBX6,[155] CBX4/CBX7,[149-150] and CDYL1/2[156] were not shown to be active in cells. Creating cellular active inhibitors that can serve as chemical probes for biological studies has therefore become a higher priority than achieving perfect paralog specificity.

The biological role of CBX8 and the mechanism in which it modulates transcription are unclear. CBX8, similar to CBX6, binds with very weak affinity to the native methylated histones substrates *in vitro*.[41] Cellular studies have shown that CBX8 does bind H3K27me3 and this interaction regulates association to chromatin *in vivo*.[243-244] Recently, association of CBX8 to chromatin was discovered to be mediated by simultaneous interactions with both DNA and H3K27me3.[244] Progress is being made to understand the complicated mechanisms of CBX8 control over gene expression, but our current knowledge is limited.

CBX6 and CBX8 are both involved in regulating stem cell differentiation. The role of CBX6 in regulating stem cell differentiation is reviewed in Chapter 3. CBX8 plays a critical role in dictating cell fate decisions and activation of differentiation genes.[245] CBX8 acts as part of the PRC1 complex and is upregulated during ESC differentiation. CBX7 in ESCs acts to maintain pluripotency and is downregulated upon increased expression of CBX8 during differentiation.[157, 178, 245] CBX8 and the other CBX polycomb paralogs are undoubtedly essential regulators of ESC differentiation. How these proteins are recruited, and why one CBX over another is utilized in PRC1, are among the many questions still unanswered.

The oncogenic role of CBX8 has been demonstrated in several malignancies including hepatocellular carcinoma (HCC),[246-247] acute myeloid leukemia (AML),[248] breast cancer[249] and esophageal carcinomas.[250] Additionally, CBX8 is overexpressed in bladder cancer,[251] gestational and non-gestational choriocarcinomas,[252] and glioblastoma multiforme.[253] The oncogenic role of CBX8 in HCC was of particular interest due to the previously described role of CBX6 in promoting aggressive forms of HCC (reviewed in Chapter 3). Increased expression of CBX8 in HCC is correlated to poor patient outcomes.[247] In vivo studies have shown CBX8 to promote the migratory, invasive and metastatic ability of HCC cells.[247] CBX8 promotes stem-cell qualities in HCC cells and this gives rise to a more aggressive form of the disease.[247]

The impact of the chemical inhibition of CBX8's chromodomain is unknown, because CBX8 inhibitors haven't been reported. Chemical inhibitors and further biological studies are needed to explore CBX8's potential as a therapeutic target. There are no potent and selective inhibitors of CBX8. Several of the inhibitors targeting CBX6 and CBX4/7 display weaker binding to CBX8 (K_d values ≥ 1.2 μM). Cell-permeable inhibitors would aid in assessing the specific functions of CBX8 and further our understanding of CBX8-driven cancers.

We sought to further explore the structure-activity relationships (SAR) of our first-generation CBX6 inhibitor as well as design a second-generation of CBX inhibitors with improved cell permeability. As we did so, we arrived at a novel class of dual-active CBX6- and CBX8-inhibiting ligands and identified the structural features that give rise to this dual selectivity. We provide preliminary biological data that demonstrate that the inhibitors have functional effects in multiple cell lines.

4.3. Results

We focused our efforts on modifications to compounds reported in Chapter 2 and 3 and chose to screen each compound against CBX1/6/7/8. This limited panel was chosen in order to most efficiently obtain data on CBX6- and CBX8-selectivity over CBX7 (typically the CBX that binds ligands most strongly) and CBX1 (as a representative of the HP1 class of chromodomains). Our previous report on CBX6 inhibition included the promiscuous inhibitor **2.4** and CBX6-selective inhibitor **3.2** that is 7–98-fold selective for CBX6 over CBX1, 7 and 8 (Figure 4.1).[155] As a first step towards making lower molecular weight compounds, we first investigated **2.5**, which is an analog of **2.4** that lacks the C-terminal Glu residue and dye-label (Figure 4.1, Table 4.1). **2.5** is 10-fold selective for CBX7 over CBX1 and 6- and 29-fold selective over CBX6 and 8 respectively.

To improve the cell-permeability and drug-like properties of our inhibitors we first focused on understanding the roles of the dye and the (–2) substituent in CBX6/8 selectivity. Substituting the alanine in compound **2.7** to a valine (**3.3**) abolished binding to CBX1 and produced an >100-fold decrease in binding to CBX7. Removal of the FITC-labeled lysine on the scaffold containing the (–2) valine residue (**3.2** compared to **3.3**) resulted in a 21-, 40- and 3-fold decrease in binding to CBX6, 7 and 8 respectively. We knew from various data in Chapters 2 and 3 that the removal of the FITC dye label makes compounds significantly less potent. From this data we can also see that the FITC removal affects CBX6 and CBX8 binding differently. Although **3.3** is still selective for CBX6 over CBX7, it is no longer selective for CBX6 over CBX8. Compound **3.3** is equipotent against CBX6/8 (IC_{50} values of ~17 μM) and is 12-fold selective over CBX7. Compound **4.1**, in which an isoleucine residue is present at the (−2) position, displayed no measurable binding to CBX1 or CBX7 (> 500 μM). Compound **4.1** displayed similar potency for CBX6 and CBX8 (IC_{50} values of 9 and 14 μM).

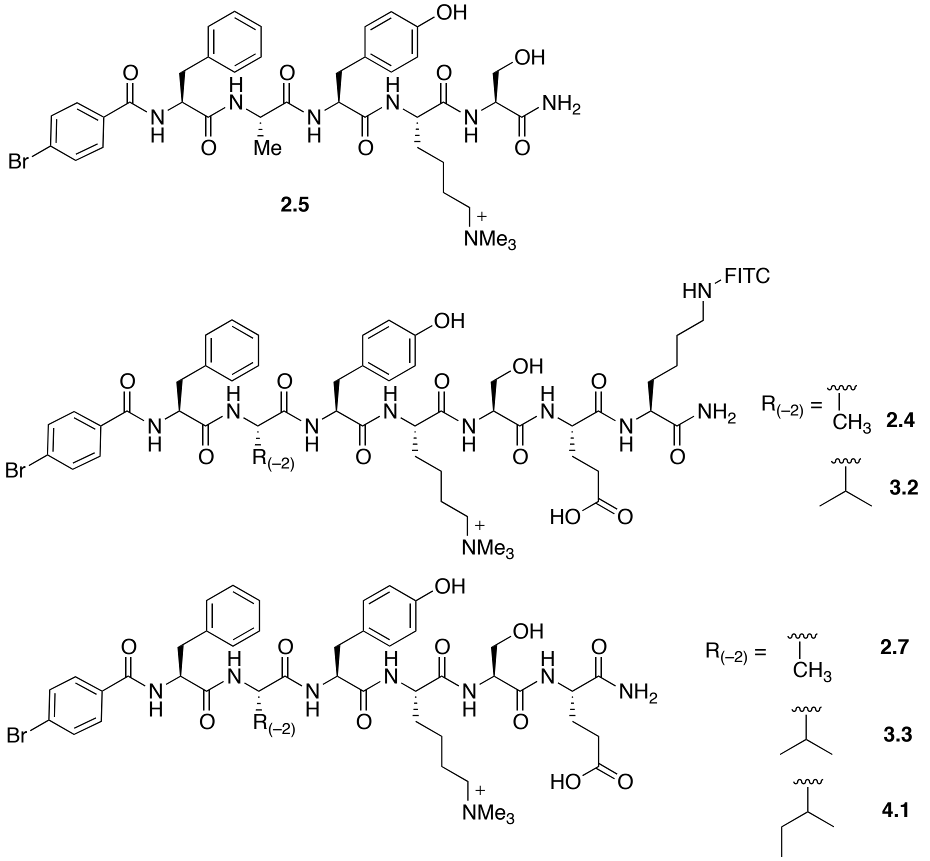
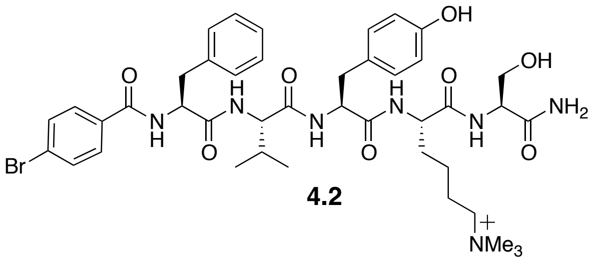


Figure 4.1. Chemical structures of compounds **2.4, 2.5, 2.7**, **3.2**, **3.3** and **4.1**.

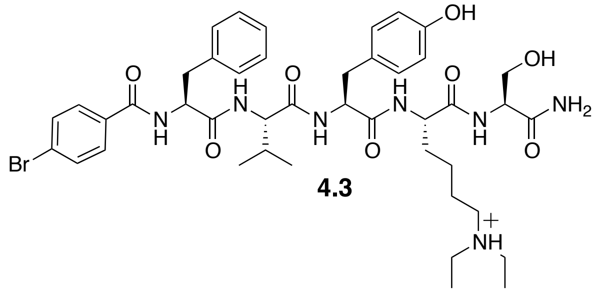
Table 4.1. IC_{50} values for compounds **2.5**, **2.7**, **3.3** and **4.1** (μM). All compounds were tested in triplicates in a competitive FP assay with **2.4** as the dye-labeled probe. Raw data and asymmetric 95% confidence intervals are reported in the supplementary Figure 4.24-Figure 4.26.

Compound	CBX1	CBX6	CBX7	CBX8
2.5	14	8	1.4	41
2.7	N.T	5	1.8	28
3.3	N.T	17	210	18
4.1	> 500	9	> 500	14

We focused next on optimization of compound **3.3** by removing the quaternary ammonium ion (Kme3) and the Glu residue, which we hypothesized were responsible for poor cell permeability. We removed the C-terminal glutamic acid residue to reduce overall charge, leading to compound **4.2**, and made an analogous compound **4.3** that swapped trimethyllysine (Kme3) for diethyllysine (Ket2) (Figure 4.2). Removal of the C-terminal Glu gave decreased binding to CBX7 and no significant change in potency was observed for CBX6/8. Replacement of Kme3 with Ket2 gave a 7-fold increase in binding to CBX6, and no change in binding to CBX8. Compound **4.3** is 5.6-fold selective for CBX6 over CBX8 and displays no binding to CBX1/7 (Table 4.2).



4.2



4.3

Figure 4.2. Structure of compounds **4.2** and Ket2 analog **4.3**.

Table 4.2. Binding affinity for compounds **4.2** and **4.3** (IC_{50} values in μM). Competitive FP assays performed in triplicate. Raw data and asymmetric 95% confidence intervals are reported in supplementary Figure 4.27, Figure 4.28.

Compound	CBX1	CBX6	CBX7	CBX8
4.2	> 500	24	> 250	18
4.3	> 500	3.2	> 500	18

We next sought to substitute the (–1) residue with a more hydrophobic leucine residue (**4.3** compared to **4.4**, Table 4.3). **4.4** showed a slight decrease in potency for CBX6 and a 4-fold increase in binding to CBX8 (Table 4.3). Neither compound showed activity against CBX1 or CBX7.

We next explored the roles of N-terminal capping (N-cap) residues on CBX6/8 selectivity, following leads from our collaborative studies of CBX8-selective inhibitors generated by the Krusemark lab.[238] On the scaffold with leucine at the (−1) position (**4.4**), we then substituted the N-cap residue with two distinct heterocycles. Substitution of the *p*-bromobenzamide group with 1*H*-pyrazole-3-carboxamide (**4.5**) resulted in a 11-fold decrease in binding to CBX6 and a 15-fold decrease in binding to CBX8 (Table 4.3). Substitution with the 5-methylisoxazole-3-carboxamide (**4.6**) gave a similar binding profile as **4.4** with CBX6/8 and showed activity against CBX7 with an IC_{50} value of 105 μM (Table 4.3). Comparison of a leucine or tyrosine residue at the (–1) position in the scaffold with the N-cap 5-methylisoxazole (**4.6** compared to **4.7**), **4.7** displayed a slight decrease in binding to CBX6, weaker binding to CBX7 and no change in binding to CBX8.

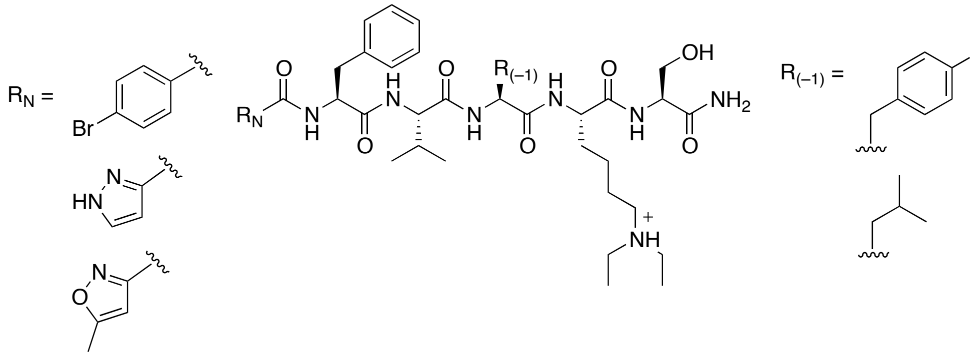
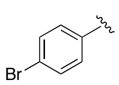
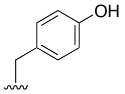
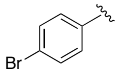
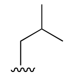
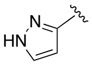
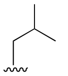
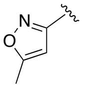
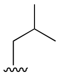
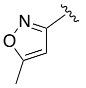
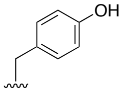


Figure 4.3. Chemical structures of compounds **4.3**-**4.7** with (–1) and N-cap substitutions. See Table 4.3 for compound numbers.

Table 4.3. Binding affinity for compounds **4.3**, **4.4-4.7** (IC_{50} values in μM). Competitive FP assays performed in triplicate. Raw data and asymmetric 95% confidence intervals are reported in supplementary Figure 4.28-Figure 4.32.

Compound	R_N	$R_{(-1)}$	CBX1	CBX6	CBX7	CBX8
4.3			> 500	3.2	> 500	18
4.4			> 300	6.6	> 250	4.5
4.5			> 300	71	162	68
4.6			> 300	12	105	7
4.7			> 500	16	>300	6

To explore the SAR in the beta-groove we combined the N-cap heterocyclic substitutions with aromatic and hydrophobic rings in the (−3) position (Figure 4.4). Replacement of the phenylalanine residue in **4.5** with a cyclohexyl group (**4.8**) showed a 2-3-fold increase in binding to CBX6, 7 and 8. The same substitution in the 5-methylisoxazole-3-carboxamide containing **4.6** to give compound **4.9**, did not significantly change binding to CBX6 and 8 but did show a decrease in binding to CBX7 (Table 4.4). Replacement of the cyclohexyl moiety with a 4-methoxyphenyl group (**4.10**) resulted in no significant change in binding to CBX6/7 and a 2-fold increase in binding to CBX8. Within the same scaffold, the addition of a pyridine-3-yl group at the (–1) position decreased binding to CBX6/7/8 by a factor of 3, 2 and 7 respectively (**4.11**, Table 4.4). The 4-methoxyphenyl in the (–3) position combined with the (–4) 5-methylisoxazole-3-carboxamide (**4.10**) provided the greatest selectivity for CBX8 and is 4.6-fold and 67-fold selective over CBX6 and CBX7.

Substitution of the isoxazole N-cap gave large improvements in potency. We synthesized compounds **4.12** and **4.13**, containing a propyl and phenyl substitution in the 5-position of the isoxazole ring (Figure 4.4, Table 4.4). The 5-propylisoxazole N-cap combined with the (–3) cyclohexyl moiety (**4.12)** displayed potent binding to CBX6 with an IC_{50} value of 2.4 μM and is 17- and 3-fold selectivity over CBX7 and CBX8. The addition of a 5-phenylisoxazole N-cap substitution with a (–3) phenyl group provided the most potent compound of the series (**4.13**, Table 4.4**). 4.13** is equipotent for CBX6 and 8 with IC_{50} values of 180 nM and is 22-fold selective over CBX7. Substitution of the methyl group in the 5-methylisoxazole N-cap with larger alkyl chains and aromatic groups improves potency due to increased engagement with the extended beta-groove of CBX6/7/8.

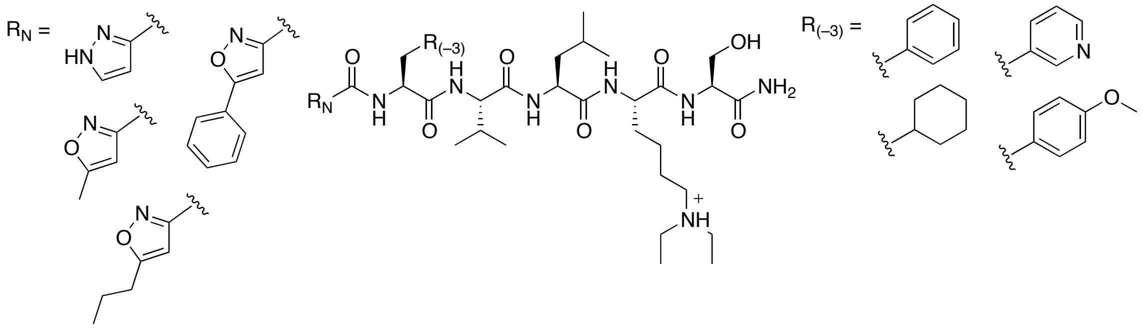
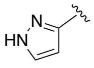
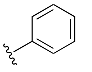
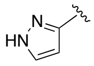
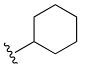
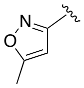
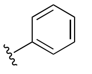
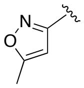
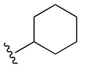
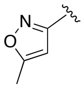
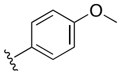
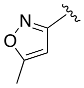
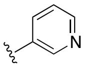
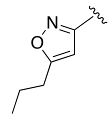
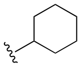
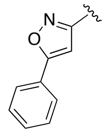
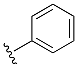


Figure 4.4. Overall chemical structures of compounds **4.5**, **4.6** and **4.8**-**4.13** with (–3) and N-cap substitutions. See Table 4.4 for compound numbers.

Table 4.4. Binding affinity for compounds **4.5**, **4.6** and **4.8-4.13** (IC_{50} values in μM). Competitive FP assays performed in triplicate. Raw data and asymmetric 95% CI are reported in supplementary Figure 4.30-Figure 4.31, and Figure 4.33-Figure 4.38.

Compound	**R_N**	**$R_{(-3)}$**	**CBX1**	**CBX6**	**CBX7**	**CBX8**
4.5			> 300	71	162	68
4.8			> 300	32	104	25
4.6			> 300	12	105	7
4.9			> 300	8	171	6
4.10			> 500	12	175	2.6
4.11			> 500	33	400	17

4.12			> 300	2.4	40	7.4
4.13			> 500	0.18	4	0.18

To visualize the binding interactions of our ligands, we carried out MD simulations with compound **4.12** (Figure 4.5). Snapshots of the protein-ligand configurations show the peptide binding in the expected conformation. We observe the Ket2 group engaging the aromatic cage, and backbone amides of the ligand hydrogen bonding with residues in CBX6 and CBX8. The 5-propylisoxazole N-cap engages the extended beta-groove and this highlights the importance of this pocket for improved potency. These simulations don’t provide any direct insight into the observed selectivity of the ligand over other CBX proteins.

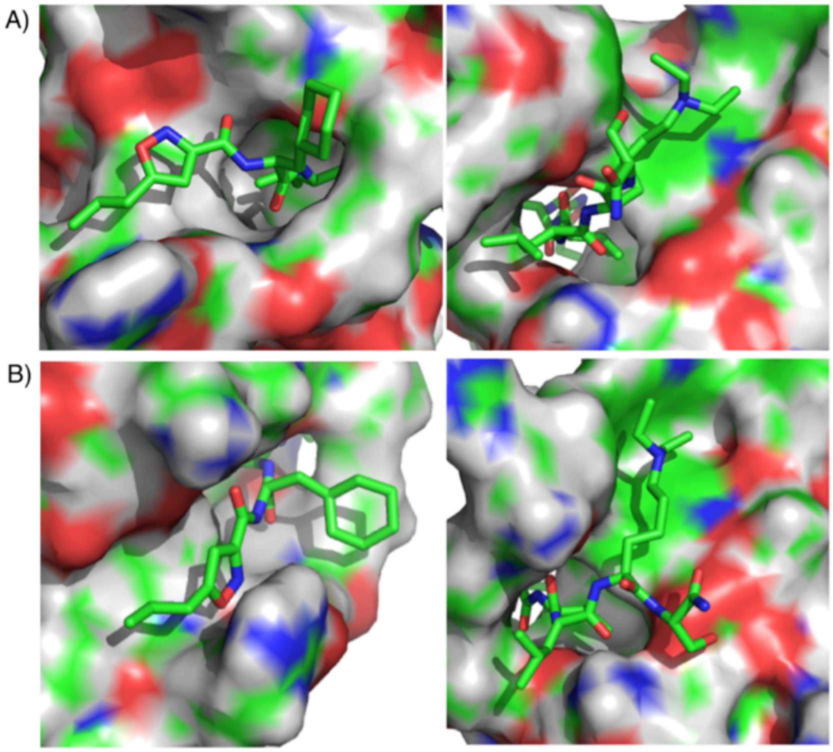


Figure 4.5. Snapshots of compound **4.12** from MD simulations with CBX6/8. A) CBX6, B) CBX8. Left panels are of the beta-groove pocket and right panels are of the aromatic cage. Snapshots are of the most commonly observed structures during each simulation.

Larger hydrophobic substitutions in the (–2) position of the ligand allow for CBX8 selective inhibitors. With new peptidic scaffolds in hand we returned to explore alkyl substitutions at the (−2) position. We synthesized analogs of compound **4.6** containing a cyclopropyl, cyclopentyl and sec-butyl group at the (−2) position (Figure 4.6). Compound **4.14** containing a cyclopropyl group was nearly equipotent for CBX6/8 and displayed between 6- and 7-fold selectivity over CBX7 (Table 4.5). Compared to the isopropyl-containing compound (**4.6**), the cyclopropyl substitution increased binding to CBX6/7/8 by a factor of 3, 5 and 2. Addition of a cyclopentyl group (**4.15)** dramatically reduced binding to CBX6/7 and decreased binding to CBX8 by a factor of 6 (IC_{50} of 18 μM, Table 4.5). **4.15** is 10- and 20-fold selective for CBX8 over CBX6 and 7 respectively. Addition of a sec-butyl group (**4.16**) further decreased binding to CBX6 and displayed no binding to CBX7 at the concentrations tested. The larger substitution in **4.16,** decreased binding to CBX8 by a factor of 3 to gave an IC_{50} value of 53 μM. All CBX polycomb paralogs tested showed decreased binding with increasing size of the substituent at the (–2) position.

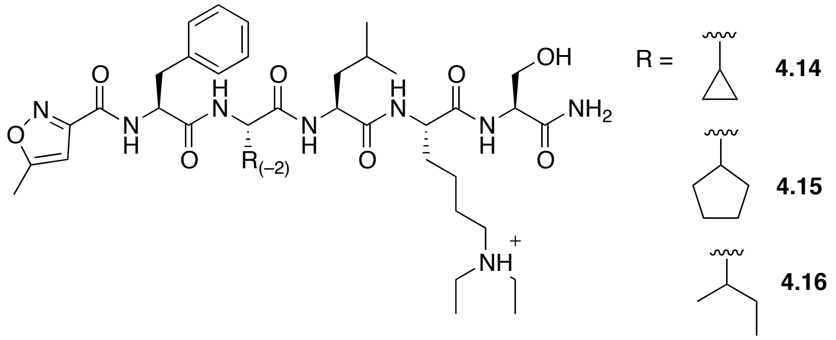
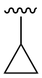
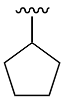
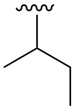


Figure 4.6. Chemical structure of peptidic ligands **4.14-4.16** with substitutions at the (–2) position.

Table 4.5. Binding affinity for compounds **4.14-4.16**. (IC_{50} values in μM). Competitive FP assays performed in triplicate. Raw data and asymmetric 95 % CI are reported in supplementary Figure 4.39-Figure 4.41.

Compound	**$R_{(-2)}$**	**CBX1**	**CBX6**	**CBX7**	**CBX8**
4.14		> 500	3.6	22	3.0
4.15		> 500	178	372	18
4.16		> 500	250	> 500	53

CBX8 was best able to tolerate larger substitutions directed into the (–2) pocket of the protein. We had previously reported the larger size of the (–2) pocket within CBX6 and CBX8 compared to other members of the polyomb paralog CBX family. The SAR reported in this chapter demonstrate the (–2) pocket of CBX8 to be capable of engaging and binding to ligands with bulkier (–2) substitutions. The (–2) pocket is nearly identical in CBX6/8 with the exception of an isoleucine residue in CBX6 that is replaced with a leucine in CBX8. Compound **4.15** containing the cyclopentyl group at the (–2) position displays promising selectivity for CBX8. We predict that combining the (–2) cyclopentyl group with other substitutions could improve the potency and selectivity of **4.15** for CBX8.

4.4. Cell-based studies

CBX6 is critical for regulating stem cell differentiation, so we used an embryonic stem cell (ESC) model to determine if our inhibitors are entering cells and engaging their intended target. RNA knockdown of CBX6 in ESCs results in downregulation of key pluripotent genes including *Rex1*, *Klf4*, *Nanog* and *Esrrb*.[163] We sought to identify whether treatment of ESCs with our inhibitors for CBX6/8 could alter the expression of the same pluripotency genes regulated by CBX6. ESCs were treated with compounds **4.3** and **4.6** for 48 hours and changes in gene expression were measured using quantitative reverse transcriptase polymerase chain reaction (qRT-PCR). We predicted that inhibition of CBX6 would produce similar changes in gene expression as was reported from the knockdown experiments. Surprisingly, we observed a 6 to 8-fold increase in expression of *Essrb* and *Klf4* following treatment with **4.3** and **4.6** (Figure 4.7A). The ESC literature on which we based this study[163] did not include a knockdown of CBX8, so it is unclear what effect CBX8 inhibition would be expected to have on these ESC genes. We conclude strongly that the compounds are entering cells. They are disrupting gene regulation pathways relevant to CBX functions, which suggests but does not demonstrate that they engage any particular target. Whether they engage CBX6 and/or CBX8 in those cells is unclear from this data.

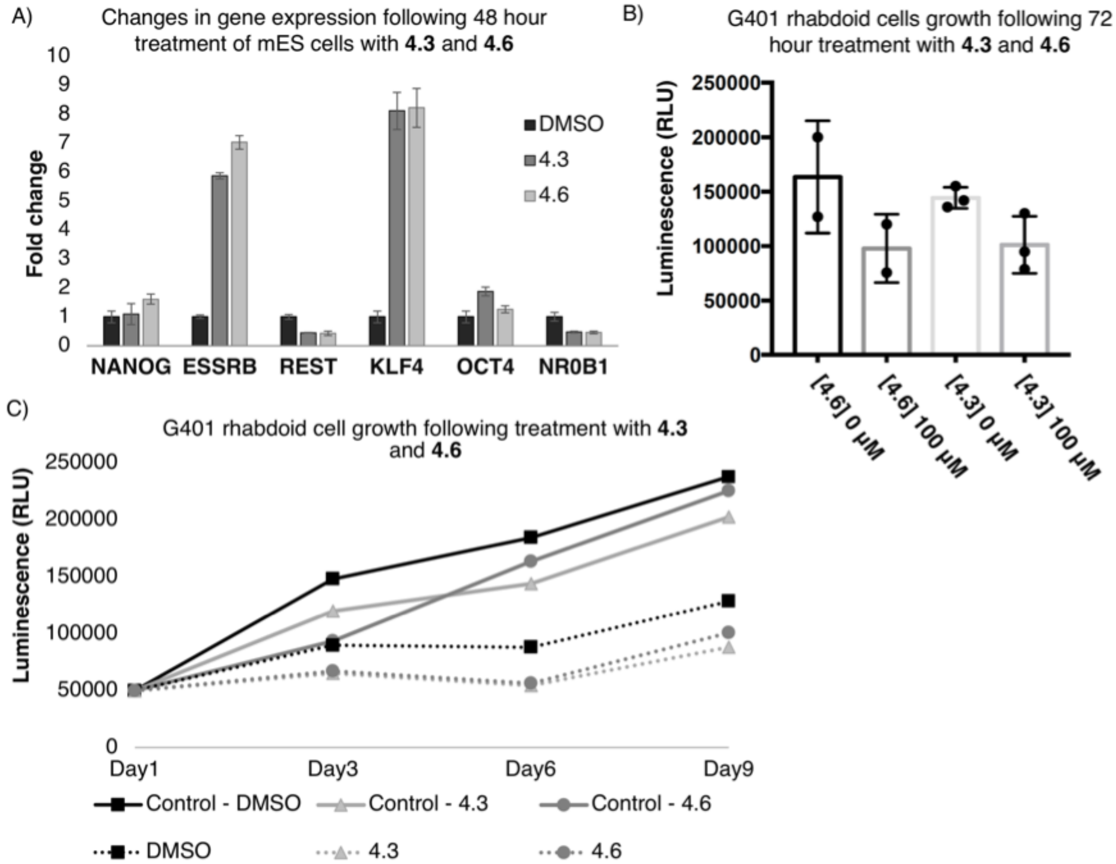


Figure 4.7. Inhibitors are cell permeable and able to engage CBX proteins. A) ESCs treated with **4.3** and **4.6** alter expression of pluripotency genes. Cells were treated with inhibitors (100 μM) for 48 hours and quantification of gene expression was measured using qRT-PCR. B) and C) G401 rhabdoid cells are affected by treatment of **4.3** and **4.6** (100 μM) and show decreased cell proliferation. Cell counts were analyzed using CellTitre-Glo following 72 hour treatment for data shown in (B) and measured following 3, 6 and 9 days for data shown in (C). Solid lines represent WT controls and dotted lines represent CBX6 KO cell lines.

CBX6 is a potential target in rhabdoid tumours.[254-255] Rhabdoid tumors are a highly aggressive pediatric malignancy that are driven by misregulated epigenetic modifications.[254] To investigate the effects of CBX inhibition in rhabdoid tumors, we measured cell proliferation of G401 rhabdoid cells following treatment with **4.3** and **4.6**. Wild-type G401 cells were treated with the inhibitors for 72 hours and cell counts were measured using CellTitre-Glo (Figure 4.7B). We next attempted treatment with inhibitors for a longer time period of up to 9 days. Cell counts were measured at day 3, 6 and 9 and cells were treated with 100 μM of compound every 3 days. Encouragingly, the growth of the rhabdoid cells was decreased in the CBX6 KO cells. Both inhibitors decreased cell proliferation in the wild-type G401 cells (Figure 4.7C). We did observe that the compounds also have an effect in the CBX6 KO cells (Figure 4.7C, dotted lines). Again, we conclude that the inhibitors are entering cells, but the exact CBX targets by which they produce their effects are not clear. These results could mean that the inhibitors are binding to other CBX proteins that contribute to the observed phenotype. The dual activity of **4.3** and **4.6** against CBX6 and CBX8 may implicate a role of CBX8 in rhabdoid cell proliferation.

4.5. Discussion

The (–2) position of the ligand is critical for selectivity within the polycomb paralog family. The ability of ligands with larger groups in the (–2) position to bind the CBX proteins depends on the peptide scaffold and neighbouring residues. Previous work (Chapter 3) demonstrated dye-labeled ligands with a valine residue in the (–2) position to be selective for CBX6. Removal of the dye from these scaffolds showed reduced selectivity for CBX6 over CBX8, however CBX6 was still better able to tolerate ligands containing bulkier groups within the (–2) pocket (compound **3.3**, **4.1**, Table 1). Larger groups in the (–2) position within the scaffolds reported in this chapter demonstrate selectivity for CBX8. The aromatic cage, hydrophobic clasp and (–2) pocket are tightly connected and residues in the ligand to the right and left of the Ket2 can change the shape and size of the (–2) pocket. Despite the highly similar sequence homology of the residues in the (–2) pocket in CBX2/6/8, our results show a clear difference in the size of this pocket depending on the chemical structure of the ligand and resulting protein-ligand interactions.

Aromatic and hydrophobic residues are well tolerated at the (–3) position of the ligand. The greatest selectivity for CBX8 within the (–3) series was achieved through the addition of a 4-methoxy phenyl moiety. Further exploration of substituted aromatic rings in the (–3) position of the ligand will aid in a better understanding of selectivity determinants.

The 5-phenylisoxazole N-cap moiety in compound **4.13** was the most potent dye-free CBX6/8 inhibitor developed to date. Replacement of the 5-methyl to the 5-phenyl group on the terminal isoxazole ring improved binding to CBX6 and 8 by a factor of 67 and 38. These results clearly demonstrate the potential for potent inhibitors to be developed through extension of the ligand to engage the proteins peptide binding beta-groove.

The cell-based studies reported demonstrate that our second generation inhibitors are able to enter the cells and engage CBX proteins. We do not yet know if the observed effects are due to engagement of CBX6 or CBX8. Next directions in cell-based studies are to use our compounds to figure out the function of CBX6 in cancer. We are currently further investigating treatment of rhabdoid tumors with our inhibitors alongside HepG2 cells as a negative control (no effect was observed in HepG2 cells following CBX6 KO).

4.6. Conclusion

CBX6 and CBX8 are exciting potential therapeutic targets for multiple aggressive malignancies including HCC. This chapter has reported the first CBX8 selective inhibitors and dual-active inhibitors for CBX6 and CBX8 that are active in cells. These are the first examples of inhibitors that bind specific CBX polycomb paralogs with no measurable activity against CBX7. The inhibitors and SAR described in this work will pave the way for future efforts in chromodomain inhibition.

CBX6 and CBX8 are poorly understood regulators of cell fate, gene expression and cancer. CBX6 is the most mysterious CBX protein and it is unknown if its actions are PRC1 dependent or are regulating transcription via alternate mechanisms. Cell permeable chemical probes will aid in unraveling the mysteries of CBX6 and elucidate the paralog-specific roles of both CBX6 and CBX8.

4.7. Experimental methods and data

4.7.1 General synthesis

All natural and unnatural amino acids and coupling agents were purchased from ChemImpex, Enamine and Sigma Aldrich. Acetylaldehyde and sodium cyanoborohydride were purchased from Sigma Aldrich and carboxylic acids used for N-terminal modifications were purchased from Enamine, ChemImpex and Sigma Aldrich. Fmoc-L-Lys(Me)$_3$-OH was purchased from GL Biochem (Shanghai).

4.7.2 General peptide synthesis methods

Peptides were either synthesized using an automated peptide synthesizer, or by manual peptide synthesis using Fmoc solid phase peptide synthesis protocols. Automated synthesis was achieved on rink amide resin (ChemImpex International) using CEM Liberty microwave synthesizer with standard Fmoc protocols using HBTU as the coupling agent. For manual peptide synthesis, Fmoc-protected rink amide resin (ChemImpex International) was swelled in CH_2Cl_2 for 20 minutes. Deprotection of Fmoc was done using 20% piperidine in DMF (10 mL, 5 minutes, 3 times) and washed with DMF (10 mL, 3 times following each incubation). Fmoc- and side chain protected amino acids as well as N-capping carboxylic acids (5 eq.) were dissolved in DMF (10 mL) with HBTU (5 eq.), and DIPEA (10 eq.), added to the Fmoc deprotected resin and/or Fmoc deprotected peptide on resin, bubbled with N_2 for one hour at room temperature. The resin was washed with DMF (10 mL, 3 times), and washed with CH_2Cl_2 (10 mL, 3 times) after the final coupling.

Peptide products were cleaved from resin with 10 mL of 95:2.5:2.5 TFA/H_2O/triisopropylsilane for 2.5 h. The solution was then concentrated in vacuo and added to cold diethyl ether to yield a crude yellow precipitate that was collected by centrifugation. Purifications by preparative HPLC were carried out using a Phenomenex Luna, 5 µm, C-18 column, 250 x 21.20 mm column. Peptides were characterized using

UPLC, HPLC, and ESI-MS and purity was determined to be >95% by analytical HPLC or analytical UPLC (exceptions are noted in the characterization data). Retention times reported arise from analytical traces done on either a Thermo Scientific C-18 column (4.6 mm x 250 mm, 5 μm), Waters UPLC-MS Acquity UPLC BEH C18 column (2.1 x 50 mm, 1.7 μM), or Shimadzu HPLC C-18 column (4.6 mm x 250 mm, 5 μm). Traces done on the Thermo or Shimadzu HPLC were at a flow rate of 1.5 mL/min, gradient running from 90:10 water (0.1 % TFA) and MeCN (0.1 % TFA) to 10:90 water (0.1 % TFA) and MeCN (0.1 % TFA) over 30-38 min. Traces done on the Waters UPLC-MS were at a flow rate of 0.5 mL/min with a gradient running from 90:10 water (0.1 % TFA) and MeCN (0.1 % TFA) to 10:90 water (0.1 % TFA) and MeCN (0.1 % TFA) over five minutes.

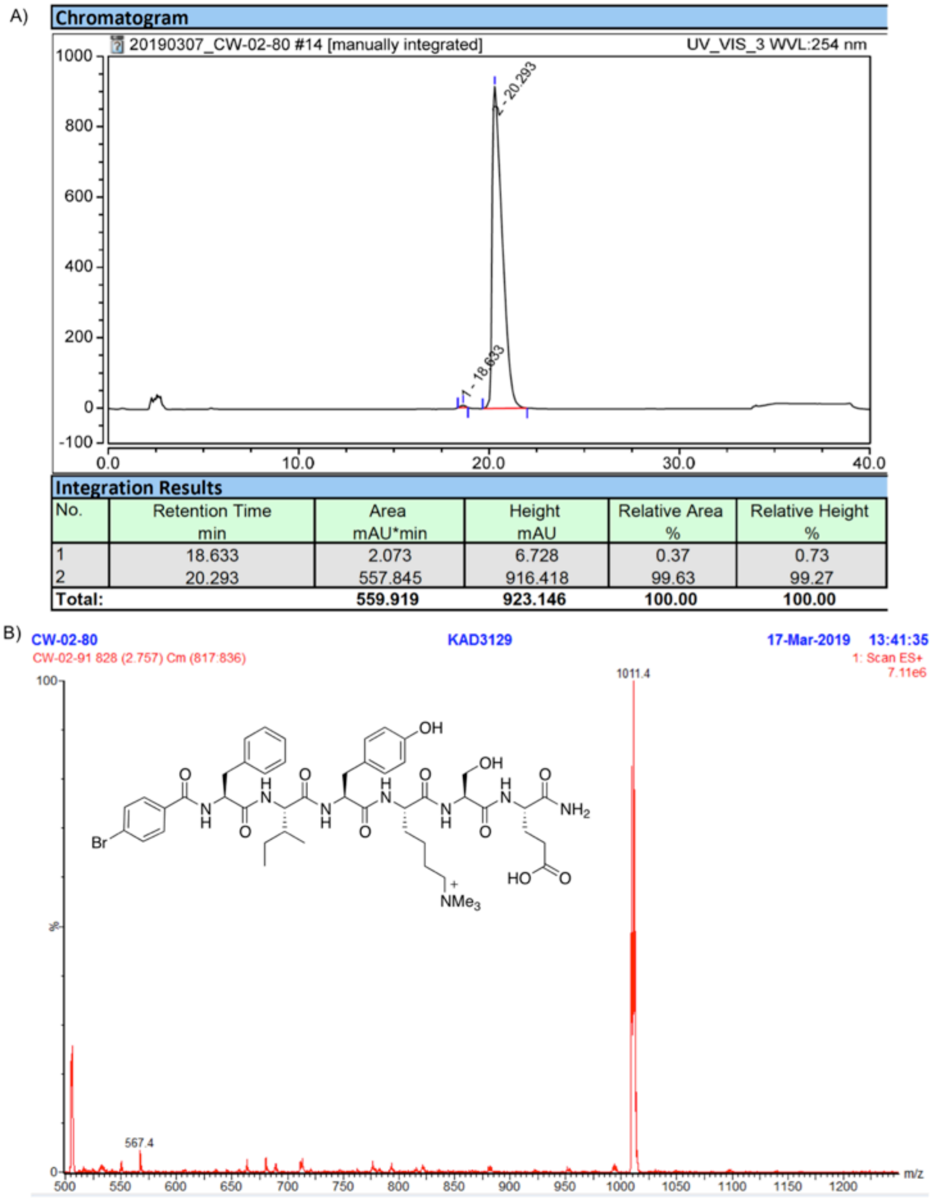
4.7.3 Synthesis of FITC labeled peptide probe for FP

For side-chain functionalization with fluorescein isothiocyanate (FITC), a side-chain (Mtt)-protected lysine was used. Synthetic methods are described in Chapter 2, Experimental methods and supplementary data section 2.8.2.

4.7.4 Synthesis of diethyl-lysine containing peptides

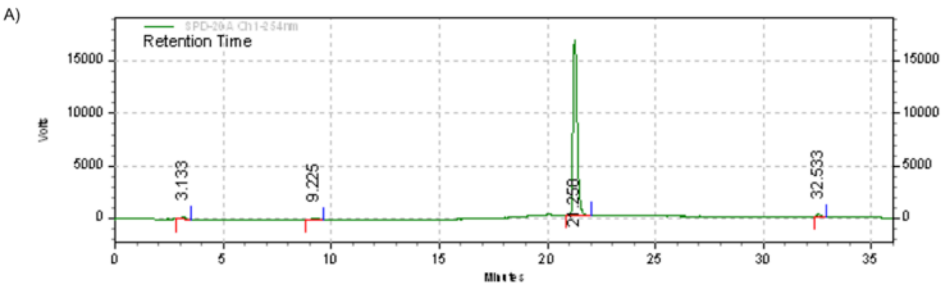
Side-chain (Mtt)-protected lysine was used in the synthesis protocols, and cleaved as described previously in Chapter 2, section 2.8.2 (treatment with 2:2:96 TFA/triisopropylsilane/CH_2Cl_2 for 20 minutes, 4 times). The resin was filtered and washed with CH_2Cl_2 (3 x 5 mL) and DMF (3 x 5 mL). A solution of acetylaldehyde (50-100 eq.), sodium cyanoborohydride (10 eq.), and 1 drop of glacial acetic acid in DMF (10 mL) was added to the peptide on resin. The solution was filtered and the resin washed with DMF (3 x 5 mL) and CH_2Cl_2 (3 x 5 mL) and air-dried. Crude peptides were dissolved in a mixture of H_2O and MeCN or DMSO, depending on solubility. Peptide cleavage from resin and purification are the same as reported above in section 4.7.3. Diethyl peptides with poor solubility were purified from DMSO stocks with a gradient running from 60:40 water (0.1 % TFA) and MeCN (0.1 % TFA) to 10:90 water (0.1 % TFA) and MeCN (0.1 % TFA) over 30-38 min.

4.7.5 Characterization data



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	18.633	2.073	6.728	0.37	0.73
2	20.293	557.845	916.418	99.63	99.27
Total:		**559.919**	**923.146**	**100.00**	**100.00**

Figure 4.8. Characterization data for compound **4.1**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. For $C_{48}H_{66}BrN_8O_{11}^{+}$: 1011.41; found: 1011.4.



SPD-20A
Ch1-254nm
Results

Retention Time	Area	Area %	Height	Height %
3.133	2620	1.12	239	1.39
9.225	2381	1.02	125	0.73
21.258	225906	96.61	16616	96.60
32.533	2935	1.26	220	1.28
Totals	233842	100.00	17200	100.00

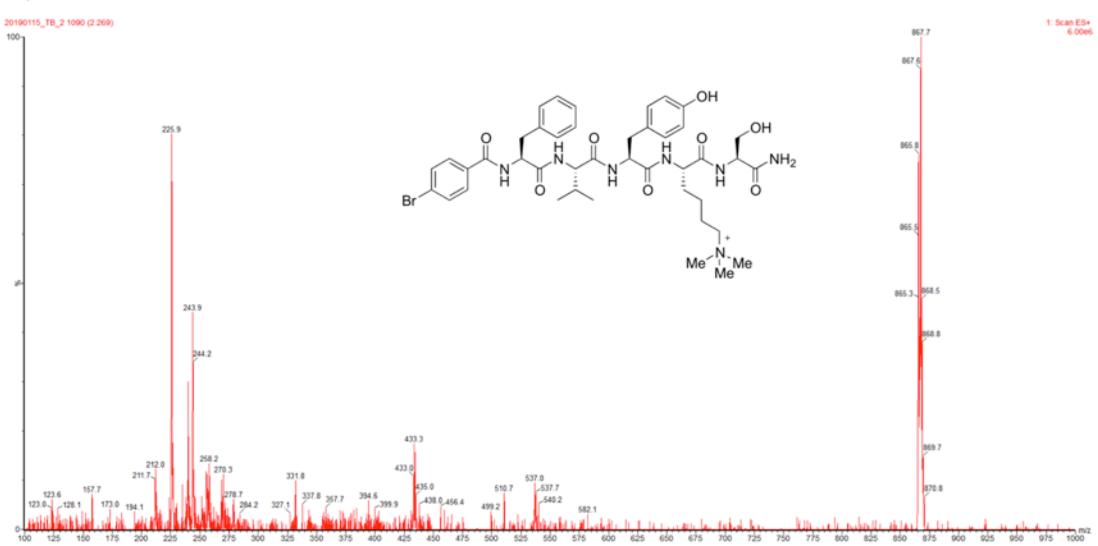


Figure 4.9. Characterization data for compound **4.2**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for $C_{42}H_{57}BrN_7O_8^+$: 867.85; found: 867.7.

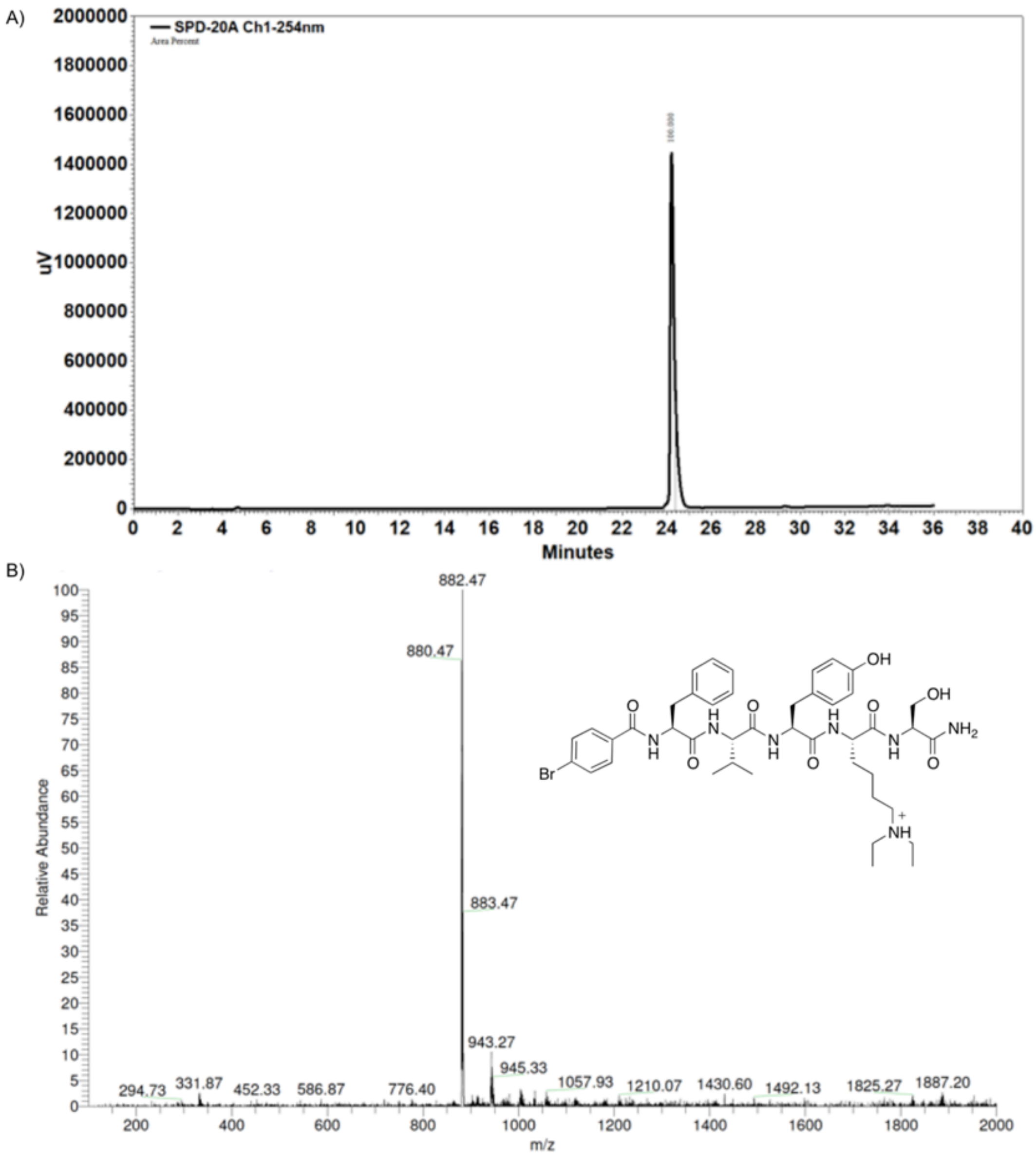


Figure 4.10. Characterization data for compound **4.3**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{43}H_{58}BrN_7O_8^+$: 880.36; found: 880.47.

