

DESIGN AND CHARACTERIZATION OF CHEMICAL PROBES TARGETING THE CHROMODOMAINS  
OF POLYCOMB REPRESSIVE COMPLEX I

Jacob I. Stuckey

A dissertation submitted to the faculty at the University of North Carolina at Chapel Hill in partial fulfillment  
of the requirements for the degree of Doctor of Philosophy in the Pharmaceutical Sciences (Chemical  
Biology & Medicinal Chemistry) program in the Eshelman School of Pharmacy.

Chapel Hill  
2016

Approved by:

Stephen V. Frye

David S. Lawrence

Albert A. Bowers

Marcey L. Waters

Brian D. Strahl

© 2016  
Jacob I. Stuckey  
ALL RIGHTS RESERVED

## **ABSTRACT**

Jacob I. Stuckey: Design and characterization of chemical probes targeting the chromodomains of Polycomb Repressive Complex I  
(Under the direction of Stephen V. Frye)

Eukaryotic DNA is wound around histone proteins to form repeating units of nucleosomes that are packaged into higher order structures to form chromosomes. Nucleosomal packaging is dynamically regulated to control the expression of the genome through regulating access of transcription factors to specific genes in response to stimuli. Critical for these processes are modifications to the N-terminal tails of histones. In particular, methylation of specific histone lysine residues has a profound impact on transcription. This “mark” recruits protein complexes that contain methyl-lysine “reader” modules that bind to the mark and facilitate transcriptional silencing or activation of the underlying gene. Mutation and misregulation of proteins that “write,” “read” and “erase” these marks play a foundational role in the development and progression of nearly all cancers.

Chemical tools that antagonize these writers, readers and erasers have been developed in recent years. These tools have been valuable in explicating the biology of these proteins, and in some cases, as therapeutic agents. Development of antagonists of methyl-lysine readers has lagged compared to both methyl-lysine writers and erasers. Herein we describe the design, development and characterization of antagonists of the chromodomain reader modules contained in Polycomb Repressive Complex I (PRC1). PRC1 contains one of five chromodomains; CBX2, -4, -6, -7 or -8. PRC1 complexes containing any one of these proteins play critical roles in normal, growth and development. Accordingly, misregulation of PRC1 chromodomains contributes to the initiation and development of numerous cancers. We developed a series of potent, peptidic antagonists of PRC1 chromodomains and an expanded set of chemical tools around one of these molecules, UNC3866, which has the highest affinity for the chromodomains of CBX4 and -7. We demonstrated the ability of UNC3866 to inhibit proliferation in a metastatic prostate cancer cells through the induction of a senescent-like state. Further optimization of UNC3866 led to a second-generation antagonist of PRC1 chromodomains with a dramatically improved residence time for CBX7

and increased cellular potency. This work has produced a valuable tool set for studying CBX biology and laid the ground work for continued efforts to evaluate and improve CBX antagonists as potential therapeutic agents.



## **DEDICATION**

To the glory of God in Christ Jesus

## **ACKNOWLEDGMENTS**

Nine years ago, I delivered a speech to my high school graduating class in which I quoted the wisdom of a respected teacher and coach who insisted that “when we crossed that stage, life was going to punch us in the face.” I continued by imploring my classmates that the difference between success and failure would be determined by those who chose to get back up when life knocks us down. What I didn’t realize when I delivered that speech so long ago, was just how relevant it was for the journey on which I was about to embark. Nothing about my life even remotely resembles what I pictured that day. Looking back at the course of my life, I am grateful to even be alive, let alone where I am today. One of my most beloved childhood heroes, Sylvester Stallone’s Rocky Balboa, put it this way, “You, me, or nobody is gonna hit as hard as life. But it ain’t about how hard you hit. It’s about how hard you can get hit and keep moving forward; how much you can take and keep moving forward.” I cannot adequately express my gratitude to my family for the many, many times that they have stood beside me and kept me moving forward through everything that life has thrown at me.

I want to thank my beautiful wife, Jennifer, who has loved, supported and pushed forward me forward thru even the worst of times. To my parents, Charles and Lari Stuckey, who have believed in me and supported me throughout my life, even thru times when I didn’t believe in myself. To my father, who constantly challenged me to think for myself and to “be my own man.” To my mother, who is perhaps the most disciplined, hard-working and strongest person I’ve met. I have learned so much from you both. I am the person I am today because of you. To my brother and his wife, Joshua and Linsey Stuckey, for their love and support thru the years. Their two beautiful daughters, Claire and Cora Stuckey, who have brought me so many much-needed smiles. Despite anyone else may say, are the two most adorable little girls in the world. To Paul, Susan and Jonathan Fleetwood, who have welcomed me into their family with open arms. To my grandparents, Charles and Wilma Stuckey, and Leon and Lola Calvin. Nothing saddens me more than to not be able to share the joy of what I have accomplished with you. You are loved and missed and your memory is source of inspiration for me on a daily basis. Last but not least, to

my dogs Lizzie, Carbon and Harley, all of whom are a constant reminder that even on my worst days, as long as I have my family, life is still pretty good.

Numerous others, in addition to my family have been instrumental in seeing me thru to where I am now. To my undergraduate adviser, Dr. Joshua Ring, whose passion and enthusiasm for organic chemistry and biochemistry was instrumental in my decision to change my major to Chemistry at the start of my final year as an undergraduate and to pursue the graduate studies in Pharmaceutical Sciences. To the faculty in the Division of Chemical Biology and Medicinal Chemistry at UNC for the immense amount of time, money and resources you invested in me in order to develop and train me as an independent scientist. To my adviser and mentor, Dr. Stephen V. Frye, who has been an absolute joy to work with. I have learned so much from you both in what you have explicitly taught me and from the example you have set as a person, leader and scientist (in that order). To Dr. Lindsey James, who has been tremendous friend and teacher throughout my graduate studies. You are a truly rare person who is both a great scientist and a great person. I owe a great deal of my success as a graduate student to you and know that you will continue to be an outstanding mentor to many graduate students to come. Maybe most of all, thank you for seeming to be genuinely interested when I felt the need to brag about my nieces and show off pictures of just how cute they are.

To my lab mates in the Center, for their guidance, patience and hard work throughout the years. Kim Barnash, who, I'm certain, made it her sole purpose on more than a few days, to keep me humble. It has been a genuine pleasure working with and learning from you. To Kyle Konze, for your friendship and support as we started our graduate careers together and somehow made it through. I want to thank Jarod Waybright for being a great person to bounce ideas off of and an even better person to grab lunch with and forget about science altogether for a little while. To Dr. Samantha Pattenden, Jacqueline Norris Drouin, Dr. Brandi Baughman and Stephanie Cholensky for all of your hard work, support and teaching that expanded my scientific expertise far beyond what I ever expected. To Dr. Brad Dickson, whose own work was critical for my success as a graduate student and who was instrumental in my transition to an independent scientist and thinker. Finally, I want to thank my committee for their time, attention and input through the years, Dr. David Lawrence, Dr. Albert Bowers, Dr. Brian Strahl and Dr. Marcey Waters.