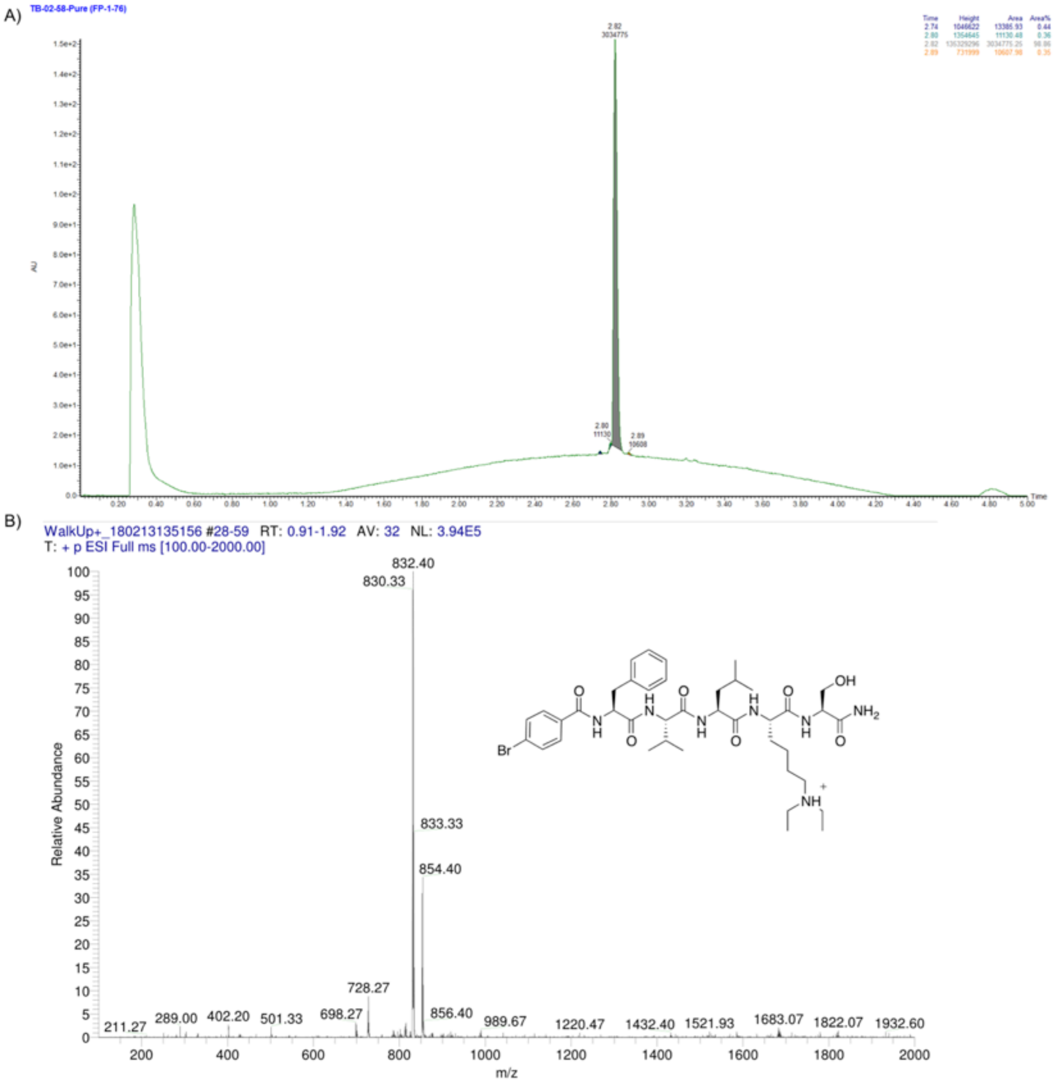
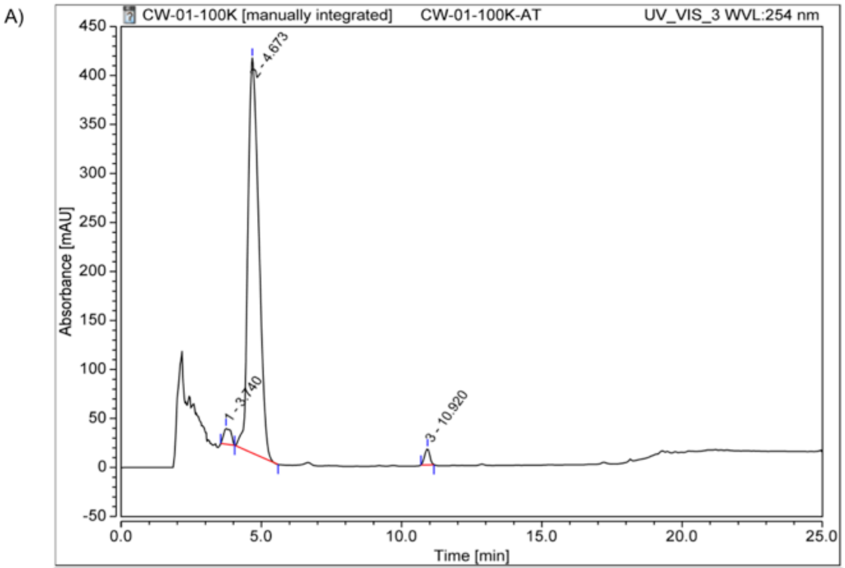


Figure 4.11. Characterization data for compound **4.4**. A) Analytical UPLC trace. Product peak is at 2.82 minutes with area of 98 %. B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{40}H_{61}BrN_7O_7^+$: 830.38; found: 830.33.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	3.740	5.061	16.235	2.69	3.73
2	4.673	179.676	403.079	95.48	92.51
3	10.920	3.445	16.411	1.83	3.77

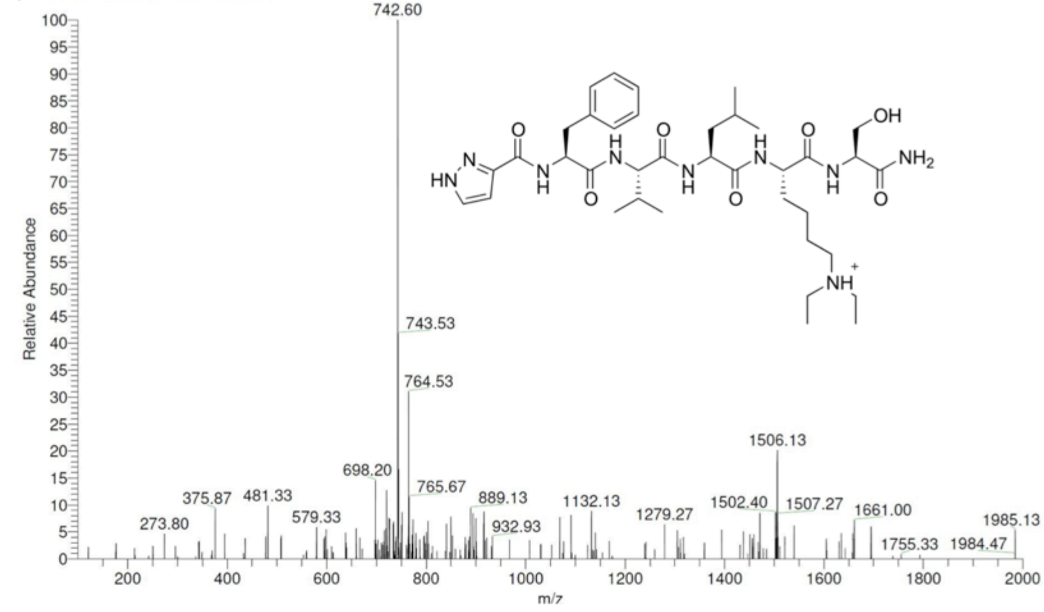
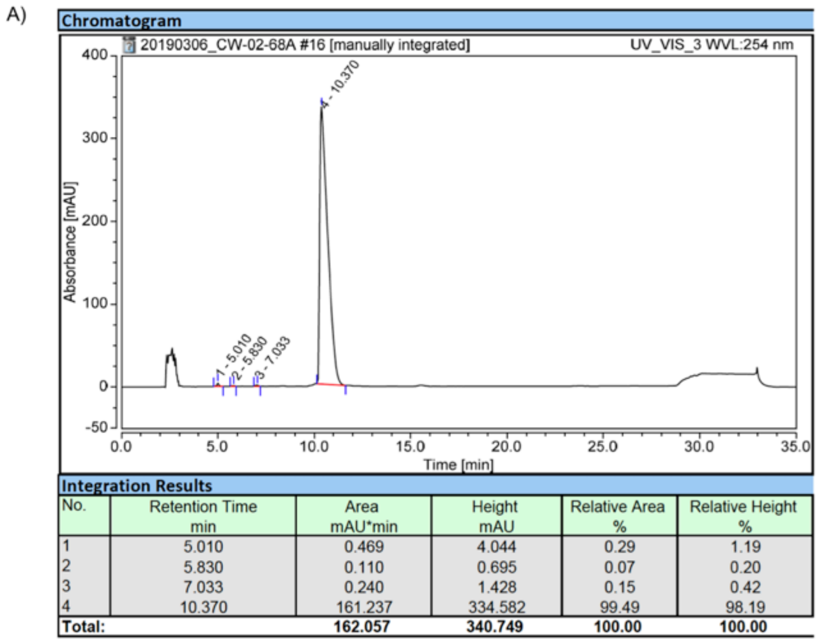


Figure 4.12. Characterization data for compound **4.5**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{37}H_{60}N_9O_7^+$: 742.46; found: 742.60.



Integration Results

No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	5.010	0.469	4.044	0.29	1.19
2	5.830	0.110	0.695	0.07	0.20
3	7.033	0.240	1.428	0.15	0.42
4	10.370	161.237	334.582	99.49	98.19
Total:		**162.057**	**340.749**	**100.00**	**100.00**

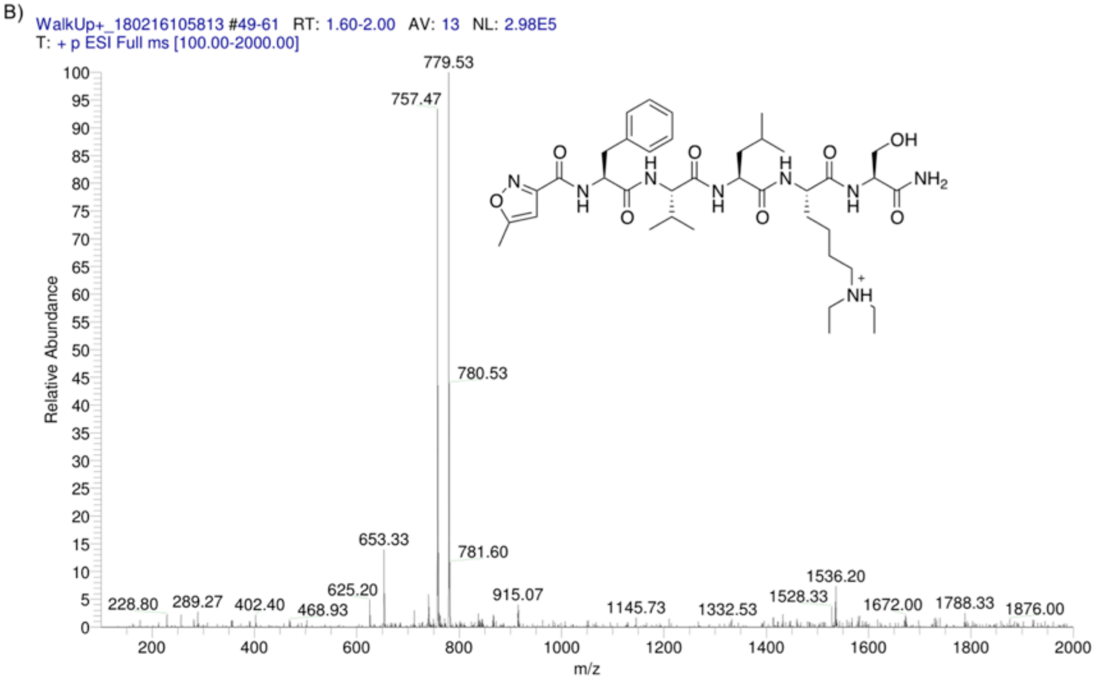
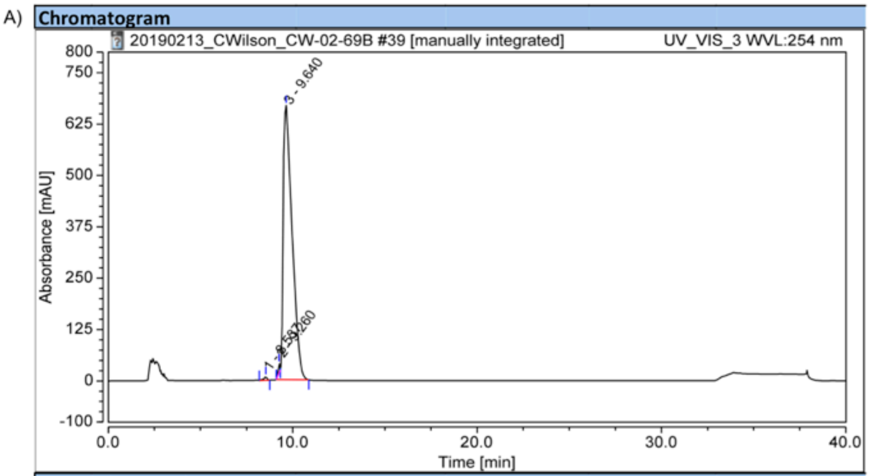


Figure 4.13. Characterization data for compound **4.6**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{38}H_{61}BrN_8O_8^+$: 757.46 ; found: 757.47



Integration Results

No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	8.537	1.663	8.473	0.44	1.21
2	9.260	2.225	22.381	0.58	3.21
3	9.640	377.021	667.116	98.98	95.58
Total:		**380.909**	**697.970**	**100.00**	**100.00**

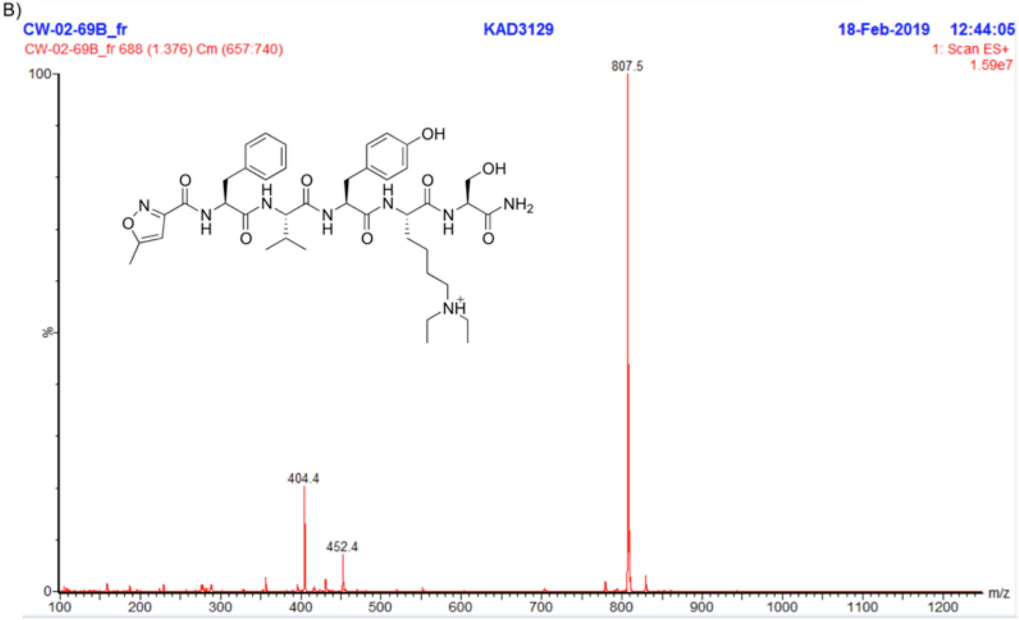
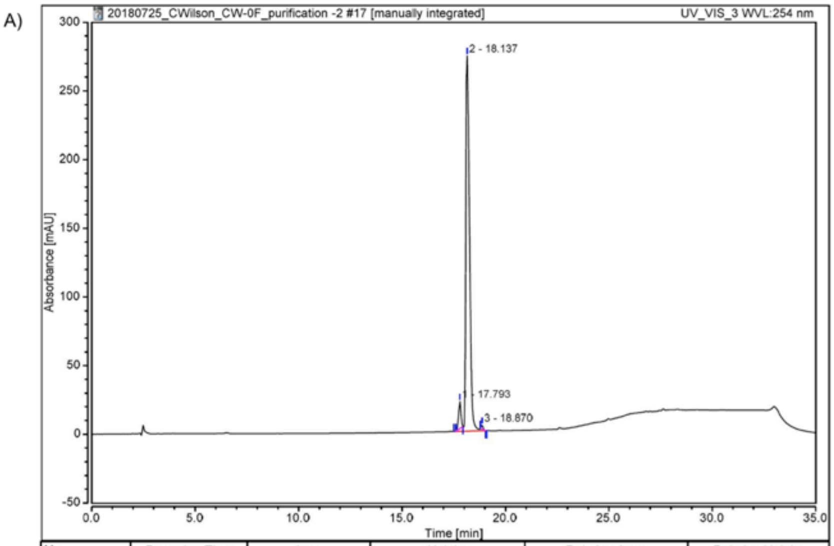


Figure 4.14. Characterization data for compound **4.7**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{41}H_{59}N_8O_9^+$: 807.44; found: 807.5.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	17.793	2.659	19.708	4.37	6.66
2	18.137	57.936	273.423	95.15	92.33
3	18.870	0.291	3.005	0.48	1.01

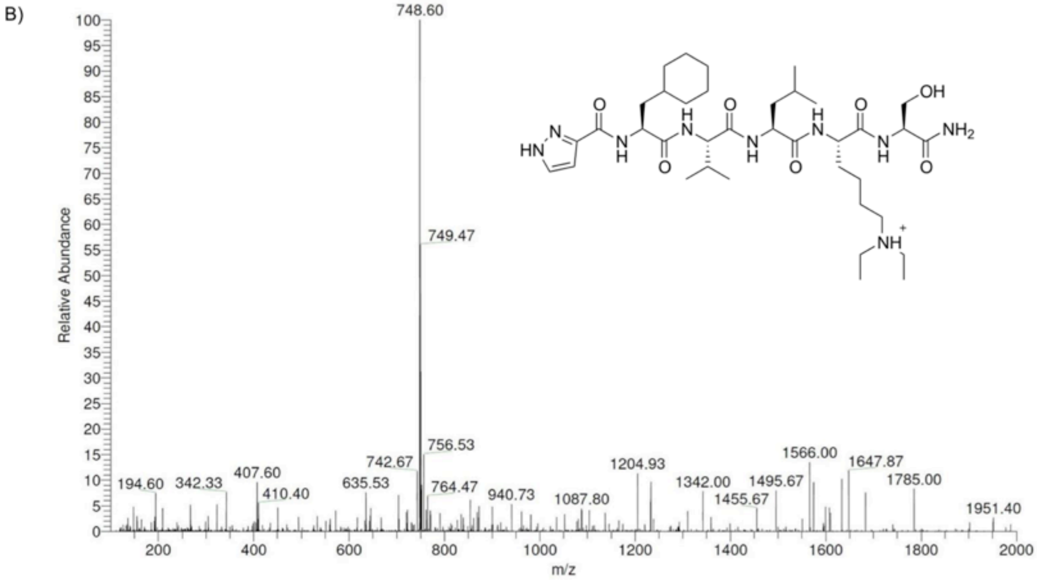
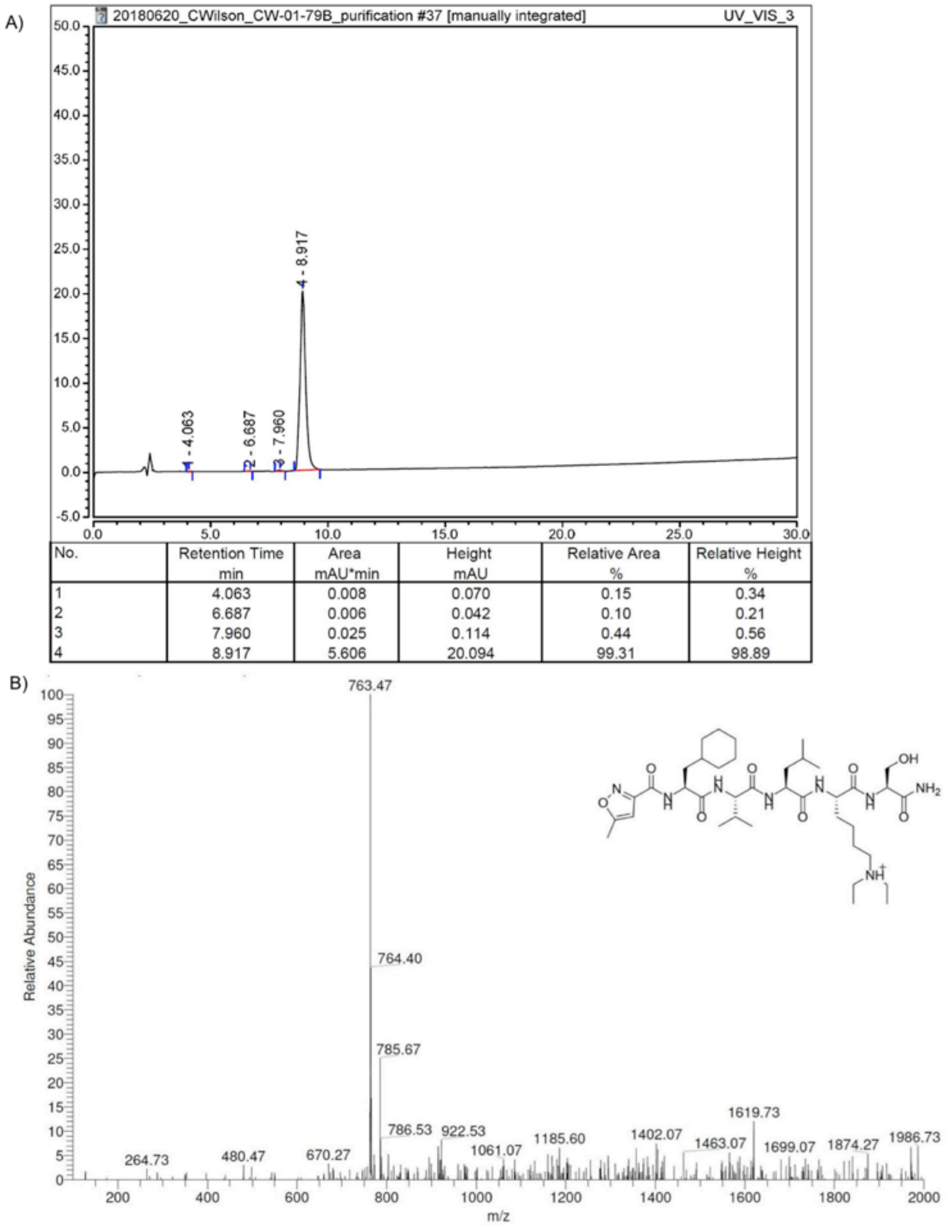
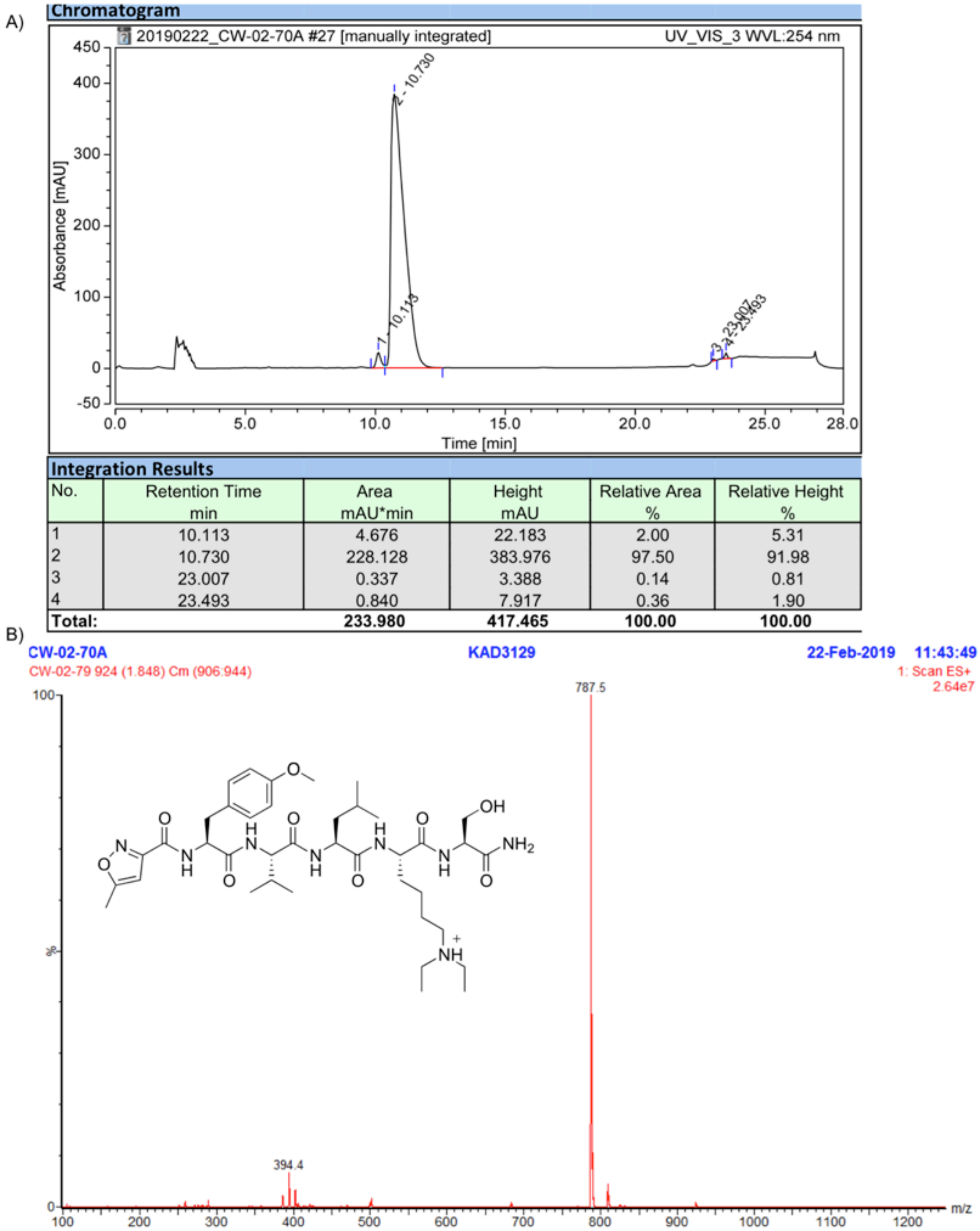


Figure 4.15. Characterization data for compound **4.8**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{37}H_{66}N_9O_7^+$: 748.51; found: 748.60.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	4.063	0.008	0.070	0.15	0.34
2	6.687	0.006	0.042	0.10	0.21
3	7.960	0.025	0.114	0.44	0.56
4	8.917	5.606	20.094	99.31	98.89

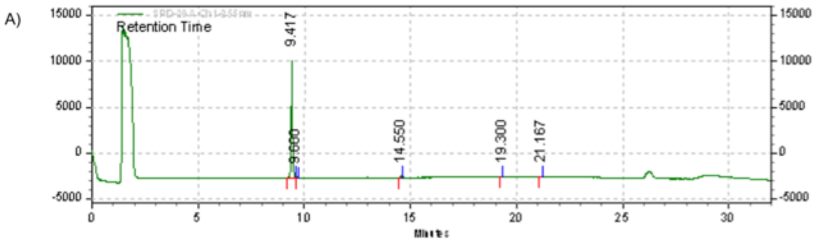
Figure 4.16. Characterization data for compound **4.9**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{38}H_{67}N_8O_8^+$: 763.51; found: 763.47.



Integration Results

No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	10.113	4.676	22.183	2.00	5.31
2	10.730	228.128	383.976	97.50	91.98
3	23.007	0.337	3.388	0.14	0.81
4	23.493	0.840	7.917	0.36	1.90
Total:		**233.980**	**417.465**	**100.00**	**100.00**

Figure 4.17. Characterization data for compound **4.10**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{39}H_{63}N_8O_9^+$: 787.47; found: 787.5.



SPD-20A
Chl-255nm
Results

Retention Time	Area	Area %	Height	Height %
9.417	53942	97.45	12752	97.67
9.600	3	0.01	0	0.00
14.550	907	1.64	210	1.61
19.300	225	0.41	55	0.42
21.167	277	0.50	39	0.30

B)

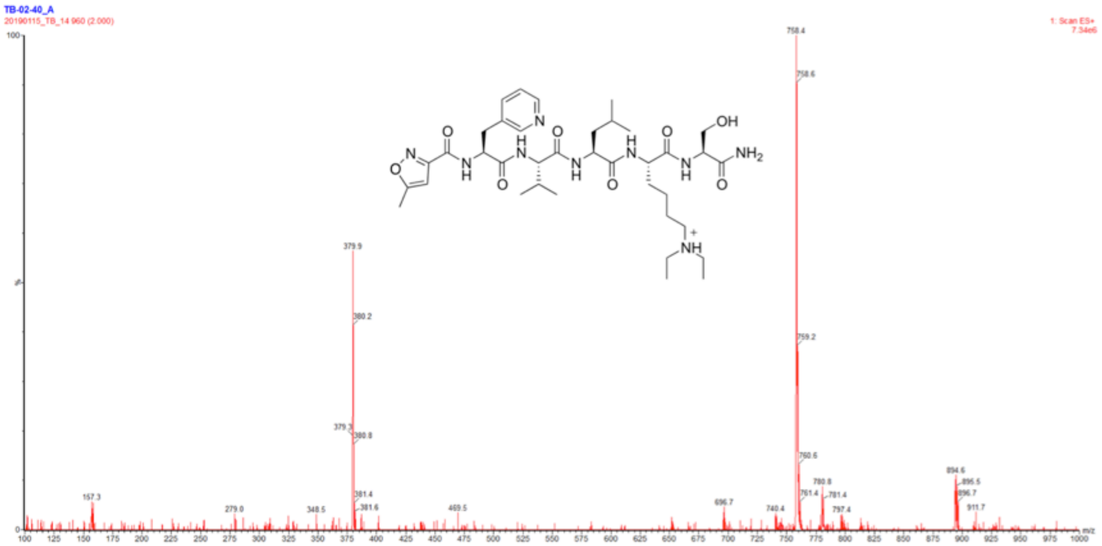


Figure 4.18. Characterization data for compound **4.11**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{37}H_{60}N_9O_8^+$: 748.46; found: 748.4.

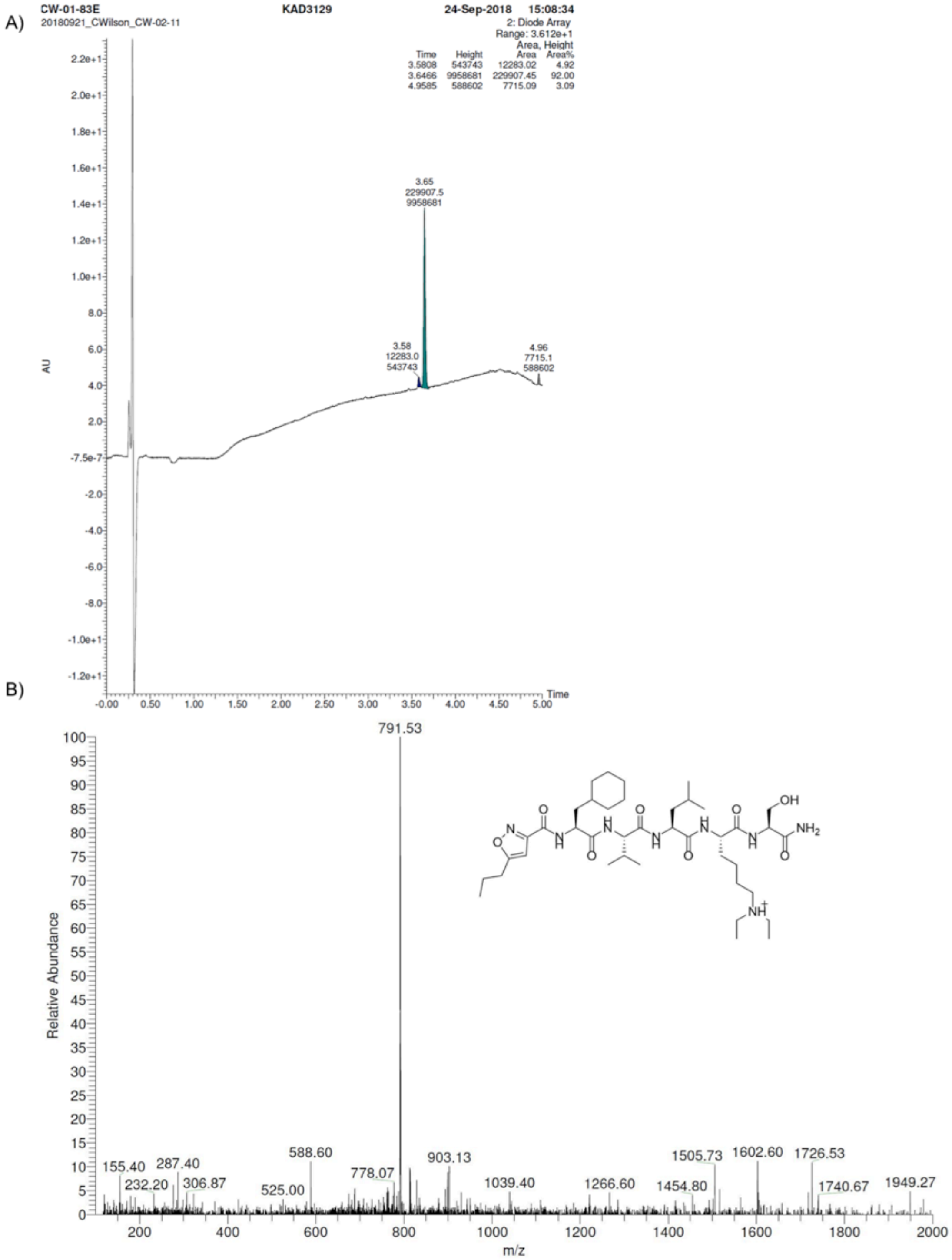


Figure 4.19. Characterization data for compound **4.12**. A) Analytical UPLC trace. Product peak is at 3.65 minutes with an area of 92%. B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{40}H_{71}N_8O_8^+$: 791.54; found: 791.53.

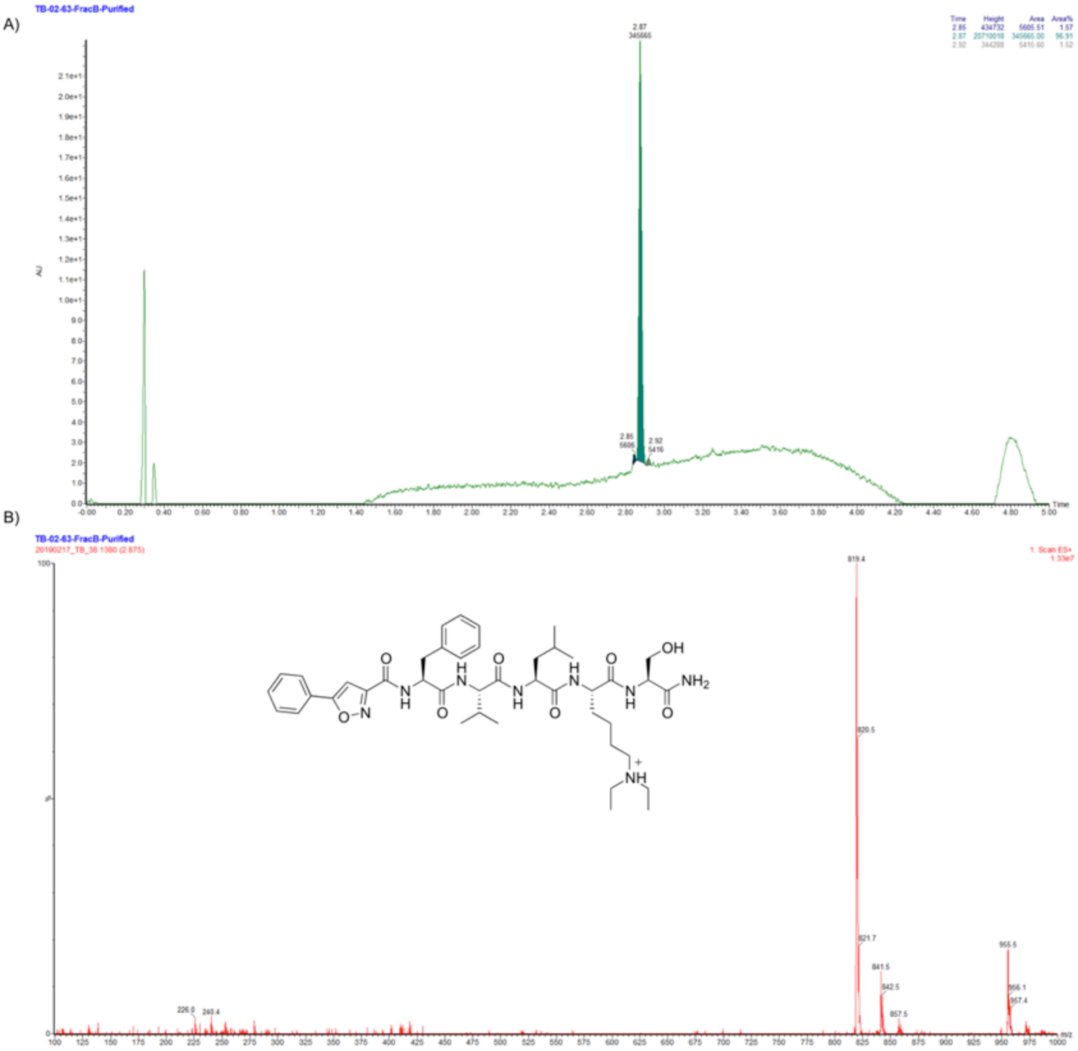
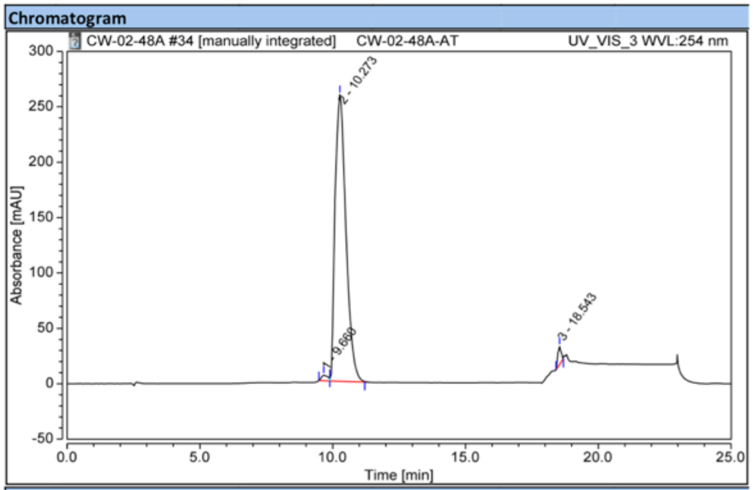


Figure 4.20. Characterization data for compound **4.13**. A) Analytical UPLC trace. Product peak is at 2.87 minutes with an area of 97%. B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{43}H_{63}N_8O_8^+$: 819.48; found: 819.4.

A)



Integration Results

No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	9.660	1.378	4.891	1.10	1.75
2	10.273	122.145	258.830	97.07	92.45
3	18.543	2.306	16.255	1.83	5.81
Total:		**125.829**	**279.976**	**100.00**	**100.00**

B)

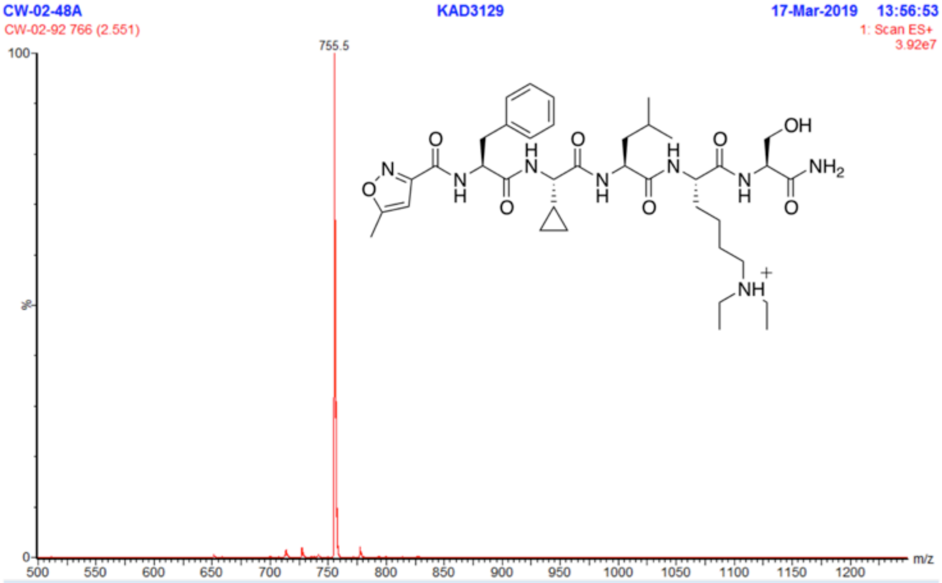
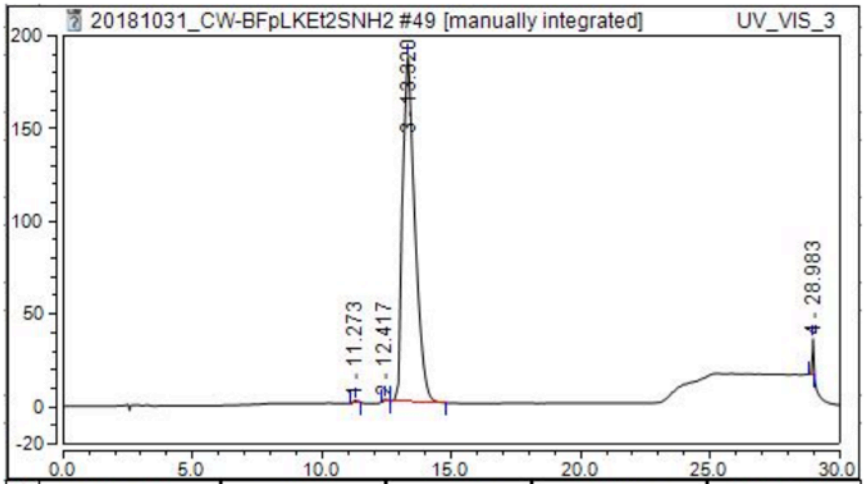


Figure 4.21. Characterization data for compound **4.14**. A) Analytical UPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{38}H_{59}N_{8}O_{8}^{+}$: 755.45; found: 755.5.

A)



No	Ret. Time min	Area mAU*min	Height mAU	Rel. Area %	Rel. Height %
1	11.273	0.304	1.483	0.29	0.71
2	12.417	0.231	1.252	0.22	0.60
3	13.320	102.349	185.733	98.98	89.37
4	28.983	0.524	19.356	0.51	9.31

B)

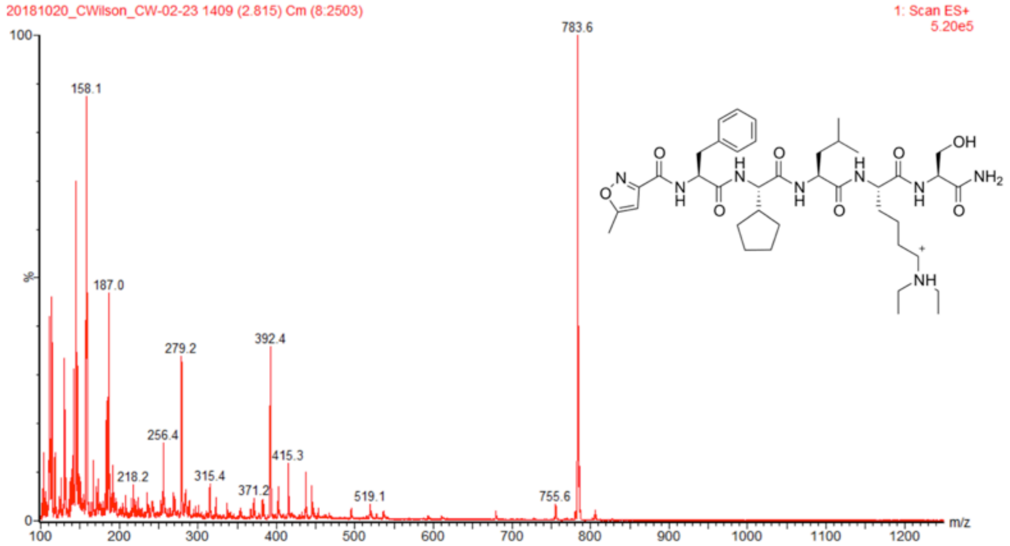
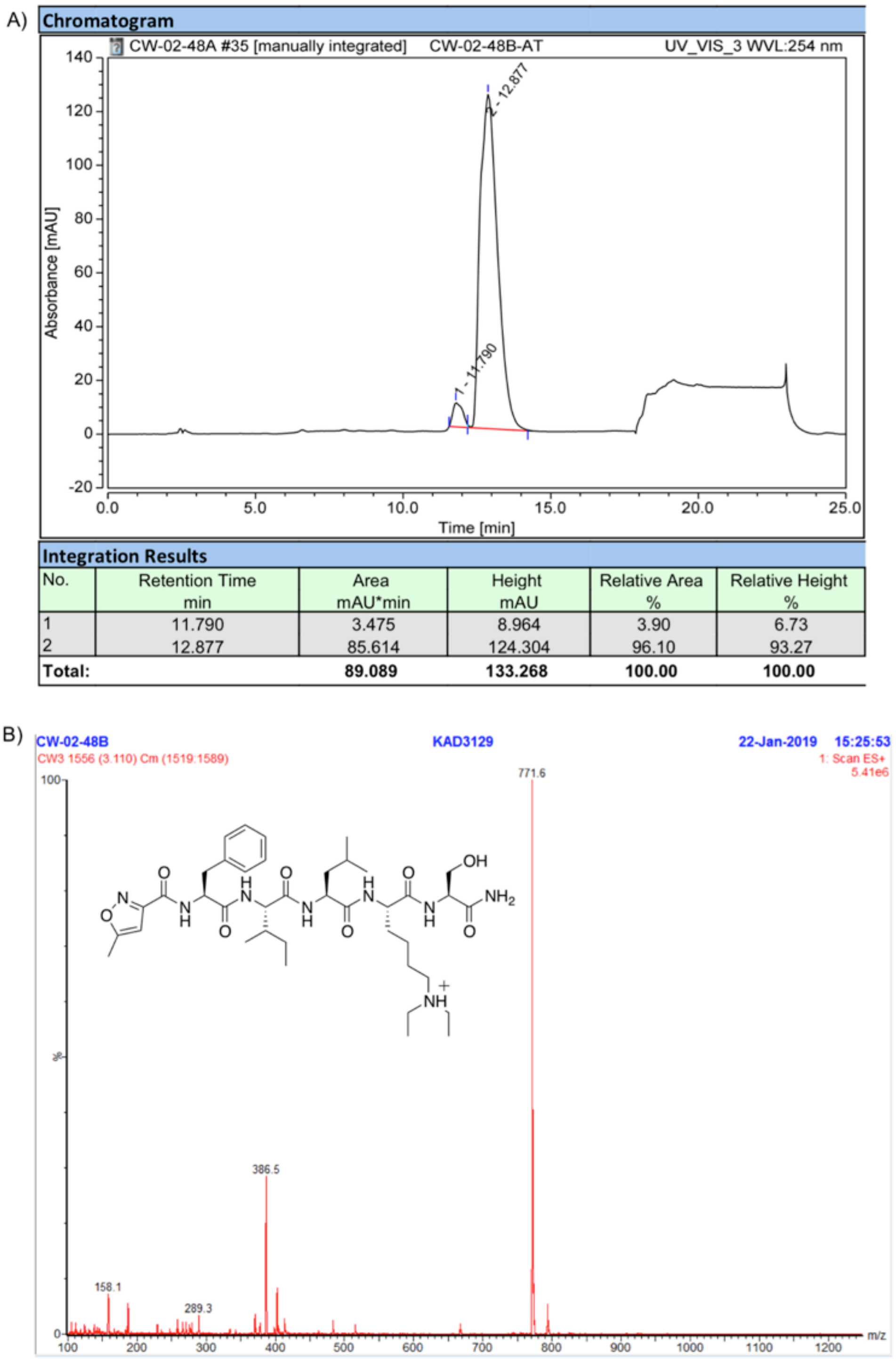


Figure 4.22. Characterization data for compound **4.15**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{40}H_{63}N_8O_8^+$: 783.48; found: 783.6.



No.	Retention Time min	Area mAU*min	Height mAU	Relative Area %	Relative Height %
1	11.790	3.475	8.964	3.90	6.73
2	12.877	85.614	124.304	96.10	93.27
Total:		**89.089**	**133.268**	**100.00**	**100.00**

Figure 4.23. Characterization data for compound **4.16**. A) Analytical HPLC trace, B) Low resolution mass spectrum. LR-ESI-MS: [*m/z*] calcd. for: $C_{39}H_{63}N_8O_8^+$: 771.48; found:

4.7.6 Protein expression and purification

Methods used are described in Chapter 2, Experimental methods and supplementary data section 2.8.6.

4.7.7 Fluorescence polarization methods

Competitive FP analysis with CBX1/6/7/8 was used to determine IC_{50} values with all compounds using **2.4** as the competitive binding probe. Methods used for competitive FP assays are described in Chapter 2, Experimental methods and supplementary data section 2.8.7. The following exceptions are described below.

All peptide inhibitors were dissolved in DMSO and then diluted into FP buffer. DMSO was kept constant at 1% in the majority of assays, with the exception of FP assays with **4.1**, **4.3**, and **4.13** (2–5% DMSO). CBX1 was used at 10 μM, CBX6 at 1 μM, CBX7 at 0.7 μM and CBX8 at 4 μM. Inhibitor concentrations varied between 0–500 μM. Following serial dilution of the inhibitor across the plate, plates were incubated for 15 minutes in the dark and then were read with the Cytation 5 cell imaging and multi-mode reader plate reader. A polarized filter cube was used with filters for excitation at 485 nm and emission at 520 nm (with a cut-off of 510 nm). All plates were read at 28°C.

Fluorescence polarization data analysis

Methods used for competitive FP assays are described in Chapter 2, Experimental methods and supplementary data section 2.8.7. Experiments were done in triplicate unless otherwise stated in raw data figures.

4.7.8 Fluorescence polarization data

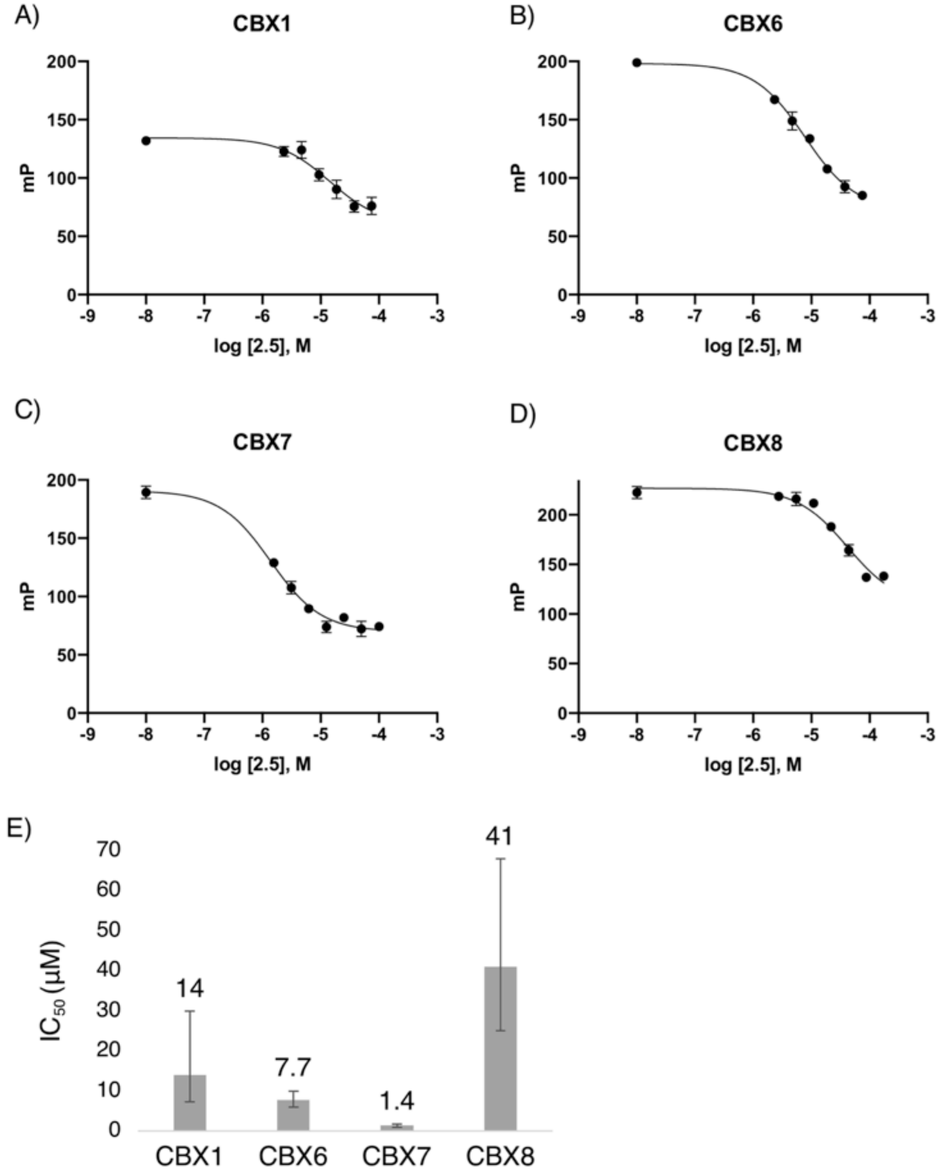


Figure 4.24. Competitive fluorescence polarization data of compound **2.5** with CBX proteins. A) CBX1 (logIC$_{50}$ = –4.845 M, 95% CI –5.137 M to -4.510 M, R^2 = 0.928), B) CBX6 (logIC$_{50}$ = –5.113 M, 95% CI –5.222 M to –5.00 M, R^2 = 0.990), C) CBX7 (logIC$_{50}$ = –5.862 M, 95% CI –5.983 M to –5.746 M, R^2 = 0.983), D) CBX8 (logIC$_{50}$ = –4.385 M 95% CI –4.593 M to –4.162 M, R^2 =0.967), E) Bar graph of IC$_{50}$ values in µM. Error bars reported as asymmetrical 95 % confidence intervals.