## TABLE OF CONTENTS

<b>LIST OF TABLES .....</b>	<b>x</b>
<b>LIST OF FIGURES .....</b>	<b>xii</b>
<b>LIST OF ABBREVIATIONS.....</b>	<b>xv</b>
<b>CHAPTER I: INTRODUCTION .....</b>	<b>1</b>
Chromatin regulation.....	1
Therapeutic interest in chromatin regulatory proteins.....	4
The tractability of chemical probe discovery for readers of histone lysine methylation .....	6
Screening for antagonists of methyl-lysine readers .....	9
Examples of previous challenges and successes of small-molecule chemical probe discovery efforts targeting methyl-lysine readers .....	11
The CBX family of chromodomains .....	14
The canonical Polycomb signaling pathway .....	16
The biological function of Pc CBX proteins.....	19
Polycomb chromodomains in cancer .....	20
Summary .....	22
<b>CHAPTER II: DEVELOPMENT OF HIGH AFFINITY LIGANDS FOR THE CBX7 CHROMODOMAIN .....</b>	<b>24</b>
Introduction .....	24
Results and Discussion .....	27
Conclusions.....	48
Methods .....	49
<b>CHAPTER III: BIOCHEMICAL AND BIOPHYSICAL CHARACTERIZATION OF CBX ANTAGONISTS .....</b>	<b>105</b>
Introduction .....	105
Results and Discussion .....	107

Conclusions.....	138
Methods .....	139
<b>CHAPTER IV: CELLULAR CHARACTERIZATION OF UNC3866 .....</b>	<b>159</b>
Introduction .....	159
Results and Discussion.....	160
Conclusions.....	178
Methods .....	179
<b>CHAPTER V: FUTURE DIRECTIONS .....</b>	<b>185</b>
Summary of Key Findings .....	185
Future Directions.....	187
Conclusions.....	194
<b>REFERENCES.....</b>	<b>195</b>

## LIST OF TABLES

Table 1.1. Examples of Kme readers of potential therapeutic relevance .....	5
Table 2.1. Exploration of sequence and length requirements for interaction with the CBX7 chromodomain .....	31
Table 2.2. Effect of select backbone methylations of UNC2868 on binding to CBX7 .....	34
Table 2.3. Exploration of Kme3 replacements for UNC2868 .....	36
Table 2.4. Initial N-terminal modifications of a norbornyl-lysine containing CBX7 antagonist.....	43
Table 2.5. N-terminal modifications of a diethyl-lysine containing CBX7 antagonist.....	44
Table 2.6. CBX7 peptide substrate used in AlphaScreen® .....	51
Table 2.7. Compound structures and characterization by LCMS .....	68
Table 3.1. Characterization of UNC3866 selectivity by AlphaScreen® or ITC against a panel of Kme reader domains .....	107
Table 3.2. Activity of UNC3866 against a panel of bromodomains as measured by $\Delta T_m$ .....	110
Table 3.3. Activity of UNC3866 against a set of demethylase enzymes as determined by AlphaScreen® .....	111
Table 3.4. Cerep selectivity profile of UNC3866 .....	112
Table 3.5. Determination of peptide length requirements for chromodomain antagonism by ALKme3S peptides .....	118
Table 3.6. Exploration of secondary and tertiary amine Kme3 replacements .....	120
Table 3.7. Effect of N-terminal variations and select lysine modifications .....	125
Table 3.8. Exploration of modifications to the phenylalanine, alanine, and C-terminus of UNC3866.....	128
Table 3.9. Kme reader AlphaScreen® proteins and their corresponding peptide substrates .....	140
Table 3.10. Biotinylated peptide substrates .....	143
Table 3.11. Anti-methyl mark antibodies.....	143
Table 3.12. Demethylase AlphaScreen® assay specifics .....	143
Table 3.13. Protein-bait ligand concentrations used for AlphaScreen® assays .....	144
Table 3.14. Data collections and refinement statistics (molecular replacement).....	148
Table 3.15. Compound structure and characterization by LCMS .....	150

Table 4.1. Mean pharmacokinetic parameters of UNC3866 in blood following a single intravenous and intraperitoneal administration to male Swiss Albino mice.....	175
Table 4.2. Primary antibodies used for western blot detection .....	181

## LIST OF FIGURES

Figure 1.1. Model of a nucleosome.....	2
Figure 1.2. Overview of the histone code hypothesis .....	3
Figure 1.3. Kme reader domain families .....	7
Figure 1.4. Comparison of surface-groove and cavity-insertion binders .....	8
Figure 1.5. Antagonists of Kme readers .....	12
Figure 1.6. Comparison of Pc and HP1 CBX chromodomains .....	15
Figure 1.7. Domain maps of Pc CBX proteins .....	16
Figure 1.8. Composition of canonical PRC1 .....	17
Figure 1.9. Canonical Polycomb Signaling pathway.....	18
Figure 2.1. Comparison of the apo, H3K9me3-bound and H3K27me3 bound structure of CBX7 .....	25
Figure 2.2. Structures of commercially available antagonists for quaternary ammonium binding proteins tested against CBX7 .....	28
Figure 2.3. Structures of top compounds from docking studies with CBX7 and the CICBDD diversity library .....	29
Figure 2.4. ITC studies with CBX7 and UNC2867 and UNC2868 .....	33
Figure 2.5. Hydrogen bonding between CBX7 and H3K9me3 .....	34
Figure 2.6. ITC studies between CBX7 and peptides with various Kme3 replacements .....	39
Figure 2.7. Comparison of CBX7 in the bound and apo conformations .....	40
Figure 2.8. Molecular dynamics simulation of the CBX7-H3K9me3 interaction .....	41
Figure 2.9 Hydrophobic contacts made by H3Q5 during association of CBX7 and H3K9me3 peptides.....	42
Figure 2.10. X-ray co-crystal structure of UNC3866 bound to CBX7 .....	47
Figure 2.11. CBX7 chemical tool kit.....	48
Figure 3.1. Characterization of UNC3866 selectivity with UNC4195.....	108
Figure 3.2. Selectivity studies with UNC3866 and its chromodomain targets .....	109
Figure 3.3. Activity of UNC3866 against a panel of methyltransferase enzymes .....	111
Figure 3.4. NK1 receptor dose-response functional assays .....	114



Figure 3.5. NK2 receptor dose-response functional assays .....	114
Figure 3.6. Opioid receptor dose-response functional assays.....	115
Figure 3.7. Vasopressin receptor dose-response functional assays .....	116
Figure 3.8. UNC3866 bound to CBX7.....	127
Figure 3.9. Co-crystal structures of UNC3866 with CBX2, CBX4, CBX7 and CBX8.....	130
Figure 3.10. Comparison of CBX2 and CBX7 in the apo and UNC3866-bound conformations .....	131
Figure 3.11 Comparison of the UNC3866 diethyl-lysine side chain electron densities from the CBX7 or CBX8 co-crystal structures.....	132
Figure 3.12 .....	134
Figure 3.13. Example SPR sensorgrams of CBX7 with each respective compound .....	135
Figure 3.14. Induced fit binding equilibrium between CBX7 and peptidic antagonists .....	136
Figure 3.15. X-ray co-crystal structure of UNC3866 bound to CBX7 .....	137
Figure 3.16 .....	138
Figure 3.17 .....	140
Figure 4.1 .....	161
Figure 4.2. UNC3866 does not bind to EED as assessed by ITC .....	162
Figure 4.3 .....	163
Figure 4.4 .....	164
Figure 4.5 .....	165
Figure 4.6 .....	166
Figure 4.7. UNC3866 exhibits low membrane permeability as measured by the bi-directional Caco-2 permeability assay .....	167
Figure 4.8. UNC4007 interacts with PRC1 chromodomains.....	168
Figure 4.9. UNC3866 and UNC4219 display no toxicity up to 100µM in PC3 and HEK293T cells as assessed by CellTiter-Glo® .....	169
Figure 4.10 .....	170
Figure 4.11 .....	171
Figure 4.12. Analysis of mRNA transcript levels from the Ink4a/ARF locus .....	172
Figure 4.13. Effect of compounds UNC3866, UNC4938, and UNC4219 on PC3 cell proliferation .....	173

Figure 4.14. Varying effects of UNC3866 on cellular proliferation .....	174
Figure 4.15. Mean blood concentration-time profiles of UNC3866 and UNC4007 following a single intravenous administration to male Swiss Albino mice .....	176
Figure 4.16. Mean blood concentration-time profiles of UNC3866 and UNC4007 following a single intraperitoneal administration to male Swiss Albino mice .....	177
Figure 4.17. Comparison of the ability of UNC4976 to antagonize CBX7 in cells compared to UNC3866 .....	178
Figure 5.1. Structures of UNC3866, UNC4219 and UNC4195 .....	185
Figure 5.2. Relationship between various modifications to our antagonists and their $K_d$ , $k_{on}$ , and $k_{off}$ for CBX7 .....	186
Figure 5.3. Summary of selectivity parameters determined from UNC3866 SAR studies .....	187
Figure 5.4. Kme3 replacement is required for inhibition of proliferation by peptidic CBX7 antagonists .....	187
Figure 5.5. van der Waals' interactions between UNC3866 and CBX7 .....	189
Figure 5.6. Strategies for side chain optimization of UNC3866 based on the UNC3866-CBX7 co-crystal structure .....	189
Figure 5.7. Strategies for improving the <i>in vivo</i> efficacy of UNC4976 .....	190
Figure 5.8 .....	191
Figure 5.9. Overlay of the CBX2 chromodomain in the apo and UNC3866-bound conformation .....	192
Figure 5.10 Fragment-based approaches for the development a CBX7 small molecule antagonist.....	193
Figure 5.11. Putative CBX7 isoforms engaged by UNC3866 .....	194

## LIST OF ABBREVIATIONS

Ac	acetylation
BET	bromodomain and extraterminal domain
CBX	chromobox homolog
CETSA	Cellular Thermal Shift Assay
CICBDD	Center for Integrative Chemical Biology and Drug Discovery
CDY	chromodomain protein, Y-linked
CDYL	chromodomain protein, Y-linked-like
DNA	deoxyribonucleic acid
EC <sub>50</sub>	concentration required for 50% efficacy
EED	embryonic ectoderm development
EZH2	enhancer of zeste homolog 2
FDA	Food & Drug Administration
FRAP	Fluorescence Recovery after Photobleaching
GPCR	G-protein coupled receptor
hib	hydroxyisobutyryl
HP1	heterochromatin protein 1
HPH	human polyhomeotic protein
H2A	histone 2A
H2AK119ub	ubiquitinyl histone 2A lysine 119
H2B	histone 2B
H3	histone 3
H3K9me3	trimethylated histone 3 lysine 9
H3K27me3	trimethylated histone 3 lysine 27
H4	histone 4
IC <sub>50</sub>	concentration to achieve 50% inhibition
IP	intraperitoneally

ITC	isothermal titration calorimetry
IV	intravenously
JQ1	Jun Qi 1
Kac	acetyl-lysine
$K_d$	dissociation constant
Kme	methyl-lysine
Kme3	trimethylated lysine
$k_{on}$	association rate
$k_{off}$	dissociation rate
M	molar
MBT	malignant brain tumor
MD	molecular dynamics
Me	methylation
mRNA	messenger ribonucleic acid
Ni	nickel
nm	nanometer
nM	nanomolar
P	phosphorylation
Pc	Polycomb
PCGF	Polycomb group family
PDB	protein databank
PHD	plant homeodomain
PRC	Polycomb Repressive Complex
PTM	posttranslational modification
RING	really interesting new gene
RNA	ribonucleic acid
RNAi	ribonucleic acid inhibitor
RT-qPCR	reverse transcriptase quantitative polymerase chain reaction

s	seconds
SAR	structure-activity relationship
SETDB1	SET Domain Bifurcated 1
SPR	surface plasmon resonance
SUMO	sumoylated
SUZ12	suppressor of zeste 12
T <sub>m</sub>	melting temperature
Ub	ubiquitinyl
UNC	University of North Carolina
μM	micromolar

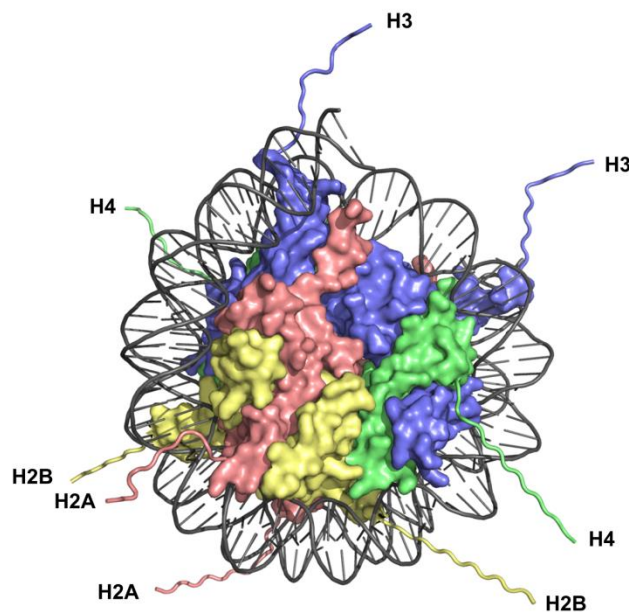
## **CHAPTER I: INTRODUCTION**

### **CHROMATIN REGULATION**

Remarkably, the entire genome of a eukaryotic organism is contained within each one its cells. This chemical information is organized and packaged into dense structures known as chromosomes. For humans, the genome is packaged into 23 pairs chromosomes that originate in a single cell following conception and eventually give rise to a person composed of more than 200 different cell types made from trillions of individual cells<sup>1</sup>. While each of these cells, with the exception of red blood cells and platelets, contains the same 23 pairs of chromosomes, each cell utilizes a mere fraction of this genetic information to accomplish its biological function. The mechanisms of genomic regulation that account for this are complex, highly evolved processes that are only beginning to be unraveled.

Initial insight into these mechanisms was gained through the study of the simpler genomes of prokaryotic organisms that lack many of the features found in eukaryotic chromosomes. These studies led to what is known as the “operon model” of genomic regulation. In this model, activator and repressor proteins, known as transcription factors, bind to specific DNA sequences in ways that are sensitive to the concentrations of metabolites and environmental nutrients. Transcription factors, metabolic sensing and positive and negative feedback loops then act together as sophisticated control networks regulating families of genes<sup>2</sup>. However, while this model partially accounts for genomic regulation in higher organisms, by itself, the operon model is insufficient for explaining the regulation of the more complicated genomes of eukaryotic organisms. This is because non-coding and coding DNA sequences have accumulated throughout evolution in higher organisms to the degree that, in humans, each individual cell contains approximately two meters of DNA that is packed within a nucleus that can range from 2-10  $\mu\text{m}$  in diameter. The spatial compaction of such a large amount of chemical matter was enabled by the simultaneous evolution of mechanisms to package and organize this vast library of chemical information. Specifically, eukaryotic DNA is stored as chromatin, which is the overall complex, of DNA, RNA and their associated proteins. Within chromatin, the nucleosome is the basic functional unit for regulating genomic