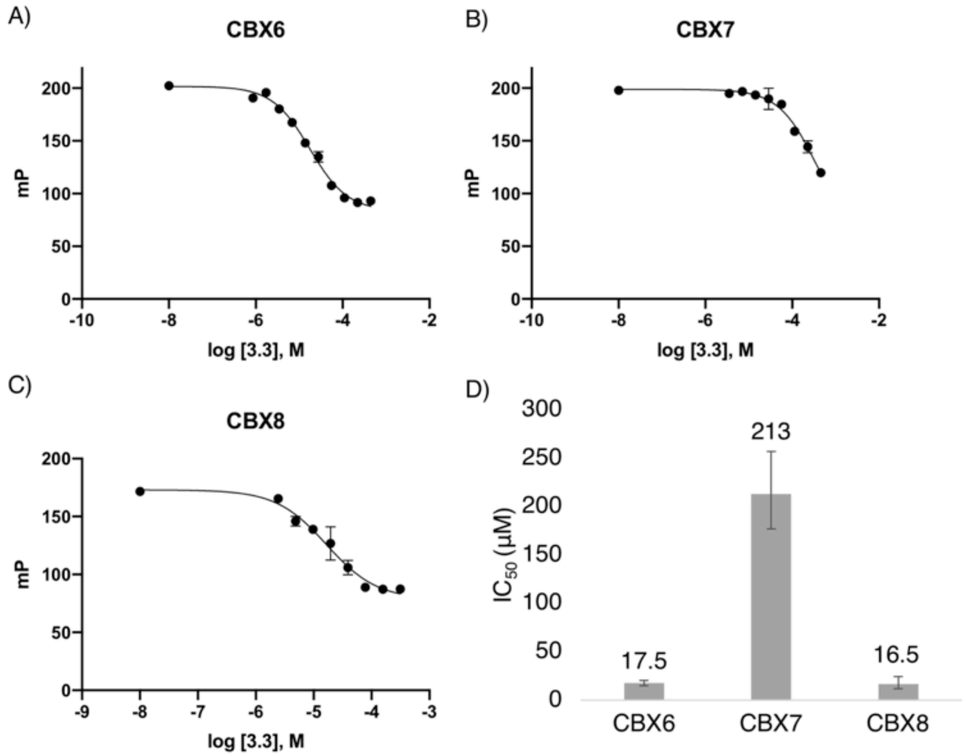


Figure 4.25. Competitive fluorescence polarization data of compound **3.3** with CBX proteins. A) CBX6 (logIC$_{50}$ = –4.756 M, 95% CI –4.834 M to –4.678 M, R^2 = 0.990), B) CBX7 (logIC$_{50}$ = –3.672 M, 95% CI –3.750 M to –3.590 M, R^2 = 0.968), C) CBX8 (logIC$_{50}$ = –4.778, 95% CI –4.949 M to –4.602 M, R^2 = 0.969), D) Bar graph of IC$_{50}$ values in μM. Error bars reported as asymmetrical 95 % confidence intervals. CBX7 was fitted constraining the mP value of the endpoint to 100 millipolarization units.

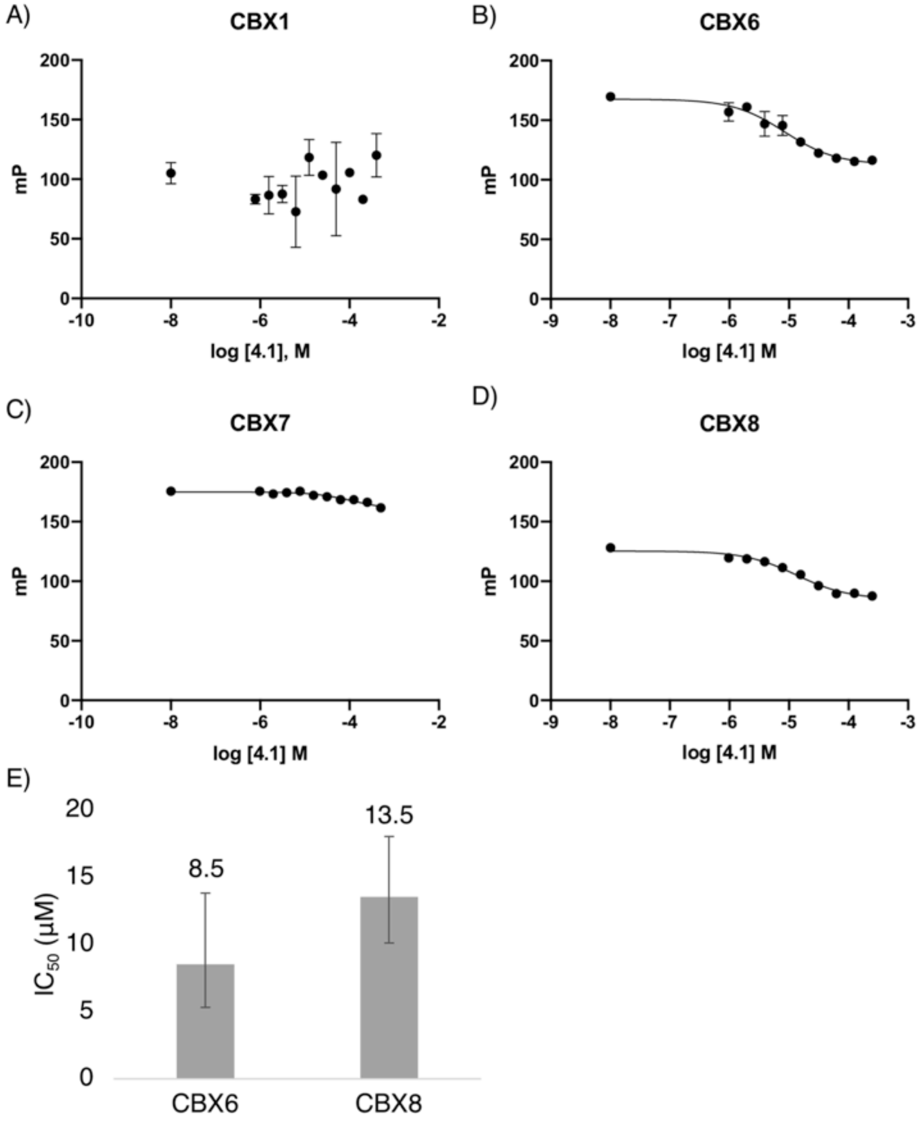


Figure 4.26. Competitive fluorescence polarization data of compound **4.1** with CBX proteins. A) CBX1 ($\log IC_{50}$ = > –3.301 M), B) CBX6 ($\log IC_{50}$ = –5.072 M, 95% CI –5.276 M to –4.861 M, $R^2 = 0.931$), C) CBX7 ($\log IC_{50}$ = > –3.301 M), D) CBX8 ($\log IC_{50}$ = –4.871 M, 95% CI –4.997 M to –4.745 M, $R^2 = 0.973$), E) Bar graph of IC_{50} values in μM. Error bars reported as asymmetrical 95 % confidence intervals.

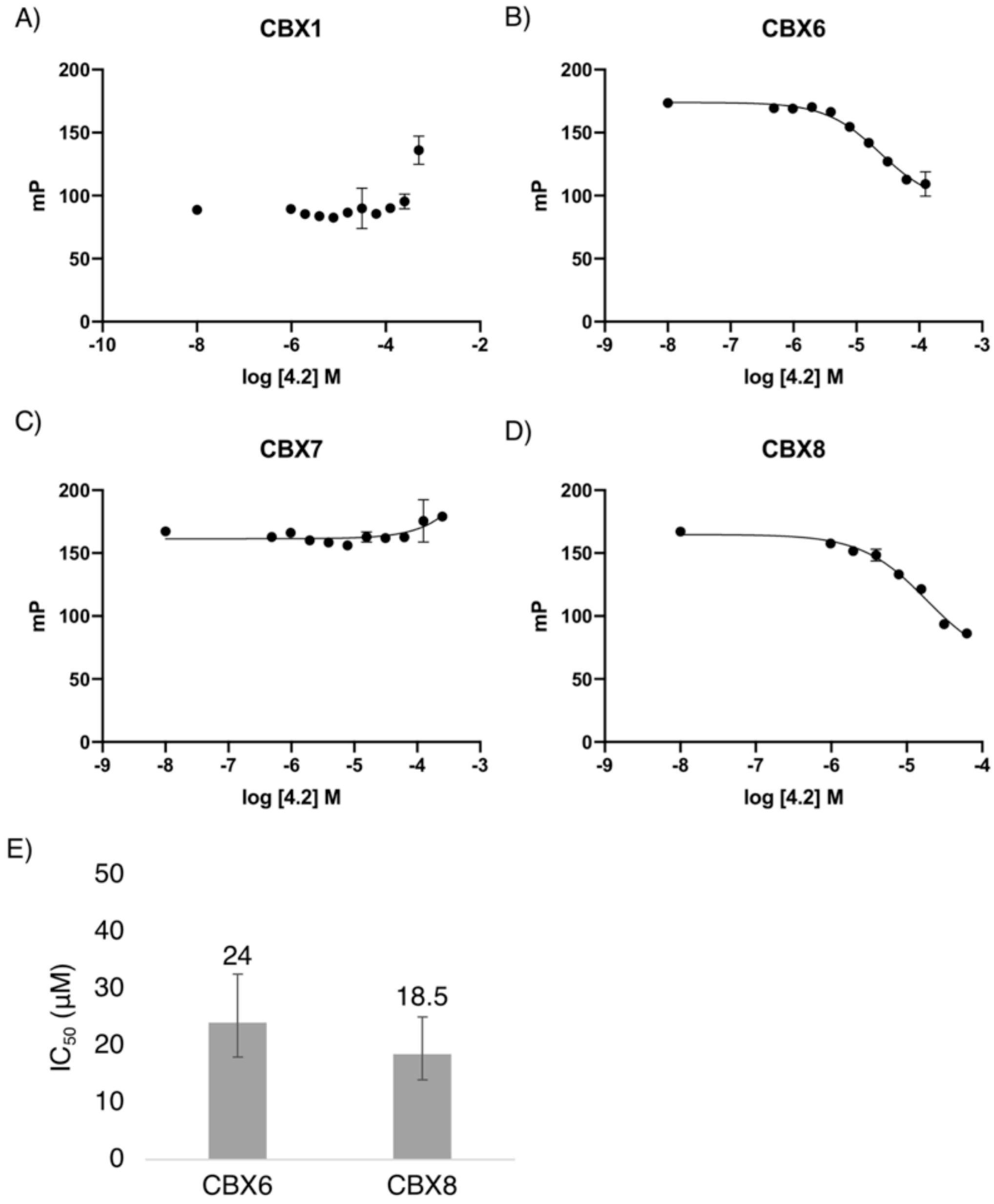


Figure 4.27. Competitive fluorescence polarization data of compound **4.2** with CBX proteins. A) CBX1 (logIC$_{50}$ = > –3.301 M), B) CBX6 (logIC$_{50}$ = –4.625 M, 95% CI –4.754 M to –4.490 M, R^2 = 0.972), C) CBX7 (logIC$_{50}$ = > –3.301 M), D) CBX8 (logIC$_{50}$ = –4.734 M, 95% CI –4.856 M to –4.4605 M, R^2 =0.984), E) Bar graph of IC$_{50}$ values in μM. Error bars reported as asymmetrical 95 % confidence intervals.

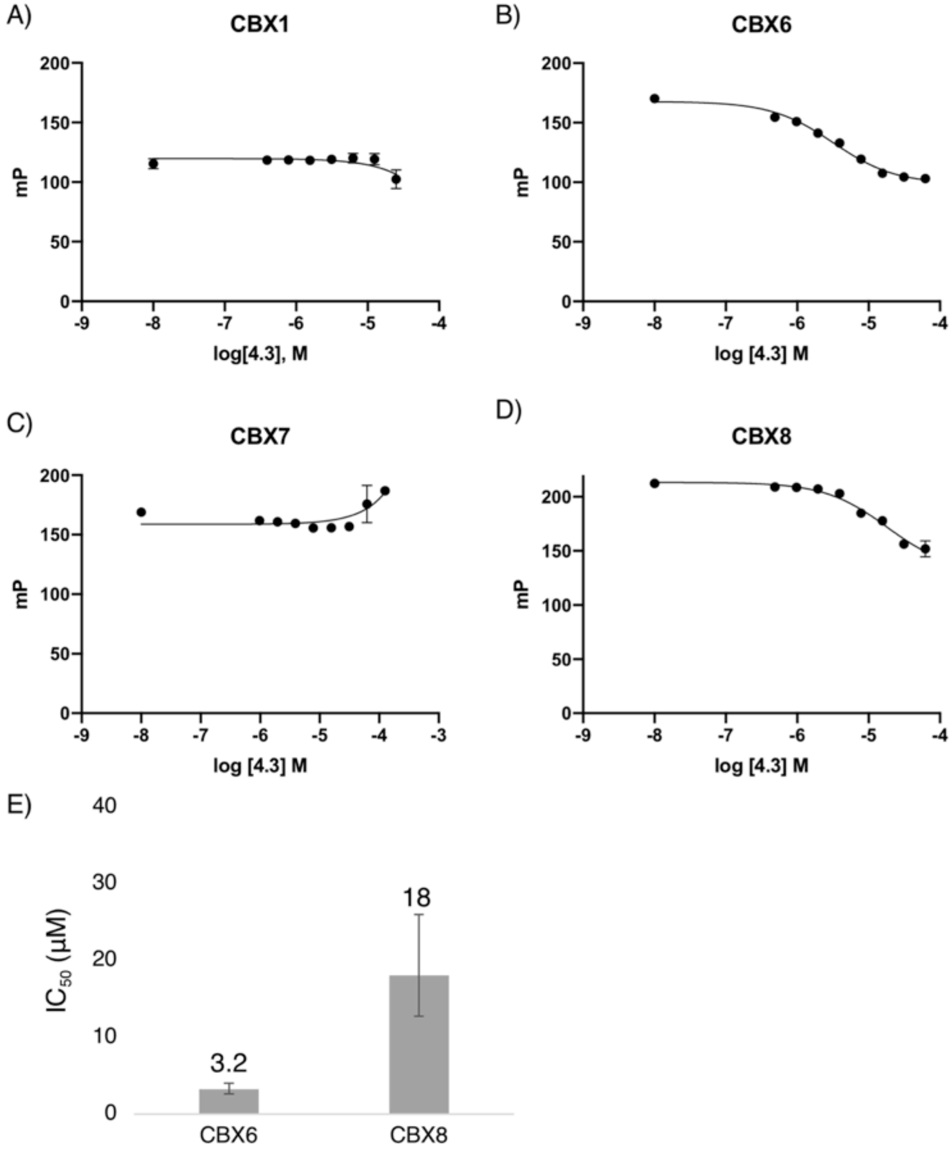


Figure 4.28. Competitive fluorescence polarization data of compound **4.3** with CBX proteins performed in triplicate. A) CBX1 (logIC$_{50}$ = > –3.69 M) B) CBX6 (logIC$_{50}$ = –5.495 M, 95% CI –5.591 M to –5.398 M, R^2 = 0.985), C) CBX7 (logIC$_{50}$ = > –3.69 M), D) CBX8 (logIC$_{50}$ = –4.746 M, 95% CI –4.896 M to –4.586 M, R^2 = 0.969), E) Bar graph of IC$_{50}$ values. Error bars reported as asymmetrical 95 % confidence intervals. Each assay was performed with 5% DMSO due to solubility issues of the inhibitor.

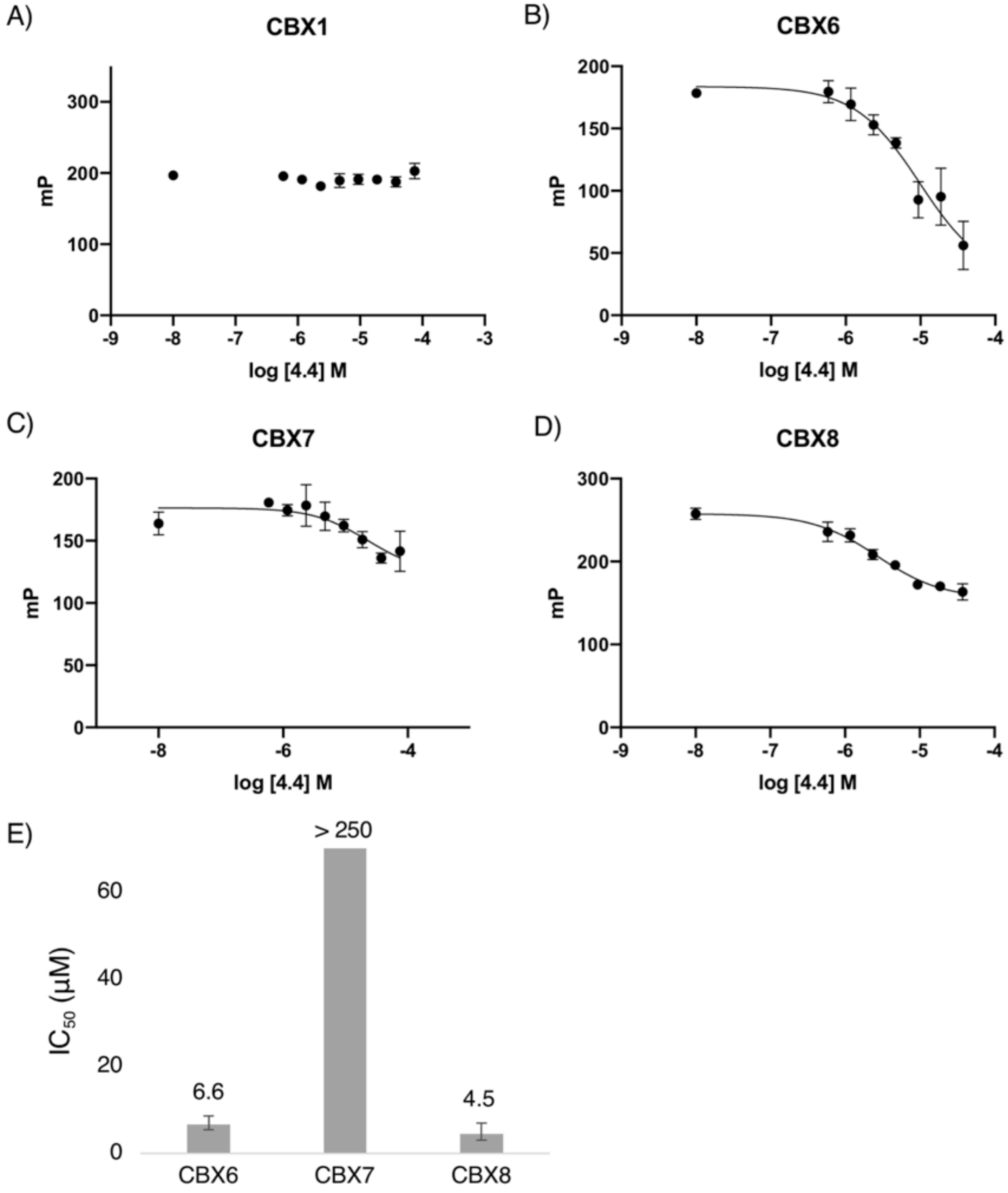


Figure 4.29. Competitive fluorescence polarization data of compound **4.4** with CBX proteins performed in triplicate. A) CBX1 (logIC_{50} = > –3.522 M) B) CBX6 (logIC_{50} = –5.178 M, 95% CI –5.235 M to –5.120 M, R^2 = 0.995), C) CBX7 (logIC_{50} = > –3.602 M), D) CBX8 (logIC_{50} = –5.342 M, 95% CI –5.525 M to –5.147 M, R^2 = 0.947), E) Bar graph of IC_{50} values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

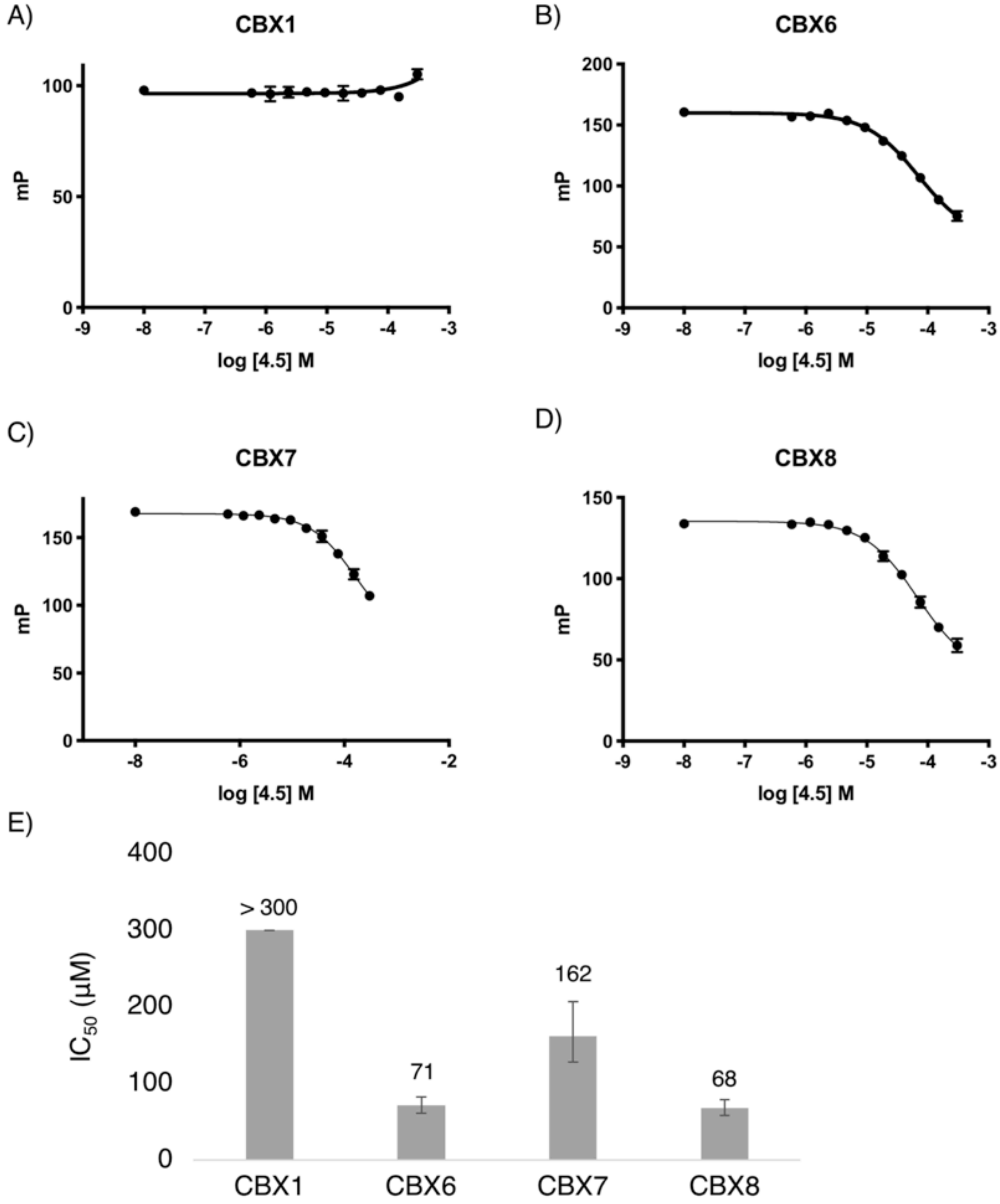


Figure 4.30. Competitive fluorescence polarization data of compound **4.5** with CBX proteins performed in triplicate. A) CBX1 (logIC$_{50}$ = > –3.522 M) B) CBX6 (logIC$_{50}$ = –4.143 M, 95% CI –4.205 M to –4080 M, R^2 = 0.994), C) CBX7 (logIC$_{50}$ = –3.791 M, 95% CI –3.892 M to –3.683 M, R^2 = 0.989), D) CBX8 (logIC$_{50}$ = –4.169 M, 95% CI –4.238 M to –4.099 M, R^2 = 0.993), E) Bar graph of IC$_{50}$ values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

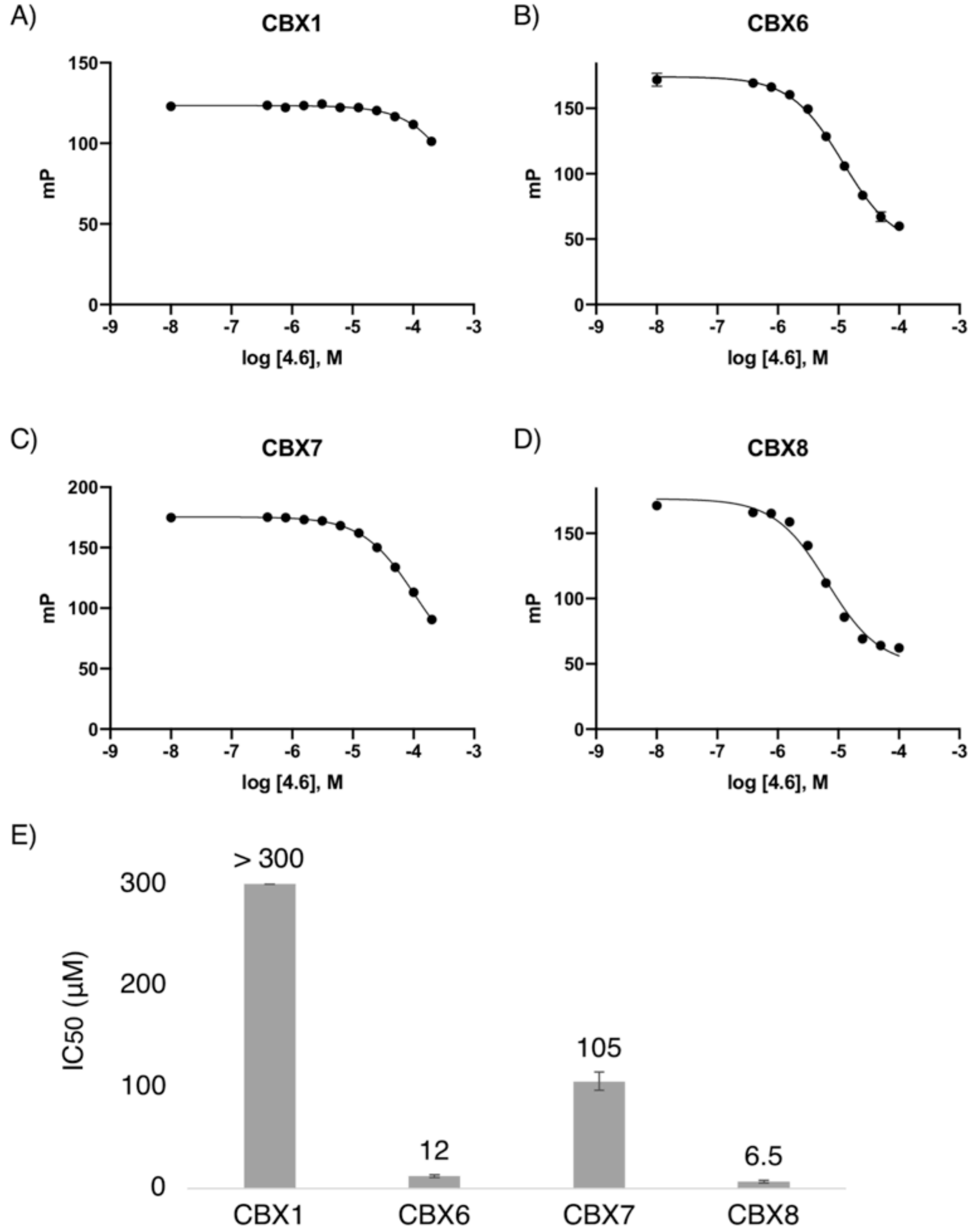


Figure 4.31. Competitive fluorescence polarization data of compound **4.6** with CBX proteins performed in triplicate. A) CBX1 ($\log IC_{50}$ = > –3.522 M) B) CBX6 ($\log IC_{50}$ = –4.982 M, 95% CI –4.983 M to –4.872 M, R^2 = 0.995), C) CBX7 ($\log IC_{50}$ = –3.978 M, 95% CI –4.013 M to –3.941 M, R^2 = 0.998), D) CBX8 ($\log IC_{50}$ = –5.186 M, 95% CI –5.265 M to –5.106 M, R^2 = 0.987), E) Bar graph of IC_{50} values (μM). Error bars reported as asymmetrical 95 % confidence intervals.

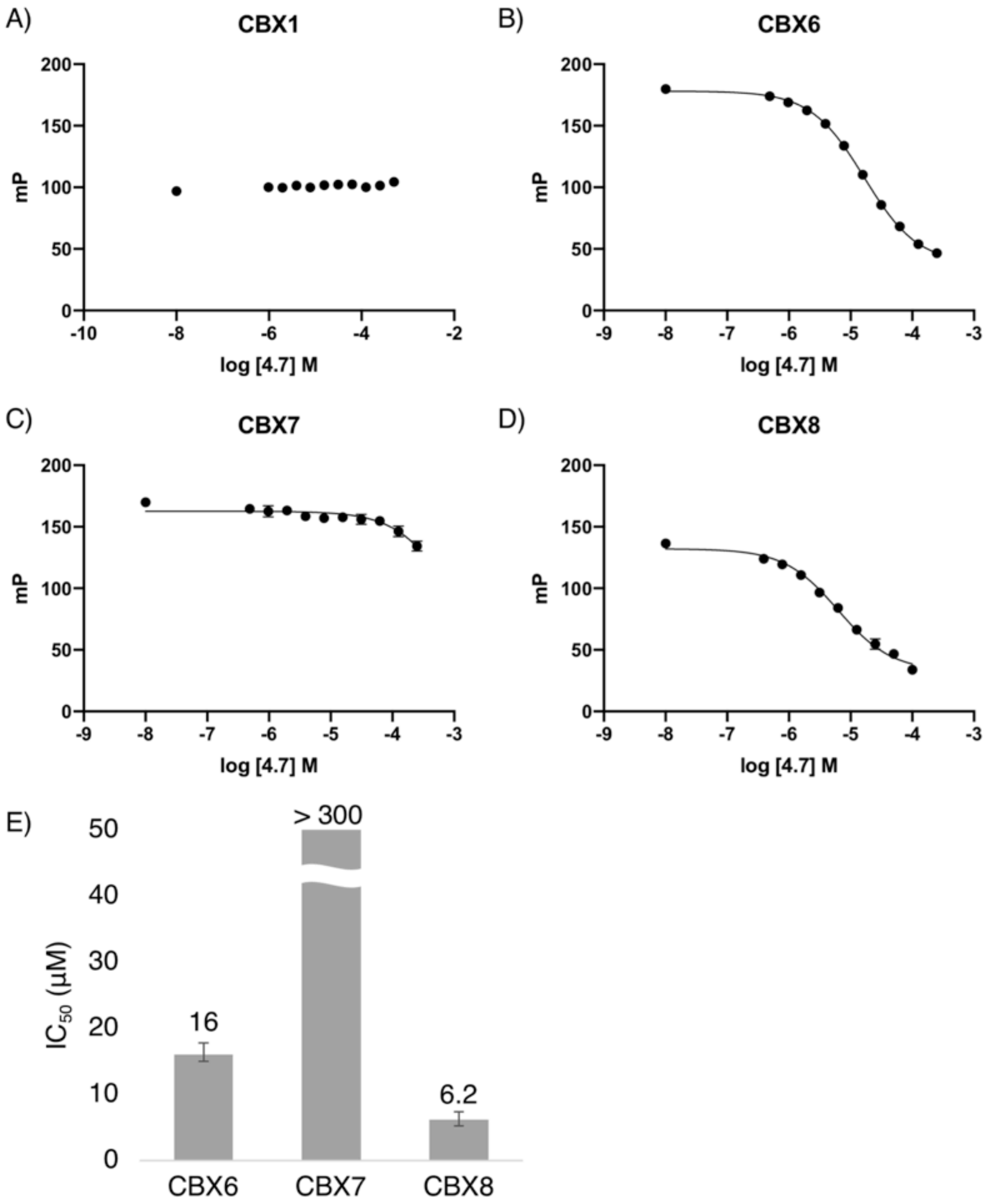


Figure 4.32. Competitive fluorescence polarization data of compound **4.7** with CBX proteins performed in triplicate. A) CBX1 ($\log IC_{50} = > -3.301$ M) B) CBX6 ($\log IC_{50} = -4.781$ M, 95% CI –4.804 M to –4.758 M, $R^2 = 0.999$), C) CBX7 ($\log IC_{50} = > -3.522$ M), D) CBX8 ($\log IC_{50} = -5.207$ M, 95% CI –5.284 M to –5.130 M, $R^2 = 0.991$), E) Bar graph of IC_{50} values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

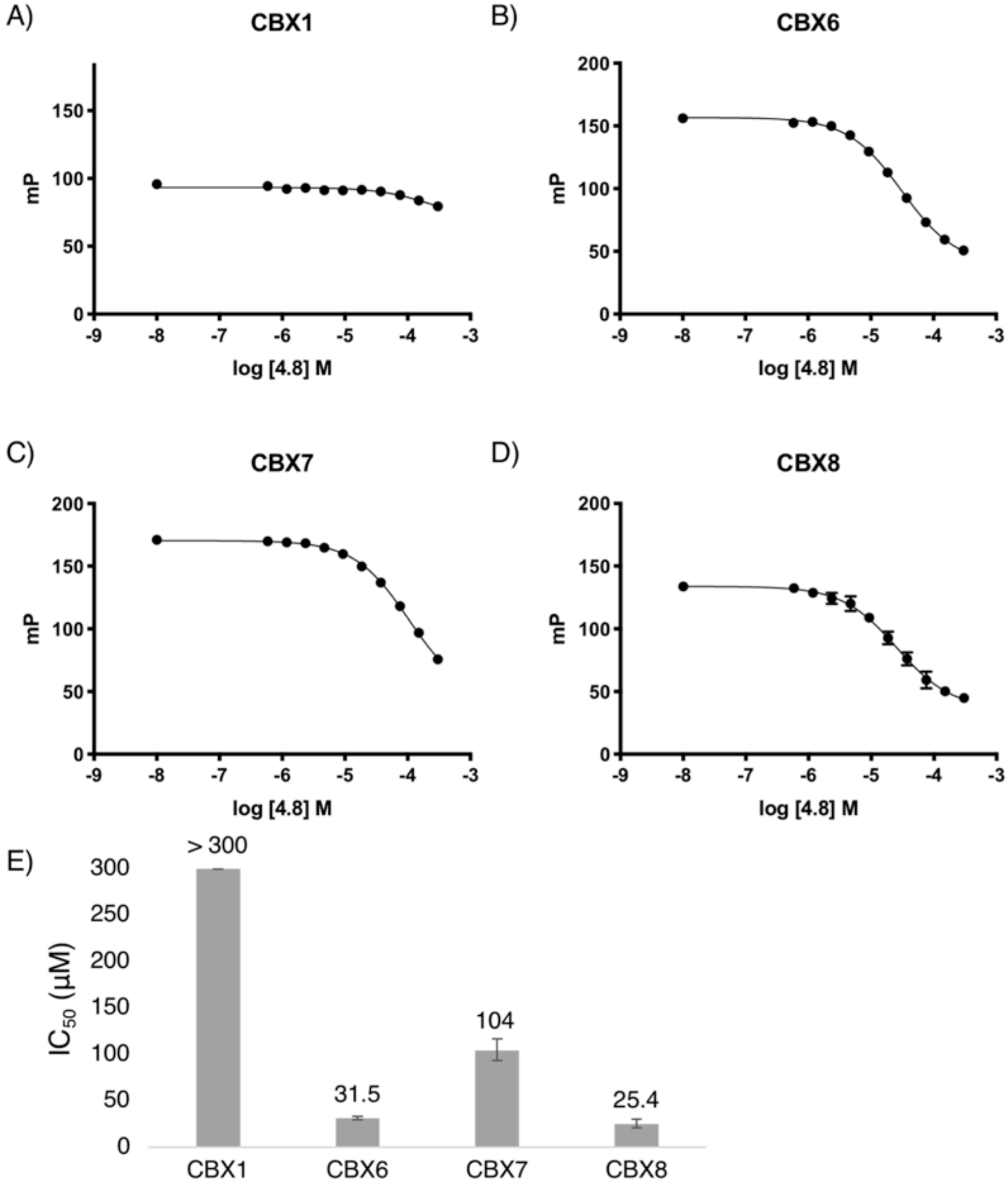


Figure 4.33. Competitive fluorescence polarization data of compound **4.8** with CBX proteins performed in triplicate. A) CBX1 (logIC$_{50}$ = > –3.522 M) B) CBX6 (logIC$_{50}$ = –4.502 M, 95% CI –4.528 M to –4.475 M, R^2 = 0.999), C) CBX7 (logIC$_{50}$ = –3.981 M, 95% CI –4.030 M to –3.932 M, R^2 = 0.997), D) CBX8 (logIC$_{50}$ = –4.594 M, 95% CI –4.673 M to –4.515 M, R^2 = 0.992), E) Bar graph of IC$_{50}$ values (μM). Error bars reported as asymmetrical 95 % confidence intervals.

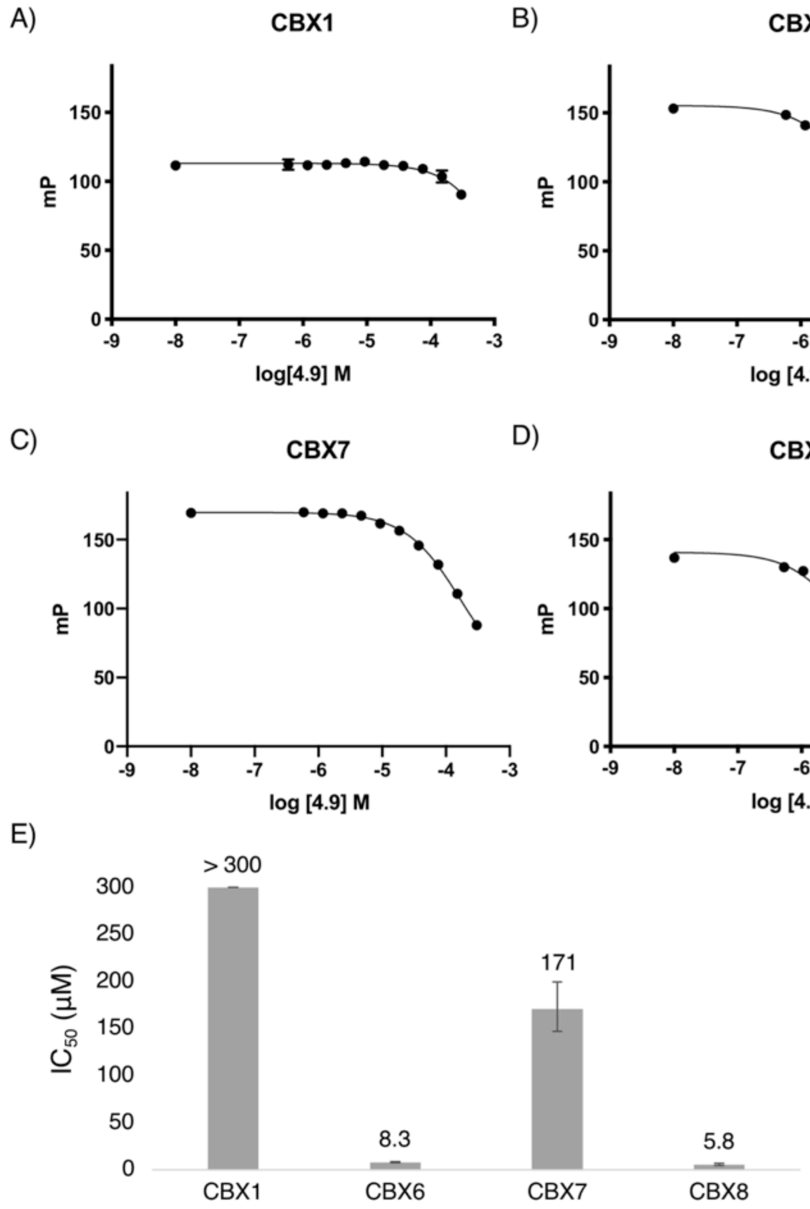


Figure 4.34. Competitive fluorescence polarization data of compound **4.9** with CBX proteins performed in triplicate. A) CBX1 (logIC_{50} = > –3.522 M) B) CBX6 (logIC_{50} = –5.080 M, 95% CI –5.120 M to –5.039 M, R^2 = 0.997), C) CBX7 (logIC_{50} = –3.676 M, 95% CI –3.833 M to –3.698 M, R^2 = 0.996), D) CBX8 (logIC_{50} = –5.234 M, 95% CI –5.317 M to –5.150 M, R^2 = 0.985), E) Bar graph of IC_{50} values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

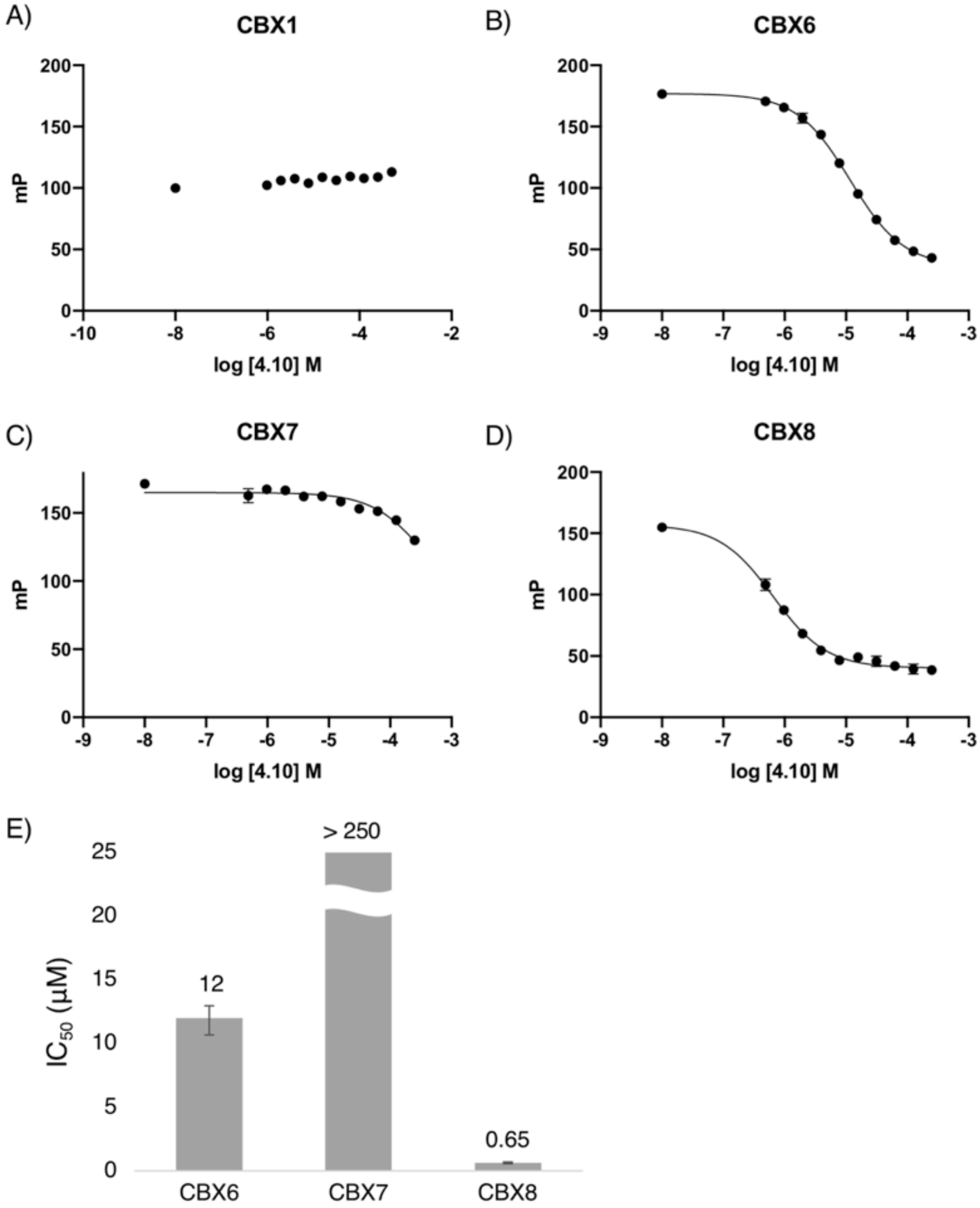


Figure 4.35. Competitive fluorescence polarization data of compound **4.10** with CBX proteins performed in triplicate. A) CBX1 (logIC$_{50}$ = > –3.301 M) B) CBX6 (logIC$_{50}$ = –4.934 M, 95% CI –4.969 M to –4.899 M, R^2 = 0.998), C) CBX7 (logIC$_{50}$ = > –3.602 M), D) CBX8 (logIC$_{50}$ = –6.189 M, 95% CI –6.239 M to –6.138 M, R^2 = 0.992), E) Bar graph of IC$_{50}$ values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

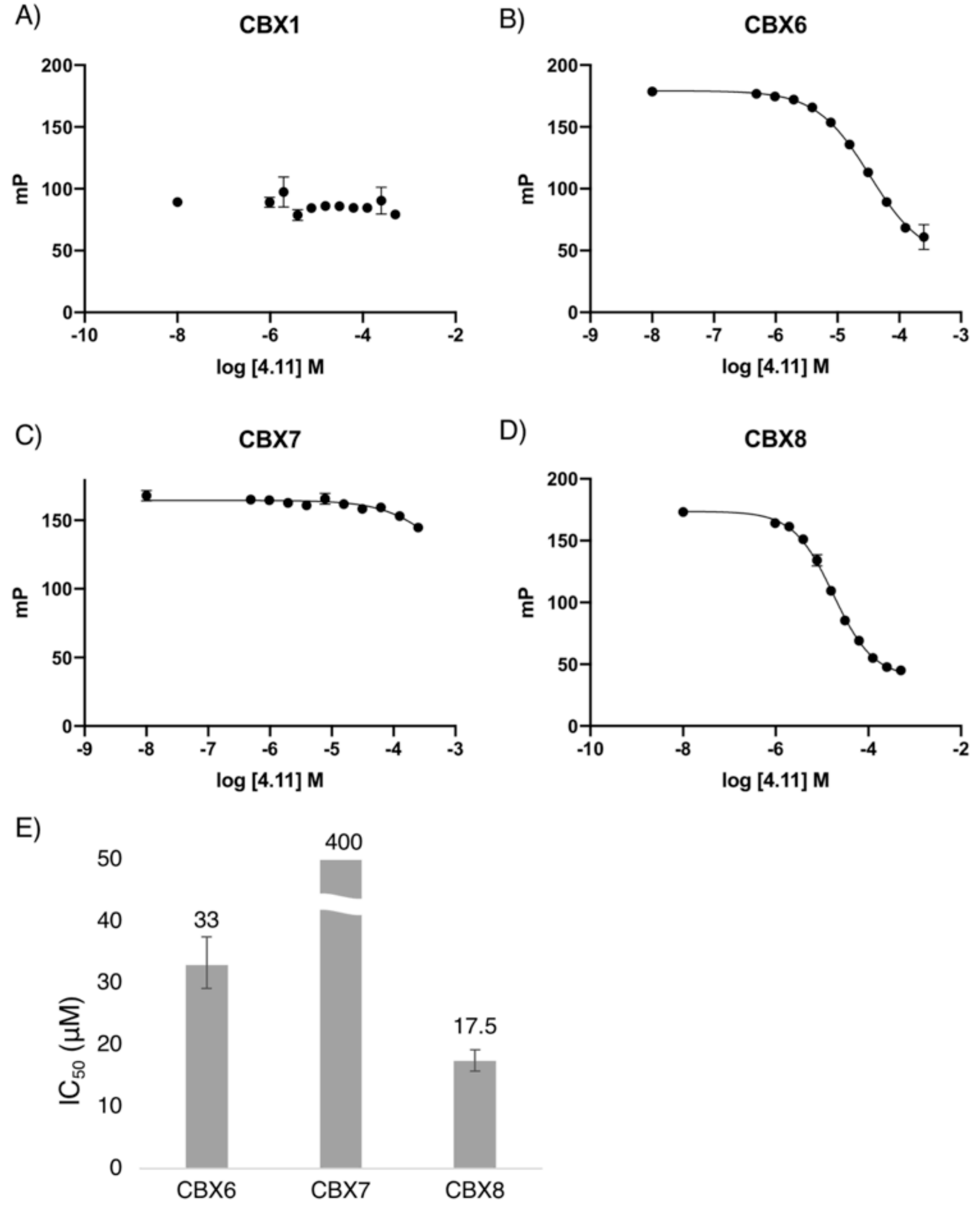


Figure 4.36. Competitive fluorescence polarization data of compound **4.11** with CBX proteins performed in duplicate. A) CBX1 (logIC$_{50}$ = > –3.301 M) B) CBX6 (logIC$_{50}$ = –4.480 M, 95% CI –4.535 M to –4.425 M, $R^2 = 0.995$), C) CBX7 (logIC$_{50}$ = > –3.602 M), D) CBX8 (logIC$_{50}$ = –4.758 M, 95% CI –4.800 M to –4.715 M, $R^2 = 0.996$), E) Bar graph of IC$_{50}$ values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

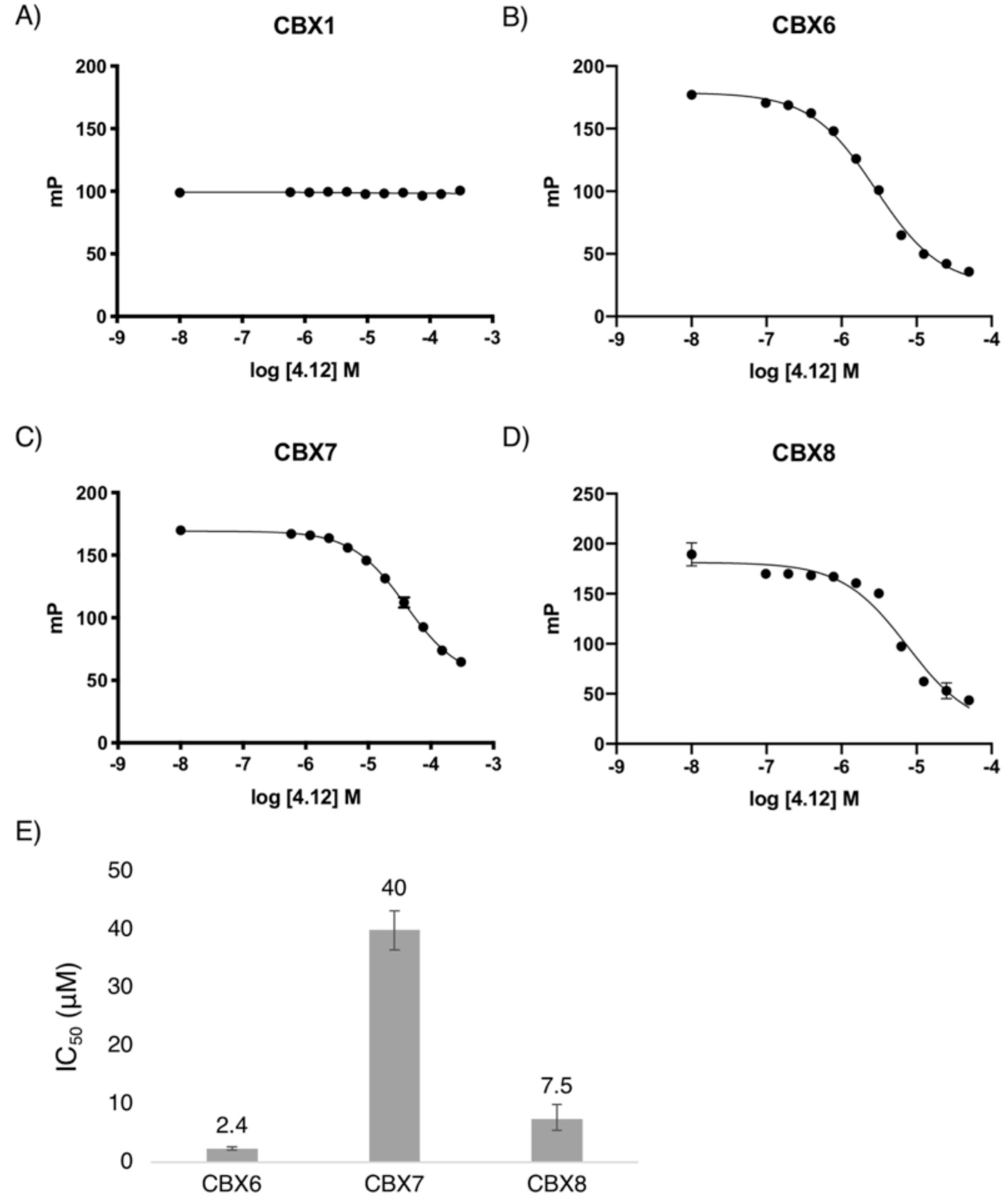


Figure 4.37. Competitive fluorescence polarization data of compound **4.12** with CBX proteins performed in triplicate. A) CBX1 (logIC_{50} = > –3.301 M) B) CBX6 (logIC_{50} = –5.546 M, 95% CI –5.592 M to –5.499 M, R^2 = 0.996), C) CBX7 (logIC_{50} = –4.401 M, 95% CI –4.438 M to –4.364 M, R^2 = 0.998), D) CBX8 (logIC_{50} = –5.125 M, 95% CI –4.989 M to –5.301 M, R^2 = 0.964), E) Bar graph of IC_{50} values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

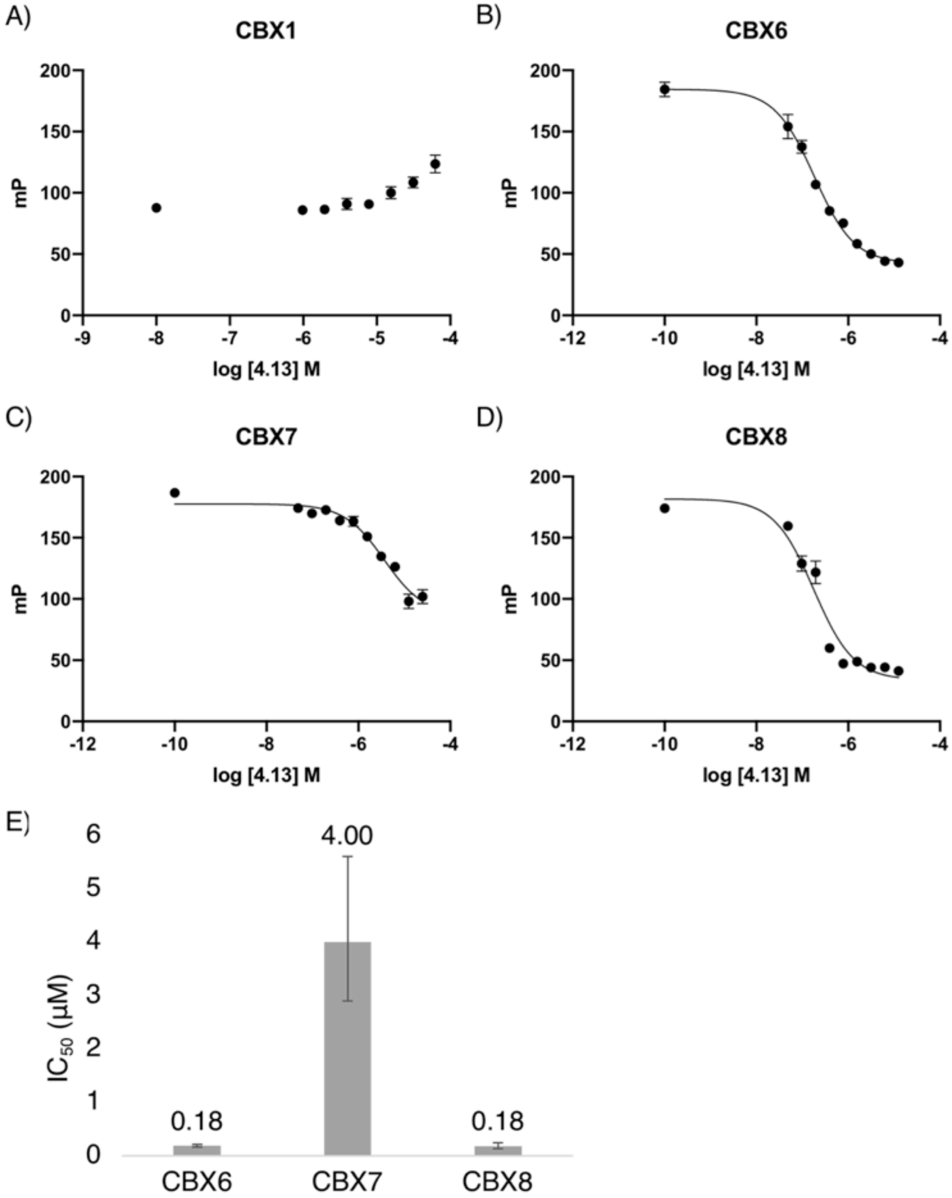


Figure 4.38. Competitive fluorescence polarization data of compound **4.13** with CBX proteins performed in triplicate. A) CBX1 (logIC_{50} = > –3.301 M) B) CBX6 (logIC_{50} = –5.546 M, 95% CI –5.592 M to –5.499 M, R^2 = 0.996), C) CBX7 (logIC_{50} = –4.401 M, 95% CI –4.438 M to –4.364 M, R^2 = 0.998), D) CBX8 (logIC_{50} = –5.125 M, 95% CI –4.989 M to –5.301 M, R^2 = 0.964), E) Bar graph of IC_{50} values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

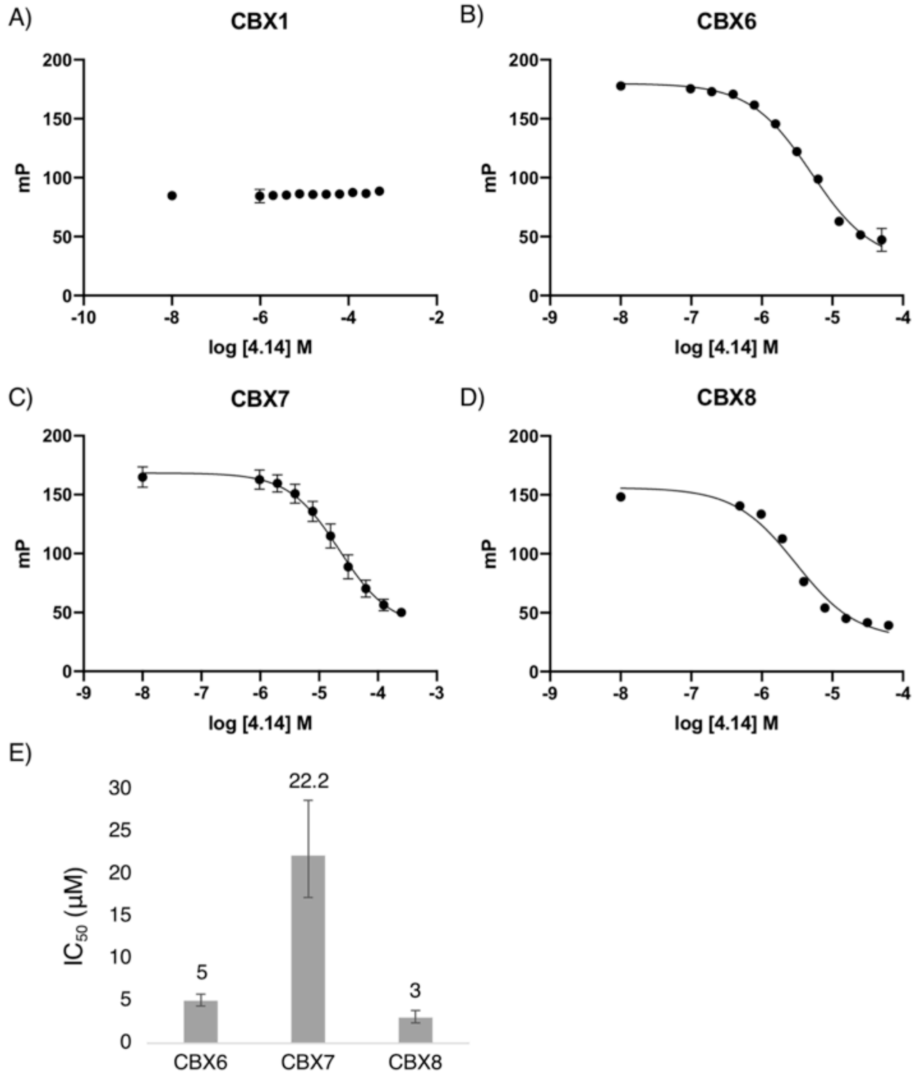


Figure 4.39. Competitive fluorescence polarization data of compound **4.14** with CBX proteins performed in triplicate. A) CBX1 (logIC$_{50}$ = > –3.301 M) B) CBX6 (logIC$_{50}$ = –5.300 M, 95% CI –5.362 M to –5.237 M, R^2 = 0.993), C) CBX7 (logIC$_{50}$ = –4.654 M, 95% CI –4.768 M to –4.538 M, R^2 = 0.977), D) CBX8 (logIC$_{50}$ = –5.518 M, 95% CI –5.620 M to –5.414 M, R^2 = 0.976), E) Bar graph of IC$_{50}$ values (μM). Error bars reported as asymmetrical 95 % confidence intervals.

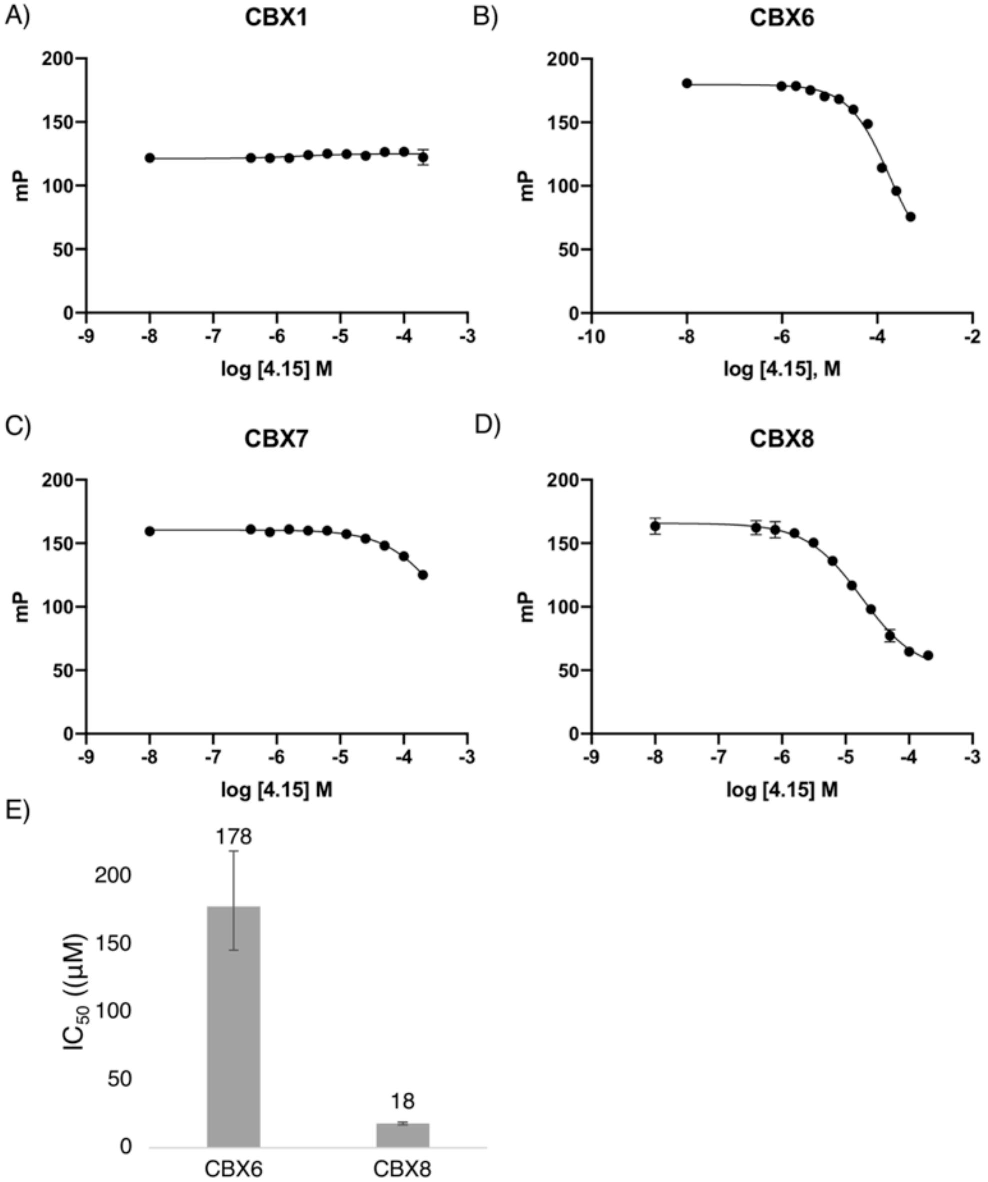


Figure 4.40. Competitive fluorescence polarization data of compound **4.15** with CBX proteins performed in triplicate. A) CBX1 ($\log IC_{50}$ = > –3.699 M) B) CBX6 ($\log IC_{50}$ = –3.750 M, 95% CI –3.837 M to –3.659 M, $R^2 = 0.989$), C) CBX7 ($\log IC_{50}$ = –3.43 M, 95% CI –3.458 M to –3.400 M, $R^2 = 0.983$), D) CBX8 ($\log IC_{50}$ = –4.751 M, 95% CI –4.824 M to –4.678 M, $R^2 = 0.989$), E) ABar graph of IC_{50} values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

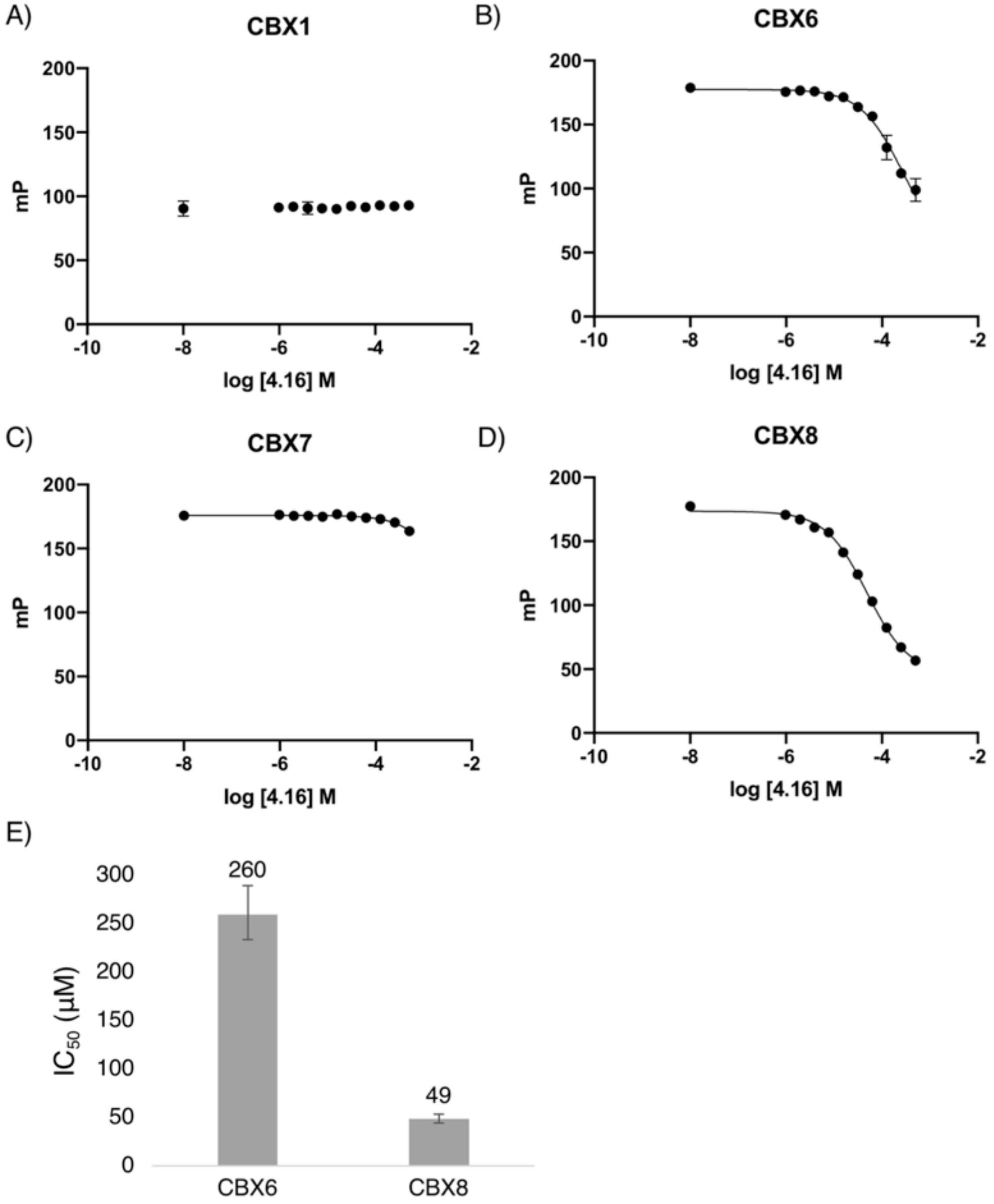


Figure 4.41. Competitive fluorescence polarization data of compound **4.16** with CBX proteins performed in triplicate. A) CBX1 (logIC$_{50}$ = > –3.301 M), B) CBX6 (logIC$_{50}$ = –3.582 M, 95% CI –3.631 M to –3.522 M, R^2 = 0.973), C) CBX7 (logIC$_{50}$ = > –3.301 M), D) CBX8 (logIC$_{50}$ = –4.315 M, 95% CI –4.356 M to –4.274 M, R^2 = 0.997), E) Bar graph of IC$_{50}$ values (µM). Error bars reported as asymmetrical 95 % confidence intervals.

4.7.9 Docking and MD methods:

3i91 crystal structure of CBX8 complexed with H3K9me3, and 3i90 crystal structure of CBX6 with H3K27me3 was used as the structures for docking. Ligand was removed, structure was cleaned, and 500 poses were generated using Autodock Vina for compound

4.12. UCSF Chimera was used to generate the initial PDBQT files for both host and inhibitors.

All molecular dynamics were done using the AMBER16 suite of programs. The generation of partial charges for non-standard residues in the inhibitors was done using a combination of Gaussian09 at the HF-6-31G* level of theory and the *residuegen* utility in AmberTools16. Simulation topologies and coordinates were set up in the xLEap graphical environment using the AMBER ff14SB force field.

For the molecular dynamics, the top five poses generated from docking were treated with minimization, temperature ramping to 300K over 200 ps, and equilibration for 10 ns at 300K in explicit solvent (approximately 6000 TIP3PBOX waters). The simulations were then run for 50 ns for data collection for a combined 250 ns of simulation time.

Each 50 ns simulation required 60 hours on 28 cores (1 node with 2 Intel Xeon E5-2680 v4 processors). Initial docking for the poses was done on a single 2.8 GHz Intel Core i5, requiring 12 hours of processing time for the generation of 500 poses for each inhibitor.

Funding for the computational studies was provided by the National Science and Engineering Research Council of Canada and the British Columbia Knowledge Development Fund. Research was performed in part using the Compute Canada and Westgrid computing resources.

4.7.10 Methods for cellular Assays

Cell proliferation assays

G401 cells were seeded at 0.1×10^6 cells/mL in 24-well flat bottom cell culture plate (#353047, Corning). Peptides **4.3** and **4.6** (100 μM) were added to the cells. Cells were grown for 72 h before trypsinized and counted. Cell suspension was homogenized by gentle pipetting and counted using Countess II (#AMQAF1000, Life Technologies). The media was then exchanged with fresh media containing DMSO, **4.3** or **4.6**. At day 6, 25 percent of cells in each well were split and transferred to new wells to avoid over-confluency, and media was replaced with fresh compounds in new media. The cell viability assay was extended to 12 days, with cell counting and fresh compound replenishment at day 3, 6, 9.

CellTiter-Glo luminescent cell viability assay

The effect of **4.3** and **4.6** on cell viability was determined using a CellTiter-Glo ATP detection system (#G7573, Promega). G401 cells were seeded in 0.1 × 10^6 cells/mL density in 96-well clear bottom white microplate (#655098, Greiner Bio-One). Cells were treated with compounds **4.3** and **4.6** for either 72 hours or 12 days (described in Figure 4.7), with fresh compounds replenishment at day 3, 6, 9 for the 12 day treatment. CellTiter-Glo reagent was added to cells, and incubated with gentle shake for 15 minutes in dim light at R.T. Luminescence was read on an GloMax® microplate reader. Luminescence was normalized to DMSO-treated groups.

Quantitative reverse transcriptase polymerase chain reaction (qRT-PCR)

mES cells were treated with **4.3** or **4.6** at 100 µM or DMSO (1%) for 48 h. Cells were harvested after 48 h for RNA extraction. After homogenization of mES cells using TRIzol reagent (Thermal Scientific), RNA was extracted from the aqueous phase in the phase separation step. RNA pellet was washed with 75% ethanol and concentrated for subsequent reverse transcription. 2 µg RNA was then converted into cDNA using Verso cDNA synthesis kit (Thermo Scientific). SYBR Green Mastermix (Thermo Scientific) was used for quantitative PCR.

Primers for mouse genes *Essrb*, *Nanog*, *Klf4*, *Rpo*, *Oct4*, *Rest* and *Nr0b1* as control are used in the qPCR.

Gene name	Forward primer (5'-3')	Reverse primer (5'-3')
Esrrb	GGCGTTCTTCAAGAGAACCA	TCCGTTTGGTGATCTCACAT
Nanog	AGGCTGATTTGGTTGGTGTCG	CCAGGAAGACCCACACTCAT
Klf4	CAGCCATGTCAGACTCGCC	GTTTTTAATCTTCGTTGACTTTGGG
Rpo	TTCATTGTGGGAGCAGAC	CAGCAGTTTCTCCAGAGC
Oct4	GAGGAGTCCCAGGACATGAA	AGATGGTGGTCTGGCTGAAC
Rest	GTGCGAACTCACACAGGAGA	AAGAGGTTTAGGCCCGTTGT
Nr0b1	TCCAGGCCATCAAGAGTTTC	ATCTGCTGGGTTCTCCACTG

Chapter 5. Installation of methyllysine mimics on phage-displayed peptide libraries

Contributions

I conceived of the experimental design and carried out the synthesis, phage experiments and data analysis. Phage related protocols were adapted from those used by Dr. Ratmir Derda at the University of Alberta, and he advised on experimental set up and analysis of data. I amplified the phage library and clone with assistance from Michael Gignac and Katrina Vizely. I conceived of the synthesis and collection of small molecule alkylating agents with synthetic assistance from Lee Treanor and Jacob McCallum. I carried out the peptide alkylation studies, and purification of peptides. I carried out and designed the phage alkylation experiments, with assistance from Katrina Vizely, Sarah Khan and Roko Nikolic. Katrina Vizely optimized the iTCEP protocol for phage alkylation. I planned and wrote the manuscript with editing help from Dr. Hof.

5.1. Foreword

My previous research has focused on rational design of inhibitors for the CBX polycombs. Several attempts of high-throughput and virtual screens against CBX proteins have been attempted, and have not yet produced potent inhibitors. My work has resulted in many potent peptidomimetic CBX inhibitors, but the methods are time-consuming and not readily translated to other methyl reader families. This motivates an innovative approach that utilizes massive chemical diversity in a way that is targeted to methyllysine reader proteins. The motivation for this work was to create a method to screen large libraries of molecules against the CBX proteins and to have the method be applicable to a large variety of methyllysine readers.

5.2. Introduction

Selectively targeting each member of the CBX family has proven challenging and time consuming. Traditional SAR studies, which involve iterative cycles of synthesizing small sets of analogs, and then testing them in order to guide subsequent designs, have yielded results at a slow pace. High-throughput approaches to studying selectivity and diversity are necessary to efficiently target these highly similar proteins. The challenges associated with targeting CBX proteins apply to many families of epigenetic reader proteins including methyllysine readers as a whole (reviewed in Chapter 1).

New methodologies are needed to successfully target methyllysine reader proteins by chemical inhibition. Combinatorial chemical libraries leverage the power of using a large and diverse chemical space in drug screening. Combinatorial approaches such as one-bead one-compound (OBOC), small-molecule combinatorial libraries, DNA-encoded chemical libraries (DELs), mRNA display and phage display have stood out as significant advancements for discovery of next-generation drugs.[256] For methyl reader proteins, two examples of the OBOC approach, involving libraries of up to 14112 compounds,[156, 210] and one example of DEL (≤ 96 compounds)[238] have been reported.

Bacteriophage ("phage") display is a powerful technique for discovery because of its use of massively diverse libraries to identify novel peptidic ligands.[257-258] Phage can be genetically encoded to express or 'display' highly variable peptide sequences on their surface coat proteins.[259] Phage display peptide libraries are similar to OBOC libraries in the sense that each phage can display one unique peptide sequence (one-phage one-peptide). Phage display libraries routinely involve $>10^8$ peptides, allowing for high-throughput screening of protein-peptide interactions.[260]

Screening of massive peptide phage libraries utilizes the direct connection between the displayed peptide on the surface of the phage and the DNA that encodes its sequence within the phage particle. Identification of phage-displayed peptide ligands for a given target is done using affinity selection by a process called biopanning.[261] Biopanning involves incubating a peptide library with a target protein and washing away any of the peptide-phage that do not bind. Phage that bind to the target protein are then eluted, amplified in bacteria and characterized by sequencing the DNA of the phage. This process has been used extensively to identify peptide ligands with specific binding selectivities and pharmacological activities.[257, 262-263]

The main limitation of genetically encoded peptide libraries is that the peptides expressed can only be encoded to contain the 20 natural amino acids. Chemical modifications to amino acid residues can overcome this limitation by adding new chemical entities.[264] Recombinant techniques allow for the creation of diverse peptide libraries and chemists have developed reactions that are biocompatible with phage to increase the diversity and functionality of these libraries.[265-267]

Despite the extensive use and popularity of phage display since its invention in 1985, chemical reactions that expand the diversity and applications of the technique are limited. Chemical reactions on phage-displayed peptides take advantage of reactive native amino acid residues such as serines, lysines, tyrosines, aspartic and glutamic acids, the N-terminus and cysteines.[265-267] Ideal labeling reactions need to be site-specific, and occur in aqueous media and under mild conditions that maintain the integrity of the phage construct. The cysteine side-chain’s soft nucleophilic thiol is a useful reactive handle because it can undergo bioorthogonal reactions with soft electrophiles. Electrophiles such as maleimides,[268-269] α-halocarbonyls,[270-272] benzyl halides[273-274] and haloarenes[275] have been used for selective reactions of cysteine on phage.

To target methyllysine reader proteins using phage display, methylated lysines would need to be present on the peptidic phage libraries. But the lack of methylation pathways in *E. coli* and in viruses mean that methyllysines can’t be directly introduced into a phage library by genetic methods. Chemical reactions on cysteine thiols have been reported to create Cys-derived methyllysine mimics on purified proteins and histones (Figure 5.1).[276] Despite this technology having >260 citations, to the best of my knowledge there are no examples of this chemistry being used outside of its originally reported context (alkylation of denatured histones and histone-derived peptides). Installation of Cys-derived methyllysine mimics on phage constructs has not been reported.

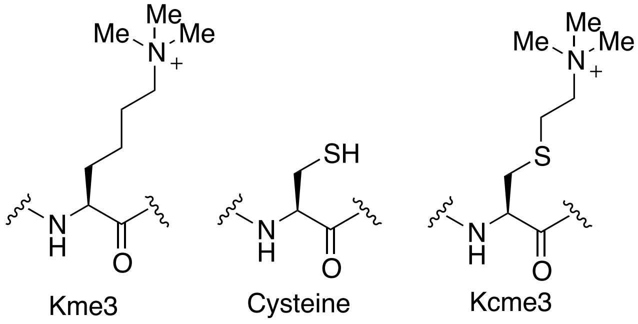


Figure 5.1. Trimethyllysine (Kme3) mimic created from cysteine residue.

To create Kcme3 peptidic inhibitors that will have increased potential as cell permeable chemical tools, we envisioned that other lysine mimics beyond Kcme3 can be synthesized and reacted with phage libraries (Figure 5.2A). During the course of this work, the installation of 4 Cys-derived methyllysine analogs (including Kcme3) was reported.[277] Similar to the initial invention,[276] this method only involves reactions on denatured histone proteins. We sought to install both Kcme3 and novel Cys-derived methyllysine mimics on peptides and phage using (2-haloethyl) amine derivatives. These alkylating agents will readily react with cysteine residues to create Kme3 mimics on peptides and peptide phage libraries.

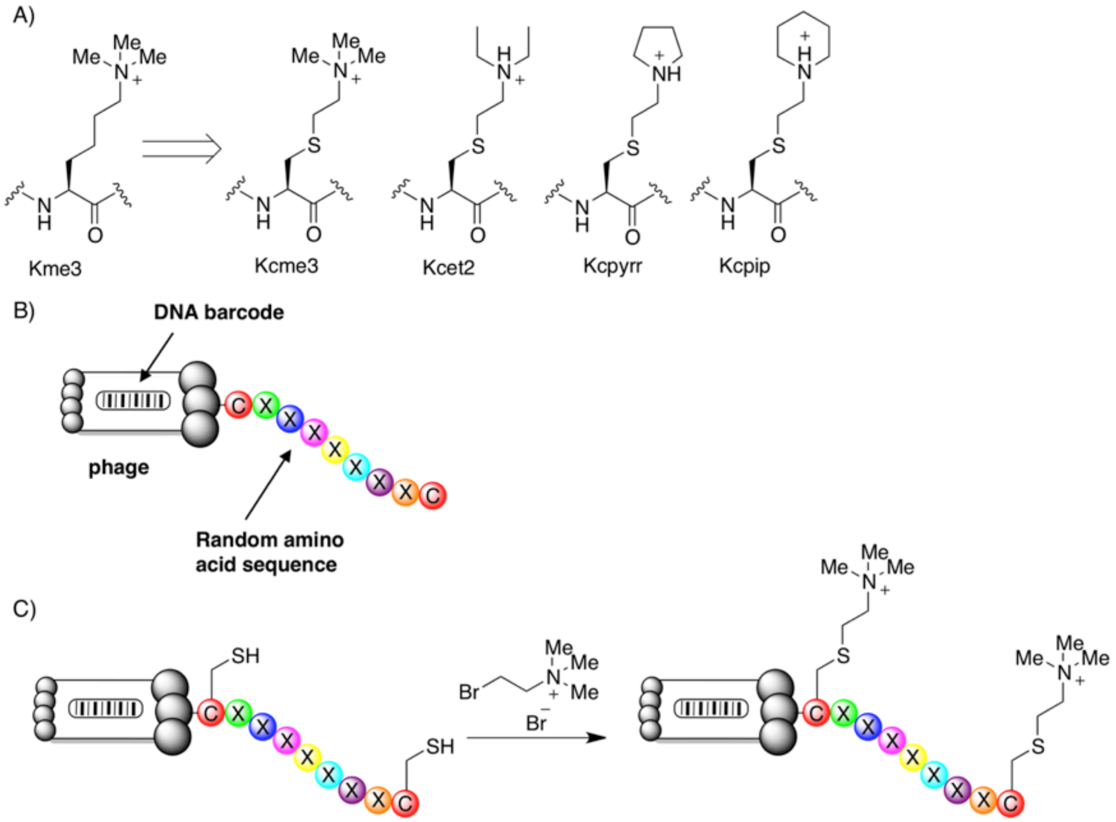


Figure 5.2. Cysteine methyllysine mimics and installation on bacteriophage. A) Cysteine methyllysine mimics including a trimethyllysine mimic (Kcme3), diethyllysine mimic (Kcet2) and a pyrrolidine lysine mimic (Kcpyr), piperidine lysine mimic (Kcpip), B) Cartoon of M13 bacteriophage with C7C peptide, C) Depiction of cysteine alkylation on C7C phage using BrNMe3.

To achieve successful installation of Kcme3 residues we chose to use a M13 Ph.D^{TM}-C7C phage display library and M13 Ph.D-C7C clone. M13 phage are one type of filamentous phage that infect *E. coli* and are able to replicate and produce new virions in the host without lysing the host cell. The minor coat proteins on both ends of the particle are essential to host infection (Figure 5.3).[259] M13 phage are commonly used for phage display because of their simple structure and genome which consists of 11 genes (Figure 5.3).[257-258] In the C7C library construct, the gene encoding the phage coat protein pIII is modified to encode for a randomized sequence of seven amino acids flanked by a cysteine residues on either end (CXXXXXXXC) (Figure 5.2B). In the C7C clone construct, a single sequence of CPARSPLEC is present. The clone is easier to use for reaction development, while making the assumption that reactivity of the single clone will represent the reactivity of the large part of the full library. Each of these phage systems are oxidized and cyclized in their disulfide form, and can be reduced to obtain two free Cys thiols for further chemistry.

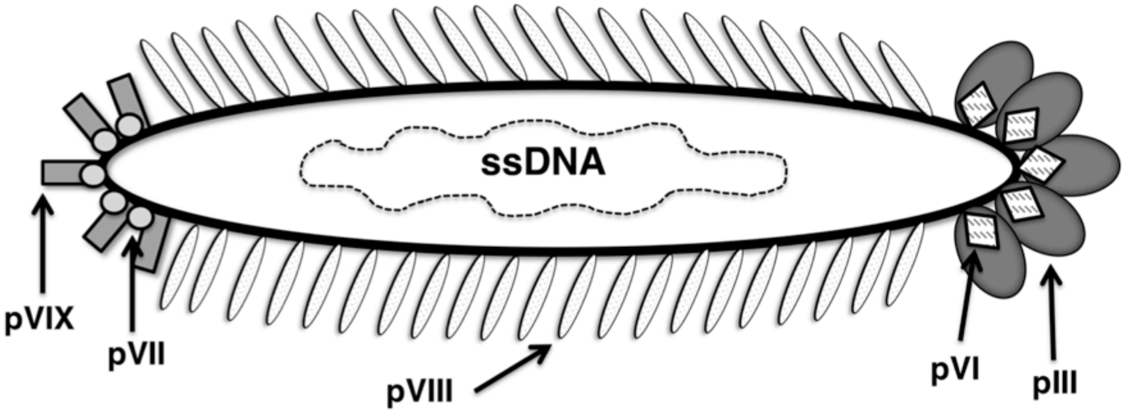


Figure 5.3. Cartoon of filamentous M13 phage particle. A single phage particle is approximately 6.5 nm in diameter and 1 μm in length. The particle is covered in many copies of the cylindrical pVIII major coat proteins. One end of the particle contains between 3-5 copies of the pVIX and PVII proteins. 3-5 copies of the pVI and pIII proteins cap the opposite end of the particle. The single stranded DNA genome is encapsulated within the particle.

This chapter will focus on my work towards creating massive libraries of phage-displayed peptidic inhibitors containing methyllysine mimics. The immediate goals of this project were to create a methodology for the reactions themselves, as well as a method for the evaluation of reaction success on phage libraries. Extensive optimizations of reaction conditions, including reagent temperature, reducing conditions, reagent concentrations, and pH were carried out in order to achieve useful reaction yields on viable phage particles, while limiting phage toxicity.

5.3. Synthesis of cysteine alkylating agents

To create post-translationally methylated peptide libraries, we designed a small library of alkylating agents consisting of small molecules with tertiary or quaternary amines attached to primary alkyl halides. Alkylating agents were used as purchased (**5.1**–**5.3**) or synthesized from (2-chloroethyl) amine starting materials (Figure 5.4). Nucleophillic substitution of the starting materials **5.2** and 1-(2-chloroethyl)piperdin-1-ium chloride with NaI achieved compounds *N,N*-diethyl-2-iodoethanaminium iodide (INEt2, **5.4)** and 1-(2-iodoethyl)piperdin-1-ium iodide (INpip, **5.5)** in yields of 60-80%.

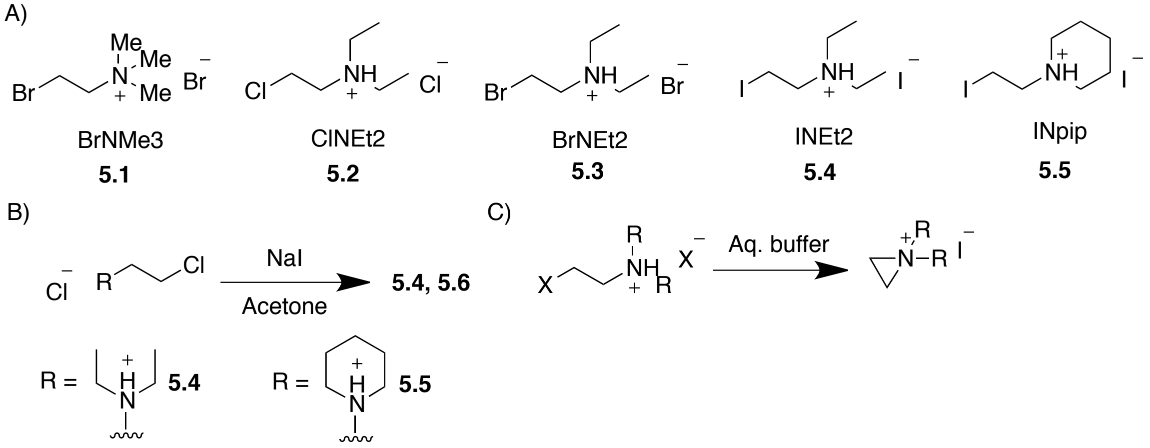


Figure 5.4. Cysteine alkylating agents for the installation of Kme3 mimics. A) Alkylating agents, B) Synthesis of (2-iodoethyl) amines, C) Cyclization of **5.2**-**5.5** in aqueous solutions to form aziridinium ions.

(2-haloethyl) amines cyclize in aqueous solution to form aziridinium ions.[278] Compounds **5.2** and **5.3** are known to undergo a intramolecular nucleophilic substitution reaction to form aziridinium ion intermediates (Figure 5.4C).[278] We predicted that compounds **5.4** and **5.5** would also cyclize to form aziridinium ions. Mass spectrometry and NMR analysis of compounds **5.4** and **5.5** indicated that each compound was the uncyclized 2-iodoethyl species when isolated from acetone or dichloromethane. Upon dissolving in phosphate buffered D_2O (pH 8), compound **5.4** cyclizes to form the aziridinium species (Figure 5.4C, Figure 5.5). Although both iodoethyl and aziridinium species are reactive towards cysteine thiols, we predict that the aziridinium ion is the active species in our reaction buffers. Compound **5.5** also cyclizes in phosphate buffered D_2O and appears as a mixture of the cyclized and uncyclized product in D_2O (Figure 5.6). 1-(2-chloroethyl)piperidine (the chlorinated analog of **5.5)** has been previously isolated as the aziridinium salt and as the dimeric piperazinium salt.[279-281]

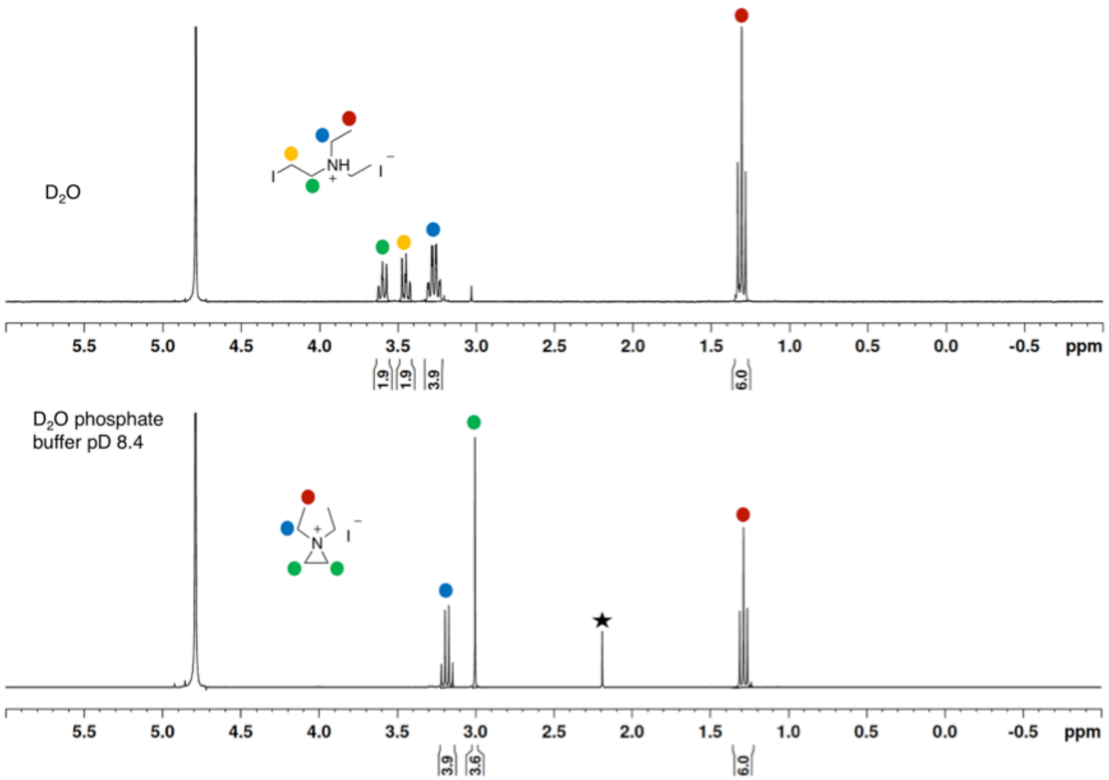


Figure 5.5. Cyclization of **5.4** in aqueous buffer to form 1,1-diethylaziridin-1-ium iodide. Top: ^{1}H NMR of **5.4** in D$_2$O. Bottom: ^{1}H NMR of cyclized **5.4** in phosphate buffered D$_2$O pD 8.4. Circles code for proton signals. Solvent impurity indicated with a black star.

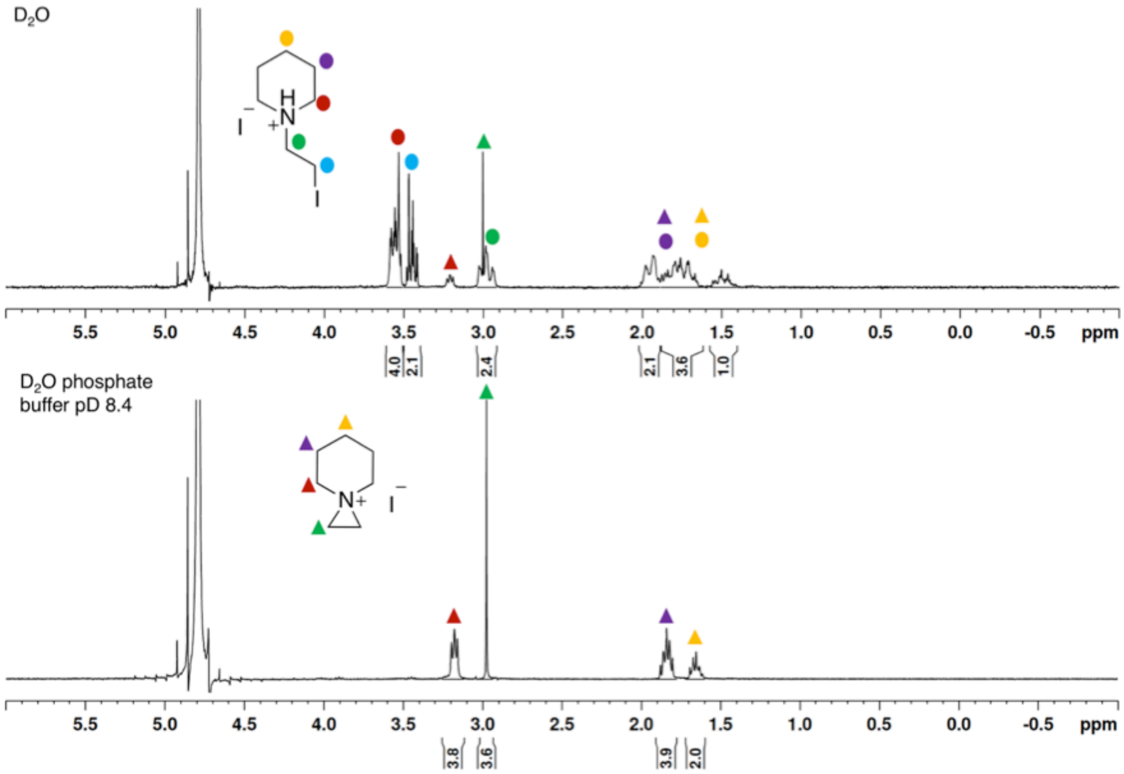


Figure 5.6. Cyclization of **5.5** in aqueous solutions to form 3-azaspiro[2.5]octan-3-ium iodide. Top: ^{1}H NMR of **5.5** in D$_2$O. A mixture of 5.5 and 3-azaspiro[2.5]octan-3-ium is observed. Bottom: ^{1}H NMR of cyclized **5.5** in phosphate buffered D$_2$O pD 8.4. Circles code for proton signals of **5.5** and triangles code for the cyclized product 3-azaspiro[2.5]octan-3-ium.

5.4. Alkylation reactions on model peptides

To test our synthetic approach towards creating cysteine Kme3 mimics we carried out alkylation reactions on model peptides using BrNMe3 (compound **5.1**, Figure 5.7). The 7-mer peptide SWCDYRC was used in a series of alkylation reactions with conditions modified from the Shokat protocol for modifying histones.[276] In an attempt to find conditions compatible with phage, we initially used a low concentration of **5.1** and low temperature. After a 30-minute reduction with TCEP at 21°C, 100 mM **5.1** was added to the solution and no reaction was observed after 1 hour and after 18 hours (supplementary Figure 5.24). We next tested the reaction using similar conditions to those previously reported.[276] The peptide was reduced for 2 hours with DTT at 37°C, followed by a 3.5 hour reaction with 200 mM **5.1** incubated at 50°C. After 3.5 hours, the mono-alkylated product (1Kcme3) was observed in a 1:1 ratio with the starting material

SWCDYRC peptide (Figure 5.8). Toxicity screens with M13 bacteriophage with 37°C and 50°C incubations in the presence of DTT and 1 M **5.1** showed that these higher-temperature conditions almost completely eliminated phage viability, showing that they're not compatible with the M13 C7C phage clone.

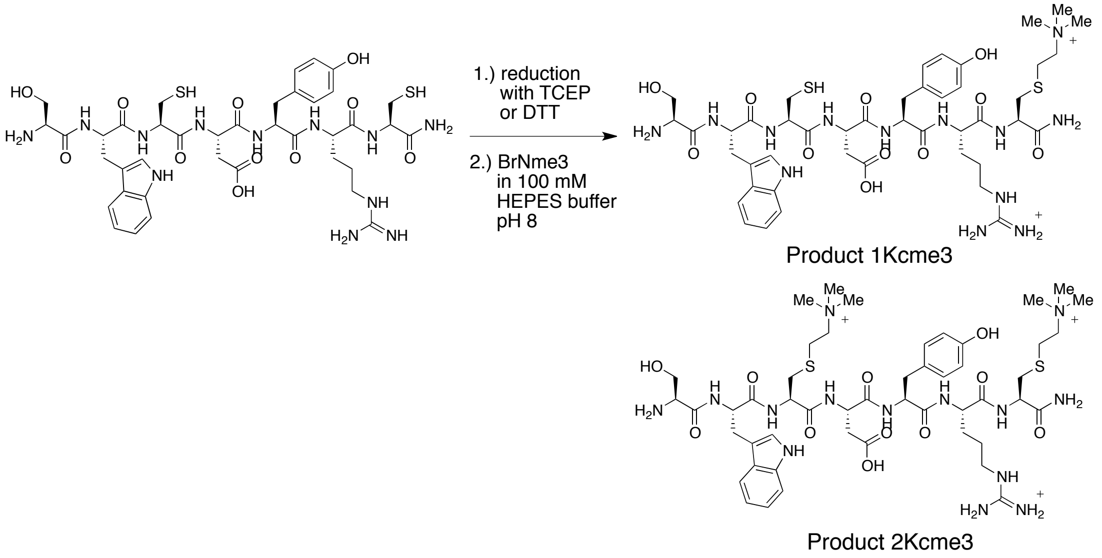


Figure 5.7. Reaction scheme of cysteine alkylation with **5.1** on model peptide SWCDYRC.

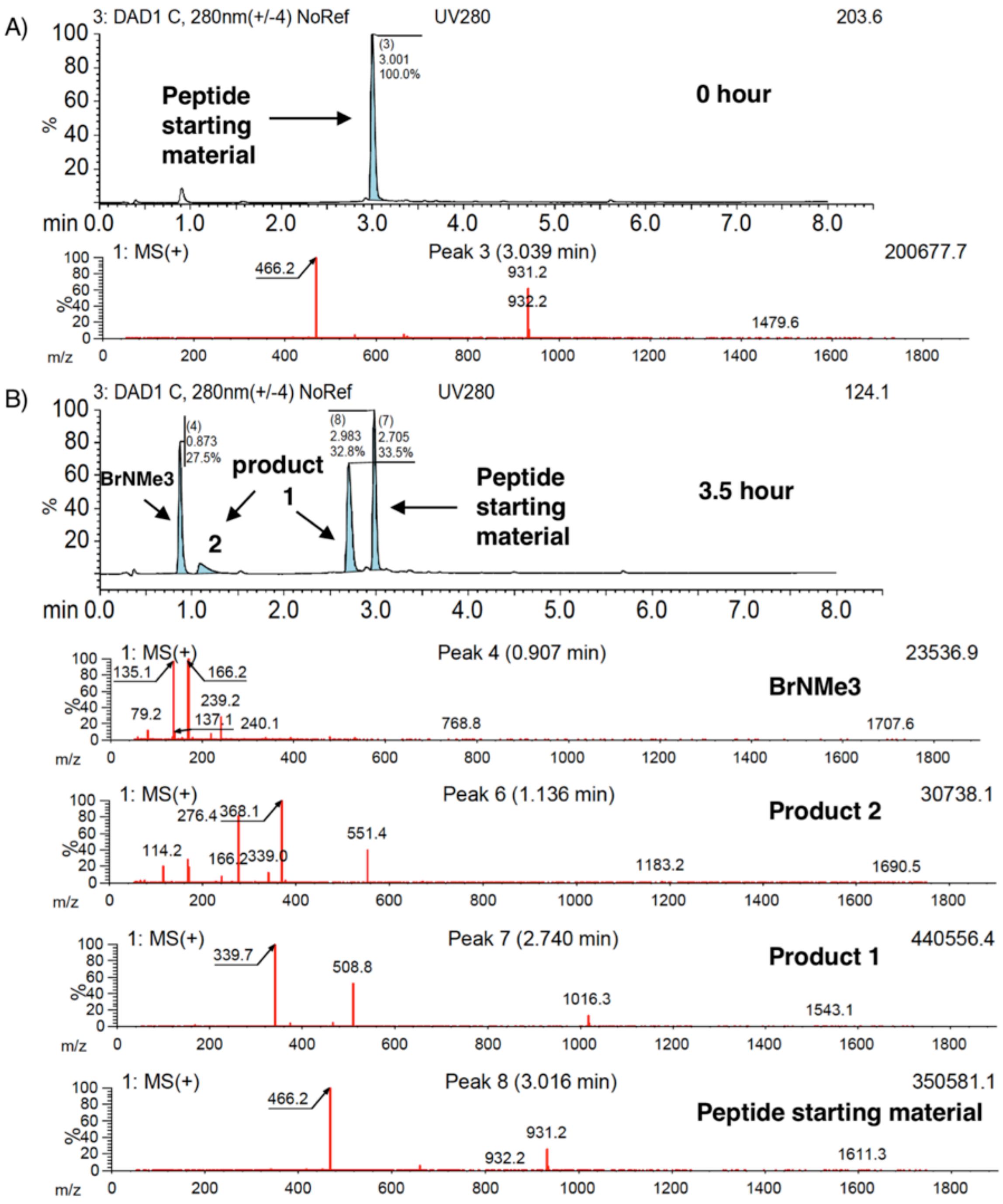


Figure 5.8. Monitoring reaction of **5.1** with SWCDYRC peptide using LCMS. SWCDYRC (2 mM) was reduced for 2 hours at 37°C with 40 mM DTT in HEPES buffer (100 mM pH 8). **5.1** was added (200 mM) and the reaction was incubated at 50°C for 3.5 hours. A) LCMS trace with UV280 and MS+ of reaction at 0 hour (reduced peptide before addition of **5.1**), B) LCMS trace with UV80 and MS+ of reaction after 3.5 hours.

Optimization of the reaction conditions resulted in the successful installation of methyllysine mimics onto the peptide, without requiring elevated temperatures. Increasing the concentration of **5.1** was critical to the success of the reaction. Reduction with DTT for 30 minutes followed by a 3.5 hour incubation with 1 M **5.1** at room temperature resulted in a mixture of product 1 (1Kcme3) and 2 (2Kcme3). The reaction mixture contained a combined 80% of product 1 and 2 and 20% starting material peptide (Figure 5.9). Addition of a catalytic amount of NaI did not significantly change the reaction yield (supplementary Figure 5.25).