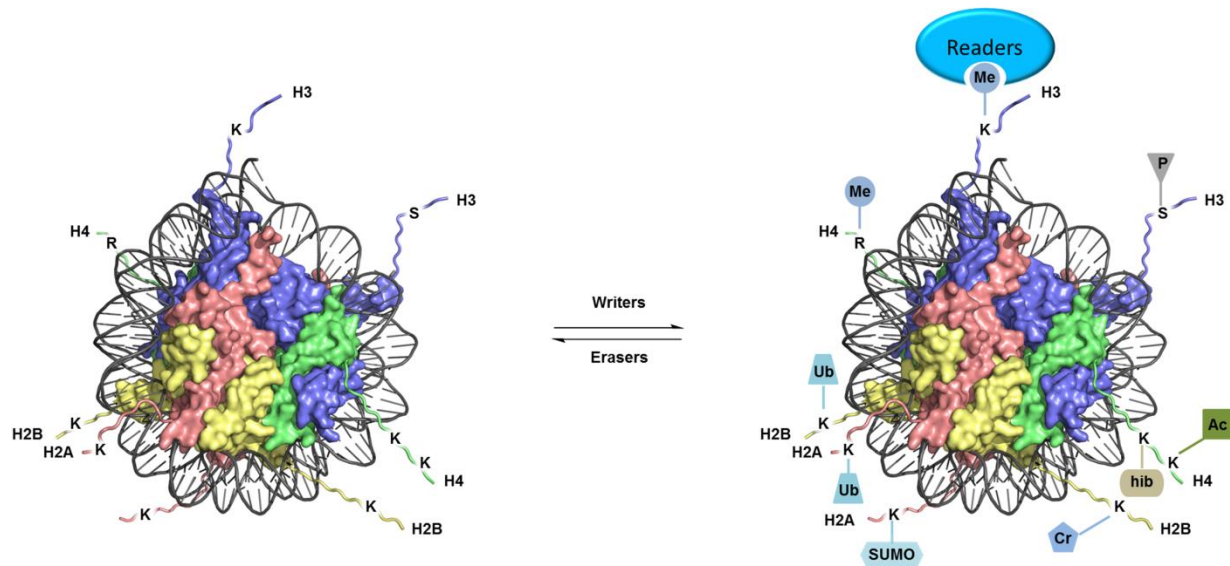
expression. A nucleosome contains 147 base pairs of DNA wound around an octamer of histone proteins, two copies each of histones H2A, H2B, H3 and H4 (**Fig. 1.1**). Nucleosomal repeats are then organized into higher order structures that makeup the overall chromatin architecture. The organization of nucleosomes into these higher-order structures controls access of transcription factors to DNA, in turn controlling the transcription of the entire genome<sup>2-4</sup>. The evolution of mechanisms to “open” and “close” chromatin enabled eukaryotic cells to dynamically regulate the exposure of specific genes to transcription factors, providing a mechanism for both differentiation into different cell types and for more precise genomic regulation in response to environmental stimuli<sup>2-4</sup>.



**Figure 1.1.** Model of a nucleosome. DNA (gray) is wrapped around an octamer of histone proteins. The histone tails of H2A (salmon), H2B (yellow), H3 (cyan), and H4 (green) are intrinsically disordered and protrude out from the nucleosome. Figure adapted from PDB 5B0Z.

Histones were once thought to contribute to genomic regulation by serving primarily as static structural elements that only passively participate in controlling the transcription of the underlying DNA sequence<sup>5</sup>. This assumption was challenged as numerous studies demonstrated that nucleosomes, especially histones, are a unique canvas for posttranslational modifications (PTMs). These PTMs occur primarily on the highly disordered N-terminal portion, or “tail”, of histones that protrude out from the

nucleosomal core and are enriched in basic amino acids<sup>2, 3, 5, 6</sup> (**Fig. 1.1**). Some of the known PTMs to histones include arginine and lysine methylation<sup>7</sup>; lysine acetylation<sup>5, 7, 8</sup>, crotonylation<sup>9</sup>, hydroxyisobutyrylation<sup>10</sup>, sumoylation<sup>11</sup>, ubiquitination<sup>12-14</sup>; and serine/threonine phosphorylation<sup>5</sup> (**Fig. 1.2**).



**Figure 1.2.** Overview of the histone code hypothesis. Posttranslational modifications installed by writers form a putative code that is recognized by readers and can be removed by erasers to regulate gene expression. There are many readers for the various histone PTMs. In the interest of simplicity, only a methyl-lysine reader is depicted. Nucleosome models were adapted from PDB 5B0Z.

Correlations established between patterns of these PTMs and transcription of the underlying gene(s) led to the proposal of the “histone code” hypothesis<sup>5</sup>, wherein individual histone PTMs are thought of as “marks” and discrete patterns of these marks are installed and then recognized within cells to bring about distinct phenotypes<sup>5</sup>. Enzymes and enzymatic complexes that install histone PTMs have come to be known as “writers” of the histone code, while enzymes that remove these PTMs have come to be known as “erasers” of the histone code. Proteins and protein complexes that recognize and bind to these PTMs are known as “readers” of the histone code and are largely responsible for affecting the downstream biology associated with a particular mark or pattern of marks<sup>6, 7</sup> (**Fig. 1.2**).



## THERAPEUTIC INTEREST IN CHROMATIN REGULATORY PROTEINS

Misregulation and mutation of the readers, writers and erasers of the histone code initiates and contributes to a variety of human disease states, including, but not limited to; neuropsychiatric diseases, inflammation, metabolic disorders, and especially cancer<sup>7, 15-26</sup>. Of the various histone marks, lysine acetylation (Kac) and methylation (Kme) are the most abundant and widely studied<sup>7</sup>. Readers, writers and erasers of these two marks consist of a large number of potentially “druggable” proteins<sup>7, 27</sup>. The term “druggability” here refers to the likelihood of modulating a specific protein with a small-molecule<sup>27</sup>. Important factors governing the “druggability” of a given protein include the hydrophobicity of the binding pocket(s) as well as the pocket’s volume and degree of enclosure<sup>28</sup>. It is important to note that, because the ability to potently target a macromolecule with a small-molecule does not guarantee the ability to generate drug-like molecule that can accomplish a desired therapeutic outcome, “druggability” is a misnomer and discussion of protein targets is best thought of in terms of “tractability for ligand discovery” and not “druggability”.

The existence of multiple chemically tractable writers, readers and erasers of lysine acetylation and methylation has given rise to intense interest in pharmacological modulation of these proteins<sup>21, 23-25, 29-34</sup>. This interest has thus far yielded five US FDA approved antagonists of lysine deacetylases with numerous more currently being investigated in clinical trials<sup>35</sup>. Additionally, at least five antagonists of histone lysine methylation and eleven antagonists of readers of lysine acetylation are currently being investigated in clinical trials<sup>36</sup>. Interest in antagonism of the erasers of lysine methylation is increasing with preliminary data suggesting that some of these molecules will soon be under investigation in clinical trials as well<sup>34, 37</sup>. However, ligand discovery efforts focused on targeting Kme reader proteins have lagged. One explanation for the hesitancy to target these proteins is over concerns of their tractability. Only a limited number of reports focused on antagonists of Kme reader proteins have surfaced in recent years, primarily from academic settings<sup>31, 38-40</sup>. Despite this initial reluctance toward discovery of ligands for Kme reader proteins, the potential utility of such molecules is vast. The therapeutic potential of many of these proteins is an area of great interest as genetic manipulation of numerous Kme readers has provided strong evidence for their therapeutic relevance<sup>7, 17-20, 22, 26, 41-57</sup> (**Table 1.1**). Chemical probes<sup>58-61</sup> that target these proteins are therefore valuable and essential tools for both addressing the tractability of

targeting different Kme readers with small molecule antagonists and for validating the therapeutic potential of this target class.

**Table 1.1.** Examples of Kme readers of potential therapeutic relevance

Reader Family	Protein	Potential therapeutic relevance
PHD Finger	JARID1A PHF23	<i>leukemia</i> <sup>26</sup>
Tandem Tudor	53BP1	<i>breast</i> <sup>50, 51</sup> <i>skin</i> <sup>52</sup> <i>colorectal</i> <sup>53</sup> <i>lung</i> <sup>51</sup>
MBT	SFMBT1	<i>breast</i> <sup>54</sup>
WD40	EED	<i>colorectal</i> <sup>55</sup> <i>breast</i> <sup>56</sup> <i>prostate</i> <sup>57</sup> <i>leukemia</i> <sup>22</sup>
	CBX4	<i>Hepatocellular carcinoma</i> <sup>47</sup>
Chromodomain	CBX7	<i>prostate</i> <sup>15, 16, 18</sup> <i>ovarian</i> <sup>48</sup> <i>leukemia</i> <sup>19</sup> <i>lymphoma</i> <sup>62</sup> <i>gastric</i> <sup>17</sup>
	CBX8	<i>MLL-AF9 leukemia</i> <sup>44</sup> <i>osteosarcoma</i> <sup>41</sup> <i>colorectal</i> <sup>49</sup>