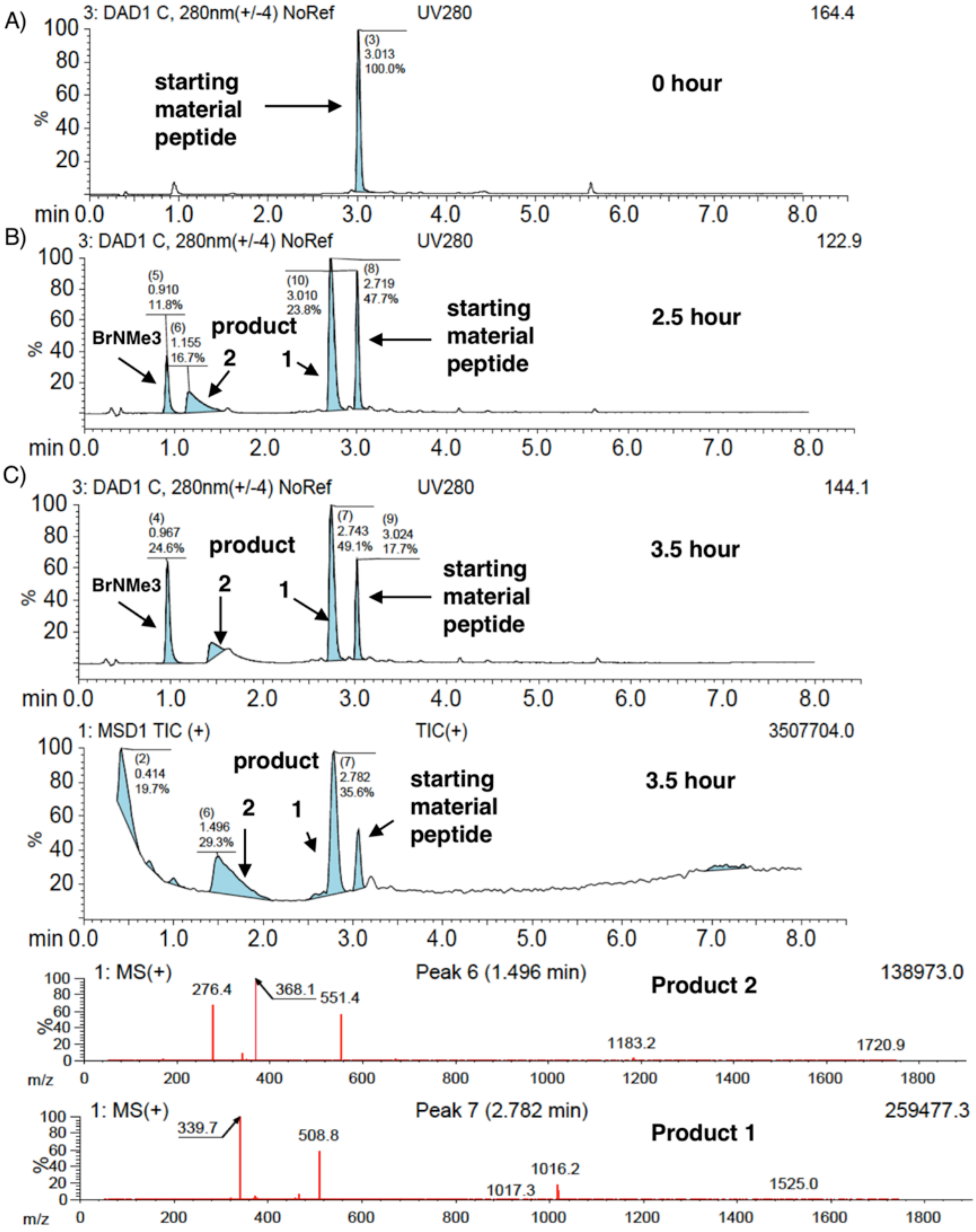


Figure 5.9. Optimized reaction of **5.1** with SWCDYRC peptide resulting in 80% of product 1 and 2. SWCDYRC (2 mM) was reduced for 30 minutes with 40 mM DTT in 100 mM HEPES buffer at room temp. **5.1** was added (1 M) and the reaction was incubated at room temp for 2.5 and 3.5 hours. A) LCMS trace with UV280 and MS+ of reaction at 0 hour (reduced peptide before addition of **5.1**), B) LCMS trace with UV80 and MS+ of reaction at 2.5 hours, C) LCMS trace with UV80 and MS+ of reaction at 3.5 hours.

We verified the alkylation of a cysteine containing peptide with compounds **5.4** and **5.5** (supplementary Figure 5.26-5.29). A cysteine-containing sequence derived from the histone 3 peptide H3K4 (termed H3Kc4) was successfully alkylated with compounds **5.4** and **5.5**, requiring significantly lower concentrations compared to **5.1** (supplementary Figure 5.28, Figure 5.29). Yields of approximately 60-100% of the alkylated cysteine peptides were observed after 4 hours. Yields of the reactions are estimates and were not quantitatively determined due to poor separation of the alkylated product from the starting material peptide (supplementary Figure 5.28, Figure 5.29).

5.5. Quantification of reaction yields on C7C phage

Synthetic optimization for the installation of Kme3 isosteres on peptide phage libraries was optimized using a single clone M13 phage with the peptide sequence CPARSPLEC (R. Derda, University of Alberta) and a Ph.D^{TM} C7C library containing 10^9 unique peptide sequences (New England Bio Labs). The clone and library are each at concentrations of 10^{12}–10^{13} pfu (plaque forming units)/mL.

The yield of each alkylation reaction was tested using an Iodoacetyl-PEG2-Biotin (BIA) pull-down experiment. Following the alkylation reaction with the Cys-reactive Kme3 mimic, cysteine residues that were not successfully alkylated readily react with the iodoacetyl group of BIA (Figure 5.10A, B). Biotinylated phage are removed from solution following incubation with streptavidin magnetic beads using a protocol adapted from previously reported methods (Figure 5.10C).[275] Reaction yields for the original alkylation were determined by titering the phage before and after capture with streptavidin and counting the plaques (Figure 5.10D). The C7C clone and library contain a LacZ reporter in the genome. Blue plaques are formed following plating in bacterial lawn with galactosidase substrate X-gal that produces a blue colour upon being processed by the LacZ gene product. We assume that all unreacted free Cys residues are successfully reacted with BIA and pulled down by streptavidin, meaning that we can use the extent of reaction with BIA to determine the success of all reactions on the phage (Figure 5.10).

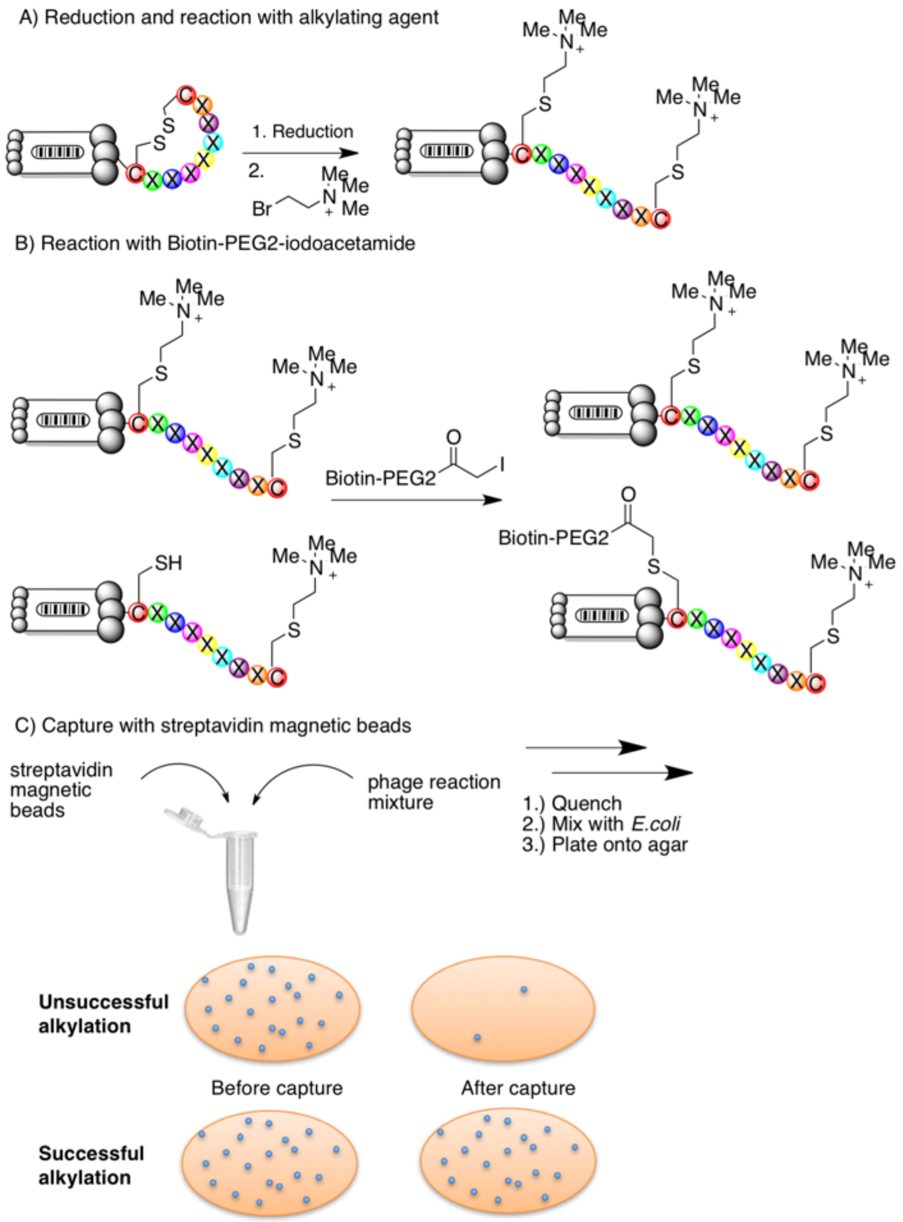


Figure 5.10. Scheme of alkylation reactions on phage and testing of percent alkylation. A) Phage are reduced and then subjected to an alkylating agent, B) Iodoacetyl-PEG2-Biotin (BIA) is added and reacts with any free un-reacted cysteine residues, C) Phage are added to streptavidin magnetic beads and any biotin labeled phage will be removed from the solution. Depiction of plaques formed from titering of phage in a unsuccessful alkylating reaction or a blank BIA reaction. These reactions would give rise to significant plaques visible for the plate representing before capture with streptavidin beads and few plaques seen for the titering after capture. Depiction of plaques formed from a successful alkylation reaction where approximately equal plaques would be present for before and after capture. If the majority of the cysteine residues were alkylated with **5.1**, they would not be biotinylated and would not be removed with streptavidin capture.

Alkylation of the phage was achieved in a two-step reaction with (1) reduction of the disulfide bonds and (2) dilution (to minimize toxicity of reducing agent) and incubation with **5.1** (or other alkylating reagent) (Figure 5.10). In addition to using the BIA pull-down for yield determination, each experiment was carried out alongside two separate controls. A 'BIA control' was used to determine the baseline for 100% capture (Figure 5.11A). The BIA control reaction is a one-pot reaction where the phage are incubated for 30 minutes with BIA and reducing agent. A high percent capture is expected from the BIA control due to the aforementioned high reactivity of the iodoacetyl group with thiols. Additionally, each experiment was conducted alongside a 'blank control' that was reduced in the same fashion as the experimental conditions and allowed to incubate for the same amount of time but without the addition of alkylating agent. The blank condition was followed by BIA capture, as normal, and therefore informed on toxicity of the alkylating agents and was used to determine reaction yields by comparing the capture of the blank to the capture of the experimental condition. Overall, we use the three samples to evaluate each reaction in the following way (Figure 5.11A, B):

- The # of phage captured in the BIA control determines how many phage were available to be alkylated in total.
- The # of streptavidin-captured phage from BIA reaction after the Blank control informs on how many phage were reduced and available for alkylation, and did not get lost due to re-oxidization of the disulfide bond during incubation.
- The # of streptavidin-captured phage after the Experimental condition determines how many phage were treated with alkylating reagent but were not successfully alkylated.

Subtraction of these amounts as shown in Figure 5.11A give us the % of phage that were lost to re-oxidation, and the % yield of phage that were successfully alkylated.

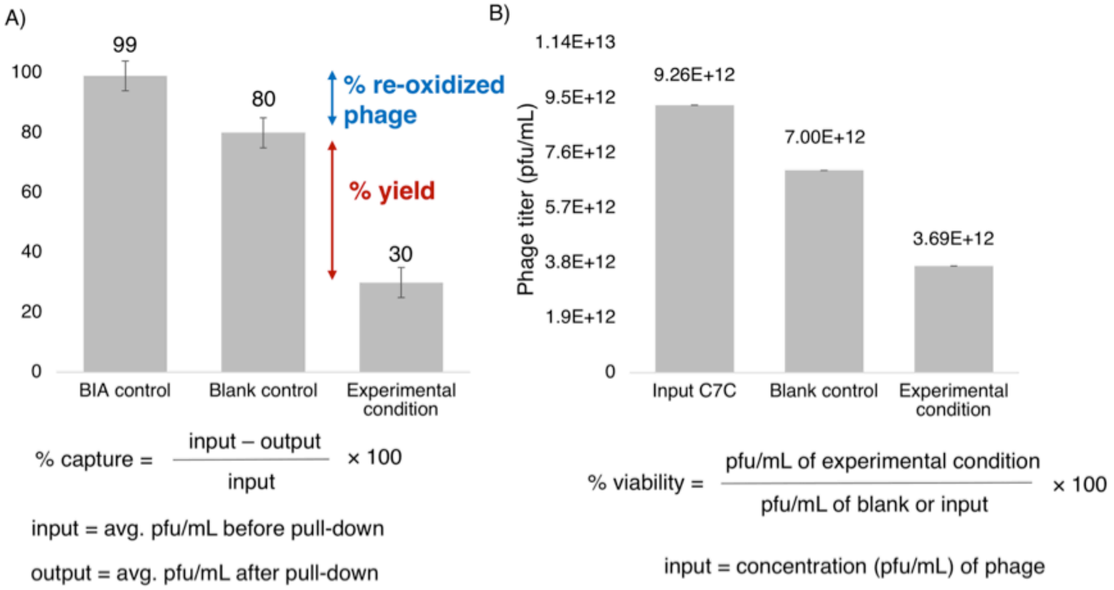


Figure 5.11. Depiction of data collected from alkylation reactions on C7C phage and determination of percent capture and viability. A) Percent capture of control and experimental conditions. Red arrow indicates reaction yield determined by the difference in capture of the control and experimental condition. Blue arrow indicates re-oxidized phage determined by the difference in capture of BIA control and blank control. Input = average pfu/mL before pull-down and output = average pfu/mL after pull-down. B) Concentration of phage (pfu/mL) is calculated for C7C phage used for each experiment. Input C7C is the concentration of phage going into each control and experiment. Percent viability is determined by the difference between the output pfu/mL of experimental condition divided by the input pfu/mL. Toxicity of each alkylating agent is measured by comparison to the blank control.

5.6. Optimization of reaction on C7C phage

Early efforts to optimize the reaction using **5.1**, **5.4** and **5.5** appeared to be successful. We observed variation in percent capture of the blank control depending on the pH and incubation time. Percent capture for the blank control ranged from 30-70% and yields of reactions were determined as the difference between the blank capture and experimental capture (Figure 5.11, Figure 5.12). Calculated yields ranged from 15-30% for alkylation reactions with **5.1**, and 40-50% for reactions with **5.4** and **5.5**. Consistent trends of increased viability with decreasing concentration were observed (Figure 5.12B).

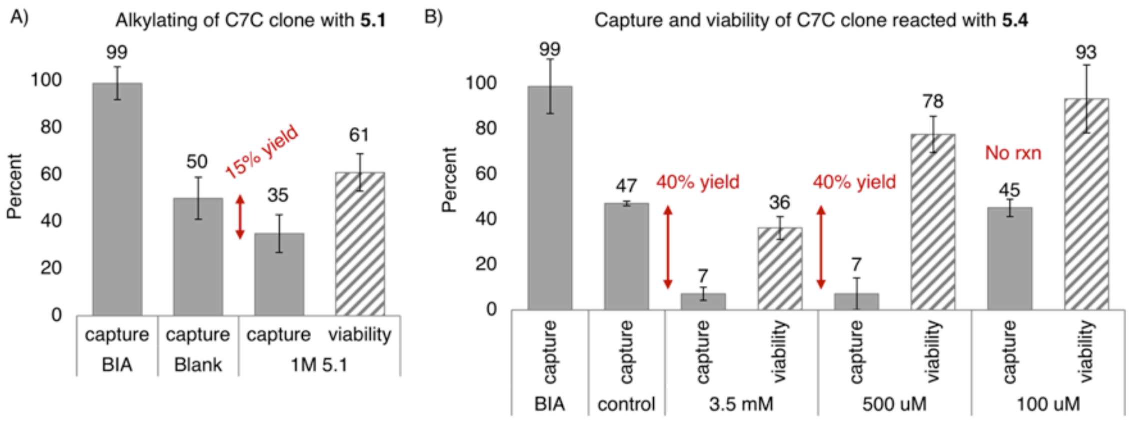


Figure 5.12. Percent capture and viability of reaction with **5.1** and **5.4** with the C7C clone. A) Phage were reduced for 30 minutes with 0.5 mM DTT in 100 mM Tris pH 8, followed by addition of 1 M of **5.1** and the reaction was incubated for 2 hours. B) Phage were reduced for 30 minutes with 0.5 mM DTT in 100 mM Tris pH 8, followed by the addition of **5.4** (100 µM-3.5 mM) and the reaction was incubated for 1 hour. Data in both A) and B) is the average of two independent experiments done on separate days, with each biotin-capture assay performed in triplicate. Error bars represent one SD.

Significant re-oxidation of the C7C phage peptides occurs during the reaction and interferes with determination of the reaction yield. Two sets of experiments were carried out in parallel where one of the control and alkylation conditions were pulsed with reducing agent during the reaction incubation (Figure 5.13). In the experiments without the pulse with DTT, 39% yield was observed compared to the blank (Figure 5.13A). In the experiments with a DTT pulse of 0.1 mM, equivalent capture of both the control and alkylation reaction were observed (Figure 5.13B). Addition of reducing agent during the incubation showed that what was observed as a reaction yield was actually re-oxidized C7C peptides that were not pulled down with the biotin-capture (Figure 5.13). Another potential explanation could be that DTT reacts with the alkylating agent and reduces the amount of alkylating agent available to react with the C7C phage peptides. These experiments demonstrated the need for more stringent reducing conditions and further optimization of reducing agent, to ensure that the C7C peptides remained reduced throughout the reaction incubation.

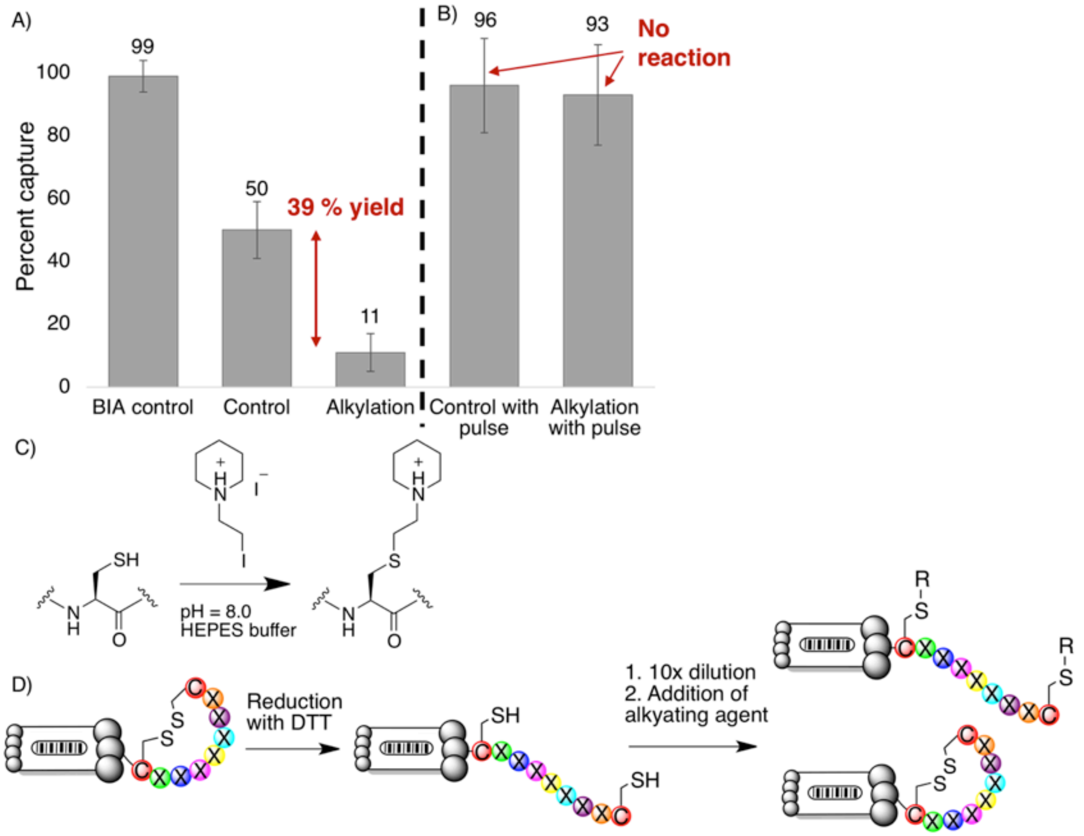


Figure 5.13. Pulsing of reducing agent during the reaction of **5.5** with C7C clone show re-oxidation of the C7C peptides and no reaction. A) Percent capture of experiment with **5.5**. B) Percent capture of experiment with further addition (pulse) of reducing agent, C) schematic of reaction with **5.5**, D) cartoon of phage showing reduction with DTT, followed by dilution and addition of alkylating agent and potential products. For experiments in A) and B) phage were reduced for 30 minutes with 0.5 mM DTT in 100 mM Tris pH 8. Following a 10× dilution, 500 µM of **5.5** was added and the reaction was incubated for 1 hour. 0.1 mM of DTT was added 30 minutes into the reaction in the conditions that were 'pulsed'. Data is an average of one experiment with titering of the pull-down performed in triplicate. Error bars are reported as % RSD.

5.7. Optimized reduction with pulsing

The pulsing method, in which reductants are added twice while in the presence of alkylating agents, led to significant toxicity that required further optimizations. Capture and viability of the C7C clone with **5.2-5.4** was tested with pulsing of reducing agent during the reaction. The C7C phage were reduced for 30-minutes (0.5 mM DTT) and incubated for 2 hours with concentrations (100 µM – 10 mM) of compounds **5.2-5.4**. At the one-hour time point the blank control and reaction condition were pulsed with 0.2 mM DTT. Compound **5.2** was toxic to phage at all concentrations tested (100 µM – 10 mM) (Figure 5.14A). Concentrations >500 µM of **5.3** abolished infectivity of the phage (Figure 5.14B). At all concentrations tested no reaction yield was observed. **5.4** did not significantly alkylate the C7C peptides at concentrations that retained phage infectivity (Figure 5.14C).

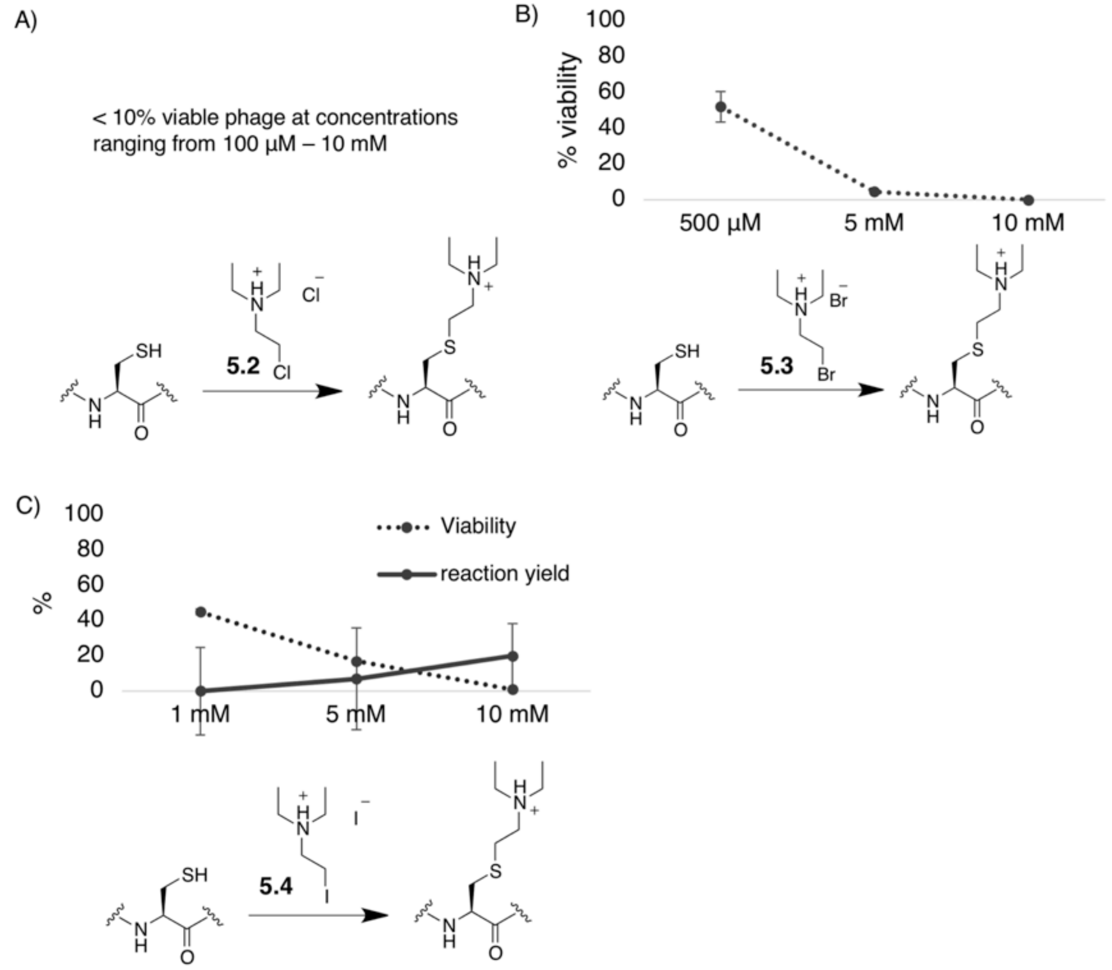


Figure 5.14. Alkylating agents **5.2**-**5.4** are toxic to phage in combination with increased amounts of reducing agent. Reaction yield and viability of C7C clone alkylated with **5.2**-**5.4** using ‘pulsing’ of reducing agent. All reactions were carried out in 100 mM HEPES buffer at pH 8. Phage were reduced with 0.5 mM DTT, followed by a 2 hour alkylation reaction with addition of 0.2 mM DTT after one hour. A) Reaction with **5.2** (100 µL - 10 mM) and resulted in less than 10% viable phage at all concentrations tested. B) Reaction with **5.3**, C) Reaction with **5.4**. Data is an average of one experiment with titering of the pull-down performed in triplicate. Error bars are reported as % RSD.

Compound **5.1** was significantly less toxic compared to **5.2**-**5.4** in reaction conditions that included further addition of reducing agent. Our initial alkylating reactions on the phage had shown viable phage with concentrations of **5.1** as high as 1 M.

Compound **5.1** was tested using the same conditions as described above for **5.2-5.4** with the exception of a DTT pulse at a lower concentration (0.1 mM). Comparison of the percent capture of the control and alkylation condition with DTT pulsing showed an approximate 20% yield. 50% viable phage were observed with pulsing with DTT compared to 85% in the reaction without pulsing (Figure 5.15). The viability of the phage with **5.1** and stringent reduction conditions motivated further optimization to increase reaction yield.

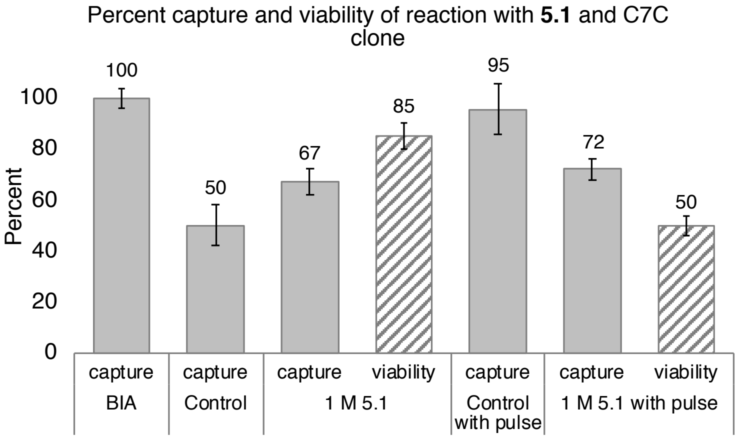


Figure 5.15. Viability of phage is retained in reactions with **5.1** and DTT pulse. Percent capture and viability of reaction of **5.1** with C7C clone were measured with and without DTT pulse. C7C phage were reduced for 30 minutes, diluted 10x and incubated for 2 hours with 1 M of **5.1** in Tris pH 8.5. The reactions were done in parallel with and without pulsing halfway during the reaction incubation with 0.1 mM DTT. Data is an average of one experiment with titering of the pull-down performed in triplicate. Error bars are reported as % RSD. Percent viability was determined compared to the pfu/mL of associated blank control.

Phage viability while under DTT/**5.1** reaction conditions was improved by raising the pH. Comparison of the C7C phage clone and library using DTT as the reducing agent showed viability to be highest at pH 6.1 (Figure 5.16A). Control BIA pull-downs using the same conditions showed that at pH 6.1 the C7C phage were not successfully reduced and alkylated by the BIA reagent (Figure 5.16B). The reaction with DTT and BIA showed increasing success with increasing pH as measured by percent capture (Figure 5.16B). This data demonstrated that reaction yields are unable to be determined when BIA and DTT are used in solutions of pH ≤ 7.

TCEP is less toxic compared to DTT, and is more effective at reducing the C7C peptides over a greater range of pH. TCEP gave higher reaction yields at pH 7 compared to DTT in the control BIA pull-downs (Figure 5.16B, D). TCEP was also found to be significantly less toxic compared to DTT (Figure 5.16C). No significant change in viability was observed by varying pH in the reaction using TCEP (Figure 5.16C). Reactions carried out with a TCEP pulse at either 0.05 mM or 0.2 mM showed decreased viability with higher concentrations of TCEP (Figure 5.16E). Determination of the percent capture of the reaction following BIA pull-down showed that the TCEP pulse at 0.05 mM was ineffective at maintaining reduced C7C peptides during the course of the reaction incubation (Figure 5.16F). Maintaining reduced C7C peptides for 1-2 hour reactions requires the addition of > 0.05 mM TCEP.

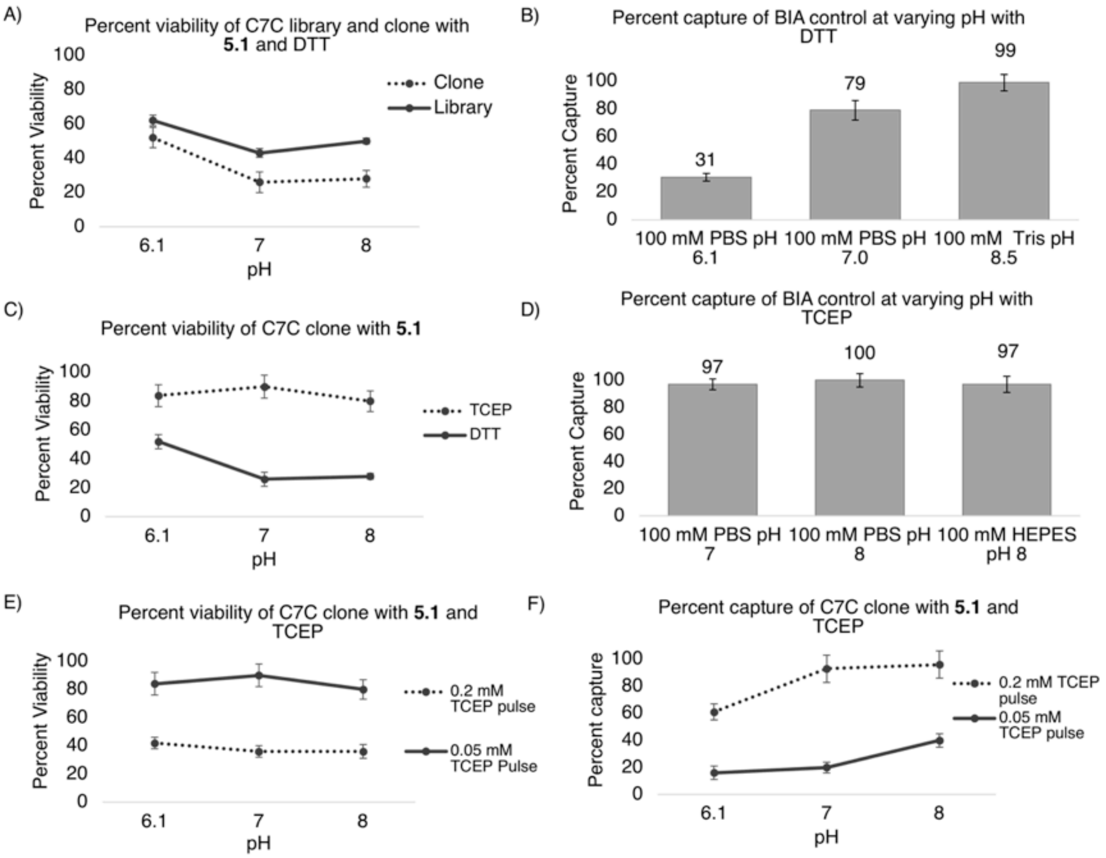


Figure 5.16. pH and reducing agent effect the viability of C7C phage and the efficacy of BIA pull-down. A) Percent viability of C7C library and clone with **5.1** and DTT, B) Percent capture after BIA pull-down control at varying pH with DTT, C) Percent viability of C7C clone with **5.1** with TCEP and DTT, D) Percent capture after BIA pull-down control at varying pH with TCEP, E) Percent viability of C7C clone with **5.1** and TCEP, F) Percent capture of reaction with C7C clone, **5.1** and TCEP. All viability tests were carried out with a 30 minute reduction with 0.5 mM reducing agent, 10-fold dilution followed by addition of **5.1** (1 M) for a one hour reaction incubation. Each reaction was pulsed with 0.05 mM reducing agent unless otherwise noted. Each BIA pull-down control was carried out as a 30 minute one-pot reaction with 0.5 mM reducing agent and 1 mM BIA. Data is an average of one experiment with titering of the pull-down performed in triplicate. Error bars are reported as % RSD.

Reduction using TCEP is superior for both alkylation yield and phage viability compared to DTT. A comparison of TCEP and DTT was done utilizing the 1.5 hour labeling incubation with 1 M **5.1** at pH 7.5 in 100 mM HEPES buffer (Figure 5.17). Each experiment was pulsed with 0.25 mM reducing agent. TCEP gave a 15% increase in reaction yield and a 20% increase in phage viability at pH 7.5 compared to the equivalent reaction with DTT. We chose to focus our future optimization efforts using TCEP as the reducing agent, with pulsing > 0.05 mM at pH > 7.

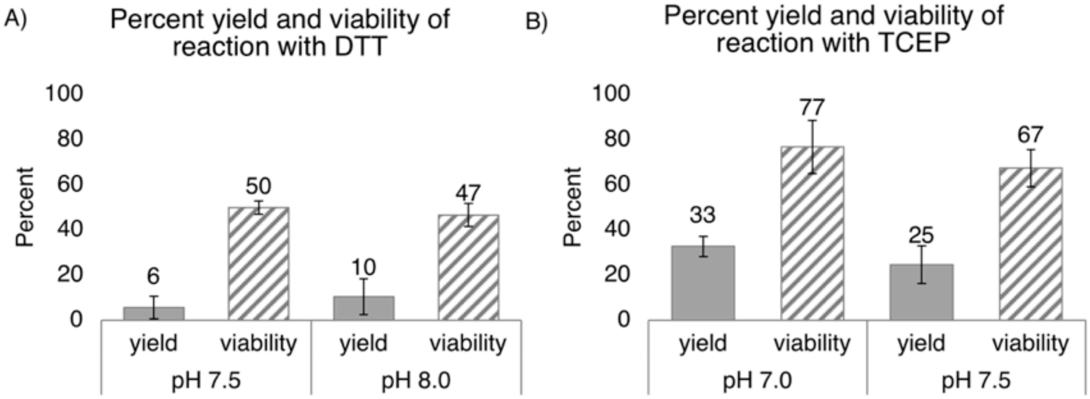


Figure 5.17. TCEP provides higher reaction yields and phage viability compared to DTT in reactions with **5.1** and C7C clone. A) Percent yield and viability of reaction with DTT at pH 7.5 and 8 with **5.1**. B) Percent yield and viability of reaction with C7C clone with TCEP and **5.1** at pH 7 and 7.5. All experiments shown in A and B underwent a 30 minute reduction with 0.5 mM reducing agent in 100 mM HEPES buffer, followed by a 10-fold dilution and addition of 1 M **5.1**. Each reaction went for 1.5 hours with the addition of 0.25 mM reducing agent halfway during the reaction. Data is the average of two independent experiments done on separate days, with each biotin-capture assay performed in triplicate. Error bars are one SD.

Higher yields of alkylation were observed by increasing time and pH. Using TCEP as the reducing agent and increasing the concentration of TCEP pulse, we observed a reaction yield of 40% at pH 7.5 after 4 hours (Figure 5.18A). We next tested different concentrations of **5.1** with a 4 hour reaction incubation, TCEP (once), and TCEP pulse with the reaction done at pH 7.5 (Figure 5.18B). Increasing concentration increased the yield while decreasing phage viability in a near linear fashion. At 1 M **5.1**, we observed a reaction yield of 35% with 54% viable phage.

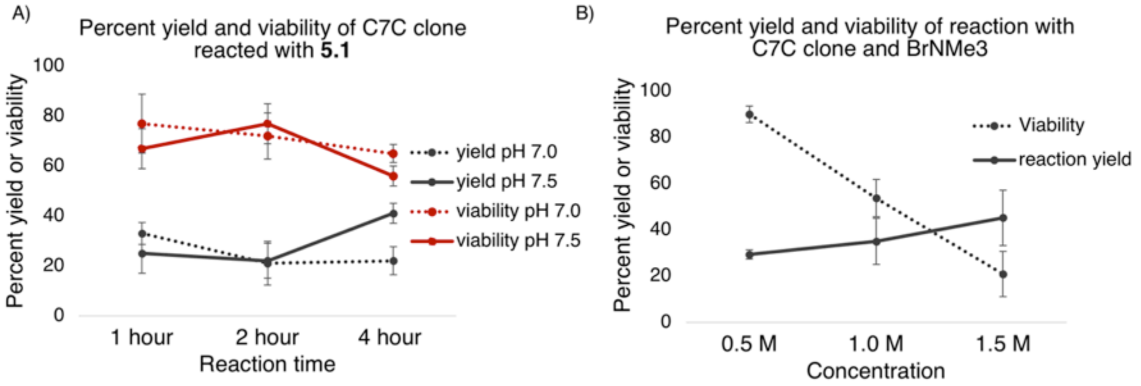


Figure 5.18. Higher reaction yields are achieved with increasing pH, reaction time and concentration of **5.1**. A) Percent yield and viability of reaction with 1, 2 and 4 hour reactions at pH 7 and 7.5 with 1 M **5.1**. Data is an average of one experiment with titering of the pull-down performed in triplicate. Error bars are reported as % RSD, B) Percent yield and viability of reaction with C7C clone and varying concentrations of **5.1** with a 4 hour reaction at pH 7.5. Data is the average of two independent experiments done on separate days, with each biotin-capture assay performed in triplicate. Error bars are one SD. All experiments shown in A and B underwent a 30 minute reduction with 0.5 mM TCEP, followed by a 10-fold dilution and addition of **5.1**. Reactions were carried out in 100 mM HEPES buffer and pulsed with 0.25 mM TCEP halfway during the reaction.

When using the optimized pH, reducing conditions, and alkylator concentration, increasing temperature decreases viability of phage and does not improve reaction yield. Two reactions were done in parallel at room temperature and at 37°C. The higher temperature showed a 6-fold decrease in phage viability with no improvement in reaction yield (Figure 5.19).

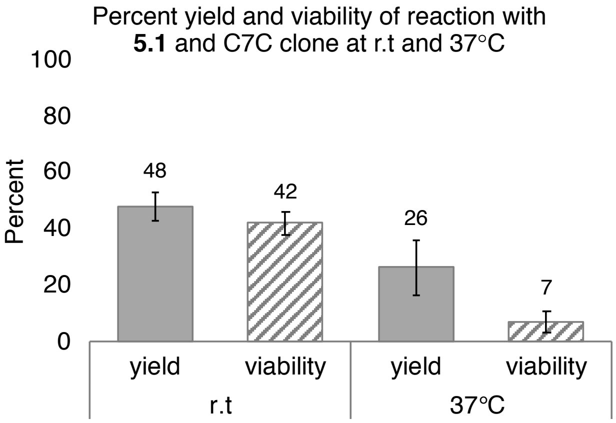


Figure 5.19. Increased temperature during reaction of C7C clone with **5.1** decreases phage viability and does not improve reaction yield. For both conditions reactions were carried out in 100 mM HEPES pH 7.5, reduced for 1 hour with 0.5 mM TCEP and incubated with 1.5 M **5.1** for 4 hours with addition of 0.25 mM TCEP after 2 hours. Data is the average of two independent experiments done on separate days, with each biotin-capture assay performed in triplicate. Error bars are one SD.

The success of the reaction with **5.1,** and the less toxic nature of **5.3** motivated the re-testing of **5.3** using the optimized conditions. We tested the reaction with **5.3** using TCEP and DTT. Incubation with **5.3** for two hours at pH 7.5 showed no significant reaction yield for both experiments (Figure 5.20). Phage viability of 50% was observed for both conditions with 500 μM treatment of **5.3**. We predicted **5.3** to have similar reactivity as **5.1** at pH 7.5. Toxicity of **5.3** limited the ability to test the reaction at higher concentrations comparable to reactions with **5.1**.

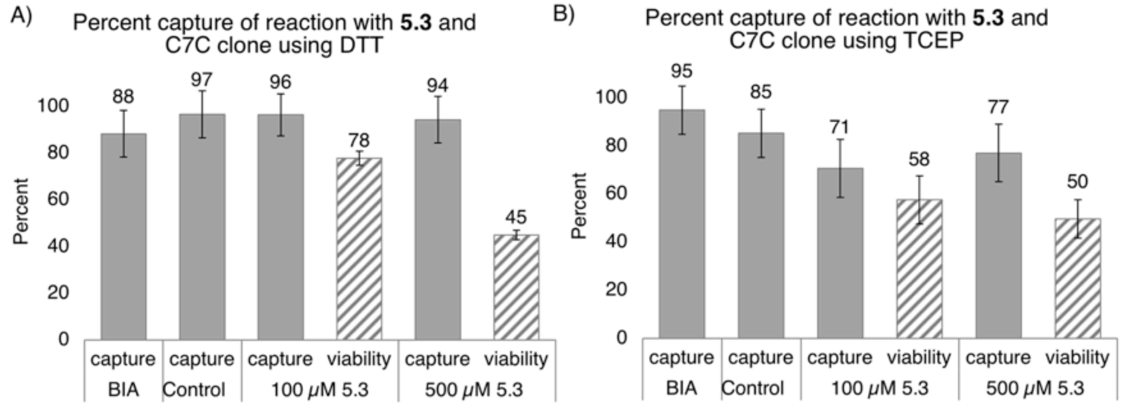


Figure 5.20. Optimized reaction conditions with **5.3** show no significant alkylation of C7C phage. A) Percent capture and viability of reaction with DTT, B) Percent capture and viability of reaction with TCEP. All reactions were incubated for 2 hours with further addition of 0.25 mM reducing agent after one hour. Data is an average of one experiment with titering of the pull-down performed in triplicate. Error bars are reported as %RSD.

5.8. Cysteine alkylation with immobilized reducing agent

Significant toxicity was observed for our alkylating agents in the presence of reducing agents. M13 phage possess four key disulfides within the phage coat pIII protein. Reduction of these buried internal disulfide bonds leads to a loss of phage viability.[282] To minimize this disruption of the phage coat structure, we carried out alkylation reactions on the phage using immobilized tris(2-carbopxyethyl)phosphine (iTCEP). iTCEP can be used to reduce the accessible disulfides on the displayed peptide without having significant impact on the internal disulfides within the PIII protein (Figure 5.21).[283] Reduction of C7C libraries with iTCEP has been reported using conditions that allowed for >90% viable phage.[283]

iTCEP was superior at reducing the disulfide bond. Optimization of reduction and biotin-capture with iTCEP provided high capture (>95%) after a 48 hour reduction (Figure 5.21B). iTCEP was effective at reducing 98% of the C7C peptides after a 48 hour incubation. An additional 20 hour incubation was tested to ensure the C7C peptides would remain reduced throughout a labeling reaction. The percent capture and viability were 96% and 95% following the 20 hour incubation (Figure 5.21B).

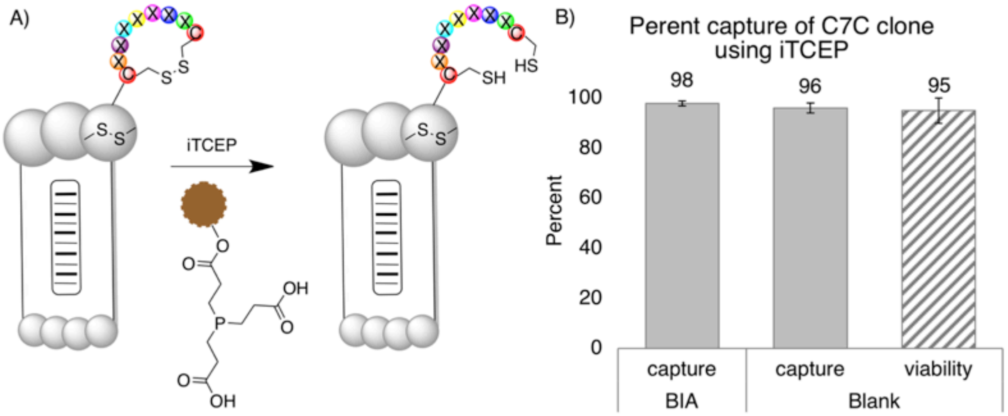


Figure 5.21. Reduction of > 90% of C7C clones is achieved using iTCEP. A) Cartoon depiction of phage particle and pIII coat proteins with internal disulfide bond and cyclized C7C peptide with disulfide bond. Reduction with iTCEP reduces the displayed peptide disulfide without reducing the internal pIII disulfides. B) Percent capture and viability of C7C clone reduced using iTCEP. Phage were reduced for 48 hours at 4°C and capture was measured following BIA pull-down. The blank condition was incubated for an additional 20 hours before BIA pull-down. Each reaction was carried out with 1.6 mM iTCEP in Tris pH 8.5. Data is the average of three independent experiments done on separate days, with each biotin-capture assay performed in triplicate. Error bars are one SD.

Alkylation reactions carried out after iTCEP reduction showed significantly reduced toxicity compared to other solution-based reducing agents. Capture experiments with phage reacted with **5.4** for one hour using solution TCEP showed no reaction to occur at concentrations that maintained viability of phage (Figure 5.14C). A series of reactions with the C7C clone using **5.4** and iTCEP demonstrated successful alkylation after a 16 hour reaction at room temperature (Figure 5.22). Concentrations below >500 μM gave yields less than 10% and concentrations greater than 3 mM resulted in low phage viability. Concentrations of **5.4** between 1-3 mM gave yields between 40-50% with 20-30% viable phage (Figure 5.22B). The experiment was repeated using 2 mM of **5.4** and the averaging of three separate experiments gave a percent yield and viability of 45% and 33% (Figure 5.22C).

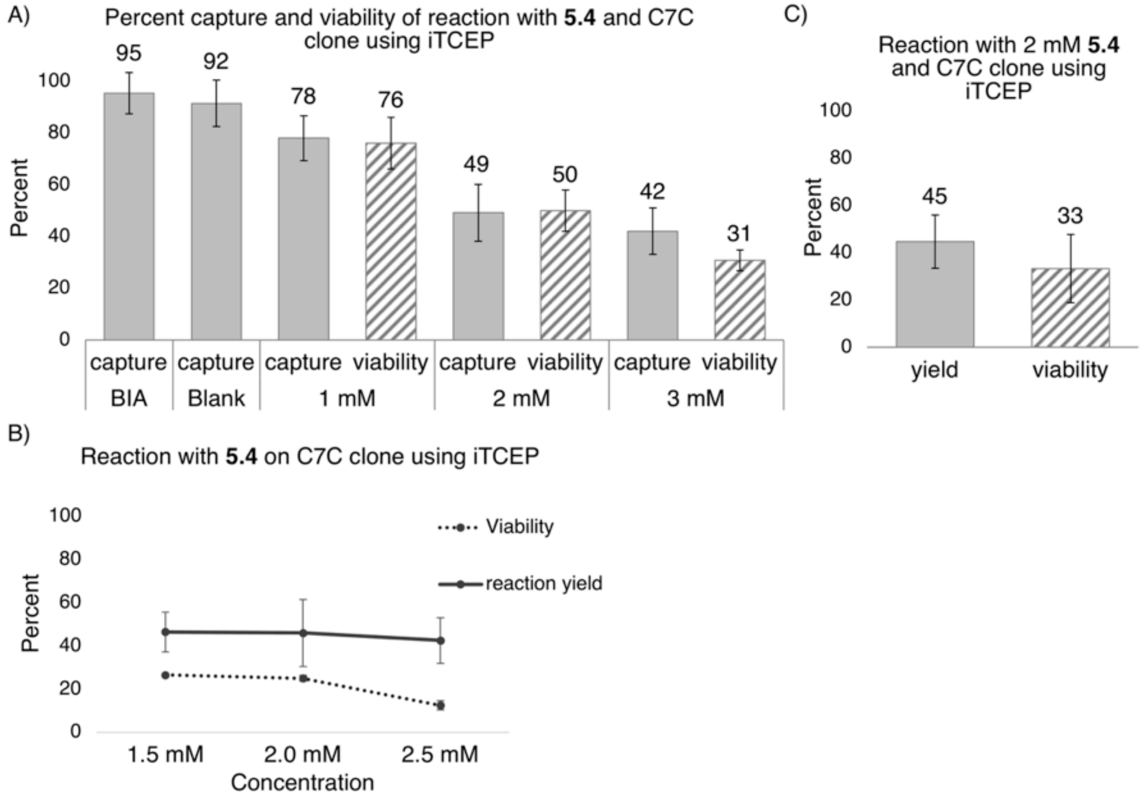


Figure 5.22. Reactions with iTCEP reduced toxicity of **5.4** and yielded successful alkylation of C7C clone. A) Percent capture of BIA control, blank control and reaction of **5.4** and C7C clone. Phage were reduced for 48 hours with iTCEP at 4°C and incubated with **5.4** for 12-16 hours at room temperature. Data is from a single experiment with the biotin-pull down titered in triplicate. Error bars represent % RSD, B) Average reaction yield and viability of **5.4.** C) Average yield and viability of the reaction of **5.4** (2 mM) and C7C phage. Data in both B and C is the average of three independent experiments done on separate days, with each biotin-capture assay performed in triplicate. Error bars are one SD.

Optimization of the reaction with **5.1** using iTCEP reduction provided high reaction yields and high levels of viable phage. Reactions with **5.1** using concentrations of 0.5-1.5 M gave reaction yields between 55-90% and 20-90% viable phage following a 24 hour reaction (Figure 5.23B). The experiment was repeated with 1 M of **5.1** and the average of two separate experiments gave a percent yield and viability of 78% and 72% (Figure 5.23C). Compared to the best reaction conditions with solution TCEP, reactions with iTCEP and 1 M **5.1** gave a 25% increase in reaction yield and 30% increase in phage viability. A similar experiment was conducted using both the C7C phage clone and library and 0.75 M **5.1** to ensure the reaction conditions were effective for alkylation of the C7C library (Figure 5.23D). A slightly lower yield was observed for the reaction with the C7C library compared to the clone with no significant difference in viability.