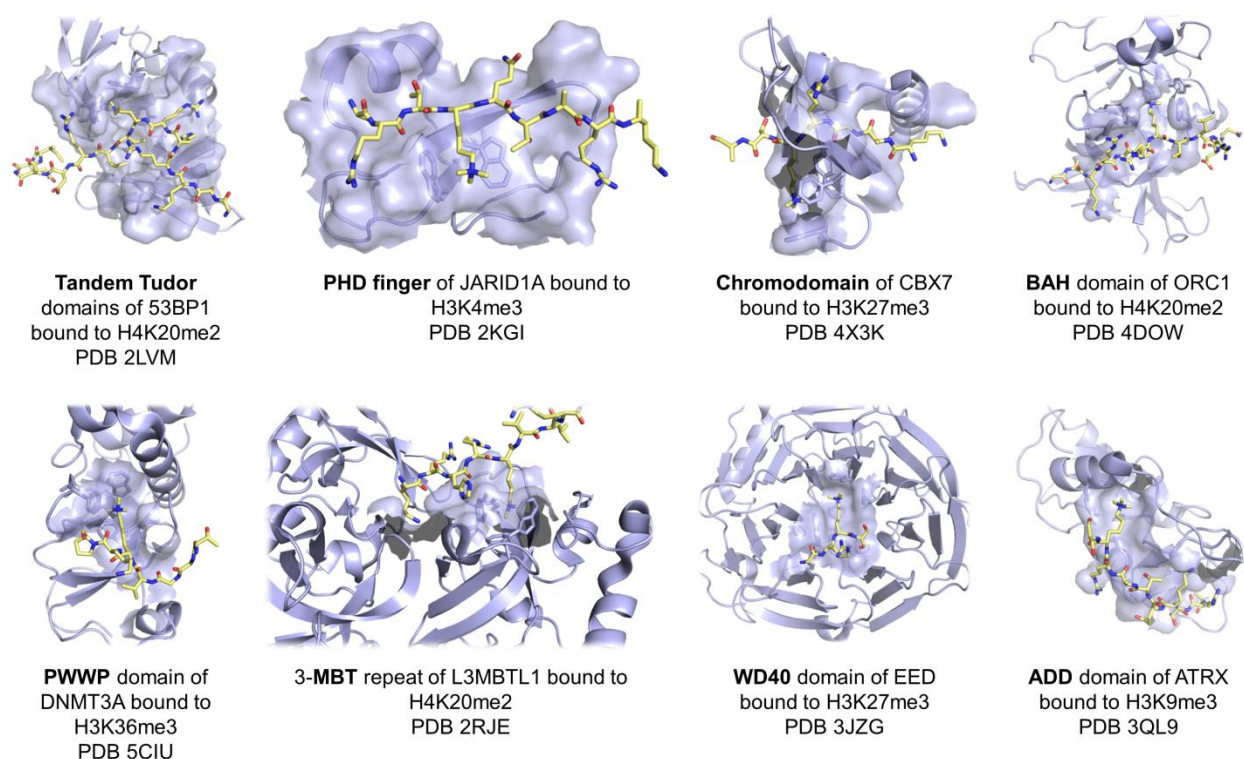
While there is currently no agreed upon, universal definition of a chemical probe, one can loosely define a chemical probe as a molecule with sufficient potency and selectivity for a given target both *in vitro* and *in vivo* (minimally cellular) to associate the molecule's phenotypic effects with modulation of the target of interest. Ideally, chemical probes are monospecific for a single protein. Pragmatically, this will

rarely be the case and as such, molecules require extensive characterization against numerous proteins targets from multiple protein families before they can be used as validated chemical probes<sup>58-61</sup>.

One of the key roles of chemical probes is to bridge the gap between target discovery and target validation. Development of potent and selective chemical probes addresses the key issue of target tractability in drug discovery and agreement between the phenotypic effects of a chemical probe and genetic manipulation of the probe's protein target provides strong validation for the target's therapeutic relevance. Thus, high quality chemical probes are essential for selecting which projects in a drug discovery portfolio should be pursued further in medicinal chemistry campaigns that seek to produce therapeutically useful molecules. Ultimately, regardless of whether they lead to novel, therapeutically relevant molecules, the development of chemical probes is a pressing area of research as probes provide the research community with an important tool that can be used to more fully interrogate and understand biological systems.

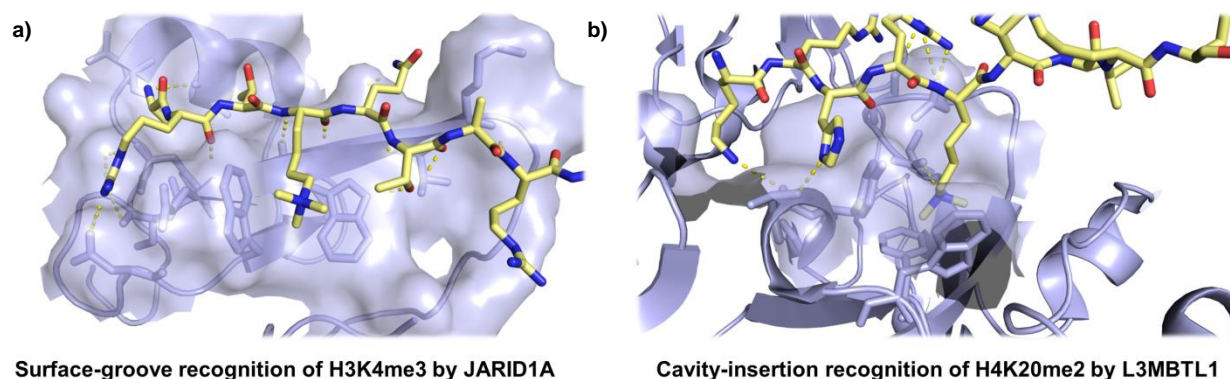
## **THE TRACTABILITY OF CHEMICAL PROBE DISCOVERY FOR READERS OF HISTONE LYSINE METHYLATION**

Questions concerning the general tractability of Kme reader domains originate, in part, from the structural diversity present in this class of proteins<sup>28, 63, 64</sup>. Reader domains of methyl-lysine binding proteins belong to one of at least 8 different families<sup>28, 63-65</sup> (**Fig. 1.3**). The central structural similarity shared amongst nearly all Kme readers is recognition of methylated-lysine through an aromatic cage. The only known exceptions at this time are the ADD domain<sup>65</sup> and certain PHD fingers that recognize unmethylated-lysine. Methyl-lysine reader aromatic cages are large enough to accommodate the hydrophobic side chain of methyl-lysine and engage the methylated amine through cation- $\pi$ , electrostatic and van der Waals interactions. Selectivity between different methylation states (mono-, di- and trimethylation) is conferred primarily by cage geometry and amino acid composition with readers of lower methylation states utilizing acidic residues that participate in a key hydrogen bond with the secondary or tertiary ammonium<sup>63, 64</sup> (**Fig. 1.3 and Fig. 1.4**)



**Figure 1.3.** Kme reader domain families. Representative members from each of the 8 Kme reader domains are shown. Reader domains (light blue) are shown as cartoon representations with their aromatic cage and cognate histone (pale yellow) shown in the stick representation. Surface representations are depicted for residues within 4Å of histones to illustrate the approximate surface areas used for binding from each of the various reader domains.

Analysis of the structural elements of the different Kme reader domains immediately surrounding the aromatic cage further illuminates the complexity of these proteins as a target class. Recognition of methylated histones is accomplished primarily through two-different binding mechanisms: i) surface-groove interactions or ii) cavity-insertion (**Figures 1.3 and 1.4**). Surface-groove readers are sensitive to both the degree of lysine methylation and the residues surrounding the modified lysine while cavity-insertion readers primarily sense the presence of specific methylation states with little preference for the surrounding sequence<sup>63, 64</sup>.



**Figure 1.4.** Comparison of surface-groove and cavity-insertion binders. **(a)** Surface-groove recognition of H3K4me3 by JARID1A. Note the large surface area of interaction, shallow surface area used and the multiple hydrogen bonds formed between JARID1A and H3 (PDB 2KGI). **(b)** Cavity –insertion recognition of H4K20me2 by L3MBTL1 (PDB 2RJE). Note the deeper cavity formed by the aromatic cage, smaller surface area of interaction, and fewer hydrogen bonds compared to the JARID1A-H3 interaction.

Each of these binding mechanisms presents unique hurdles for the development of potent ligands for these proteins. While cavity-insertion readers are attractive for the design of small-molecules because of their enclosed binding pockets, these pockets are relatively small with few moieties available for forming stable interactions<sup>66</sup>. In contrast, while surface-groove Kme readers tend to bind more tightly to their histone substrates compared to cavity-insertion binders, they are unfavorable for ligand discovery efforts because they utilize large, open areas of their surface for binding histones<sup>26, 67, 68</sup>. The challenges in targeting each of these classes of Kme readers were quantified in a study utilizing a computational analysis comparing the histone binding sites of multiple families of Kme readers and bromodomains (Kac readers). These studies quantitatively scored the “druggability” of these proteins and concluded that methyl-lysine binding proteins are, on average, considerably less tractable for ligand discovery efforts than bromodomains, which have been effectively targeted in multiple small-molecule ligand discovery campaigns<sup>28</sup>.

## SCREENING FOR ANTAGONISTS OF METHYL-LYSINE READERS

While Kme readers appear to be less tractable as a target class compared to Kac readers, individual families within the Kme reader class of proteins do appear to be promising for the development of potent small molecules<sup>28</sup>. Additionally the therapeutic potential of many of these proteins warrants further investigation into the tractability of Kme readers. In order to explore the tractability of any target class, robust, cost-effective high-throughput screening methods are required. This was one of the most significant initial challenges in targeting Kme reader domains. As opposed to enzymatic protein targets, which offer the advantage of catalyzing a chemical reaction that can be monitored for the disappearance of substrate(s) or the appearance of product(s), Kme reader domains perform their biological function through simple binding, a considerably more difficult event to monitor than chemical reactions.

Monitoring for blockade of binding interactions is typically accomplished through assays which utilize tagged versions of one binding partner that undergoes a measurable change when it has been displaced from its binding partner. Traditional high-throughput displacement assays such as fluorescence polarization have been applied to screen for methyl-lysine antagonists<sup>38, 40</sup>, but the weak nature of the reader-histone interactions requires the use of relatively large amounts of protein. This is problematic primarily for two reasons. First, the large amount of protein required places practical limits on the number of compounds that can be screened against the desired target, thereby decreasing the likelihood of discovering a useful hit molecule. Second, it precludes simultaneous screening against other Kme readers. Screening against multiple targets with similar biological functions is known as target class screening and is advantageous because it both provides knowledge of the selectivity profile of compounds for closely related proteins and enables the serendipitous discovery of antagonists for Kme readers that were not the subject of the original screen.

The challenges in developing a high-throughput platform targeting Kme readers were largely overcome through the use of AlphaScreen® technology. AlphaScreen® is a bead-based, proximity assay that measures the transfer of singlet oxygen between a donor and acceptor bead. The assay functions through the use of tagged Kme readers domains (usually a His-tag). This tag is recognized by Ni<sup>2+</sup>, which coats the singlet oxygen acceptor bead. The streptavidin coating on the singlet oxygen donor bead causes these beads bind to selectively bind to the biotinylated histone substrate of the reader domain.

When the histone is engaged with the reader domain, the donor and acceptor beads are brought into close proximity with each other. Donor beads are then excited with 680 nm  $\lambda$  light, generating singlet oxygen that is in close enough proximity to diffuse to the acceptor bead and stimulate the acceptor to generate an emission event in the 520-620 nm  $\lambda$  range. The presence of multiple streptavidin and  $\text{Ni}^{2+}$  molecules on the surface of the donor and acceptor beads creates an avidity effect that overcomes the weak nature of histone-reader interactions, allowing for a dramatic increase in assay sensitivity and a decrease in both required protein and cost<sup>69</sup>. Antagonists of the interaction displace the peptide from the reader domain, preventing transfer of singlet oxygen, which can only diffuse distances less than 200 nm, between the donor and acceptor beads. Monitoring for a dose-dependent loss of the emission event in the 520-620 nm  $\lambda$  range then enables the identification of Kme reader antagonists.

The advantages of AlphaScreen® technology in screening make it a valuable tool for identifying antagonists of Kme reader domains, but this assay format is not without its drawbacks. Compounds that can chelate the acceptor beads or absorb light in the wavelengths important for the assay will show inhibition in AlphaScreen® even if they do not physically interact with the protein of interest. Additionally, as with most other high-throughput assays, AlphaScreen® technology is limited in its ability to distinguish between true, selective binders and non-specific aggregators. Therefore, this platform must always be used in conjunction with orthogonal biophysical techniques to distinguish between tractable hit molecules and assay artifacts.

The issues associated with AlphaScreen® can largely be addressed through the use of orthogonal biophysical techniques as secondary assays. The gold standard of biophysical techniques for characterizing protein-ligand interactions is isothermal titration calorimetry<sup>70</sup> (ITC). A typical ITC experiment measures the energy released when titrating a ligand into a solution of the protein target of interest. The results of this experiment provide the stoichiometry, affinity, enthalpy and entropy of the interaction, thus providing a complete thermodynamic picture of the binding equilibrium<sup>71</sup>. ITC is indispensable for distinguishing between the assay artifacts, aggregators and true binders that emerge from AlphaScreen®. However, because of the large amounts of both compound and protein required, there is a practical limit on the number of compounds that ITC can be used to evaluate. While both AlphaScreen® and ITC have their own unique limitations, pairing the high-throughput and cost-effective

AlphaScreen® platform with ITC as a secondary assay produces a powerful combination for identifying and characterizing antagonists of Kme reader domains.

In addition to ITC, surface plasmon resonance (SPR) is another validated biophysical technique that can provide valuable information concerning protein-ligand interactions. SPR is medium-throughput assay format where the protein target of interest is immobilized to a gold surface. Compounds are then be flowed across the protein-coated surface and binding of the compound is quantified by measuring the changes in the refractive index of the solvent layer near the surface<sup>72</sup>. Like ITC, this assay can be used to measure binding affinity, but it also can be used to characterize the kinetics of the interaction. Combining AlphaScreen®, ITC and SPR allows for the identification of histone competitive antagonists with complete characterization of both the thermodynamic and kinetic parameters of the protein-antagonist interaction; information that is invaluable for the design of improved compounds.