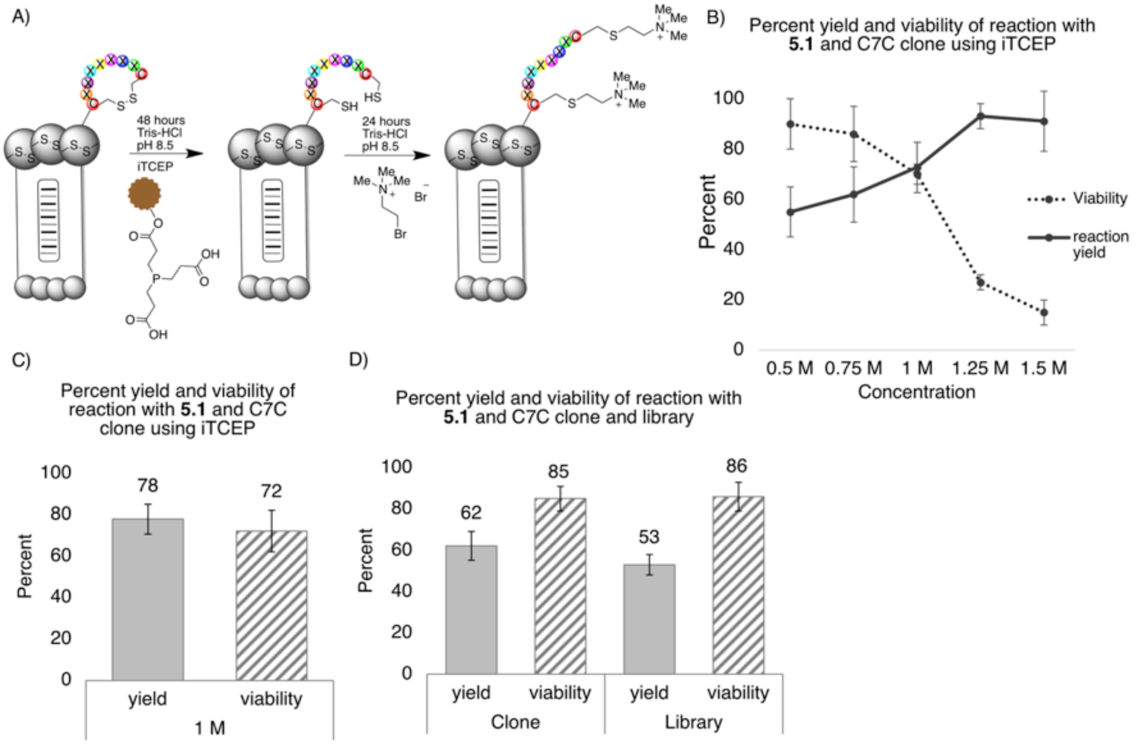


Figure 5.23. High yields and viability were achieved using iTCEP in the reaction with **5.1** and C7C clone. A) Cartoon schematic of iTCEP reduction and alkylation of C7C peptides with **5.1**. B) Percent yield and viability of reaction with **5.1** and C7C clone. Data is from a single experiment with the biotin-pull down titered in triplicate. Error bars represent % RSD, C) Average reaction yield and viability with C7C clone and 1 M treatment of **5.1**. D) Average reaction yield and viability with C7C clone and library with 0.75 M treatment of **5.1**. In all experiments phage were reduced for 48 hours with iTCEP at 4°C and incubated with **5.1** for 24 hours at room temperature. Data in C) and D) are the average of two independent experiments done on separate days, with each biotin-capture assay performed in triplicate. Error bars are one SD.

5.9. Discussion

A significant achievement in this work was first identifying that re-oxidation to form C7C disulfides is a major competing side reaction, and then overcoming the re-oxidation of the C7C phage to keep the phage in reduced form during the alkylation reaction. Stringent reduction conditions were required to prevent the rapid intramolecular cyclization. One way to minimize this challenge would be to use a phage construct expressing a single cysteine-containing peptide. In this work, using phage libraries containing two cysteines was necessary because unpaired cysteine residues are extremely rare in bacteriophage combinatorial peptide libraries. Studies on the sequence diversity of M13 Ph.D-12^{TM} peptide phage display libraries show that very few cysteine residues are present, and even fewer sequences exist that contain an odd number of cysteines (less than 1%).[284] Commercial combinatorial peptide phage display libraries with single cysteine containing sequences do not exist but libraries with paired cysteines such as those used in this work are available.

Any procedure to chemically modify Ph.D C7C phage must allow the phage to retain their ability to infect bacteria. Peptide sequences expressed on the phage are identified following amplification and sequencing of the phage DNA. This is necessary if the modified libraries are to be used in panning experiments with target proteins. Therefore, chemical modifications of the C7C phage need to both be high yielding and have low phage toxicity.

Despite extensive experimentation, conditions for successful alkylation of C7C phage using **5.2** and **5.3** were not identified. The high toxicity of **5.2–5.5** with increased concentration of reducing agent was discouraging for further testing and optimization. At concentrations of the alkylating agents that allowed the phage to retain the ability to infect *E. coli*, no significant reaction yields were observed.

The reasons for the increased toxicity of **5.2–5.5** are unknown, but are presumably because of the fact that they are analogs of nitrogen half mustards[285] and are capable of alkylating DNA.[286] The difference in toxicity between **5.1** and **5.2–5.5** may be due to the ability of the latter to cyclize spontaneously and form highly reactive aziridinium ions. This might explain why the alkylations of phage with **5.2–5.5** required so much lower concentrations than **5.1**, which can only exist as a primary alkyl bromide (a less active alkylating agent). But it also probably explains the increased toxicity levels of **5.2–5.5**. Both sulfur and nitrogen mustards have been shown to alkylate DNA in several bacteriophage constructs, however half mustards were shown to be less effective at alkylation of the phage DNA.[287-288] Future efforts to install methyllysine mimics of lower methylation states should investigate alternative reactive handles and synthetic strategies.

The use of iTCEP reduced toxicity enough to allow for identification of conditions to alkylate C7C peptides on phage with compound **5.4**. Reaction yields of

45% with 20% viable phage were observed in our best conditions. Testing of **5.5** alkylation using iTCEP reductions was not carried out in this work but we predict that **5.5** may be similarly able to alkylate the C7C peptides with the conditions described. Toxicity of **5.4** was a significant challenge but future optimization efforts may overcome this limitation and improve yield and viability.

Optimization of reducing agent, concentration, pH and time resulted in the successful reaction of **5.1** with the C7C phage peptides. Repeated experiments consistently demonstrated reaction yields of 35-50% with 40-60% viable phage using TCEP. Further improvements were achieved through reactions using iTCEP. The reduced toxicity of iTCEP allowed for longer reaction times while maintaining the reduced C7C peptides. iTCEP allowed for reaction yields as high as 80% with 70% viable phage. The high yield and viability of phage following reaction with **5.1** allows for the creation of a diverse library of C7C peptides. This reaction can provide $>2.5 \times 10^{11}$ phage and $\sim 2.5 \times 10^{8}$ unique peptide sequences containing Kcme3.

Methyllysine readers are capable of binding to cysteine methyllysine mimics and we predict this methodology will be a useful tool for the discovery of methyllysine reader antagonists. One recent study analyzed the binding of Kme3 peptides and their analogous Kcme3 peptides to a selection of reader proteins.[289] Replacement of Kme3 with Kcme3 in the H3K9me3 peptide sequence resulted in a 5-fold decrease in binding to the chromodomain of Su(Var)205 as determined by ITC. Despite this small loss in binding, the Kcme3 analog still had good affinity relative to many methyl reader-peptide complexes ($K_d = 1.6\ \mu M$). The PHD finger of ING1 tolerated a Kcme3 analog without loss in affinity.[289] The high biological interest in PHD fingers, combined with almost complete lack of success found by targeting them through traditional medicinal chemistry methods, makes them an ideal set of proteins to be targeted by the methyllysine mimic phage display libraries developed in this work.[109]

5.10. Conclusion

Successful alkylation of cysteine containing peptides and C7C phage was achieved through systematic optimization of multiple different reaction parameters. Alkylation of cysteine residues on C7C peptide phage constructs with 2-bromo-*N,N,N*-

trimethylethaninium bromide (**5.1**) is an effective strategy for the installation of methyllysine mimics.

Methyllysine reader proteins are challenging targets. We lack chemical tools to drive biological discoveries on the critical roles these proteins play in driving disease. The combined synthetic and genetic approach developed in this work allows for massive and diverse genetically encoded peptide phage libraries to be used for specific targeting of methyllysine reader proteins. Future work with Kcme3 functionalized C7C libraries in panning experiments will be used to identify potent and selective peptide inhibitors of methyllysine reader proteins.

5.11. Experimental Methods and Supplementary Data

5.11.1 General Information

All chemicals for synthesis of alkylating agents were purchased from Sigma Aldrich or Thermo Fisher Scientific. 2-bromo-*N,N,N*-trimethylethaninium bromide (**5.1**), 2-chloro-*N,N*-diethylethanaminium chloride (**5.2**) and 2-bromo-N,N-diethylethanaminium bromide (**5.3**) were used as purchased from Sigma Aldrich. Reducing agents DTT and TCEP were purchased from Chem Impex and Sigma Aldrich. The di-cysteine peptide used for alkylation reactions (SWCDYRC) was donated by the Derda lab.[275] The M13 Ph.D C7C phage clone was also generously donated to us by the Derda lab. The M13 C7C library was purchased from New England Biolabs (Ph.D.^{TM}-C7C library). Iodoacetyl-PEG2-Biotin was purchased from Thermo Scientific (21334). Streptavidin functionalized magnetic beads (Streptavidin MagneSphere® paramagnetic particles) were purchased from Promega. The cysteine H3 peptide (H3C4 - Ac-ARTCQTARKSTGGKA) was purchased from Genscript. The immobilized TCEP used was Pierce™ disulfide reducing gel purchased from Thermo Fisher Scientific (77712).

5.11.2 Synthesis and characterization of alkylating agents

Compounds 5.4 and 5.5

To a solution of the respective (2-chloroethyl) amine (0.1 g, 1 eq) in acetone (2 mL), NaI (5 eq) was added and the reaction mixture was heated to 60°C and refluxed with stirring overnight (18 hr). The acetone was removed under reduced pressure and the crude mixture was re-suspended in DCM and resulting precipitate was removed by filtration. The DCM in the filtrate was removed under reduced pressure to give the (2-iodoethyl) amines **5.4** and **5.5** in 70-80 % yield.

5.4

N,N-diethyl-2-iodoethanaminium iodide (compound **5.4**): MP: 136 º C. ATR-IR (neat): 2977w, 2912s, 2755m, 2715m, 2672s, 2479w, 1445s, 1417s, 1391m, 1355m, 1190s, 1167s, 1093m, 1066w, 1030s, 961w, 944w, 879w, 834w, 804m, 732m, 504w, 458m, 418w, 408w. ^{1}H NMR ($CDCl_3$, 300 MHz): 10.66 (br s, 1H), 3.62 (m, 2H), 3.47 (m, 2H), 3.24 (q, J = 7 Hz, 4H), 1.49 (t, J = 7 Hz, 6H). ^{13}C NMR ($CDCl_3$, 300 MHz): 54.4, 47.0, 8.75, –8.75. LR-ESI-MS: [*m/z*] calcd. for $C_6H_{15}IN^+$: 228.02; found: 228.07.

5.5

1-(2-iodoethyl)piperdin-1-ium iodide (compound **5.5**): MP: 189°C, ATR-IR (neat): 2941m, 2906s, 2721s, 2675s, 2539m, 1456s, 1416s, 1328w, 1281w, 1261m, 1208m, 1176m, 1132w, 1078m, 1053s, 1019s, 950s, 938s, 897m, 852m, 801m, 563w, 544m, 451w, 428m, 418w, 404w. ^{1}H NMR ($CDCl_3$, 300 MHz): δ 3.71 (m, 2H), 3.40 (m, 2H), 2.75 (br s, 1H), 2.36 (br s, 1H), 1.93 (s, br, 2H), 1.55 (s, br, 8H). ^{13}C NMR (MeOD, 300 MHz): 60.1, 54.1, 24.2, 22.6, –8.4. LR-ESI-MS: [*m/z*] calcd. for $C_7H_{15}IN^+$: 240.02; found: 240.13.

5.11.3 General procedure for reaction between 5.1 and cysteine containing peptides

All reactions were carried out in a total volume of 100 μL in 100 mM HEPES buffer pH 8. The peptide (SWCDYRC) was reduced using either DTT or TCEP for 30-120 minutes. **5.1** was added to the solution and the reaction was incubated protected from light. Specific conditions, reaction times and concentrations are indicated in Figure 5.8, Figure 5.9, Figure 5.24 and Figure 5.25. LCMS data was obtained using an Agilent Technologies 6130 LCMS. A gradient going from 10% acetonitrile/90% H_2O to 90% acetonitrile/10% H_2O over 8 minutes was used with a flow rate was 0.5 mL/min. All solvents contained 0.1% trifluoroacetic acid.

5.11.4 General procedure for reaction of 5.4 and 5.5 with cysteine containing peptides

A 4 mM stock of H3Kc4 peptide (Figure 5.26) was dissolved in DMF. 7.75 µL of this stock was diluted into 69.75 µL of H_2O. 20 µL of a 500 mM HEPES buffer (pH 8) was added to the solution. 2.5 µL of a 400 mM stock solution of TCEP was added last. Reduction of the peptide was allowed to proceed at room temperature for one hour. 10 µL of the reduction reaction was diluted into 33 µL of H_2O and an additional 20 µL of 500 mM HEPES buffer was added. From a 55 mM stock of the respective alkylating agent, 37 µL was added to the reaction bringing the total reaction volume to 100 µL. The reaction was allowed to proceed at room temperature protected from light. Reaction progress was monitored using a Waters UPLC-MS. An Acquity UPLC BEH C18 column (2.1 x 50 mm, 1.7 µM particles) was used with a LC method going from 5% acetonitrile/90% H_2O to 50% acetonitrile/10% H_2O over five minutes at a flow rate of 0.5 mL/min. All solvents contained 0.1% trifluoroacetic acid. Both starting material and product eluted at 3.7 minutes.

5.11.5 LCMS characterization of alkylated peptides

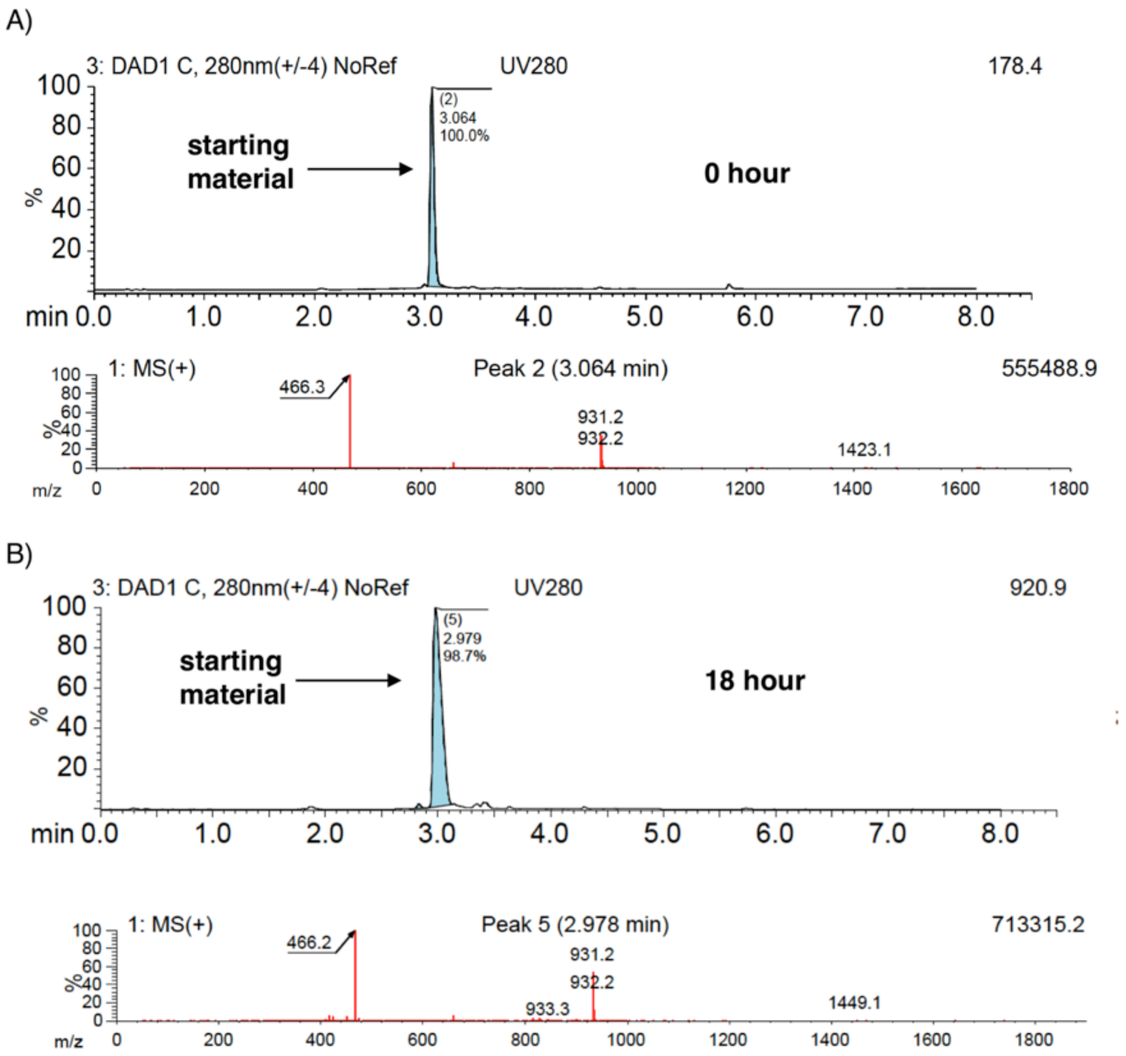


Figure 5.24. Reaction of **5.1** (100 mM) with SWCDYRC peptide showed no reaction yield after 18 hours. SWCDYRC (2 mM) was reduced for 30 minutes with 35 mM TCEP in 100 mM HEPES pH 8. **5.1** was added (100 mM) and the reaction was incubated at 37°C for 18 hours. A) LCMS trace with UV280 and MS+ of reaction at 0 hour, B) LCMS trace with UV80 and MS+ of reaction at 18 hours.

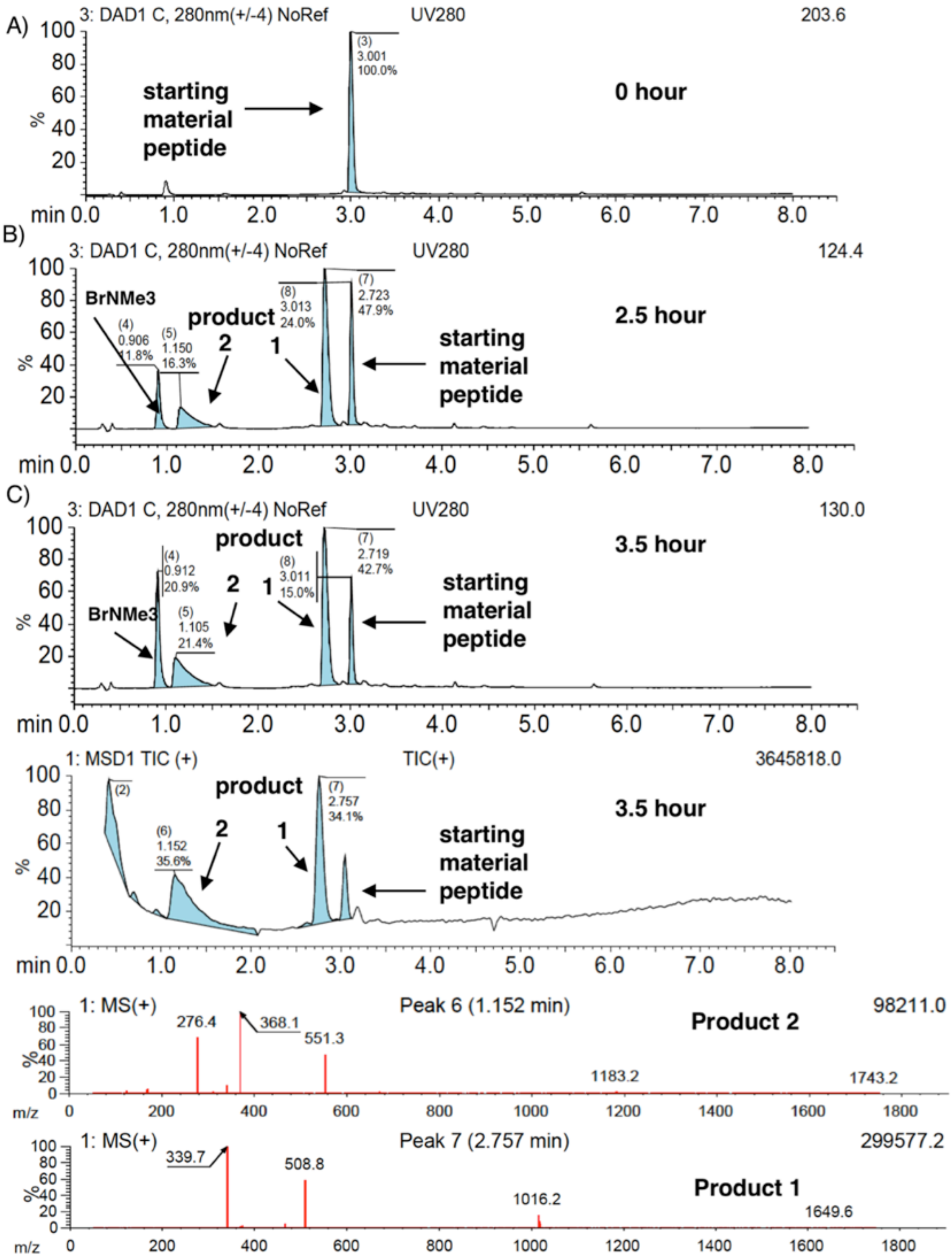


Figure 5.25. Reaction monitoring of **5.1** and NaI with SWCDYRC peptide. SWCDYRC (2 mM) was reduced for 30 minutes with 40 mM DTT in 100 mM HEPES buffer pH 8. **5.1** (1 M) was added along with NaI (2 mM) and the reaction was incubated at room temp for 3.5 hours. A) LCMS trace with UV280 and MS+ of reaction at 0 hour (reduced peptide before addition of **5.1**, B) LCMS trace with UV80 and MS+ of reaction at 2.5 hours, C) LCMS trace with UV80 and MS+ of reaction at 3.5 hours

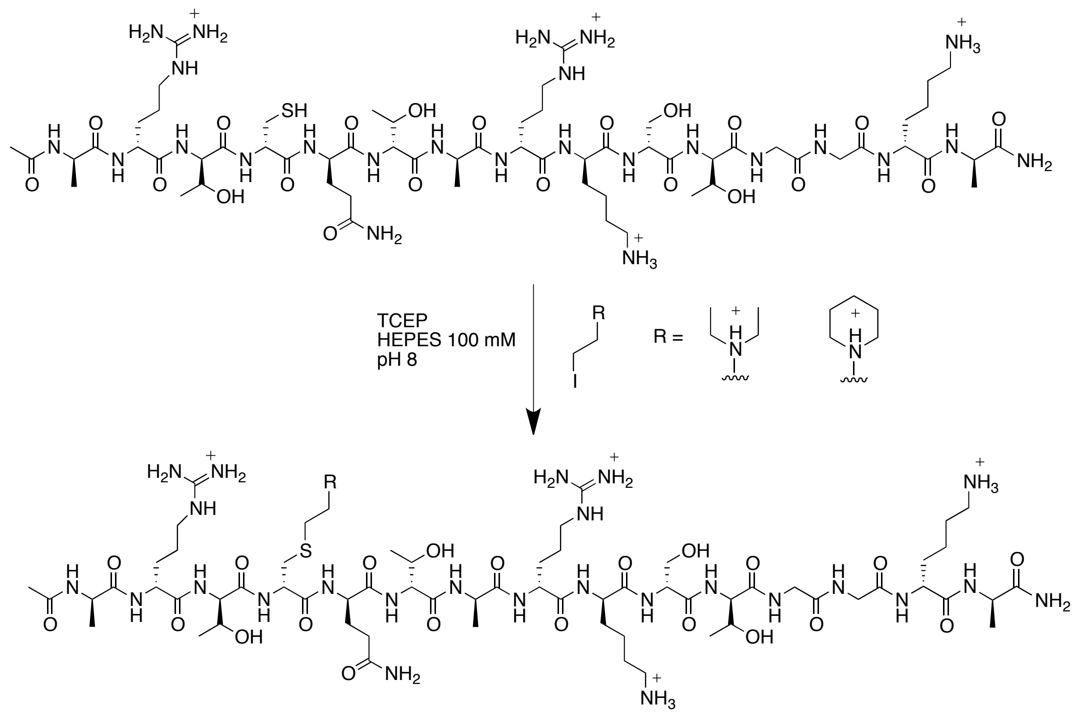


Figure 5.26. Reaction scheme of alkylation of H3Kc4 peptide (Ac-ARTCQTARKSTGGKA) with compounds **5.4** and **5.5**.

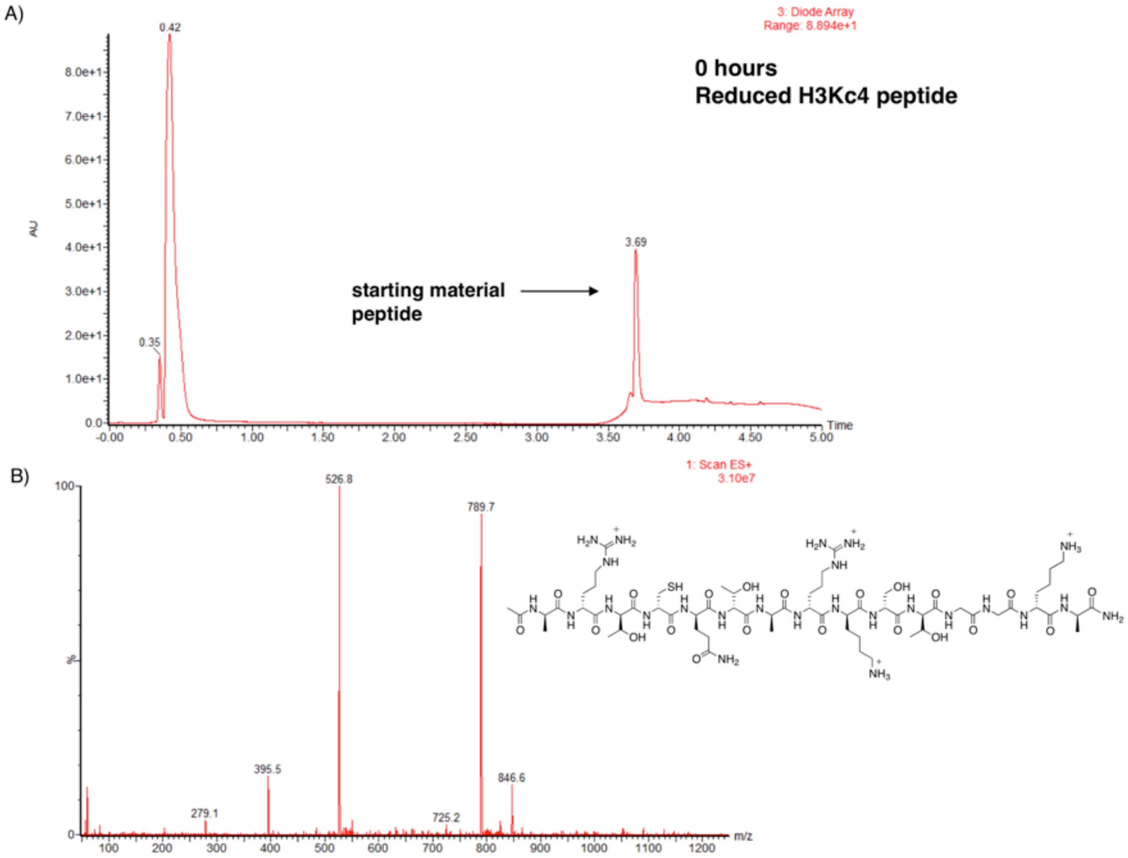


Figure 5.27. UPLC trace of H3Kc4 peptide in 100 mM HEPES buffer pH 8 with 10 mM TCEP. A) UV trace of H3Kc4, B) MS ES+ trace of H3Kc4.

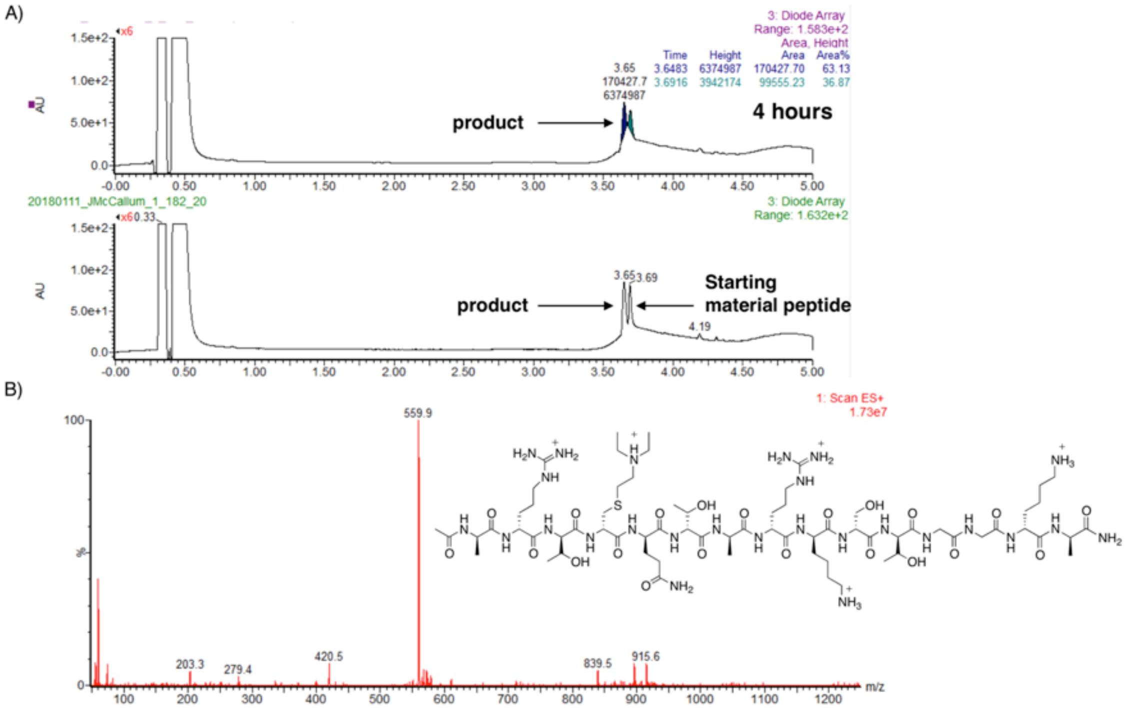


Figure 5.28. UPLC trace of reaction of **5.4** with H3Kc4 peptide in 100 mM HEPES pH 8. H3Kc4 was reduced for 1 hour with 10 mM TCEP prior to the addition of **5.4**. A) UV trace of reaction at 4 hours, B) MS ES+ trace of product peak.

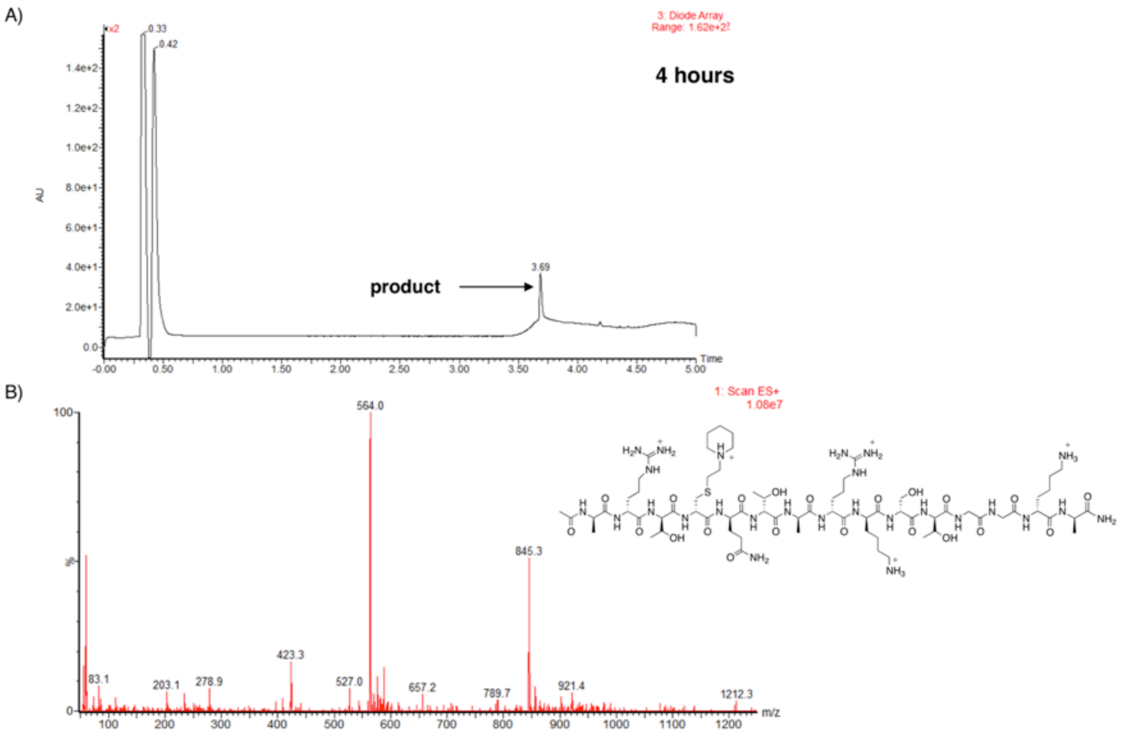


Figure 5.29. UPLC trace of reaction of **5.5** with H3Kc4 peptide in 100 mM HEPES pH 8. H3Kc4 was reduced for 1 hour with 10 mM TCEP prior to the addition of **5.5**. A) UV trace of reaction after 4 hours, B) MS ES+ trace of product peak.

5.11.6 General procedures for reduction and alkylation of Ph.D.™ C7C phage

General procedure for modification of phage with alkylating agents

All reductions and reactions were done in a total volume of 100 μL. 10 μL of a solution of C7C phage at a concentration of ~1×10^{12} pfu/mL was added to 87.5 μL of of aqueous buffer (pH followed by addition of 2.5 μL of 20 mM TCEP or DTT. The solution was incubated for 30-60 minutes at room temperature. 10 μL of the reduced phage solution was added to a new tube containing aqueous buffer. Alkylating agents were added from stock solutions and the total volume was adjusted to 100 μL. The reactions were incubated in the dark for the duration of the reaction time. At the end of the reaction incubation, 1 μL of BIA from a 100 mM stock (dissolved in DMF) was added to the reaction mixture. The reaction was incubated for another 30 minutes in the dark. The reaction was quenched by diluting 10^6 times in 1×PBS. The diluted and quenched reaction was split into two. Half of the reaction was combed with *E. coli* K12 and titered using X-Gal/IPTG LB plates. Plates are incubated over night at 37°C. The other half of the reaction was used to carry out the biotin pull-down using streptavidin magnetic beads.

5.11.7 General procedure for capture and titre of phage reaction solution following biotin pull-down using streptavidin magnetic beads

10 μL of Streptavidin MagneSphere® paramagnetic particles were used for each individual control or experimental condition. Beads were washed 3 times with 1×PBS with 100 μL of buffer used for each 10 μL of beads. Wash buffer was removed by placing the tubes on a magnetic stand for 30-60 seconds and pipetting out the supernatant. Beads were blocked with 100 μL of 2% BSA in 1×PBS for 15 minutes on a vortex. Blocking solution was removed and the beads were incubated with 100 μL of diluted C7C phage and allowed to mix vigorously on a vortex for 20 minutes. A magnetic stand was again used to separate the beads from the supernatant. The supernatant was removed and combined with *E. coli* K12 and titered using X-Gal/IPTG LB plates. Plates are incubated over night at 37°C.

5.11.8 Optimized procedure for alkylation of C7C phage with 5.1 and solution TCEP

Phage titers of 10^{11} pfu/mL were reduced using 0.5 mM TCEP for 30 minutes in 100 mM HEPES pH 7.5. 10 μL of the reduced phage solution was diluted 10-fold into a solution containing 1 M **5.1** and 100 mM HEPES pH 7.5. The reaction was incubated in the dark for 4 hours. To maintain reduction of the disulfide bonds, the solution was 'pulsed' halfway through the alkylation step with 0.2 mM TCEP.

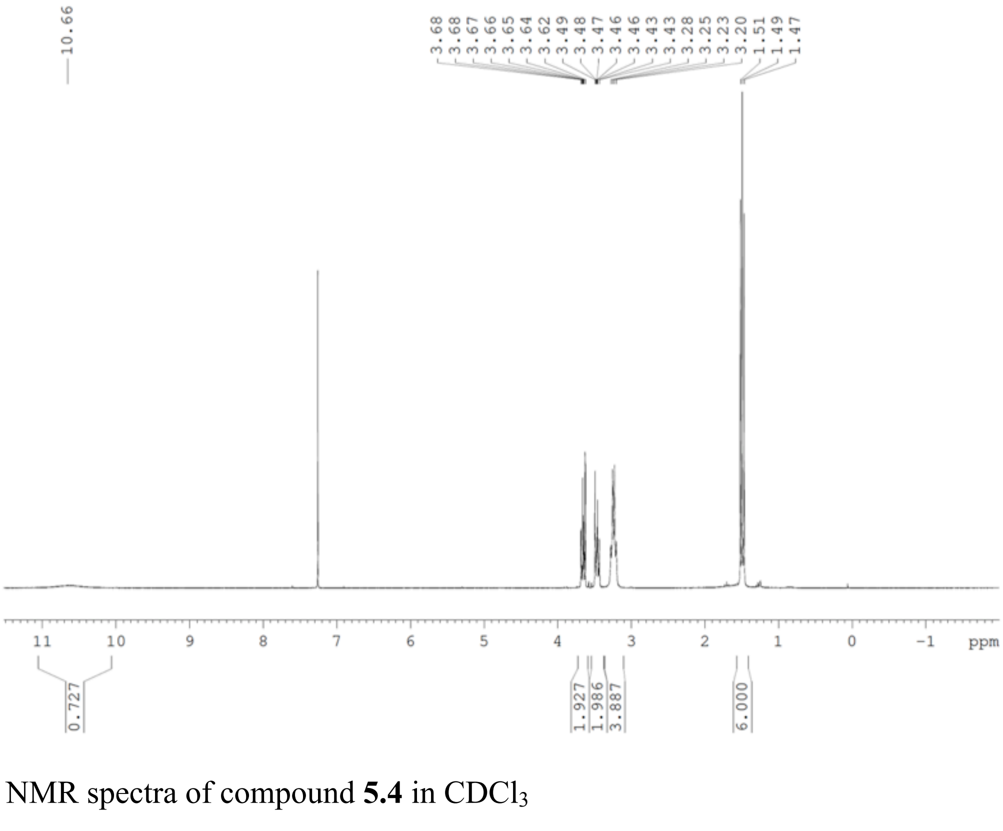
5.11.9 General procedure for reduction of phage with iTCEP

Immobilized TCEP (iTCEP) reduction and alkylation were carried out in three steps. Reduction using phage titers of 10^{7}-10^{11} pfu/mL was done using 1.6 mM iTCEP beads in 100 mM Tris (pH 8.5) for 48 hours at 4°C with slow spinning on a rotator. Alkylating agent was added directly to the reduction reaction and the reaction tube was incubated for 12–24 hours on a rotator at room temperature. The phage peptide was then biotinylated with 1 mM Iodoacetyl-PEG2-Biotin (BIA) for 30 minutes at room temperature on the rotator. Incubations with alkylating agent and BIA were done in the dark. The reaction solution was removed following spinning down the beads at 2.7 RPM for 60 seconds. The beads were washed with 100ul of sterile distilled water and this was added to the supernatant reaction solution. The phage were precipitated out of the reaction solution with 60% (v/v) of PEG-NaCl on ice for two hours. The phage particles were then resuspended in 1×PBS and diluted accordingly (10^5–10^9) depending on the original concentration of phage.

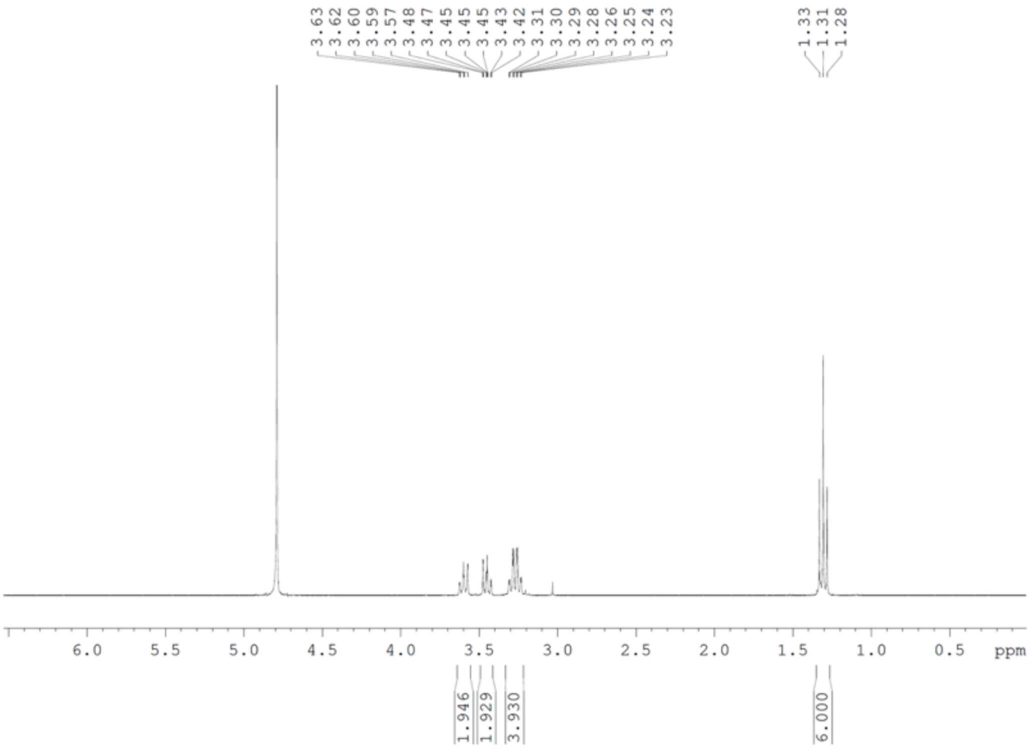
5.11.10 Optimized procedure for alkylation of C7C phage with 5.1 and solution iTCEP

Phage titers of 10^8–10^{11} pfu/mL were reduced with 1.6 mM iTCEP beads in 100 mM Tris (pH 8.5) for 48 hours at 4°C with slow spinning on a rotator. 1 M of **5.1** was added directly to the reduction reaction and the reaction tube was incubated for 24 hours on a rotator at room temperature.

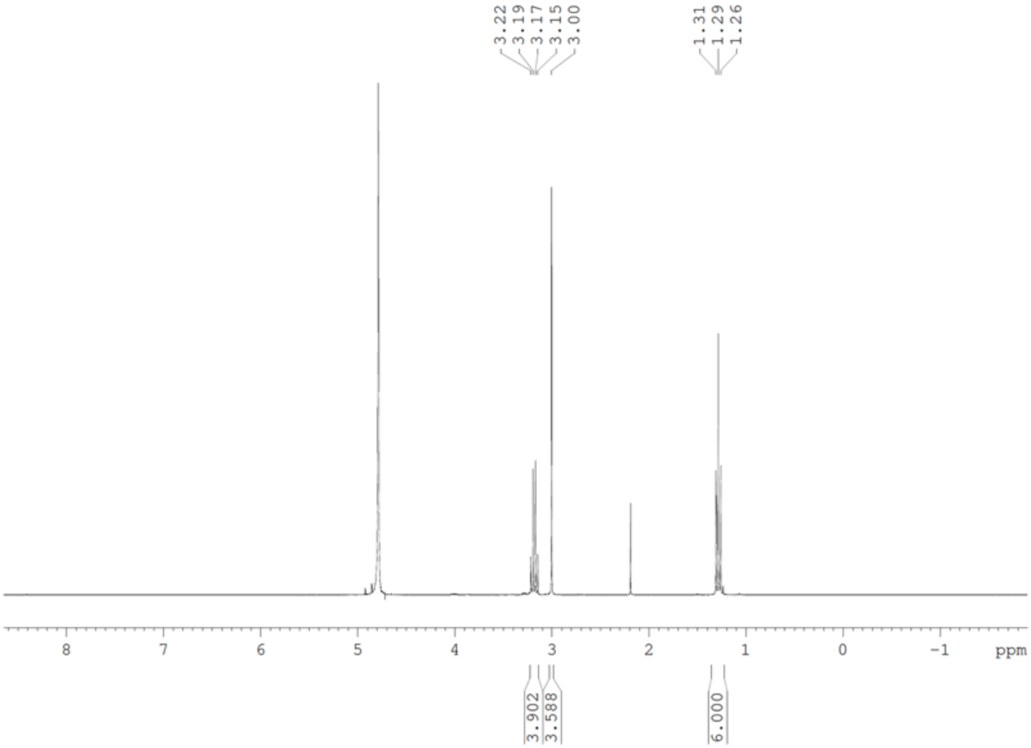
5.11.11 ^{1}H NMR Spectra



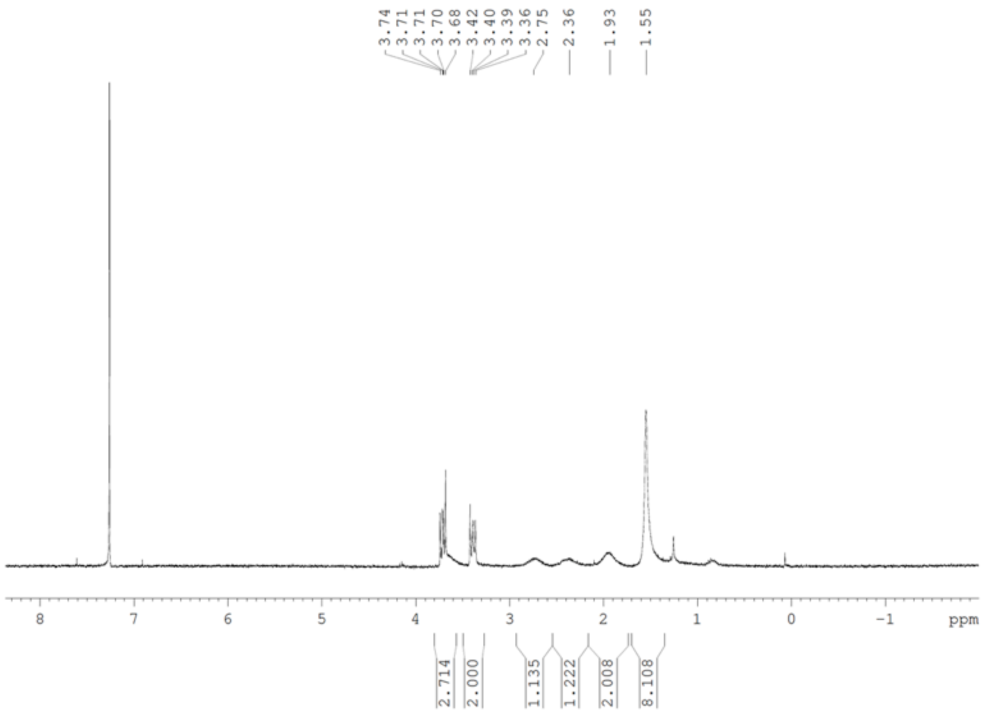
^{1}H NMR spectra of compound **5.4** in $CDCl_3$



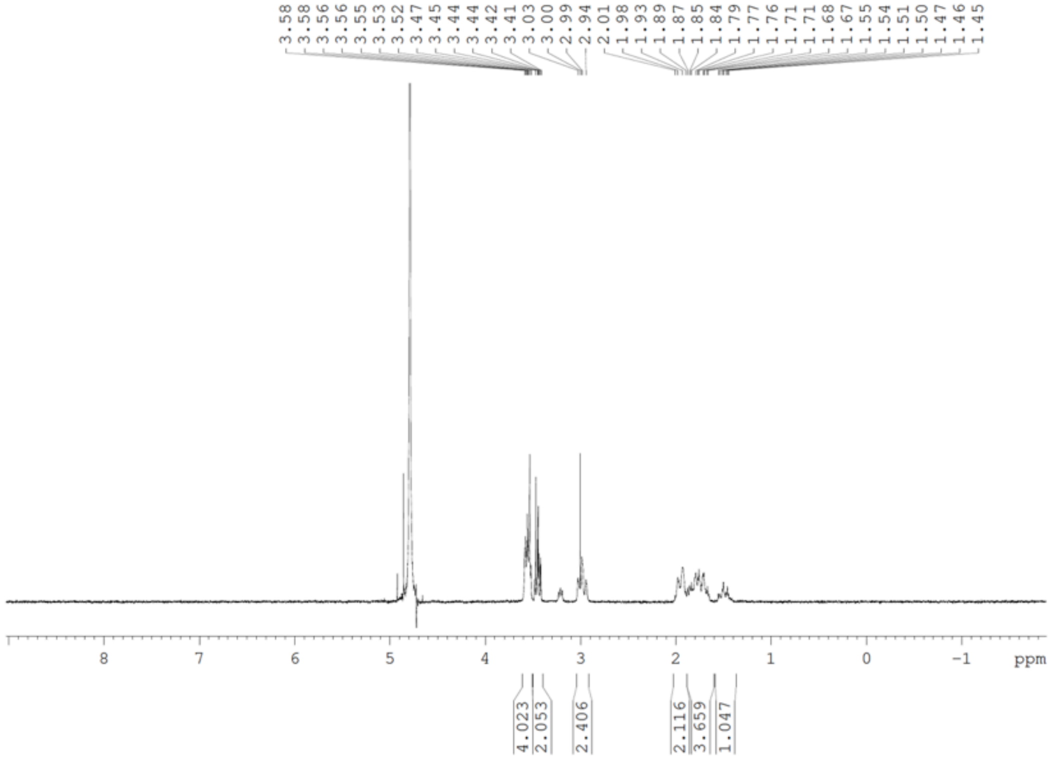
^{1}H NMR spectra of compound **5.4** in D_2O



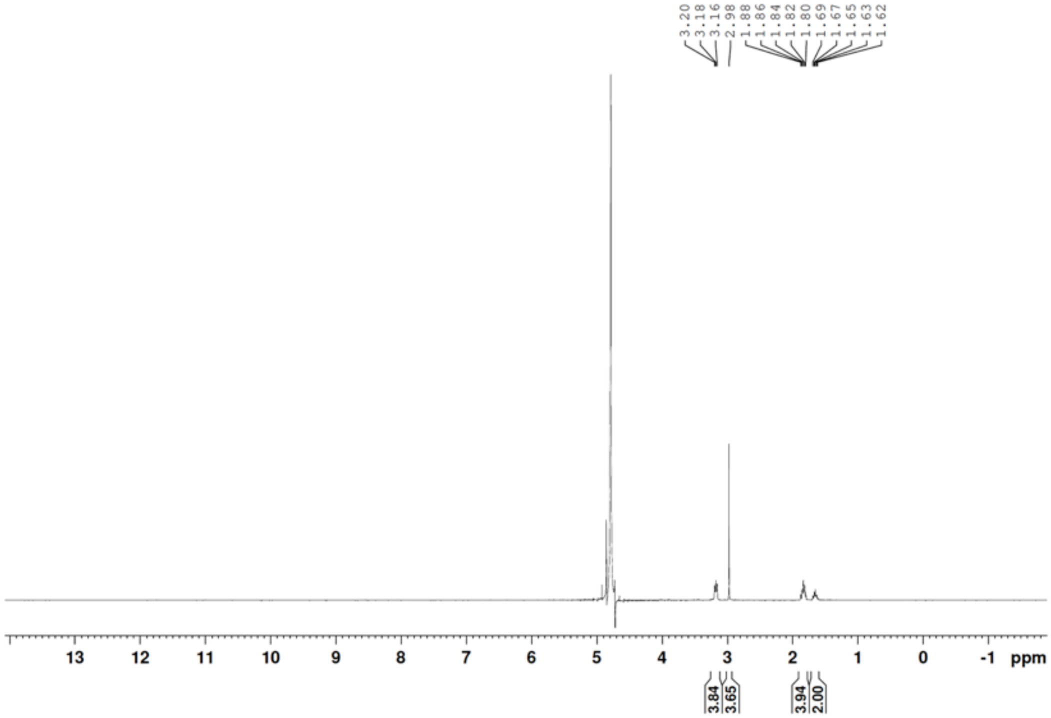
^{1}H NMR spectra of compound **5.4** in D_2O phosphate buffer pD 8.4