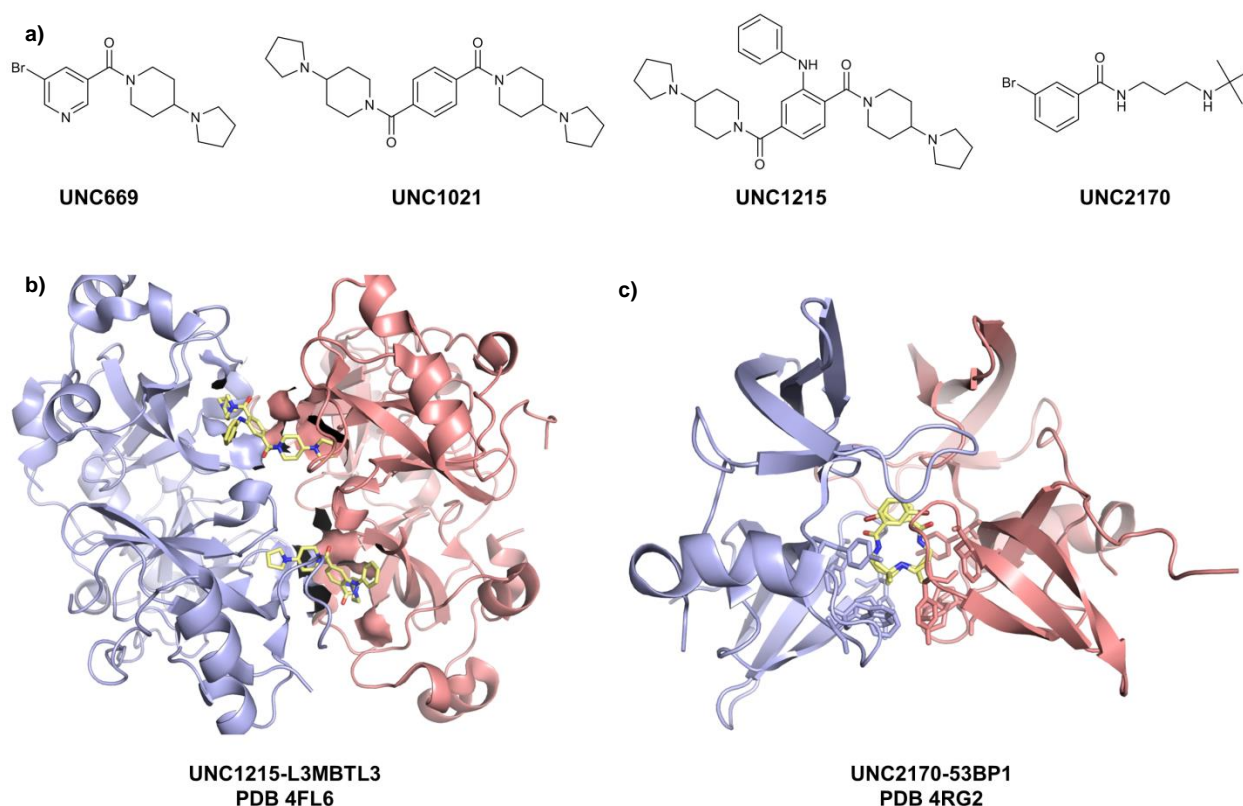
## **EXAMPLES OF PREVIOUS CHALLENGES AND SUCCESSES OF SMALL-MOLECULE CHEMICAL PROBE DISCOVERY EFFORTS TARGETING METHYL-LYSINE READERS**

Initial reports of Kme reader antagonists focused on small-molecules that targeted the cavity-insertion binder, L3MBTL1. This protein recognizes H4K20me1 and H4K20me2 through a 3-domain repeat of MBT domains<sup>66, 73, 74</sup>. These efforts ultimately led to the development UNC669 (**Fig. 1.5a**), which possesses a 4-(pyrrolidin-1-yl)piperidin-1-yl moiety as a conformationally-constrained methyl-lysine mimic that specifically engages the second MBT domain of L3MBTL1 with modest affinity (ITC  $K_d = 5 \mu\text{M}$ )<sup>74</sup>. Disappointingly, the lack of significant functionalities surrounding the aromatic cage has thus far prevented the development of submicrolar antagonists of L3MBTL1.

In addition to L3MBTL1, UNC669 also interacts, with similar affinities, with the 3-MBT domain repeat of L3MBTL3 and the tandem Tudor domains of 53BP1. Like L3MBTL1, both L3MBTL3 and 53BP1 are H4K20me2 readers. SAR efforts around UNC669 targeting 53BP1, led to UNC1021 (**Fig. 1.5a**), which possesses a second 4-(pyrrolidin-1-yl)piperidin-1-yl moiety around a slightly different aromatic core than UNC669. While UNC1021 does not possess increased affinity for 53BP1 as intended, it was serendipitously discovered through target class screening that UNC1021 potentially antagonizes the MBT repeats of L3MBTL3 (AlphaScreen®  $\text{IC}_{50} = 0.04 \mu\text{M}$ ). Continued SAR studies led to the identification of



UNC1215 (**Fig. 5a**, AlphaScreen®  $IC_{50} = 0.04 \mu M$ , ITC  $K_d = 0.12 \mu M$ ), a cellularly active chemical probe of L3MBTL3 and the first chemical probe for any Kme binding protein. The mechanism for the increased potency of these dibasic compounds for L3MBTL3 was uncovered through the use of X-ray crystallography. These studies revealed that L3MBTL3 forms a 2:2 dimeric complex with UNC1215. Here the lack of significant interactions with residues outside of the aromatic cage is overcome by additional contacts formed between each L3MBTL3 monomer as well as the interaction between a second reader pocket from each L3MBTL3 monomer and the additional constrained methyl-lysine mimic of a second molecule of UNC1215 (**Fig. 5b**)<sup>75</sup>.



**Figure 1.5.** Antagonists of Kme readers. **(a)** Structures of L3MBTL1 and L3MBTL3 antagonists UNC669, UNC1021 and UNC1215. **(b)** Crystal structure of 2 molecules of UNC1215 (pale yellow) bound to a homodimer of the 3-MBT repeats of L3MBTL3 (one monomer is shown in light blue and the other in salmon). **(c)** Crystal structure of UNC2170 (pale yellow) bound to a homodimer of tandem Tudor domains of 53BP1 (one monomer is shown in light blue and one in salmon). UNC2170 can exist in two different conformations within the 53BP1 homodimer, both of which are shown in the image.

While UNC1021 and its derivatives failed to improve 53BP1 potency as intended, continued target class screening efforts eventually identified a fragment-like ligand, UNC2170 (**Fig. 1.5a**), with low micromolar potency for 53BP1. This molecule utilizes a *t*-butyl amine to interact with the aromatic cage of 53BP1. The identification of a second, distinct small molecule antagonist of 53BP1 in addition to UNC669 and its subsequent derivatives was encouraging; especially considering that 53BP1 recognizes H4K20me2 with an open and shallow surface-groove. While the regions surrounding the aromatic cage of 53BP1 present more attractive functionalities for targeting with small-molecule ligands than L3MBTL1 or L3MBTL3, SAR studies revealed minimal tolerance for modification to UNC2170. X-ray crystallography and ITC studies revealed that the tandem Tudor domain of 53BP1, similar to L3MBTL3 and UNC1215, dimerizes around UNC2170 (**Fig. 5c**). However, the complex formed between 53BP1 and UNC2170 is instead a 1:2 complex (1 molecule of UNC2170 to 2 molecules of 53BP1). The dimeric complex of 53BP1 formed around UNC2170 leaves little steric tolerance for any modifications to the compound, explaining the lack of tractable SAR around this compound<sup>76</sup>.

Difficulties targeting the MBT repeats of L3MBTL1 exemplify the difficulties in targeting cavity-insertion binders<sup>63, 66, 73, 74</sup> and despite multiple seemingly tractable hit molecules starting points for targeting 53BP1; this protein has so far proven intractable for the development of potent small molecules<sup>76</sup>. In contrast, L3MBTL3 has proven highly tractable for the design and discovery of potent antagonists. In fact, there are now multiple series of compounds capable of targeting the MBT domains of L3MBTL3<sup>75, 77, 78</sup>. However, the origin of these compounds was primarily serendipitous in nature<sup>75</sup>. Taken together, these examples highlight the challenge of rationally identifying and designing potent antagonists of Kme reader proteins.