^{1}H NMR spectra of compound **5.5** in $CDCl_3$

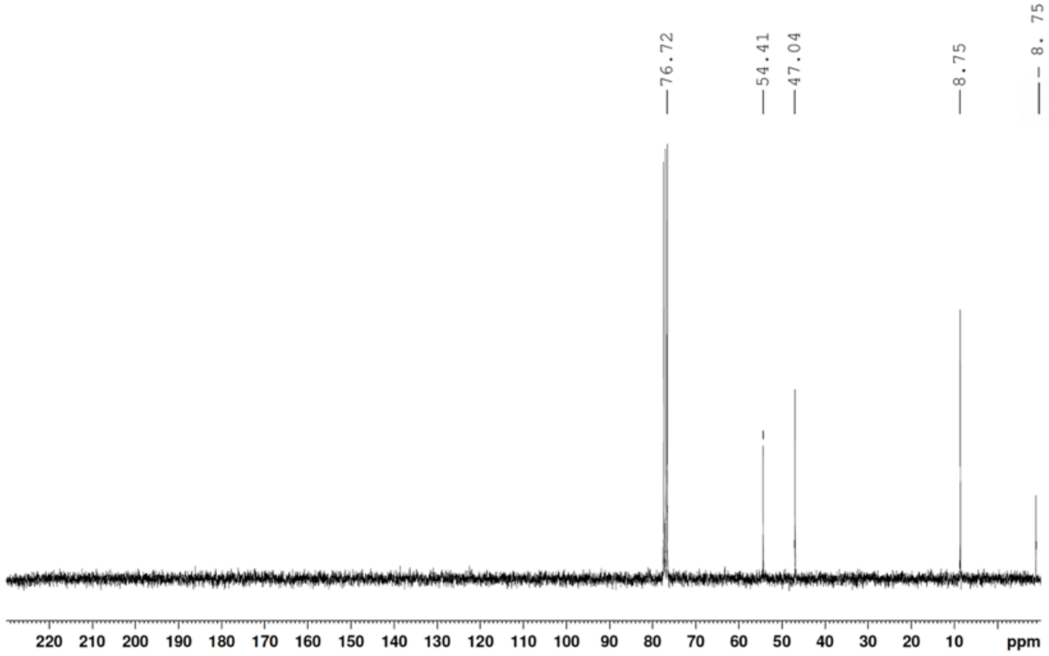


^{1}H NMR spectra of compound **5.5** in D_2O

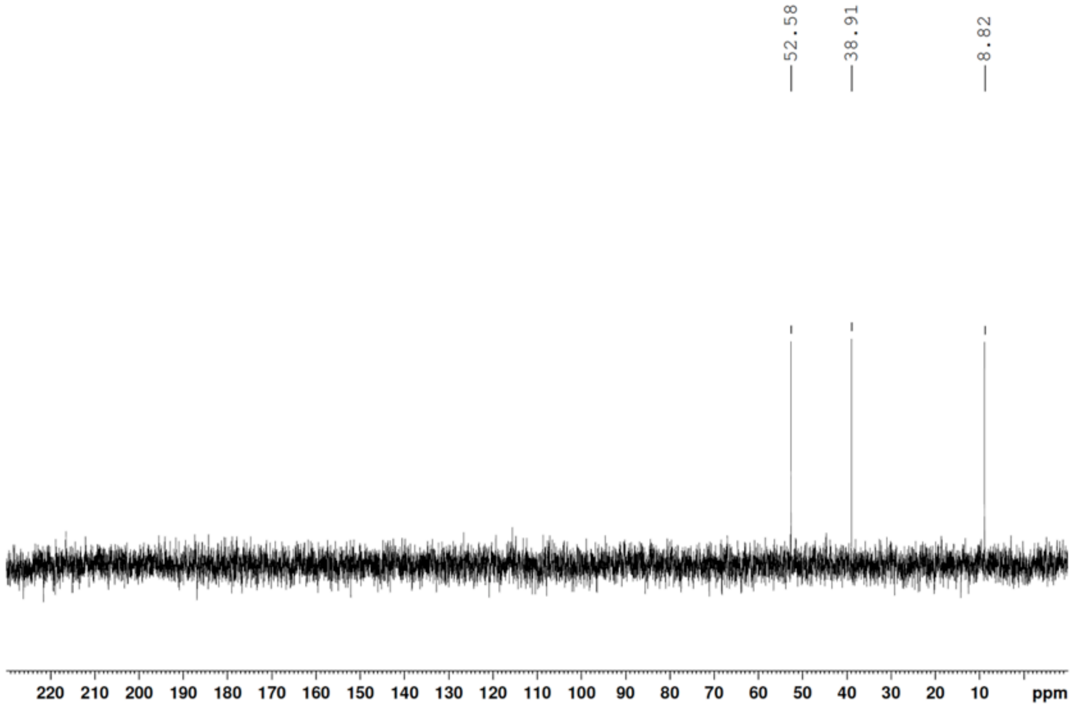


^{1}H NMR spectra of compound **5.5** in D_2O phosphate buffer pD 8.4

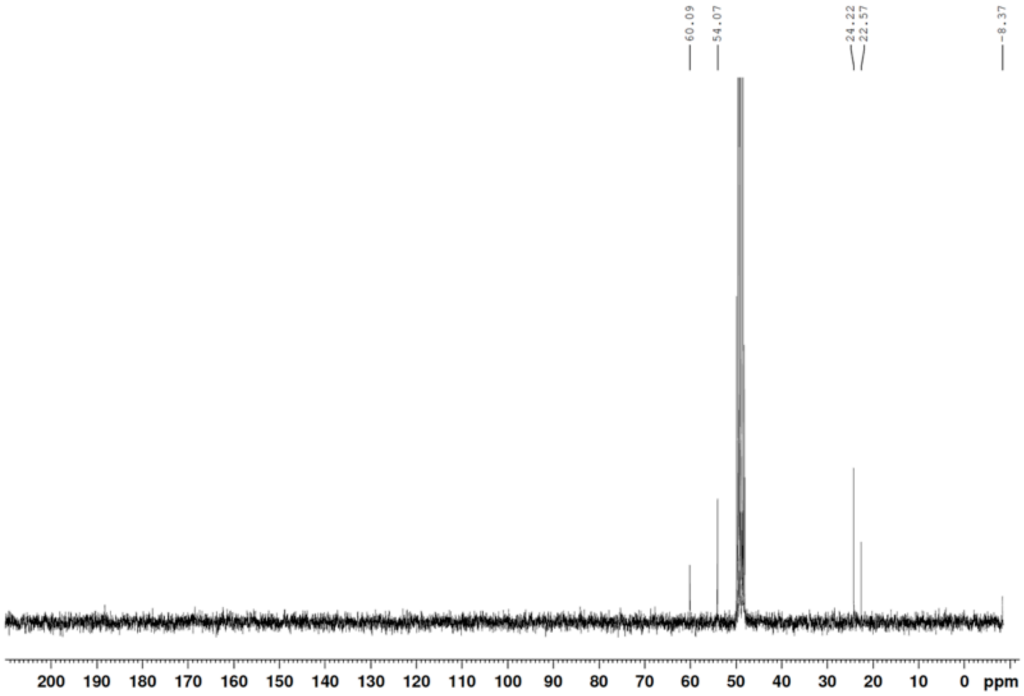
5.11.12 ^{13}C NMR spectra



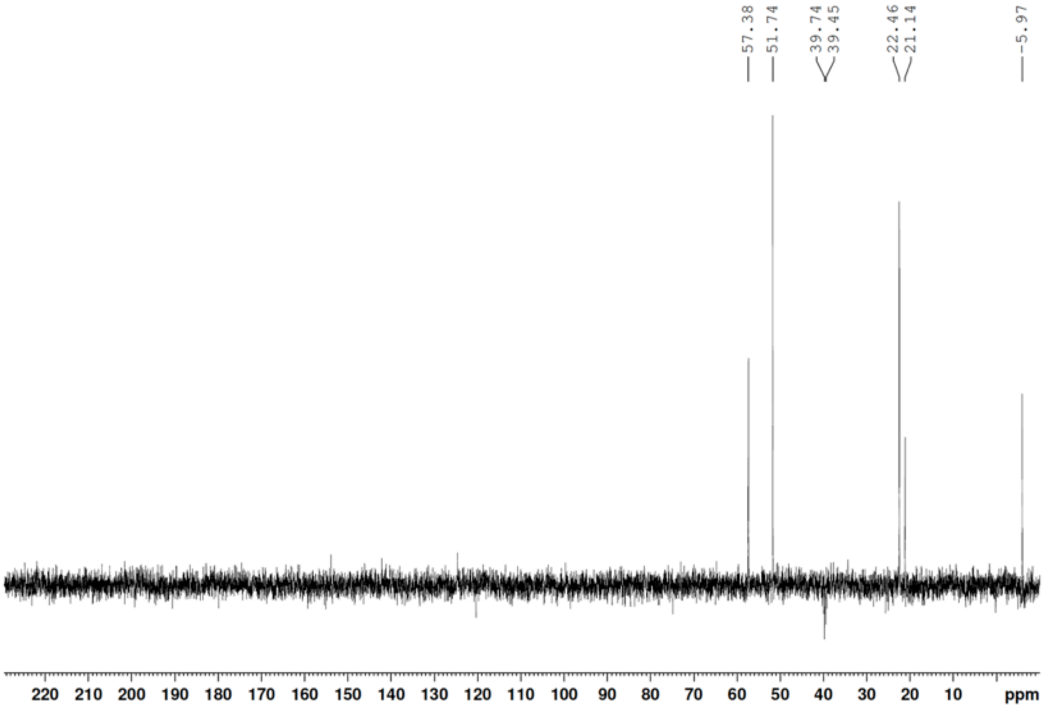
^{13}C NMR spectra of compound **5.4** in $CDCl_3$



^{13}C NMR spectra of compound **5.4** in D_2O phosphate buffer pD 8



^{13}C NMR spectra of compound **5.5** in MeOD



^{13}C DEPT-135 NMR spectra of compound **5.5** in DMSO-d$_6$

Chapter 6. Context for advances in this thesis

6.1. Introduction

Despite increased attention into the biology, inhibition of chromodomain methyllysine readers is in its infancy. At the start of my thesis, there was one report of inhibitors for CBX4/7,[149] followed by a report of small molecule CBX7 inhibitors in 2015.[153] Following my publication on the first CBX6 inhibitor,[155] there have been five reports of inhibitors for CBX4/7,[150-152, 154, 238] and one for the closely related chromodomain containing CDYL1/2 proteins.[156] Screening efforts against the chromodomain of M-phase phosphoprotein 8 (MPP8) and Heterochromatin protein 1-beta (HP1-beta) have been undertaken but ultimately were not reported in peer-reviewed literature.[290-291] Although progress is being made, there is a significant need for more chemical tools to study CBX proteins and other chromodomains.

CBX proteins have attracted the most interest among chromodomain-containing proteins. Efforts to inhibit CBX proteins are increasing in number because of the growing body of literature demonstrating these proteins to be master regulators and drivers of disease.[196, 292] At the time of publication of my CBX6 inhibitors, there were no publications on the biology of CBX6. Since 2017, CBX6 has been shown to regulate stem cell differentiation and to be a driver of aggressive hepatocellular carcinoma (HCC).[163, 170] Similarly, CBX8 was scarcely understood at the outset of my thesis research, but was recently shown to promote development and progression of HCC.[246] Numerous reports have indicated the oncogenic role of CBX8 in leukemia,[248] bladder cancer,[251] breast cancer[249] and other malignancies.[214, 248, 250, 293-295]

My thesis has focused on the discovery of CBX antagonists and on the development of new methodologies for targeting these proteins. In Chapters 2, 3 and 4 I described my efforts to make potent and selective CBX inhibitors and highlighted the challenges in achieving selective inhibition. In Chapter 5, I described my work on chemically modifying genetically encoded bacteriophage libraries for high-throughput discovery of ligands for methyllysine reader proteins. This chapter will summarize my progress in CBX inhibition and discuss the main conclusions from my research along with ongoing and proposed future directions.

6.2. Peptide-driven synthesis of the first non-CBX7 antagonists

My efforts to inhibit CBX proteins has largely focused on rational design and one-at-a-time synthesis of peptidic inhibitors. I have investigated and discovered key structural differences that exist in the peptide binding groove within the CBX family and this lead to my discovery of the first non-CBX7 antagonist.[155]

My pan-specific promiscuous dye-labeled inhibitors were serendipitously found to bind to all CBX proteins with high affinity (Chapter 2). They have both improved testing of new ligands within our own lab and have driven new discoveries in the hands of our collaborators. They offer improved testing for two reasons: the ability to measure more potent inhibitors, and the ability to measure inhibitors against multiple CBX proteins using the same probe ligand under the same conditions. One fundamental limit of competitive FP assays is that they can only measure potencies for inhibitors that bind more weakly than the probe ligand with which they are competing.[296] The increased potency of **2.4** (0.02–0.85 µM against the range of CBX1/2/4/6/7/8) versus its predecessor H3K27me3 (110 µM to unmeasurable binding against the range of CBX1/4/6/7/8) enabled the study of the highly potent compounds that are reported in Chapters 4. The promiscuity of probe **2.4** across the same set of CBX's allowed us to establish a "full-panel" CBX FP assay that we used to test the affinity of individual ligands against up to 6 different CBX proteins. Without this pan-specific ligand, the IC_{50} values determined for different proteins using distinct probes under distinct conditions would not have been directly comparable to each other, and therefore would not have been useful for determining patterns of selectivity within this class of proteins. The new assays powered by 2.4 have aided this thesis work, as well as the discovery of new ligands targeting CBX7/8 both within our own lab and in collaborators' labs that are not included in this thesis.[150, 238] Compound **2.4** has also enabled collaborations with two pharma companies for screening of fragments and small molecules against the CBX family.

My second-generation dual active inhibitors for CBX6 and CBX8 display improved potency and cell permeability. To overcome the limitations of my first generation CBX inhibitors, I designed and synthesized a library of second-generation inhibitors that are shorter in length, lack a permanent cation, and contain more unnatural amino acid residues. This work has resulted in the first CBX8-selective inhibitors, and dual-active inhibitors for CBX6 and 8 that have unprecedented potency. These inhibitors show promise in early cell-based studies and their activity in cells will allow for further research into the phenotypes of CBX6/8 inhibition. The fact that both CBX6 and CBX8 seem to play distinct roles in promoting HCC,[163, 170, 246] makes this dual inhibition approach especially interesting.

6.3. Future directions and on-going work on CBX peptidic inhibitors

Future work will focus on improving potency and selectivity of the inhibitors and on gaining greater structural understanding through MD simulations and crystallization of protein-ligand complexes. Ongoing work is focusing on expanding the library of inhibitors. Structural motifs currently being designed and synthesized are based on the combined data of this work, and DELs and MD simulations carried out by our collaborators. Selective ligands for CBX6 and CBX8 will require larger hydrophobic substitutions in the (–2) position in combination with N-cap substitutions that further engage the proteins' extended binding grooves.

Additional ongoing work is focusing on using the peptidic inhibitors I have developed in further cell-based studies. Testing of inhibitors **4.3** and **4.6** with longer time periods (12 days) are predicted to give increased observable phenotypic changes. We predict that several cell turnovers will be required to see changes to the epigenome and associated phenotypic changes. Previous studies on the effects of CBX7 inhibition in PC3 cells were carried out over 6 days with re-treatment of the inhibitor at day 3.[151]

Our studies will include longer treatments with inhibitors using ESCs and HCC cells. Investigations on the morphology, proliferation and expression of pluripotency markers in ESCs with our inhibitors will aid in understanding the role of CBX6 and 8 in controlling stem cell differentiation. Similarly, we hope to identify the potential of CBX6 and CBX8 as therapeutic targets through inhibiting these proteins in wild type and KO

HCC cells lines. Biotinylated analogs of lead inhibitors (including **4.3** and **4.6**) will be used in pull-down experiments to verify engagement with CBX6/8 and the associated chromatin modifying multi-protein complex PRC1.

The weak affinity of CBX6 and CBX8 to H3K27me3 and H3K9me3 brings to question whether these proteins bind and recognize other methylation marks. In related work, we are aiming to discover currently unknown methyllysine-containing binding partners of the CBX proteins. Ligands described in this work are currently being used in competitive pull-down experiments with cell lysates, in which the ligand's selective inhibition of one or two CBX proteins can be used in comparative proteomics studies to generate lists of new candidate proteins that bind to a given CBX in the cell. Through proteomic analysis of the pull-down samples in the absence and presence of a selective CBX6 inhibitor, we hope to identify novel binding partners of CBX6.

6.4. Massive methyllysine modified genetically encoded peptide libraries

To improve the ability to discovery ligands for CBX proteins, I designed and evaluated a synthetic methodology for the installation of methyllysine mimics on cysteine containing peptides and bacteriophage peptide libraries. Synthetic optimization resulted in the successful alkylation of cysteine containing peptides with (2-haloethyl) amines designed to mimic methyllysine residues. Extensive optimization on C7C peptide phage constructs allowed for the successful installation of Kme3 mimics.

This work described in Chapter 5 is the first report of a post-translational methylated peptide phage library. I report conditions for reactions with 2-bromo-*N,N,N*-trimethylethaninium bromide that provide $>2.5 \times 10^{11}$ phage and $\sim 2.5 \times 10^{8}$ unique peptide sequences containing. The methodology I developed can be used in a synthetic chemistry-driven adaptation of traditional phage display for the screening of millions of peptide-based compounds. Strategies that allow for diversity and high throughput screening are critical for successfully targeting the highly similar CBX proteins.

6.5. Future directions for methyllysine modified phage display libraries

Ongoing and future work with the methyllysine modified phage libraries will focus on biopanning experiments with a selection of methyllysine reader proteins. To demonstrate the usefulness of methyllysine modified phage libraries in finding new peptide-based inhibitors of methyllysine readers, I propose testing of the Kcme3 functionalized phage libraries on the PHD finger of ING1. Currently there are two reports of inhibitors (IC_{50} values of 26-40 μM and K_d of 7.5 mM)[207, 297] for two of the 99 PHD fingers in the human genome.[64] PHD fingers are hard to target with high-throughput screens and some of the few, weak, fragment-like inhibitors developed do not directly target the methyllyisne recognition motif.[207, 297] We have expressed and purified ING1 in our lab (following published protocols),[298] and will use the protein in panning experiments with Kcme3 phage display libraries.

To overcome the challenges with selective inhibition of CBX proteins, future work will utilize competitive-panning experiments to find selective binders for one CBX over other family members. My on-going efforts to achieve this goal have included demonstrating that CBX proteins can be immobilized and retain their ability to bind to Kme3 peptides. Adding unbound competitive CBX proteins into the panning experiments will allow for CBX6/8 selective sequences to be selected. This will be done by immobilization of for example, CBX6, onto a 96-well plate, blocking, and then incubation with phage in the presence of free CBX7/8. Unbound CBX7/8 will allow peptide sequences that bind strongly to these proteins to be washed away and not bound to the immobilized CBX6. Non-specific binders can be ruled out by comparing the CBX6 immobilized wells to the control BSA wells. The comparison of selected/counter-selected and control wells using bioinformatics analyses will identify CBX6 selective peptide sequences. Protocols for extraction of the phage DNA, PCR and quantification of DNA yields have been successfully optimized for our chemically modified C7C phage.

Peptidic inhibitors developed from this methodology have potential to be lead scaffolds for further development of selective inhibitors of polycomb paralogs and other methyllysine reader proteins. These compounds can serve as new chemical tools needed to learn the therapeutic potential of methyllysine reader protein inhibition. As I've described above, the discovery of a strong-binding peptidic inhibitor improves assays and therefore can also power new screens that can ultimately lead to more potent, more drug-like inhibitors. This boot-strapping approach is critical to the progress of many drug development programs. The greatest power of the new technology is that a single phage-displayed library of methyllysine mimics could provide the starting points (moderate-to-strong peptidic ligands) for dozens of different methyllysine readers.

6.6. Conclusions

The research summarized in this thesis has provided potent inhibitors for studying the CBX methyllysine reader proteins. These inhibitors have improved screening for the discovery of new ligands, increased our structural understanding of each protein, and are new tools for investigations of the phenotypes generated by inhibition. Our current understanding of epigenetic regulation of gene expression by methyllysine readers is akin to the tip of the iceberg. The chemical antagonists reported in this thesis are tools for driving new chemical biology and medicinal chemistry.

In spite of this progress, significant challenges still remain for creating potent and selective inhibitors of methyllysine readers. Improved screening, new methyllysine isosteres, and greater chemical diversity are needed to better differentiate between the structurally similar readers within certain families. My reported method for the creation of massive methyllysine containing peptide libraries represents a novel strategy to overcome these aforementioned challenges.

Bibliography

1. Delcuve, G. P.; Rastegar, M.; Davie, J. R., Epigenetic control. *J. Cell. Physiol.* **2009,** *219* (2), 243-50.

2. Schübeler, D., Function and information content of DNA methylation. *Nature* **2015,** *517*, 321.

3. Gros, C.; Fahy, J.; Halby, L.; Dufau, I.; Erdmann, A.; Gregoire, J.-M.; Ausseil, F.; Vispé, S.; Arimondo, P. B., DNA methylation inhibitors in cancer: Recent and future approaches. *Biochimie* **2012,** *94* (11), 2280-2296.

4. Rothbart, S. B.; Strahl, B. D., Interpreting the language of histone and DNA modifications. *Biochim. Biophys. Acta* **2014,** *1839* (8), 627-643.

5. Venkatesh, S.; Workman, J. L., Histone exchange, chromatin structure and the regulation of transcription. *Nat. Rev. Mol. Cell Biol.* **2015,** *16*, 178.

6. Bannister, A. J.; Kouzarides, T., Regulation of chromatin by histone modifications. *Cell Res.* **2011,** *21* (3), 381-395.

7. Strahl, B. D.; Allis, C. D., The language of covalent histone modifications. *Nature* **2000,** *403* (6765), 41-5.

8. Turner, B. M., Cellular memory and the histone code. *Cell* **2002,** *111* (3), 285-91.

9. Jenuwein, T.; Allis, C. D., Translating the histone code. *Science* **2001,** *293* (5532), 1074-80.

10. Prabakaran, S.; Lippens, G.; Steen, H.; Gunawardena, J., Post-translational modification: nature's escape from genetic imprisonment and the basis for dynamic information encoding. *Wiley Interdiscip Rev Syst Biol Med* **2012,** *4* (6), 565-83.

11. Margueron, R.; Reinberg, D., Chromatin structure and the inheritance of epigenetic information. *Nat. Rev. Genet.* **2010,** *11*, 285.

12. Janssen, K. A.; Sidoli, S.; Garcia, B. A., Recent Achievements in Characterizing the Histone Code and Approaches to Integrating Epigenomics and Systems Biology. *Methods Enzymol.* **2017,** *586*, 359-378.

13. Greer, E. L.; Shi, Y., Histone methylation: a dynamic mark in health, disease and inheritance. *Nat. Rev. Genet.* **2012,** *13* (5), 343-57.

14. Kouzarides, T., Chromatin modifications and their function. *Cell* **2007,** *128* (4), 693-705.

15. Black, Joshua C.; Van Rechem, C.; Whetstine, Johnathan R., Histone Lysine Methylation Dynamics: Establishment, Regulation, and Biological Impact. *Mol. Cell.* **2012,** *48* (4), 491-507.

16. Zhang, X.; Wen, H.; Shi, X., Lysine methylation: beyond histones. *Acta. Biochim. Biophys. Sin.* **2012,** *44* (1), 14-27.

17. Clarke, S. G., Protein methylation at the surface and buried deep: thinking outside the histone box. *Trends Biochem. Sci.* **2013,** *38* (5), 243-252.

18. Carlson, S. M.; Gozani, O., Emerging technologies to map the protein methylome. *J. Mol. Biol.* **2014,** *426* (20), 3350-62.

19. Dillon, S. C.; Zhang, X.; Trievel, R. C.; Cheng, X., The SET-domain protein superfamily: protein lysine methyltransferases. *Genome Biology* **2005,** *6* (8), 1-10.

20. Metzger, E.; Schule, R., The expanding world of histone lysine demethylases. *Nat. Struct. Mol. Biol.* **2007,** *14* (4), 252-254.

21. Murray, K., The Occurrence of iε-N-Methyl Lysine in Histones. *Biochemistry* **1964,** *3* (1), 10-15.

22. Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T., Regulation of chromatin structure by site-specific histone H3 methyltransferases. *Nature* **2000,** *406* (6796), 593-9.

23. Feng, Q.; Wang, H.; Ng, H. H.; Erdjument-Bromage, H.; Tempst, P.; Struhl, K.; Zhang, Y., Methylation of H3-lysine 79 is mediated by a new family of HMTases without a SET domain. *Curr. Biol.* **2002,** *12* (12), 1052-8.

24. Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y., Histone demethylation mediated by the nuclear amine oxidase homolog LSD1. *Cell* **2004,** *119* (7), 941-53.

25. Walport, L. J.; Hopkinson, R. J.; Schofield, C. J., Mechanisms of human histone and nucleic acid demethylases. *Curr. Opin. Chem. Biol.* **2012,** *16* (5-6), 525-34.

26. Kaniskan, H. U.; Konze, K. D.; Jin, J., Selective inhibitors of protein methyltransferases. *J. Med. Chem.* **2015,** *58* (4), 1596-629.

27. Thinnes, C. C.; England, K. S.; Kawamura, A.; Chowdhury, R.; Schofield, C. J.; Hopkinson, R. J., Targeting histone lysine demethylases — Progress, challenges, and the future. *BBA-Gene. Regul. Mech.* **2014,** *1839* (12), 1416-1432.

28. Hojfeldt, J. W.; Agger, K.; Helin, K., Histone lysine demethylases as targets for anticancer therapy. *Nat. Rev. Drug. Discov.* **2013,** *12* (12), 917-930.

29. Shah, R. N.; Grzybowski, A. T.; Cornett, E. M.; Johnstone, A. L.; Dickson, B. M.; Boone, B. A.; Cheek, M. A.; Cowles, M. W.; Maryanski, D.; Meiners, M. J.; Tiedemann, R. L.; Vaughan, R. M.; Arora, N.; Sun, Z.-W.; Rothbart, S. B.; Keogh, M.-C.; Ruthenburg, A. J., Examining the Roles of H3K4 Methylation States with Systematically Characterized Antibodies. *Mol. Cell.* **2018,** *72* (1), 162-177.e7.

30. Vastenhouw, N. L.; Schier, A. F., Bivalent histone modifications in early embryogenesis. *Curr. Opin. Cell Biol.* **2012,** *24* (3), 374-86.

31. Matsumura, Y.; Nakaki, R.; Inagaki, T.; Yoshida, A.; Kano, Y.; Kimura, H.; Tanaka, T.; Tsutsumi, S.; Nakao, M.; Doi, T.; Fukami, K.; Osborne, Timothy F.; Kodama, T.; Aburatani, H.; Sakai, J., H3K4/H3K9me3 Bivalent Chromatin Domains Targeted by Lineage-Specific DNA Methylation Pauses Adipocyte Differentiation. *Mol. Cell.* **2015,** *60* (4), 584-596.

32. Fischle, W.; Tseng, B. S.; Dormann, H. L.; Ueberheide, B. M.; Garcia, B. A.; Shabanowitz, J.; Hunt, D. F.; Funabiki, H.; Allis, C. D., Regulation of HP1-chromatin binding by histone H3 methylation and phosphorylation. *Nature* **2005,** *438* (7071), 1116-22.

33. Yun, M.; Wu, J.; Workman, J. L.; Li, B., Readers of histone modifications. *Cell Res.* **2011,** *21* (4), 564-578.

34. Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J., How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. *Nat. Struct. Mol. Biol.* **2007,** *14* (11), 1025-40.

35. Maurer-Stroh, S.; Dickens, N. J.; Hughes-Davies, L.; Kouzarides, T.; Eisenhaber, F.; Ponting, C. P., The Tudor domain 'Royal Family': Tudor, plant Agenet, Chromo, PWWP and MBT domains. *Trends Biochem. Sci.* **2003,** *28* (2), 69-74.

36. Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L., Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. *Proc. Natl. Acad. Sci. U.S.A.* **2007,** *104* (27), 11184-8.
37. Kamps, J. J. A. G.; Huang, J.; Poater, J.; Xu, C.; Pieters, B. J. G. E.; Dong, A.; Min, J.; Sherman, W.; Beuming, T.; Matthias Bickelhaupt, F.; Li, H.; Mecinovic, J., Chemical basis for the recognition of trimethyllysine by epigenetic reader proteins. *Nat. Commun.* **2015,** *6*, 1-12.
38. Beaver, J. E.; Waters, M. L., Molecular Recognition of Lys and Arg Methylation. *ACS Chemical Biology* **2016,** *11* (3), 643-653.
39. Li, H.; Fischle, W.; Wang, W.; Duncan, E. M.; Liang, L.; Murakami-Ishibe, S.; Allis, C. D.; Patel, D. J., Structural basis for lower lysine methylation state-specific readout by MBT repeats of L3MBTL1 and an engineered PHD finger. *Mol. Cell.* **2007,** *28* (4), 677-91.
40. Botuyan, M. V.; Lee, J.; Ward, I. M.; Kim, J.-E.; Thompson, J. R.; Chen, J.; Mer, G., Structural Basis for the Methylation State-Specific Recognition of Histone H4-K20 by 53BP1 and Crb2 in DNA Repair. *Cell* **2006,** *127* (7), 1361-1373.
41. Kaustov, L.; Ouyang, H.; Amaya, M.; Lemak, A.; Nady, N.; Duan, S.; Wasney, G. A.; Li, Z.; Vedadi, M.; Schapira, M.; Min, J.; Arrowsmith, C. H., Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* **2011,** *286* (1), 521-9.
42. Musselman, C. A.; Lalonde, M.-E.; Côté, J.; Kutateladze, T. G., Perceiving the epigenetic landscape through histone readers. *Nat. Struct. Mol. Biol.* **2012,** *19* (12), 1218-1227.
43. Mansfield, R. E.; Musselman, C. A.; Kwan, A. H.; Oliver, S. S.; Garske, A. L.; Davrazou, F.; Denu, J. M.; Kutateladze, T. G.; Mackay, J. P., Plant homeodomain (PHD) fingers of CHD4 are histone H3-binding modules with preference for unmodified H3K4 and methylated H3K9. *J Biol Chem* **2011,** *286* (13), 11779-91.
44. Ruthenburg, A. J.; Allis, C. D.; Wysocka, J., Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark. *Mol. Cell.* **2007,** *25* (1), 15-30.
45. Lee, J.-S.; Smith, E.; Shilatifard, A., The Language of Histone Crosstalk. *Cell* **2010,** *142* (5), 682-685.
46. Ruthenburg, A. J.; Li, H.; Patel, D. J.; Allis, C. D., Multivalent engagement of chromatin modifications by linked binding modules. *Nat. Rev. Mol. Cell Biol.* **2007,** *8* (12), 983-94.
47. Wang, Z.; Patel, D. J., Combinatorial readout of dual histone modifications by paired chromatin-associated modules. *J. Biol. Chem.* **2011,** *286* (21), 18363-8.
48. Adams-Cioaba, M. A.; Min, J., Structure and function of histone methylation binding proteins. *Biochem. Cell Biol.* **2009,** *87* (1), 93-105.
49. Müller, C.; Leutz, A., Chromatin remodeling in development and differentiation. *Curr. Opin. Genet. Dev.* **2001,** *11* (2), 167-174.
50. Chi, P.; Allis, C. D.; Wang, G. G., Covalent histone modifications--miswritten, misinterpreted and mis-erased in human cancers. *Nat. Rev. Cancer* **2010,** *10* (7), 457-69.
51. Albert, M.; Helin, K., Histone methyltransferases in cancer. *Semin. Cell. Dev. Biol.* **2010,** *21* (2), 209-20.

52. Varier, R. A.; Timmers, H. T. M., Histone lysine methylation and demethylation pathways in cancer. *BBA-Rev. Cancer* **2011,** *1815* (1), 75-89.

53. Wang, G. G.; Song, J.; Wang, Z.; Dormann, H. L.; Casadio, F.; Li, H.; Luo, J.-L.; Patel, D. J.; Allis, C. D., Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **2009,** *459* (7248), 847-851.

54. Wang, G. G.; Allis, C. D.; Chi, P., Chromatin remodeling and cancer, part I: covalent histone modifications. *Trends Mol. Med.* **2007,** *13* (9), 363-372.

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

56. Bonasio, R.; Lecona, E.; Reinberg, D., MBT domain proteins in development and disease. *Semin. Cell. Dev. Biol.* **2010,** *21* (2), 221-230.

57. Ali, M.; Hom, R. A.; Blakeslee, W.; Ikenouye, L.; Kutateladze, T. G., Diverse functions of PHD fingers of the MLL/KMT2 subfamily. *BBA-Mol. Cell Res.* **2014,** *1843* (2), 366-371.

58. Bracken, A. P.; Helin, K., Polycomb group proteins: navigators of lineage pathways led astray in cancer. *Nat. Rev. Cancer* **2009,** *9* (11), 773-784.

59. Jacobs, S. A.; Khorasanizadeh, S., Structure of HP1 chromodomain bound to a lysine 9-methylated histone H3 tail. *Science* **2002,** *295* (5562), 2080-3.

60. Fischle, W.; Wang, Y.; Jacobs, S. A.; Kim, Y.; Allis, C. D.; Khorasanizadeh, S., Molecular basis for the discrimination of repressive methyl-lysine marks in histone H3 by Polycomb and HP1 chromodomains. *Genes Dev.* **2003,** *17* (15), 1870-81.

61. Min, J.; Zhang, Y.; Xu, R. M., Structural basis for specific binding of Polycomb chromodomain to histone H3 methylated at Lys 27. *Genes Dev.* **2003,** *17* (15), 1823-8.

62. Pena, P. V.; Davrazou, F.; Shi, X.; Walter, K. L.; Verkhusha, V. V.; Gozani, O.; Zhao, R.; Kutateladze, T. G., Molecular mechanism of histone H3K4me3 recognition by plant homeodomain of ING2. *Nature* **2006,** *442* (7098), 100-3.

63. Taverna, S. D.; Ilin, S.; Rogers, R. S.; Tanny, J. C.; Lavender, H.; Li, H.; Baker, L.; Boyle, J.; Blair, L. P.; Chait, B. T.; Patel, D. J.; Aitchison, J. D.; Tackett, A. J.; Allis, C. D., Yng1 PHD finger binding to H3 trimethylated at K4 promotes NuA3 HAT activity at K14 of H3 and transcription at a subset of targeted ORFs. *Mol. Cell.* **2006,** *24* (5), 785-96.

64. Liu, L.; Zhen, X. T.; Denton, E.; Marsden, B. D.; Schapira, M., ChromoHub: a data hub for navigators of chromatin-mediated signalling. *Bioinformatics* **2012,** *28* (16), 2205-6.

65. Gao, C.; Herold, J. M.; Kireev, D.; Wigle, T.; Norris, J. L.; Frye, S., Biophysical probes reveal a "compromise" nature of the methyl-lysine binding pocket in L3MBTL1. *J. Am. Chem. Soc.* **2011,** *133* (14), 5357-62.

66. Santiago, C.; Nguyen, K.; Schapira, M., Druggability of methyl-lysine binding sites. *J. Comput. Aided Mol. Des.* **2011,** *25* (12), 1171-8.

67. Kalakonda, N.; Fischle, W.; Boccuni, P.; Gurvich, N.; Hoya-Arias, R.; Zhao, X.; Miyata, Y.; Macgrogan, D.; Zhang, J.; Sims, J. K.; Rice, J. C.; Nimer, S. D., Histone H4 lysine 20 monomethylation promotes transcriptional repression by L3MBTL1. *Oncogene* **2008,** *27* (31), 4293-304.

68. Nady, N.; Krichevsky, L.; Zhong, N.; Duan, S.; Tempel, W.; Amaya, M. F.; Ravichandran, M.; Arrowsmith, C. H., Histone Recognition by Human Malignant Brain Tumor Domains. *J. Mol. Biol.* **2012,** *423* (5), 702-718.

69. Trojer, P.; Li, G.; Sims, R. J., 3rd; Vaquero, A.; Kalakonda, N.; Boccuni, P.; Lee, D.; Erdjument-Bromage, H.; Tempst, P.; Nimer, S. D.; Wang, Y. H.; Reinberg, D., L3MBTL1, a histone-methylation-dependent chromatin lock. *Cell* **2007,** *129* (5), 915-28.
70. West, L. E.; Roy, S.; Lachmi-Weiner, K.; Hayashi, R.; Shi, X.; Appella, E.; Kutateladze, T. G.; Gozani, O., The MBT repeats of L3MBTL1 link SET8-mediated p53 methylation at lysine 382 to target gene repression. *J. Biol. Chem.* **2010,** *285* (48), 37725-32.
71. Wismar, J.; Loffler, T.; Habtemichael, N.; Vef, O.; Geissen, M.; Zirwes, R.; Altmeyer, W.; Sass, H.; Gateff, E., The Drosophila melanogaster tumor suppressor gene lethal(3)malignant brain tumor encodes a proline-rich protein with a novel zinc finger. *Mech. Dev.* **1995,** *53* (1), 141-54.
72. Gurvich, N.; Perna, F.; Farina, A.; Voza, F.; Menendez, S.; Hurwitz, J.; Nimer, S. D., L3MBTL1 polycomb protein, a candidate tumor suppressor in del(20q12) myeloid disorders, is essential for genome stability. *Proc. Natl. Acad. Sci. U.S.A.* **2010,** *107* (52), 22552-7.
73. Perna, F.; Gurvich, N.; Hoya-Arias, R.; Abdel-Wahab, O.; Levine, R. L.; Asai, T.; Voza, F.; Menendez, S.; Wang, L.; Liu, F.; Zhao, X.; Nimer, S. D., Depletion of L3MBTL1 promotes the erythroid differentiation of human hematopoietic progenitor cells: possible role in 20q- polycythemia vera. *Blood* **2010,** *116* (15), 2812-21.
74. Mao, W.; Salzberg, A. C.; Uchigashima, M.; Hasegawa, Y.; Hock, H.; Watanabe, M.; Akbarian, S.; Kawasawa, Y. I.; Futai, K., Activity-Induced Regulation of Synaptic Strength through the Chromatin Reader L3mbtl1. *Cell Rep.* **2018,** *23* (11), 3209-3222.
75. Northcott, P. A.; Nakahara, Y.; Wu, X.; Feuk, L.; Ellison, D. W.; Croul, S.; Mack, S.; Kongkham, P. N.; Peacock, J.; Dubuc, A.; Ra, Y. S.; Zilberberg, K.; McLeod, J.; Scherer, S. W.; Sunil Rao, J.; Eberhart, C. G.; Grajkowska, W.; Gillespie, Y.; Lach, B.; Grundy, R.; Pollack, I. F.; Hamilton, R. L.; Van Meter, T.; Carlotti, C. G.; Boop, F.; Bigner, D.; Gilbertson, R. J.; Rutka, J. T.; Taylor, M. D., Multiple recurrent genetic events converge on control of histone lysine methylation in medulloblastoma. *Nat. Genet.* **2009,** *41* (4), 465-72.
76. Honda, H.; Takubo, K.; Oda, H.; Kosaki, K.; Tazaki, T.; Yamasaki, N.; Miyazaki, K.; Moore, K. A.; Honda, Z.; Suda, T.; Lemischka, I. R., Hemp, an mbt domain-containing protein, plays essential roles in hematopoietic stem cell function and skeletal formation. *Proc. Natl. Acad. Sci. U. S. A.* **2011,** *108* (6), 2468-73.
77. Chen, C.; Nott, T. J.; Jin, J.; Pawson, T., Deciphering arginine methylation: Tudor tells the tale. *Nat. Rev. Mol. Cell Biol.* **2011,** *12* (10), 629-42.
78. Pek, J. W.; Anand, A.; Kai, T., Tudor domain proteins in development. *Development* **2012,** *139* (13), 2255-66.
79. Tang, J.; Cho, N. W.; Cui, G.; Manion, E. M.; Shanbhag, N. M.; Botuyan, M. V.; Mer, G.; Greenberg, R. A., Acetylation Limits 53BP1 Association with Damaged Chromatin to Promote Homologous Recombination. *Nat. Struct. Mol. Biol.* **2013,** *20* (3), 317-325.
80. Mallette, F. A.; Mattiroli, F.; Cui, G.; Young, L. C.; Hendzel, M. J.; Mer, G.; Sixma, T. K.; Richard, S., RNF8- and RNF168-dependent degradation of KDM4A/JMJD2A triggers 53BP1 recruitment to DNA damage sites. *EMBO J.* **2012,** *31* (8), 1865-78.

81. Lu, R.; Wang, G. G., Tudor: a versatile family of histone methylation 'readers'. *Trends Biochem. Sci.* **2013,** *38* (11), 546-555.

82. Yang, N.; Wang, W.; Wang, Y.; Wang, M.; Zhao, Q.; Rao, Z.; Zhu, B.; Xu, R.-M., Distinct mode of methylated lysine-4 of histone H3 recognition by tandem tudor-like domains of Spindlin1. *Proc. Natl. Acad. Sci. U. S. A.* **2012,** *109* (44), 17954-17959.

83. Nady, N.; Lemak, A.; Walker, J. R.; Avvakumov, G. V.; Kareta, M. S.; Achour, M.; Xue, S.; Duan, S.; Allali-Hassani, A.; Zuo, X.; Wang, Y. X.; Bronner, C.; Chedin, F.; Arrowsmith, C. H.; Dhe-Paganon, S., Recognition of multivalent histone states associated with heterochromatin by UHRF1 protein. *J Biol Chem* **2011,** *286* (27), 24300-11.

84. Cheng, J.; Yang, Y.; Fang, J.; Xiao, J.; Zhu, T.; Chen, F.; Wang, P.; Li, Z.; Yang, H.; Xu, Y., Structural insight into coordinated recognition of trimethylated histone H3 lysine 9 (H3K9me3) by the plant homeodomain (PHD) and tandem tudor domain (TTD) of UHRF1 (ubiquitin-like, containing PHD and RING finger domains, 1) protein. *J Biol Chem* **2013,** *288* (2), 1329-39.

85. Arita, K.; Isogai, S.; Oda, T.; Unoki, M.; Sugita, K.; Sekiyama, N.; Kuwata, K.; Hamamoto, R.; Tochio, H.; Sato, M.; Ariyoshi, M.; Shirakawa, M., Recognition of modification status on a histone H3 tail by linked histone reader modules of the epigenetic regulator UHRF1. *Proc Natl Acad Sci U S A* **2012,** *109* (32), 12950-5.

86. Shibata, A.; Barton, O.; Noon, A. T.; Dahm, K.; Deckbar, D.; Goodarzi, A. A.; Lobrich, M.; Jeggo, P. A., Role of ATM and the damage response mediator proteins 53BP1 and MDC1 in the maintenance of G(2)/M checkpoint arrest. *Mol. Cell. Biol.* **2010,** *30* (13), 3371-83.

87. Escribano-Diaz, C.; Orthwein, A.; Fradet-Turcotte, A.; Xing, M.; Young, J. T.; Tkac, J.; Cook, M. A.; Rosebrock, A. P.; Munro, M.; Canny, M. D.; Xu, D.; Durocher, D., A cell cycle-dependent regulatory circuit composed of 53BP1-RIF1 and BRCA1-CtIP controls DNA repair pathway choice. *Mol. Cell.* **2013,** *49* (5), 872-83.

88. Daley, J. M.; Sung, P., 53BP1, BRCA1, and the Choice between Recombination and End Joining at DNA Double-Strand Breaks. *Mol. Cell. Biol.* **2014,** *34* (8), 1380-1388.

89. Wu, J.; Lu, L. Y.; Yu, X., The role of BRCA1 in DNA damage response. *Protein & cell* **2010,** *1* (2), 117-23.

90. Neff, R. T.; Senter, L.; Salani, R., BRCA mutation in ovarian cancer: testing, implications and treatment considerations. *Ther Adv Med Oncol* **2017,** *9* (8), 519-531.

91. Godet, I.; Gilkes, D. M., BRCA1 and BRCA2 mutations and treatment strategies for breast cancer. *Integr. Cancer Sci. Ther.* **2017,** *4* (1), 10.15761/ICST.1000228.

92. Bouwman, P.; Aly, A.; Escandell, J. M.; Pieterse, M.; Bartkova, J.; van der Gulden, H.; Hiddingh, S.; Thanasoula, M.; Kulkarni, A.; Yang, Q.; Haffty, B. G.; Tommiska, J.; Blomqvist, C.; Drapkin, R.; Adams, D. J.; Nevanlinna, H.; Bartek, J.; Tarsounas, M.; Ganesan, S.; Jonkers, J., 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. *Nat. Struct. Mol. Biol.* **2010,** *17* (6), 688-695.

93. Bi, J.; Huang, A.; Liu, T.; Zhang, T.; Ma, H., Expression of DNA damage checkpoint 53BP1 is correlated with prognosis, cell proliferation and apoptosis in colorectal cancer. *Int. J. Clin. Exp. Pathol.* **2015,** *8* (6), 6070-6082.

94. Ward, I. M.; Minn, K.; van Deursen, J.; Chen, J., p53 Binding protein 53BP1 is required for DNA damage responses and tumor suppression in mice. *Mol. Cell. Biol.* **2003,** *23* (7), 2556-63.

95. Zhang, P.; Cong, B.; Yuan, H.; Chen, L.; Lv, Y.; Bai, C.; Nan, X.; Shi, S.; Yue, W.; Pei, X., Overexpression of spindlin1 induces metaphase arrest and chromosomal instability. *J. Cell. Physiol.* **2008,** *217* (2), 400-8.
96. Yue, W.; Sun, L. Y.; Li, C. H.; Zhang, L. X.; Pei, X. T., [Screening and identification of ovarian carcinomas related genes]. *Chin. J. Cancer.* **2004,** *23* (2), 141-5.
97. Wang, J. X.; Zeng, Q.; Chen, L.; Du, J. C.; Yan, X. L.; Yuan, H. F.; Zhai, C.; Zhou, J. N.; Jia, Y. L.; Yue, W.; Pei, X. T., SPINDLIN1 promotes cancer cell proliferation through activation of WNT/TCF-4 signaling. *Mol. Cancer Res.* **2012,** *10* (3), 326-35.
98. Franz, H.; Greschik, H.; Willmann, D.; Ozretic, L.; Jilg, C. A.; Wardelmann, E.; Jung, M.; Buettner, R.; Schule, R., The histone code reader SPIN1 controls RET signaling in liposarcoma. *Oncotarget* **2015,** *6* (7), 4773-89.
99. Wang, W.; Chen, Z.; Mao, Z.; Zhang, H.; Ding, X.; Chen, S.; Zhang, X.; Xu, R.; Zhu, B., Nucleolar protein Spindlin1 recognizes H3K4 methylation and stimulates the expression of rRNA genes. *EMBO reports* **2011,** *12* (11), 1160-6.
100. Bronner, C.; Krifa, M.; Mousli, M., Increasing role of UHRF1 in the reading and inheritance of the epigenetic code as well as in tumorogenesis. *Biochem. Pharmacol.* **2013,** *86* (12), 1643-9.
101. Alhosin, M.; Omran, Z.; Zamzami, M. A.; Al-Malki, A. L.; Choudhry, H.; Mousli, M.; Bronner, C., Signalling pathways in UHRF1-dependent regulation of tumor suppressor genes in cancer. *J. Exp. Clin. Cancer Res.* **2016,** *35* (1), 174.
102. Unoki, M.; Nishidate, T.; Nakamura, Y., ICBP90, an E2F-1 target, recruits HDAC1 and binds to methyl-CpG through its SRA domain. *Oncogene* **2004,** *23* (46), 7601-10.
103. Jin, W.; Chen, L.; Chen, Y.; Xu, S. G.; Di, G. H.; Yin, W. J.; Wu, J.; Shao, Z. M., UHRF1 is associated with epigenetic silencing of BRCA1 in sporadic breast cancer. *Breast Cancer Res. Treat.* **2010,** *123* (2), 359-73.
104. Jeanblanc, M.; Mousli, M.; Hopfner, R.; Bathami, K.; Martinet, N.; Abbady, A. Q.; Siffert, J. C.; Mathieu, E.; Muller, C. D.; Bronner, C., The retinoblastoma gene and its product are targeted by ICBP90: a key mechanism in the G1/S transition during the cell cycle. *Oncogene* **2005,** *24* (49), 7337-45.
105. Sidhu, H.; Capalash, N., UHRF1: The key regulator of epigenetics and molecular target for cancer therapeutics. *Tumour Biol* **2017,** *39* (2), 1010428317692205.
106. Geng, Y.; Gao, Y.; Ju, H.; Yan, F., Diagnostic and prognostic value of plasma and tissue ubiquitin-like, containing PHD and RING finger domains 1 in breast cancer patients. *Cancer science* **2013,** *104* (2), 194-9.
107. Ge, M.; Gui, Z.; Wang, X.; Yan, F., Analysis of the UHRF1 expression in serum and tissue for gastric cancer detection. *Biomarkers* **2015,** *20* (3), 183-8.
108. Unoki, M.; Kelly, J. D.; Neal, D. E.; Ponder, B. A.; Nakamura, Y.; Hamamoto, R., UHRF1 is a novel molecular marker for diagnosis and the prognosis of bladder cancer. *Br. J. Cancer* **2009,** *101* (1), 98-105.
109. Baker, L. A.; Allis, C. D.; Wang, G. G., PHD fingers in human diseases: disorders arising from misinterpreting epigenetic marks. *Mutat. Res.* **2008,** *647* (1-2), 3-12.
110. Musselman, C. A.; Kutateladze, T. G., PHD Fingers: Epigenetic Effectors and Potential Drug Targets. *Mol. Interventions* **2009,** *9* (6), 314-323.

111. Sanchez, R.; Zhou, M. M., The PHD finger: a versatile epigenome reader. *Trends Biochem. Sci.* **2011,** *36* (7), 364-72.
112. Fiedler, M.; Sanchez-Barrena, M. J.; Nekrasov, M.; Mieszczanek, J.; Rybin, V.; Muller, J.; Evans, P.; Bienz, M., Decoding of methylated histone H3 tail by the Pygo-BCL9 Wnt signaling complex. *Mol. Cell.* **2008,** *30* (4), 507-18.
113. Miller, T. C.; Rutherford, T. J.; Johnson, C. M.; Fiedler, M.; Bienz, M., Allosteric remodelling of the histone H3 binding pocket in the Pygo2 PHD finger triggered by its binding to the B9L/BCL9 co-factor. *J. Mol. Biol.* **2010,** *401* (5), 969-84.
114. Andrews, Phillip G. P.; He, Z.; Popadiuk, C.; Kao, Kenneth R., The transcriptional activity of Pygopus is enhanced by its interaction with cAMP-response-element-binding protein (CREB)-binding protein. *Biochem. J* **2009,** *422* (3), 493-501.
115. Chen, J.; Luo, Q.; Yuan, Y.; Huang, X.; Cai, W.; Li, C.; Wei, T.; Zhang, L.; Yang, M.; Liu, Q.; Ye, G.; Dai, X.; Li, B., Pygo2 associates with MLL2 histone methyltransferase and GCN5 histone acetyltransferase complexes to augment Wnt target gene expression and breast cancer stem-like cell expansion. *Mol. Cell. Biol.* **2010,** *30* (24), 5621-35.
116. Gu, B.; Sun, P.; Yuan, Y.; Moraes, R. C.; Li, A.; Teng, A.; Agrawal, A.; Rheaume, C.; Bilanchone, V.; Veltmaat, J. M.; Takemaru, K.; Millar, S.; Lee, E. Y.; Lewis, M. T.; Li, B.; Dai, X., Pygo2 expands mammary progenitor cells by facilitating histone H3 K4 methylation. *J. Cell Biol.* **2009,** *185* (5), 811-26.
117. Gu, B.; Watanabe, K.; Dai, X., Pygo2 regulates histone gene expression and H3 K56 acetylation in human mammary epithelial cells. *Cell cycle* **2012,** *11* (1), 79-87.
118. Nair, M.; Nagamori, I.; Sun, P.; Mishra, D. P.; Rheaume, C.; Li, B.; Sassone-Corsi, P.; Dai, X., Nuclear regulator Pygo2 controls spermiogenesis and histone H3 acetylation. *Dev. Biol.* **2008,** *320* (2), 446-55.
119. Belenkaya, T. Y.; Han, C.; Standley, H. J.; Lin, X.; Houston, D. W.; Heasman, J.; Lin, X., pygopus Encodes a nuclear protein essential for wingless/Wnt signaling. *Development* **2002,** *129* (17), 4089-101.
120. Jessen, S.; Gu, B.; Dai, X., Pygopus and the Wnt signaling pathway: A diverse set of connections. *BioEssays* **2008,** *30* (5), 448-456.
121. Gu, B.; Watanabe, K.; Sun, P.; Fallahi, M.; Dai, X., Chromatin Effector Pygo2 Mediates Wnt-Notch Crosstalk to Suppress Luminal/Alveolar Potential of Mammary Stem and Basal Cells. *Cell Stem Cell* **2013,** *13* (1), 48-61.
122. Andrews, P. G.; Lake, B. B.; Popadiuk, C.; Kao, K. R., Requirement of Pygopus 2 in breast cancer. *Int. J. Oncol.* **2007,** *30* (2), 357-63.
123. Yu, H.; Jiang, Y.; Liu, L.; Shan, W.; Chu, X.; Yang, Z.; Yang, Z.-Q., Integrative genomic and transcriptomic analysis for pinpointing recurrent alterations of plant homeodomain genes and their clinical significance in breast cancer. *Oncotarget* **2017,** *8* (8), 13099-13115.
124. Popadiuk, C. M.; Xiong, J.; Wells, M. G.; Andrews, P. G.; Dankwa, K.; Hirasawa, K.; Lake, B. B.; Kao, K. R., Antisense Suppression of Pygopus2 Results in Growth Arrest of Epithelial Ovarian Cancer. *Clin. Cancer Res.* **2006,** *12* (7), 2216-2223.
125. Yang, L.; Lin, C.; Jin, C.; Yang, J. C.; Tanasa, B.; Li, W.; Merkurjev, D.; Ohgi, K. A.; Meng, D.; Zhang, J.; Evans, C. P.; Rosenfeld, M. G., lncRNA-dependent mechanisms of androgen-receptor-regulated gene activation programs. *Nature* **2013,** *500*, 598-602.

126. Lu, X.; Pan, X.; Wu, C. J.; Zhao, D.; Feng, S.; Zang, Y.; Lee, R.; Khadka, S.; Amin, S. B.; Jin, E. J.; Shang, X.; Deng, P.; Luo, Y.; Morgenlander, W. R.; Weinrich, J.; Lu, X.; Jiang, S.; Chang, Q.; Navone, N. M.; Troncoso, P.; DePinho, R. A.; Wang, Y. A., An In Vivo Screen Identifies PYGO2 as a Driver for Metastatic Prostate Cancer. *Cancer Res* **2018,** *78* (14), 3823-3833.

127. Zhou, C.; Zhang, Y.; Dai, J.; Zhou, M.; Liu, M.; Wang, Y.; Chen, X.-Z.; Tang, J., Pygo2 functions as a prognostic factor for glioma due to its up-regulation of H3K4me3 and promotion of MLL1/MLL2 complex recruitment. *Sci. Rep.* **2016,** *6*, 22066.

128. Harmeyer, K. M.; Facompre, N. D.; Herlyn, M.; Basu, D., JARID1 Histone Demethylases: Emerging Targets in Cancer. *Trends Cancer* **2017,** *3* (10), 713-725.

129. Schapira, M.; Tyers, M.; Torrent, M.; Arrowsmith, C. H., WD40 repeat domain proteins: a novel target class? *Nat Rev. Drug. Discov.* **2017,** *16*, 773-786.

130. Margueron, R.; Justin, N.; Ohno, K.; Sharpe, M. L.; Son, J.; Drury, W. J.; Voigt, P.; Martin, S.; Taylor, W. R.; De Marco, V.; Pirrotta, V.; Reinberg, D.; Gamblin, S. J., Role of the polycomb protein Eed in the propagation of repressive histone marks. *Nature* **2009,** *461* (7265), 762-767.

131. Jiao, L.; Liu, X., Structural basis of histone H3K27 trimethylation by an active polycomb repressive complex 2. *Science* **2015,** *350* (6258), aac4383.

132. Xu, C.; Bian, C.; Yang, W.; Galka, M.; Ouyang, H.; Chen, C.; Qiu, W.; Liu, H.; Jones, A. E.; MacKenzie, F.; Pan, P.; Li, S. S.-C.; Wang, H.; Min, J., Binding of different histone marks differentially regulates the activity and specificity of polycomb repressive complex 2 (PRC2). *Proc. Natl. Acad. Sci. U. S. A.* **2010,** *107* (45), 19266-19271.

133. Sanulli, S.; Justin, N.; Teissandier, A.; Ancelin, K.; Portoso, M.; Caron, M.; Michaud, A.; Lombard, B.; da Rocha, S. T.; Offer, J.; Loew, D.; Servant, N.; Wassef, M.; Burlina, F.; Gamblin, S. J.; Heard, E.; Margueron, R., Jarid2 Methylation via the PRC2 Complex Regulates H3K27me3 Deposition during Cell Differentiation. *Mol. Cell.* **2015,** *57* (5), 769-783.

134. Song, R.; Wang, Z.-D.; Schapira, M., Disease Association and Druggability of WD40 Repeat Proteins. *J. Proteome Res.* **2017,** *16* (10), 3766-3773.

135. Shi, Y.; Wang, X.-x.; Zhuang, Y.-w.; Jiang, Y.; Melcher, K.; Xu, H. E., Structure of the PRC2 complex and application to drug discovery. *Acta Pharmacol. Sin.* **2017,** *38* (7), 963-976.

136. Kim, K. H.; Roberts, C. W. M., Targeting EZH2 in cancer. *Nat. Med.* **2016,** *22*, 128-134.

137. Kim, W.; Bird, G. H.; Neff, T.; Guo, G.; Kerenyi, M. A.; Walensky, L. D.; Orkin, S. H., Targeted disruption of the EZH2–EED complex inhibits EZH2-dependent cancer. *Nat. Chem. Biol.* **2013,** *9*, 643.

138. Bannister, A. J.; Zegerman, P.; Partridge, J. F.; Miska, E. A.; Thomas, J. O.; Allshire, R. C.; Kouzarides, T., Selective recognition of methylated lysine 9 on histone H3 by the HP1 chromo domain. *Nature* **2001,** *410* (6824), 120-4.

139. Lachner, M.; O'Carroll, D.; Rea, S.; Mechtler, K.; Jenuwein, T., Methylation of histone H3 lysine 9 creates a binding site for HP1 proteins. *Nature* **2001,** *410* (6824), 116-20.

140. Eissenberg, J. C., Structural biology of the chromodomain: Form and function. *Gene* **2012,** *496* (2), 69-78.

141. Dialynas, G. K.; Vitalini, M. W.; Wallrath, L. L., Linking Heterochromatin Protein 1 (HP1) to cancer progression. *Mutat. Res.* **2008,** *647* (1-2), 13-20.

142. Di Croce, L.; Helin, K., Transcriptional regulation by Polycomb group proteins. *Nat. Struct. Mol. Biol.* **2013,** *20*, 1147-1155.

143. Gil, J.; Bernard, D.; Martinez, D.; Beach, D., Polycomb CBX7 has a unifying role in cellular lifespan. *Nat. Cell Biol.* **2004,** *6* (1), 67-72.

144. Yap, K. L.; Li, S.; Muñoz-Cabello, A. M.; Raguz, S.; Zeng, L.; Mujtaba, S.; Gil, J.; Walsh, M. J.; Zhou, M.-M., Molecular Interplay of the Non-coding RNA ANRIL and Methylated Histone H3 Lysine 27 by Polycomb CBX7 in Transcriptional Silencing of INK4a. *Mol. Cell.* **2010,** *38* (5), 662-674.

145. Caron, C.; Pivot-Pajot, C.; van Grunsven, L. A.; Col, E.; Lestrat, C.; Rousseaux, S.; Khochbin, S., Cdyl: a new transcriptional co-repressor. *EMBO Rep.* **2003,** *4* (9), 877-882.

146. Kim, J.; Daniel, J.; Espejo, A.; Lake, A.; Krishna, M.; Xia, L.; Zhang, Y.; Bedford, M. T., Tudor, MBT and chromo domains gauge the degree of lysine methylation. *EMBO Rep.* **2006,** *7* (4), 397-403.

147. Vermeulen, M.; Eberl, H. C.; Matarese, F.; Marks, H.; Denissov, S.; Butter, F.; Lee, K. K.; Olsen, J. V.; Hyman, A. A.; Stunnenberg, H. G.; Mann, M., Quantitative Interaction Proteomics and Genome-wide Profiling of Epigenetic Histone Marks and Their Readers. *Cell* **2010,** *142* (6), 967-980.

148. Fischle, W.; Franz, H.; Jacobs, S. A.; Allis, C. D.; Khorasanizadeh, S., Specificity of the Chromodomain Y Chromosome Family of Chromodomains for Lysine-methylated ARK(S/T) Motifs. *J. Biol. Chem.* **2008,** *283* (28), 19626-19635.

149. Simhadri, C.; Daze, K. D.; Douglas, S. F.; Quon, T. T.; Dev, A.; Gignac, M. C.; Peng, F.; Heller, M.; Boulanger, M. J.; Wulff, J. E.; Hof, F., Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J. Med. Chem.* **2014,** *57* (7), 2874-83.

150. Simhadri, C.; Gignac, M. C.; Anderson, C. J.; Milosevich, N.; Dheri, A.; Prashar, N.; Flemmer, R. T.; Dev, A.; Henderson, T. G.; Douglas, S. F.; Wulff, J. E.; Hof, F., Structure–Activity Relationships of Cbx7 Inhibitors, Including Selectivity Studies against Other Cbx Proteins. *ACS Omega* **2016,** *1* (4), 541-551.

151. Stuckey, J. I.; Dickson, B. M.; Cheng, N.; Liu, Y.; Norris, J. L.; Cholensky, S. H.; Tempel, W.; Qin, S.; Huber, K. G.; Sagum, C.; Black, K.; Li, F.; Huang, X.-P.; Roth, B. L.; Baughman, B. M.; Senisterra, G.; Pattenden, S. G.; Vedadi, M.; Brown, P. J.; Bedford, M. T.; Min, J.; Arrowsmith, C. H.; James, L. I.; Frye, S. V., A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. *Nat. Chem. Biol.* **2016,** *12* (3), 180-187.

152. Stuckey, J. I.; Simpson, C.; Norris-Drouin, J. L.; Cholensky, S. H.; Lee, J.; Pasca, R.; Cheng, N.; Dickson, B. M.; Pearce, K. H.; Frye, S. V.; James, L. I., Structure–Activity Relationships and Kinetic Studies of Peptidic Antagonists of CBX Chromodomains. *J. Med. Chem.* **2016,** *59* (19), 8913-8923.

153. Ren, C.; Morohashi, K.; Plotnikov, A. N.; Jakoncic, J.; Smith, S. G.; Li, J.; Zeng, L.; Rodriguez, Y.; Stojanoff, V.; Walsh, M.; Zhou, M. M., Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem. Biol.* **2015,** *22* (2), 161-8.

154. Ren, C.; Smith, S. G.; Yap, K. L.; Li, S.; Li, J.; Mezei, M.; Vincek, A. S.; Aguilo, F.; Walsh, M. J.; Zhou, M.-M., Structure-Guided Discovery of Selective Antagonists for

the Chromodomain of Polycomb Repressive Protein CBX7. *ACS Med. Chem. Lett.* **2016,** *7* (6), 601-5.

155. Milosevich, N.; Gignac, M. C.; McFarlane, J.; Simhadri, C.; Horvath, S.; Daze, K. D.; Croft, C. S.; Dheri, A.; Quon, T. T. H.; Douglas, S. F.; Wulff, J. E.; Paci, I.; Hof, F., Selective Inhibition of CBX6: A Methyllysine Reader Protein in the Polycomb Family. *ACS Med. Chem. Lett.* **2016,** *7* (2), 139-144.

156. Barnash, K. D.; Lamb, K. N.; Stuckey, J. I.; Norris, J. L.; Cholensky, S. H.; Kireev, D. B.; Frye, S. V.; James, L. I., Chromodomain Ligand Optimization via Target-Class Directed Combinatorial Repurposing. *ACS Chemical Biology* **2016,** *11* (9), 2475-2483.

157. Morey, L.; Pascual, G.; Cozzuto, L.; Roma, G.; Wutz, A.; Benitah, S. A.; Di Croce, L., Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell Stem Cell* **2012,** *10* (1), 47-62.

158. Simon, J. A.; Kingston, R. E., Mechanisms of Polycomb gene silencing: knowns and unknowns. *Nat. Rev. Mol. Cell Biol.* **2009,** *10* (10), 697-708.

159. Bernard, D.; Martinez-Leal, J. F.; Rizzo, S.; Martinez, D.; Hudson, D.; Visakorpi, T.; Peters, G.; Carnero, A.; Beach, D.; Gil, J., CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. *Oncogene* **2005,** *24* (36), 5543-51.

160. Aguilo, F.; Zhou, M. M.; Walsh, M. J., Long noncoding RNA, polycomb, and the ghosts haunting INK4b-ARF-INK4a expression. *Cancer Res* **2011,** *71* (16), 5365-9.

161. Li, Q.; Wang, X.; Lu, Z.; Zhang, B.; Guan, Z.; Liu, Z.; Zhong, Q.; Gu, L.; Zhou, J.; Zhu, B.; Ji, J.; Deng, D., Polycomb CBX7 directly controls trimethylation of histone H3 at lysine 9 at the p16 locus. *PloS one* **2010,** *5* (10), e13732.

162. Zhang, X. W.; Zhang, L.; Qin, W.; Yao, X. H.; Zheng, L. Z.; Liu, X.; Li, J.; Guo, W. J., Oncogenic role of the chromobox protein CBX7 in gastric cancer. *J. Exp. Clin. Cancer Res.* **2010,** *29*, 114.

163. Santanach, A.; Blanco, E.; Jiang, H.; Molloy, K. R.; Sansó, M.; LaCava, J.; Morey, L.; Di Croce, L., The Polycomb group protein CBX6 is an essential regulator of embryonic stem cell identity. *Nat. Commun.* **2017,** *8* (1), 1235.

164. Pallante, P.; Forzati, F.; Federico, A.; Arra, C.; Fusco, A., Polycomb protein family member CBX7 plays a critical role in cancer progression. *Am. J. Cancer Res.* **2015,** *5* (5), 1594-601.

165. Forzati, F.; Federico, A.; Pallante, P.; Abbate, A.; Esposito, F.; Malapelle, U.; Sepe, R.; Palma, G.; Troncone, G.; Scarfo, M.; Arra, C.; Fedele, M.; Fusco, A., CBX7 is a tumor suppressor in mice and humans. *J Clin. Investig.* **2012,** *122* (2), 612-23.

166. Shinjo, K.; Yamashita, Y.; Yamamoto, E.; Akatsuka, S.; Uno, N.; Kamiya, A.; Niimi, K.; Sakaguchi, Y.; Nagasaka, T.; Takahashi, T.; Shibata, K.; Kajiyama, H.; Kikkawa, F.; Toyokuni, S., Expression of chromobox homolog 7 (CBX7) is associated with poor prognosis in ovarian clear cell adenocarcinoma via TRAIL-induced apoptotic pathway regulation. *Int. J. Cancer* **2014,** *135* (2), 308-18.

167. Scott, C. L.; Gil, J.; Hernando, E.; Teruya-Feldstein, J.; Narita, M.; Martinez, D.; Visakorpi, T.; Mu, D.; Cordon-Cardo, C.; Peters, G.; Beach, D.; Lowe, S. W., Role of the chromobox protein CBX7 in lymphomagenesis. *Proc. Natl. Acad. Sci. U. S. A.* **2007,** *104* (13), 5389-94.

168. Klauke, K.; Radulović, V.; Broekhuis, M.; Weersing, E.; Zwart, E.; Olthof, S.; Ritsema, M.; Bruggeman, S.; Wu, X.; Helin, K.; Bystrykh, L.; de Haan, G., Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nat. Cell. Biol.* **2013,** *15*, 353-362.

169. Bernstein, E.; Duncan, E. M.; Masui, O.; Gil, J.; Heard, E.; Allis, C. D., Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell. Biol.* **2006,** *26* (7), 2560-9.

170. Zheng, H.; Jiang, W.-h.; Tian, T.; Tan, H.-s.; Chen, Y.; Qiao, G.-l.; Han, J.; Huang, S.-y.; Yang, Y.; Li, S.; Wang, Z.-g.; Gao, R.; Ren, H.; Xing, H.; Ni, J.-s.; Wang, L.-H.; Ma, L.-j.; Zhou, W.-p., CBX6 overexpression contributes to tumor progression and is predictive of a poor prognosis in hepatocellular carcinoma. *Oncotarget* **2017,** *8* (12), 18872-18884.

171. Deng, H.; Guan, X.; Gong, L.; Zeng, J.; Zhang, H.; Chen, M. Y.; Li, G., CBX6 is negatively regulated by EZH2 and plays a potential tumor suppressor role in breast cancer. *Sci. Rep.* **2019,** *9* (1), 197.

172. Zhang, Y.; Yang, X.; Gui, B.; Xie, G.; Zhang, D.; Shang, Y.; Liang, J., Corepressor protein CDYL functions as a molecular bridge between polycomb repressor complex 2 and repressive chromatin mark trimethylated histone lysine 27. *J. Biol. Chem.* **2011,** *286* (49), 42414-25.

173. Liu, S.; Yu, H.; Liu, Y.; Liu, X.; Zhang, Y.; Bu, C.; Yuan, S.; Chen, Z.; Xie, G.; Li, W.; Xu, B.; Yang, J.; He, L.; Jin, T.; Xiong, Y.; Sun, L.; Liu, X.; Han, C.; Cheng, Z.; Liang, J.; Shang, Y., Chromodomain Protein CDYL Acts as a Crotonyl-CoA Hydratase to Regulate Histone Crotonylation and Spermatogenesis. *Mol. Cell.* **2017,** *67* (5), 853-866.e5.

174. Enguix-Riego, M. V.; Torroglosa, A.; Fernandez, R. M.; Moya-Jimenez, M. J.; de Agustin, J. C.; Antinolo, G.; Borrego, S., Identification of different mechanisms leading to PAX6 down-regulation as potential events contributing to the onset of Hirschsprung disease. *Sci. Rep.* **2016,** *6*, 21160.

175. Qin, R.; Cao, S.; Lyu, T.; Qi, C.; Zhang, W.; Wang, Y., CDYL Deficiency Disrupts Neuronal Migration and Increases Susceptibility to Epilepsy. *Cell Rep.* **2017,** *18* (2), 380-390.

176. Vandamme, J.; Volkel, P.; Rosnoblet, C.; Le Faou, P.; Angrand, P. O., Interaction proteomics analysis of polycomb proteins defines distinct PRC1 complexes in mammalian cells. *Mol. Cell Proteomics* **2011,** *10* (4), M110.002642.

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

178. O'Loghlen, A.; Munoz-Cabello, A. M.; Gaspar-Maia, A.; Wu, H. A.; Banito, A.; Kunowska, N.; Racek, T.; Pemberton, H. N.; Beolchi, P.; Lavial, F.; Masui, O.; Vermeulen, M.; Carroll, T.; Graumann, J.; Heard, E.; Dillon, N.; Azuara, V.; Snijders, A. P.; Peters, G.; Bernstein, E.; Gil, J., MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell Stem Cell* **2012,** *10* (1), 33-46.

179. Pemberton, H.; Anderton, E.; Patel, H.; Brookes, S.; Chandler, H.; Palermo, R.; Stock, J.; Rodriguez-Niedenfuhr, M.; Racek, T.; de Breed, L.; Stewart, A.; Matthews, N.; Peters, G., Genome-wide co-localization of Polycomb orthologs and their effects on gene expression in human fibroblasts. *Genome Biol* **2014,** *15* (2), R23.

180. Overhoff, M. G.; Garbe, J. C.; Koh, J.; Stampfer, M. R.; Beach, D. H.; Bishop, C. L., Cellular senescence mediated by p16INK4A-coupled miRNA pathways. *Nucleic Acids Res.* **2014,** *42* (3), 1606-18.
181. Wu, H. A.; Balsbaugh, J. L.; Chandler, H.; Georgilis, A.; Zullow, H.; Shabanowitz, J.; Hunt, D. F.; Gil, J.; Peters, G.; Bernstein, E., Mitogen-activated protein kinase signaling mediates phosphorylation of polycomb ortholog Cbx7. *J Biol Chem* **2013,** *288* (51), 36398-408.
182. Kawaguchi, T.; Machida, S.; Kurumizaka, H.; Tagami, H.; Nakayama, J.-i., Phosphorylation of CBX2 controls its nucleosome-binding specificity. *The Journal of Biochemistry* **2017,** *162* (5), 343-355.
183. Cao, R.; Zhang, Y., The functions of E(Z)/EZH2-mediated methylation of lysine 27 in histone H3. *Current Opinion in Genetics & Development* **2004,** *14* (2), 155-164.
184. Wang, H.; Wang, L.; Erdjument-Bromage, H.; Vidal, M.; Tempst, P.; Jones, R. S.; Zhang, Y., Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **2004,** *431* (7010), 873-8.
185. Cao, R.; Tsukada, Y.; Zhang, Y., Role of Bmi-1 and Ring1A in H2A ubiquitylation and Hox gene silencing. *Mol. Cell.* **2005,** *20* (6), 845-54.
186. Simon, Jeffrey A.; Kingston, Robert E., Occupying Chromatin: Polycomb Mechanisms for Getting to Genomic Targets, Stopping Transcriptional Traffic, and Staying Put. *Molecular Cell* **2013,** *49* (5), 808-824.
187. Blackledge, N. P.; Rose, N. R.; Klose, R. J., Targeting Polycomb systems to regulate gene expression: modifications to a complex story. *Nature reviews. Molecular cell biology* **2015,** *16* (11), 643-649.
188. Copeland, R. A.; Solomon, M. E.; Richon, V. M., Protein methyltransferases as a target class for drug discovery. *Nat Rev. Drug. Discov.* **2009,** *8* (9), 724-732.
189. Knutson, S. K.; Warholic, N. M.; Wigle, T. J.; Klaus, C. R.; Allain, C. J.; Raimondi, A.; Porter Scott, M.; Chesworth, R.; Moyer, M. P.; Copeland, R. A.; Richon, V. M.; Pollock, R. M.; Kuntz, K. W.; Keilhack, H., Durable tumor regression in genetically altered malignant rhabdoid tumors by inhibition of methyltransferase EZH2. *Proc. Natl. Acad. Sci. U.S.A.* **2013,** *110* (19), 7922-7927.
190. Arrowsmith, C. H.; Bountra, C.; Fish, P. V.; Lee, K.; Schapira, M., Epigenetic protein families: a new frontier for drug discovery. *Nat Rev. Drug. Discov.* **2012,** *11* (5), 384-400.
191. Khan, O.; La Thangue, N. B., HDAC inhibitors in cancer biology: emerging mechanisms and clinical applications. *Immunol. Cell Biol.* **2012,** *90* (1), 85-94.
192. Filippakopoulos, P.; Knapp, S., Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev. Drug. Discov.* **2014,** *13* (5), 337-356.
193. Prince, H. M.; Bishton, M. J.; Harrison, S. J., Clinical Studies of Histone Deacetylase Inhibitors. *Clin. Cancer Res.* **2009,** *15* (12), 3958-3969.
194. Dekker, F. J.; van den Bosch, T.; Martin, N. I., Small molecule inhibitors of histone acetyltransferases and deacetylases are potential drugs for inflammatory diseases. *Drug Discov. Today* **2014,** *19* (5), 654-660.
195. Manzo, F.; Tambaro, F. P.; Mai, A.; Altucci, L., Histone acetyltransferase inhibitors and preclinical studies. *Expert Opin. Ther. Pat.* **2009,** *19* (6), 761-774.
196. Milosevich, N.; Hof, F., Chemical Inhibitors of Epigenetic Methyllysine Reader Proteins. *Biochemistry* **2016,** *55* (11), 1570-1583.

197. Hajduk, P. J.; Huth, J. R.; Tse, C., Predicting protein druggability. *Drug Discov Today* **2005,** *10* (23-24), 1675-82.

198. Herold, J. M.; James, L. I.; Korboukh, V. K.; Gao, C.; Coil, K. E.; Bua, D. J.; Norris, J. L.; Kireev, D. B.; Brown, P. J.; Jin, J.; Janzen, W. P.; Gozani, O.; Frye, S. V., Structure-activity relationships of methyl-lysine reader antagonists. *Med. Chem. Comm.* **2012,** *3* (1), 45-51.

199. James, L. I.; Barsyte-Lovejoy, D.; Zhong, N.; Krichevsky, L.; Korboukh, V. K.; Herold, J. M.; MacNevin, C. J.; Norris, J. L.; Sagum, C. A.; Tempel, W.; Marcon, E.; Guo, H.; Gao, C.; Huang, X. P.; Duan, S.; Emili, A.; Greenblatt, J. F.; Kireev, D. B.; Jin, J.; Janzen, W. P.; Brown, P. J.; Bedford, M. T.; Arrowsmith, C. H.; Frye, S. V., Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nat. Chem. Biol.* **2013,** *9* (3), 184-91.

200. James, L. I.; Korboukh, V. K.; Krichevsky, L.; Baughman, B. M.; Herold, J. M.; Norris, J. L.; Jin, J.; Kireev, D. B.; Janzen, W. P.; Arrowsmith, C. H.; Frye, S. V., Small-molecule ligands of methyl-lysine binding proteins: optimization of selectivity for L3MBTL3. *J. Med. Chem.* **2013,** *56* (18), 7358-71.

201. Perfetti, M. T.; Baughman, B. M.; Dickson, B. M.; Mu, Y.; Cui, G.; Mader, P.; Dong, A.; Norris, J. L.; Rothbart, S. B.; Strahl, B. D.; Brown, P. J.; Janzen, W. P.; Arrowsmith, C. H.; Mer, G.; McBride, K. M.; James, L. I.; Frye, S. V., Identification of a Fragment-like Small Molecule Ligand for the Methyl-lysine Binding Protein, 53BP1. *ACS Chem. Biol.* **2015,** *10* (4), 1072-1081.

202. Wagner, T.; Greschik, H.; Burgahn, T.; Schmidtkunz, K.; Schott, A.-K.; McMillan, J.; Baranauskienė, L.; Xiong, Y.; Fedorov, O.; Jin, J.; Oppermann, U.; Matulis, D.; Schüle, R.; Jung, M., Identification of a small-molecule ligand of the epigenetic reader protein Spindlin1 via a versatile screening platform. *Nucleic Acids Res.* **2016,** *44* (9), e88-e88.

203. Sweis, R. F.; Pliushchev, M.; Brown, P. J.; Guo, J.; Li, F.; Maag, D.; Petros, A. M.; Soni, N. B.; Tse, C.; Vedadi, M.; Michaelides, M. R.; Chiang, G. G.; Pappano, W. N., Discovery and Development of Potent and Selective Inhibitors of Histone Methyltransferase G9a. *ACS Med. Chem. Lett.* **2014,** *5* (2), 205-209.

204. Bae, N.; Viviano, M.; Su, X.; Lv, J.; Cheng, D.; Sagum, C.; Castellano, S.; Bai, X.; Johnson, C.; Khalil, M. I.; Shen, J.; Chen, K.; Li, H.; Sbardella, G.; Bedford, M. T., Developing Spindlin1 small-molecule inhibitors by using protein microarrays. *Nat. Chem. Biol.* **2017,** *13* (7), 750-756.

205. Houliston, R. S.; Lemak, A.; Iqbal, A.; Ivanochko, D.; Duan, S.; Kaustov, L.; Ong, M. S.; Fan, L.; Senisterra, G.; Brown, P. J.; Wang, Y. X.; Arrowsmith, C. H., Conformational dynamics of the TTD-PHD histone reader module of the UHRF1 epigenetic regulator reveals multiple histone-binding states, allosteric regulation, and druggability. *J Biol Chem* **2017,** *292* (51), 20947-20959.

206. Senisterra, G.; Zhu, H. Y.; Luo, X.; Zhang, H.; Xun, G.; Lu, C.; Xiao, W.; Hajian, T.; Loppnau, P.; Chau, I.; Li, F.; Allali-Hassani, A.; Atadja, P.; Oyang, C.; Li, E.; Brown, P. J.; Arrowsmith, C. H.; Zhao, K.; Yu, Z.; Vedadi, M., Discovery of Small-Molecule Antagonists of the H3K9me3 Binding to UHRF1 Tandem Tudor Domain. *SLAS Discov* **2018,** *23* (9), 930-940.

207. Miller, T. C.; Rutherford, T. J.; Birchall, K.; Chugh, J.; Fiedler, M.; Bienz, M., Competitive binding of a benzimidazole to the histone-binding pocket of the Pygo PHD finger. *ACS Chem. Biol.* **2014,** *9* (12), 2864-74.

208. Qi, W.; Zhao, K.; Gu, J.; Huang, Y.; Wang, Y.; Zhang, H.; Zhang, M.; Zhang, J.; Yu, Z.; Li, L.; Teng, L.; Chuai, S.; Zhang, C.; Zhao, M.; Chan, H.; Chen, Z.; Fang, D.; Fei, Q.; Feng, L.; Feng, L.; Gao, Y.; Ge, H.; Ge, X.; Li, G.; Lingel, A.; Lin, Y.; Liu, Y.; Luo, F.; Shi, M.; Wang, L.; Wang, Z.; Yu, Y.; Zeng, J.; Zeng, C.; Zhang, L.; Zhang, Q.; Zhou, S.; Oyang, C.; Atadja, P.; Li, E., An allosteric PRC2 inhibitor targeting the H3K27me3 binding pocket of EED. *Nat. Chem. Biol.* **2017,** *13* (4), 381-388.

209. He, Y.; Selvaraju, S.; Curtin, M. L.; Jakob, C. G.; Zhu, H.; Comess, K. M.; Shaw, B.; The, J.; Lima-Fernandes, E.; Szewczyk, M. M.; Cheng, D.; Klinge, K. L.; Li, H. Q.; Pliushchev, M.; Algire, M. A.; Maag, D.; Guo, J.; Dietrich, J.; Panchal, S. C.; Petros, A. M.; Sweis, R. F.; Torrent, M.; Bigelow, L. J.; Senisterra, G.; Li, F.; Kennedy, S.; Wu, Q.; Osterling, D. J.; Lindley, D. J.; Gao, W.; Galasinski, S.; Barsyte-Lovejoy, D.; Vedadi, M.; Buchanan, F. G.; Arrowsmith, C. H.; Chiang, G. G.; Sun, C.; Pappano, W. N., The EED protein-protein interaction inhibitor A-395 inactivates the PRC2 complex. *Nat. Chem. Biol.* **2017,** *13* (4), 389-395.

210. Barnash, K. D.; The, J.; Norris-Drouin, J. L.; Cholensky, S. H.; Worley, B. M.; Li, F.; Stuckey, J. I.; Brown, P. J.; Vedadi, M.; Arrowsmith, C. H.; Frye, S. V.; James, L. I., Discovery of Peptidomimetic Ligands of EED as Allosteric Inhibitors of PRC2. *ACS Comb. Sci.* **2017,** *19* (3), 161-172.

211. Liu, H.; Li, Z.; Li, L., The molecular selectivity of UNC3866 inhibitor for Polycomb CBX7 protein from molecular dynamics simulation. *Comput. Biol. Chem.* **2018,** *74*, 339-346.

212. Arrowsmith, C. H.; Audia, J. E.; Austin, C.; Baell, J.; Bennett, J.; Blagg, J.; Bountra, C.; Brennan, P. E.; Brown, P. J.; Bunnage, M. E.; Buser-Doepner, C.; Campbell, R. M.; Carter, A. J.; Cohen, P.; Copeland, R. A.; Cravatt, B.; Dahlin, J. L.; Dhanak, D.; Edwards, A. M.; Frederiksen, M.; Frye, S. V.; Gray, N.; Grimshaw, C. E.; Hepworth, D.; Howe, T.; Huber, K. V. M.; Jin, J.; Knapp, S.; Kotz, J. D.; Kruger, R. G.; Lowe, D.; Mader, M. M.; Marsden, B.; Mueller-Fahrnow, A.; Muller, S.; O'Hagan, R. C.; Overington, J. P.; Owen, D. R.; Rosenberg, S. H.; Ross, R.; Roth, B.; Schapira, M.; Schreiber, S. L.; Shoichet, B.; Sundstrom, M.; Superti-Furga, G.; Taunton, J.; Toledo-Sherman, L.; Walpole, C.; Walters, M. A.; Willson, T. M.; Workman, P.; Young, R. N.; Zuercher, W. J., The promise and peril of chemical probes. *Nat. Chem. Biol.* **2015,** *11* (8), 536-541.

213. Ma, R. G.; Zhang, Y.; Sun, T. T.; Cheng, B., Epigenetic regulation by polycomb group complexes: focus on roles of CBX proteins. *Journal of Zhejiang University. Science. B* **2014,** *15* (5), 412-28.

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

215. Herold, J. M.; Wigle, T. J.; Norris, J. L.; Lam, R.; Korboukh, V. K.; Gao, C.; Ingerman, L. A.; Kireev, D. B.; Senisterra, G.; Vedadi, M.; Tripathy, A.; Brown, P. J.; Arrowsmith, C. H.; Jin, J.; Janzen, W. P.; Frye, S. V., Small-molecule ligands of methyl-lysine binding proteins. *J. Med. Chem.* **2011,** *54* (7), 2504-11.

216. Filippakopoulos, P.; Qi, J.; Picaud, S.; Shen, Y.; Smith, W. B.; Fedorov, O.; Morse, E. M.; Keates, T.; Hickman, T. T.; Felletar, I.; Philpott, M.; Munro, S.;

McKeown, M. R.; Wang, Y.; Christie, A. L.; West, N.; Cameron, M. J.; Schwartz, B.; Heightman, T. D.; La Thangue, N.; French, C. A.; Wiest, O.; Kung, A. L.; Knapp, S.; Bradner, J. E., Selective inhibition of BET bromodomains. *Nature* **2010,** *468* (7327), 1067-1073.
217. Miyoshi S, O. S., Iwata K, Hikawa H, Sugaraha K., Antitumor Agent. . *Mitsubishi Tanabe Pharma Corporation* **2009**, 1-37.
218. Filippakopoulos, P.; Knapp, S., Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev. Drug. Discov.* **2014,** *13*, 337.
219. Shortt, J.; Ott, C. J.; Johnstone, R. W.; Bradner, J. E., A chemical probe toolbox for dissecting the cancer epigenome. *Nature Reviews Cancer* **2017,** *17*, 160.
220. Theodoulou, N. H.; Tomkinson, N. C. O.; Prinjha, R. K.; Humphreys, P. G., Clinical progress and pharmacology of small molecule bromodomain inhibitors. *Curr. Opin. Chem. Biol.* **2016,** *33*, 58-66.
221. Wójcik, P.; Berlicki, Ł., Peptide-based inhibitors of protein–protein interactions. *Bioorg. Med. Chem. Lett.* **2016,** *26* (3), 707-713.
222. London, N.; Raveh, B.; Schueler-Furman, O., Druggable protein–protein interactions – from hot spots to hot segments. *Curr. Opin. Chem. Biol.* **2013,** *17* (6), 952-959.
223. Nevola, L.; Giralt, E., Modulating protein–protein interactions: the potential of peptides. *Chem. Commun.* **2015,** *51* (16), 3302-3315.
224. Lau, J. L.; Dunn, M. K., Therapeutic peptides: Historical perspectives, current development trends, and future directions. *Biorg. Med. Chem.* **2018,** *26* (10), 2700-2707.
225. Di, L., Strategic approaches to optimizing peptide ADME properties. *The AAPS journal* **2014,** *17* (1), 134-143.
226. Lea, W. A.; Simeonov, A., Fluorescence polarization assays in small molecule screening. *Expert Opin. Drug Discov.* **2011,** *6* (1), 17-32.
227. Hu, S.; Xie, Z.; Qian, J.; Blackshaw, S.; Zhu, H., Functional protein microarray technology. *Wiley Interdiscip Rev Syst Biol Med* **2011,** *3* (3), 255-268.
228. Senthilkumar, R.; Mishra, R. K., Novel motifs distinguish multiple homologues of Polycomb in vertebrates: expansion and diversification of the epigenetic toolkit. *BMC genomics* **2009,** *10*, 549.
229. Tardat, M.; Albert, M.; Kunzmann, R.; Liu, Z.; Kaustov, L.; Thierry, R.; Duan, S.; Brykczynska, U.; Arrowsmith, C. H.; Peters, A. H., Cbx2 Targets PRC1 to Constitutive Heterochromatin in Mouse Zygotes in a Parent-of-Origin-Dependent Manner. *Mol. Cell* **2015**.
230. Ning, G.; Huang, Y. L.; Zhen, L. M.; Xu, W. X.; Jiao, Q.; Yang, F. J.; Wu, L. N.; Zheng, Y. Y.; Song, J.; Wang, Y. S.; Xie, C.; Peng, L., Transcriptional expressions of Chromobox 1/2/3/6/8 as independent indicators for survivals in hepatocellular carcinoma patients. *Aging* **2018,** *10* (11), 3450-3473.
231. Hu, Q.; Wu, W.; Zeng, A.; Yu, T.; Shen, F.; Nie, E.; Wang, Y.; Liu, N.; Zhang, J.; You, Y., Polycomb group expression signatures in the malignant progression of gliomas. *Oncology letters* **2017,** *13* (4), 2583-2590.
232. El-Serag, H. B.; Rudolph, K. L., Hepatocellular carcinoma: epidemiology and molecular carcinogenesis. *Gastroenterology* **2007,** *132* (7), 2557-76.
233. Torre, L. A.; Bray, F.; Siegel, R. L.; Ferlay, J.; Lortet-Tieulent, J.; Jemal, A., Global cancer statistics, 2012. *CA: a cancer journal for clinicians* **2015,** *65* (2), 87-108.

234. Hard, R.; Li, N.; He, W.; Ross, B.; Mo, G. C. H.; Peng, Q.; Stein, R. S. L.; Komives, E.; Wang, Y.; Zhang, J.; Wang, W., Deciphering and engineering chromodomain-methyllysine peptide recognition. *Science advances* **2018,** *4* (11), eaau1447.
235. Bernstein, E.; Duncan, E. M.; Masui, O.; Gil, J.; Heard, E.; Allis, C. D., Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol. Cell Biol.* **2006,** *26* (7), 2560-9.
236. Simhadri, C.; Daze, K. D.; Douglas, S. F.; Quon, T. T. H.; Dev, A.; Gignac, M. C.; Peng, F. N.; Heller, M.; Boulanger, M. J.; Wulff, J. E.; Hof, F., Chromodomain Antagonists That Target the Polycomb-Group Methyllysine Reader Protein Chromobox Homo log 7 (CBX7). *J. Med. Chem.* **2014,** *57* (7), 2874-2883.
237. Huang, X., Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J. Biomol. Screen.* **2003,** *8* (1), 34-8.
238. Denton, K. E.; Wang, S.; Gignac, M. C.; Milosevich, N.; Hof, F.; Dykhuizen, E. C.; Krusemark, C. J., Robustness of In Vitro Selection Assays of DNA-Encoded Peptidomimetic Ligands to CBX7 and CBX8. *SLAS Discov.* **2018,** *23* (5), 417-428.
239. Nguyen, H. H.; Park, J.; Kang, S.; Kim, M., Surface plasmon resonance: a versatile technique for biosensor applications. *Sensors (Basel, Switzerland)* **2015,** *15* (5), 10481-510.
240. Tardat, M.; Albert, M.; Kunzmann, R.; Liu, Z.; Kaustov, L.; Thierry, R.; Duan, S.; Brykczynska, U.; Arrowsmith, C. H.; Peters, A. H., Cbx2 targets PRC1 to constitutive heterochromatin in mouse zygotes in a parent-of-origin-dependent manner. *Mol. Cell.* **2015,** *58* (1), 157-71.
241. Martin, Y. C., A Bioavailability Score. *J. Med. Chem.* **2005,** *48* (9), 3164-3170.
242. Gleeson, M. P., Generation of a Set of Simple, Interpretable ADMET Rules of Thumb. *J. Med. Chem.* **2008,** *51* (4), 817-834.
243. Zhen, C. Y.; Tatavosian, R.; Huynh, T. N.; Duc, H. N.; Das, R.; Kokotovic, M.; Grimm, J. B.; Lavis, L. D.; Lee, J.; Mejia, F. J.; Li, Y.; Yao, T.; Ren, X., Live-cell single-molecule tracking reveals co-recognition of H3K27me3 and DNA targets polycomb Cbx7-PRC1 to chromatin. *eLife* **2016,** *5*.
244. Connelly, K. E.; Weaver, T. M.; Alpsoy, A.; Gu, B. X.; Musselman, C. A.; Dykhuizen, E. C., Engagement of DNA and H3K27me3 by the CBX8 chromodomain drives chromatin association. *Nucleic Acids Res.* **2018,** *47* (5), 2289-2305.
245. Creppe, C.; Palau, A.; Malinverni, R.; Valero, V.; Buschbeck, M., A Cbx8-Containing Polycomb Complex Facilitates the Transition to Gene Activation during ES Cell Differentiation. *PLOS Genetics* **2014,** *10* (12), e1004851.
246. Zhang, C. Z.; Chen, S.-L.; Wang, C.-H.; He, Y.-F.; Yang, X.; Xie, D.; Yun, J.-P., CBX8 exhibits oncogenic activity via AKT/β-Catenin activation in hepatocellular carcinoma. *Cancer Res.* **2017**, canres.0700.2017.
247. Tang, B.; Tian, Y.; Liao, Y.; Li, Z.; Yu, S.; Su, H.; Zhong, F.; Yuan, G.; Wang, Y.; Yu, H.; Tomlinson, S.; Qiu, X.; He, S., CBX8 exhibits oncogenic properties and serves as a prognostic factor in hepatocellular carcinoma. *Cell Death & Disease* **2019,** *10* (2), 52.

248. Tan, J.; Jones, M.; Koseki, H.; Nakayama, M.; Muntean, A. G.; Maillard, I.; Hess, J. L., CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. *Cancer Cell* **2011,** *20* (5), 563-75.
249. Chung, C.-Y.; Sun, Z.; Mullokandov, G.; Bosch, A.; Qadeer, Z. A.; Cihan, E.; Rapp, Z.; Parsons, R.; Aguirre-Ghiso, J. A.; Farias, E. F.; Brown, B. D.; Gaspar-Maia, A.; Bernstein, E., Cbx8 Acts Non-canonically with Wdr5 to Promote Mammary Tumorigenesis. *Cell Rep.* **2016,** *16* (2), 472-486.
250. Xiao, W.; Ou, C.; Qin, J.; Xing, F.; Sun, Y.; Li, Z.; Qiu, J., CBX8, a novel DNA repair protein, promotes tumorigenesis in human esophageal carcinoma. *Int. J. Clin. Exp. Pathol.* **2014,** *7* (8), 4817-4826.
251. Yuan, G. J.; Chen, X.; Lu, J.; Feng, Z. H.; Chen, S. L.; Chen, R. X.; Wei, W. S.; Zhou, F. J.; Xie, D., Chromobox homolog 8 is a predictor of muscle invasive bladder cancer and promotes cell proliferation by repressing the p53 pathway. *Cancer science* **2017,** *108* (11), 2166-2175.
252. Mello, J. B.; Ramos Cirilo, P. D.; Michelin, O. C.; Custodio Domingues, M. A.; Cunha Rudge, M. V.; Rogatto, S. R.; Maesta, I., Genomic profile in gestational and non-gestational choriocarcinomas. *Placenta* **2017,** *50*, 8-15.
253. Li, G.; Warden, C.; Zou, Z.; Neman, J.; Krueger, J. S.; Jain, A.; Jandial, R.; Chen, M., Altered expression of polycomb group genes in glioblastoma multiforme. *PloS one* **2013,** *8* (11), e80970.
254. Geller, J. I.; Roth, J. J.; Biegel, J. A., Biology and Treatment of Rhabdoid Tumor. *Crit Rev Oncog* **2015,** *20* (3-4), 199-216.
255. Gadd, S.; Sredni, S. T.; Huang, C. C.; Perlman, E. J., Rhabdoid tumor: gene expression clues to pathogenesis and potential therapeutic targets. *Lab. Invest.* **2010,** *90* (5), 724-38.
256. Liu, R.; Li, X.; Lam, K. S., Combinatorial chemistry in drug discovery. *Curr. Opin. Chem. Biol.* **2017,** *38*, 117-126.
257. Hamzeh-Mivehroud, M.; Alizadeh, A. A.; Morris, M. B.; Church, W. B.; Dastmalchi, S., Phage display as a technology delivering on the promise of peptide drug discovery. *Drug Discov Today* **2013,** *18* (23-24), 1144-57.
258. Pande, J.; Szewczyk, M. M.; Grover, A. K., Phage display: Concept, innovations, applications and future. *Biotechnol. Adv.* **2010,** *28* (6), 849-858.
259. Rahbarnia, L.; Farajnia, S.; Babaei, H.; Majidi, J.; Veisi, K.; Ahmadzadeh, V.; Akbari, B., Evolution of phage display technology: from discovery to application. *J. Drug Target.* **2017,** *25* (3), 216-224.
260. Krumpe, L. R.; Mori, T., Potential of phage-displayed peptide library technology to identify functional targeting peptides. *Expert opinion on drug discovery* **2007,** *2* (4), 525-525.
261. Wu, C.-H.; Liu, I.-J.; Lu, R.-M.; Wu, H.-C., Advancement and applications of peptide phage display technology in biomedical science. *J Biomed Sci* **2016,** *23* (1), 8.
262. Brissette, R.; Prendergast, J. K.; Goldstein, N. I., Identification of cancer targets and therapeutics using phage display. *Curr Opin Drug Discov Devel* **2006,** *9* (3), 363-9.
263. Nixon, A. E.; Sexton, D. J.; Ladner, R. C., Drugs derived from phage display: from candidate identification to clinical practice. *mAbs* **2014,** *6* (1), 73-85.
264. Heinis, C.; Winter, G., Encoded libraries of chemically modified peptides. *Curr. Opin. Chem. Biol.* **2015,** *26*, 89-98.

265. Ng, S.; Tjhung, K. F.; Paschal, B. M.; Noren, C. J.; Derda, R., Chemical posttranslational modification of phage-displayed peptides. *Methods Mol Biol* **2015,** *1248*, 155-72.

266. Mohan, K.; Weiss, G. A., Chemically Modifying Viruses for Diverse Applications. *ACS chemical biology* **2016,** *11* (5), 1167-1179.

267. Ng, S.; Jafari, M. R.; Derda, R., Bacteriophages and Viruses as a Support for Organic Synthesis and Combinatorial Chemistry. *ACS Chemical Biology* **2012,** *7* (1), 123-138.

268. Santoso, B.; Lam, S.; Murray, B. W.; Chen, G., A simple and efficient maleimide-based approach for peptide extension with a cysteine-containing peptide phage library. *Bioorg. Med. Chem. Lett.* **2013,** *23* (20), 5680-3.

269. Khan, A. R.; Williams, K. A.; Boggs, J. M.; Deber, C. M., Accessibility and dynamics of Cys residues in Bacteriophage IKe and M13 major coat protein mutants. *Biochemistry* **1995,** *34* (38), 12388-97.

270. Tokunaga, Y.; Azetsu, Y.; Fukunaga, K.; Hatanaka, T.; Ito, Y.; Taki, M., Pharmacophore generation from a drug-like core molecule surrounded by a library peptide via the 10BASEd-T on bacteriophage T7. *Molecules* **2014,** *19* (2), 2481-96.

271. Jafari, M. R.; Deng, L.; Kitov, P. I.; Ng, S.; Matochko, W. L.; Tjhung, K. F.; Zeberoff, A.; Elias, A.; Klassen, J. S.; Derda, R., Discovery of Light-Responsive Ligands through Screening of a Light-Responsive Genetically Encoded Library. *ACS Chemical Biology* **2014,** *9* (2), 443-450.

272. Bellotto, S.; Chen, S.; Rentero Rebollo, I.; Wegner, H. A.; Heinis, C., Phage Selection of Photoswitchable Peptide Ligands. *J. Am. Chem. Soc.* **2014,** *136* (16), 5880-5883.

273. Heinis, C.; Rutherford, T.; Freund, S.; Winter, G., Phage-encoded combinatorial chemical libraries based on bicyclic peptides. *Nat. Chem. Biol.* **2009,** *5* (7), 502-7.

274. Diderich, P.; Bertoldo, D.; Dessen, P.; Khan, M. M.; Pizzitola, I.; Held, W.; Huelsken, J.; Heinis, C., Phage Selection of Chemically Stabilized α-Helical Peptide Ligands. *ACS Chemical Biology* **2016,** *11* (5), 1422-1427.

275. Kalhor-Monfared, S.; Jafari, M. R.; Patterson, J. T.; Kitov, P. I.; Dwyer, J. J.; Nuss, J. M.; Derda, R., Rapid biocompatible macrocyclization of peptides with decafluoro-diphenylsulfone. *Chemical Science* **2016**.

276. Simon, M. D.; Chu, F.; Racki, L. R.; de la Cruz, C. C.; Burlingame, A. L.; Panning, B.; Narlikar, G. J.; Shokat, K. M., The site-specific installation of methyl-lysine analogs into recombinant histones. *Cell* **2007,** *128* (5), 1003-12.

277. Pieters, B.; Hintzen, J. C. J.; Grobben, Y.; Al Temimi, A. H. K.; Kamps, J.; Mecinovic, J., Installation of Trimethyllysine Analogs on Intact Histones via Cysteine Alkylation. *Bioconjug Chem* **2019**.

278. Dermer, O. C.; Ham, G. E., Ethylenimine and other aziridines : chemistry and applications. **1969**.

279. Heymann, H., Methoden der Organischen Chemie (Houben-Weyl). *J. Am. Chem. Soc.* **1958,** *80* (4), 1012-1012.

280. Crist, D. R.; Leonard, N. J., Small Charged Heterocycles. *Angew Chem Int Ed Eng* **1969,** *8* (12), 962-974.

281. Leonard, N. J.; Paukstelis, J. V., SMALL CHARGED RINGS. VII. INTERCONVERSION OF SUBSTITUTED BETA-CHLOROETHYLAMINES AND AZIRIDINIUM SALTS. *J. Org. Chem.* **1965,** *30*, 821-5.
282. Kather, I.; Bippes, C. A.; Schmid, F. X., A stable disulfide-free gene-3-protein of phage fd generated by in vitro evolution. *J. Mol. Biol.* **2005,** *354* (3), 666-78.
283. Jafari, M. R.; Lakusta, J.; Lundgren, R. J.; Derda, R., Allene Functionalized Azobenzene Linker Enables Rapid and Light-Responsive Peptide Macrocyclization. *Bioconjug Chem* **2016,** *27* (3), 509-14.
284. Rodi, D. J.; Soares, A. S.; Makowski, L., Quantitative Assessment of Peptide Sequence Diversity in M13 Combinatorial Peptide Phage Display Libraries. *J. Mol. Biol.* **2002,** *322* (5), 1039-1052.
285. Sartori, M. F., New Developments in the Chemistry of War Gases. *Chem. Rev.* **1951,** *48* (2), 225-257.
286. Povirk, L. F.; Shuker, D. E., DNA damage and mutagenesis induced by nitrogen mustards. *Mut Res Rev Genet Toxicol* **1994,** *318* (3), 205-226.
287. Lawley, P. D.; Lethbridge, J. H.; Edwards, P. A.; Shooter, K. V., Inactivation of bacteriophage T7 by mono- and difunctional sulphur mustards in relation to cross-linking and depurination of bacteriophage DNA. *J. Mol. Biol.* **1969,** *39* (1), 181-98.
288. Yamamoto, N.; Naito, T.; Shimkin, M. B., Mechanism of Inactivation of DNA and RNA Bacteriophages by Alkylating Agents in Vitro. *Cancer Res.* **1966,** *26* (11 Part 1), 2301-2306.
289. Seeliger, D.; Soeroes, S.; Klingberg, R.; Schwarzer, D.; Grubmuller, H.; Fischle, W., Quantitative assessment of protein interaction with methyl-lysine analogues by hybrid computational and experimental approaches. *ACS Chem Biol* **2012,** *7* (1), 150-4.
290. National Center for Biotechnology Information. PubChem Identifier. AID = 488946. https://pubchem.ncbi.nlm.nih.gov/assay/488946.
291. National Center for Biotechnology Information. PubChem Identifier. AID = 540317. https://pubchem.ncbi.nlm.nih.gov/bioassay/540317.
292. Milosevich, N.; Warmerdam, Z.; Hof, F., Structural aspects of small-molecule inhibition of methyllysine reader proteins. *Future Med Chem* **2016,** *8* (13), 1681-702.
293. Yang, S.; Liu, W.; Li, M.; Wen, J.; Zhu, M.; Xu, S., Insulin-Like Growth Factor-1 Modulates Polycomb Cbx8 Expression and Inhibits Colon Cancer Cell Apoptosis. *Cell Biochem Biophys* **2015,** *71* (3), 1503-1507.
294. Tang, J.; Wang, G.; Zhang, M.; Li, F.-y.; Sang, Y.; Wang, B.; Hu, K.; Wu, Y.; Luo, R.; Liao, D.; Cao, J.; Wang, X.; Wang, L.; Zhang, R.; Zhang, X.; Deng, W.-g.; Xie, D.; Xu, R.-h.; Kang, T., Paradoxical role of CBX8 in proliferation and metastasis of colorectal cancer. *Oncotarget* **2014,** *5* (21), 10778-10790.
295. Song, X.; Tang, T.; Li, C.; Liu, X.; Zhou, L., CBX8 and CD96 Are Important Prognostic Biomarkers of Colorectal Cancer. *Med Sci Monit* **2018,** *24*, 7820-7827.
296. Huang, X., Fluorescence polarization competition assay: the range of resolvable inhibitor potency is limited by the affinity of the fluorescent ligand. *J Biomol Screen* **2003,** *8* (1), 34-8.
297. Wagner, E. K.; Nath, N.; Flemming, R.; Feltenberger, J. B.; Denu, J. M., Identification and characterization of small molecule inhibitors of a plant homeodomain finger. *Biochemistry* **2012,** *51* (41), 8293-306.

298. Pena, P. V.; Hom, R. A.; Hung, T.; Lin, H.; Kuo, A. J.; Wong, R. P.; Subach, O. M.; Champagne, K. S.; Zhao, R.; Verkhusha, V. V.; Li, G.; Gozani, O.; Kutateladze, T. G., Histone H3K4me3 binding is required for the DNA repair and apoptotic activities of ING1 tumor suppressor. *J. Mol. Biol.* **2008,** *380* (2), 303-12.