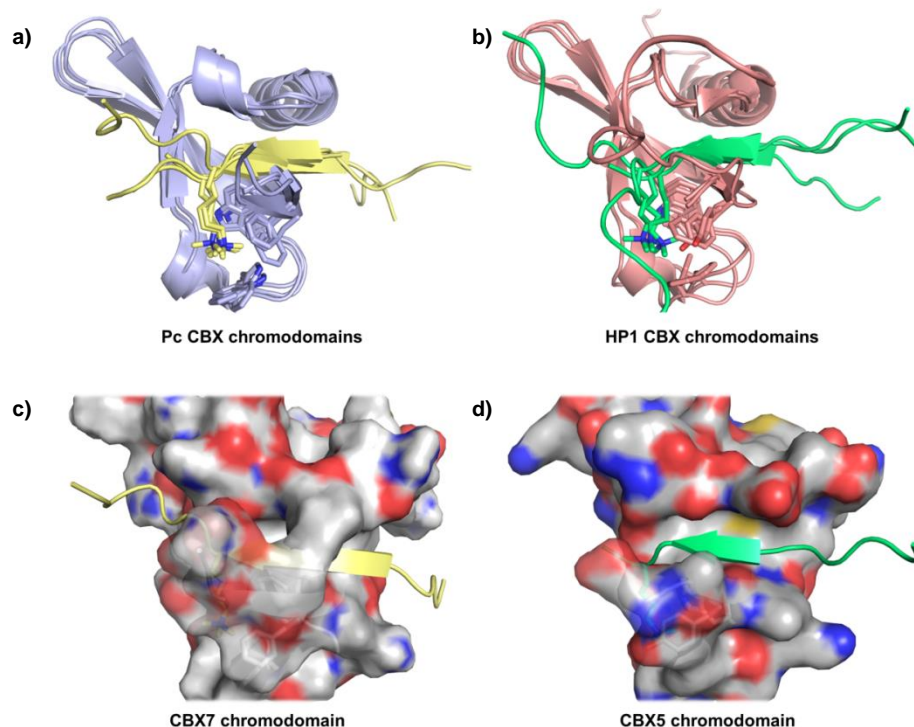
While the discovery of UNC1215 proved that some Kme readers can, in fact, be tractable macromolecular targets, UNC1215 has been of limited utility in answering any significant biological questions. The biological role of L3MBLT3 is relatively unknown, making it challenging to test specific hypotheses with UNC1215. Additionally, UNC1215 does not appear to have any observable phenotypic or morphological effects on cells that suggest a cellular role for L3MBTL3. In time, UNC1215 may prove to be a useful tool, but the current lack of utility of UNC1215 has led to an increased focus on ligand-discovery efforts for Kme readers that are more readily correlated with specific phenotypes in biological

systems. Focusing on these readers is more likely to produce molecules that will address pressing biological questions.

## THE CBX FAMILY OF CHROMODOMAINS

One Kme reader family that has been linked to multiple specific phenotypes of interest is the chromodomain family of reader proteins. The chromodomain is a trimethyl-lysine binding structural motif that consists of three beta strands packed against a C-terminal alpha helix that mediates protein-protein interactions. Chromodomains are believed to target their associated protein and protein complexes to specific sites on chromatin. They are essential for proper genomic regulation in numerous organisms spanning fungi, plants and animals<sup>79</sup>. In mammals, perhaps the best studied chromodomain-containing proteins are the Chromobox Homolog (CBX) proteins<sup>15-20, 41-49, 62, 68, 80-93</sup>.

There are eight mammalian CBX proteins that are divided into two different families, the Heterochromatin 1 (HP1; CBX1, -3 and -5) and Polycomb (Pc; CBX2, -4, -6, -7 and -8) families<sup>68, 79</sup>. Members of both of these protein families possess an N-terminal chromodomain that selectively bind to ARKme3S motifs. The structures of CBX chromodomains have been thoroughly characterized and have been shown to recognize ARKme3S motifs through a surface-groove recognition mechanism<sup>68</sup> (**Fig. 1.6**). However, despite their high degree of structural similarity and binding preferences, there are differences in the binding preferences between these two families of chromodomains. HP1 chromodomains selectively bind to the ARKme3S motif of H3K9me3 while Pc chromodomains are generally capable of binding to the ARKme3 motifs of both H3K9me3 and H3K27me3<sup>68, 94</sup>. The differences in the binding preferences between these two families of chromodomains despite their high degree of structural similarity is believed to be mediated by the more polar surface of HP1 chromodomains relative to the lipophilic surface of Pc chromodomains. The polar surface of HP1 chromodomains more favorably interacts with the polar side chains surrounding the ARKS motif of H3K9me3 while the lipophilicity of Pc chromodomains increases their promiscuity between these marks<sup>68</sup> (**Fig. 1.6c, d and Table 1.2**)



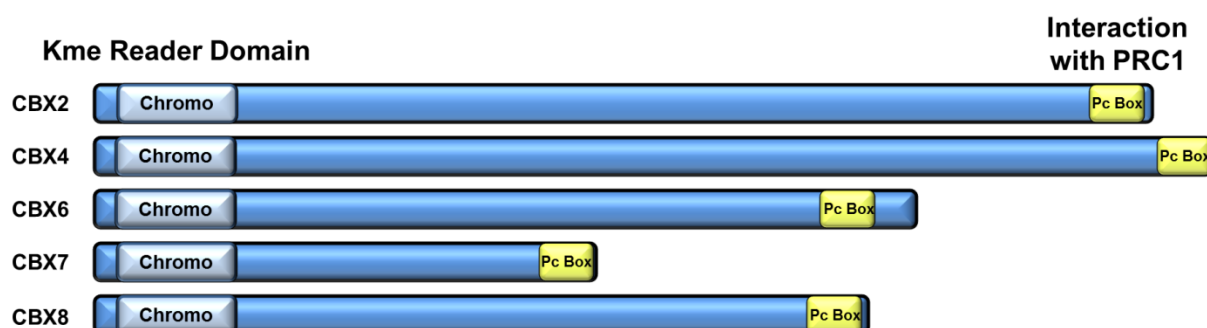
**Figure 1.6.** Comparison of Pc and HP1 CBX chromodomains. **(a)** Overlay of the chromodomains of CBX2, -6, 7 and -8 when bound to H3K9me3 or H3K27me3 (there is currently no published structure of CBX4 bound to H3). **(b)** Overlay of the chromodomains of CBX1, -3 and -5 chromodomains bound to H3K9me3. **(c)** Surface representation of CBX7 bound to H3K27me3. **(d)** Surface representation of CBX5 bound to H3K9me3. Note the similar folds for both Pc and HP1 chromodomains. For **(c)** and **(d)**, regions of gray represent lipophilic surfaces while blue and red indicate polar areas that are electron deficient (blue) and electron rich (red). Note the greater polarity of CBX5 **(d)**, resulting in a more open binding groove compared to CBX7 **(c)**.

**Table 1.2.** Comparison between Pc and CBX chromodomains

	Chromodomain Family	
	Polycomb (Pc)	Heterochromatin Protein1 (HP1)
<b>Members</b>	CBX2, -4, -6, -7 and -8	CBX1, -3 and -5
<b>Histone substrate</b>	H3K9me3 and H3K27me3	H3K9me3 selective
<b>Binding affinity</b>	Approximate $K_d$ 's between 50-100 $\mu$ M	Approximate $K_d$ 's between 5-40 $\mu$ M
<b>Nature of binding surface</b>	Lipophilic	Polar

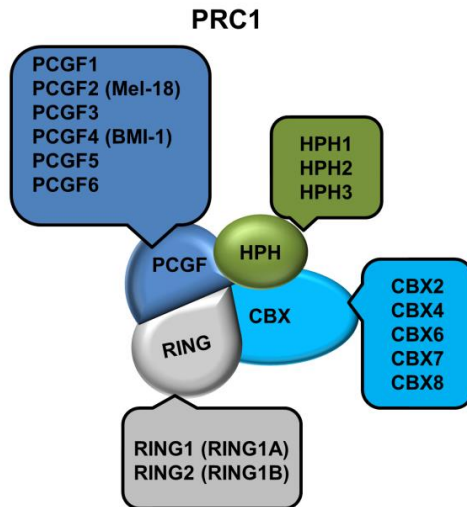
The more hydrophobic and enclosed nature of Pc chromodomains suggests that these proteins are more tractable for ligand discovery efforts, a fact that is reiterated by their “druggability” scores wherein three out of four Pc CBX chromodomains analyzed were scored as “druggable” (CBX2, -6 and-7) while only one out of three HP1 chromodomains (CBX1) was scored as “druggable”. The only chromodomain not analyzed in this study was the chromodomain of CBX4, for which there is no structure available of the chromodomain bound to H3<sup>28</sup>.

## THE CANONICAL POLYCOMB SIGNALING PATHWAY



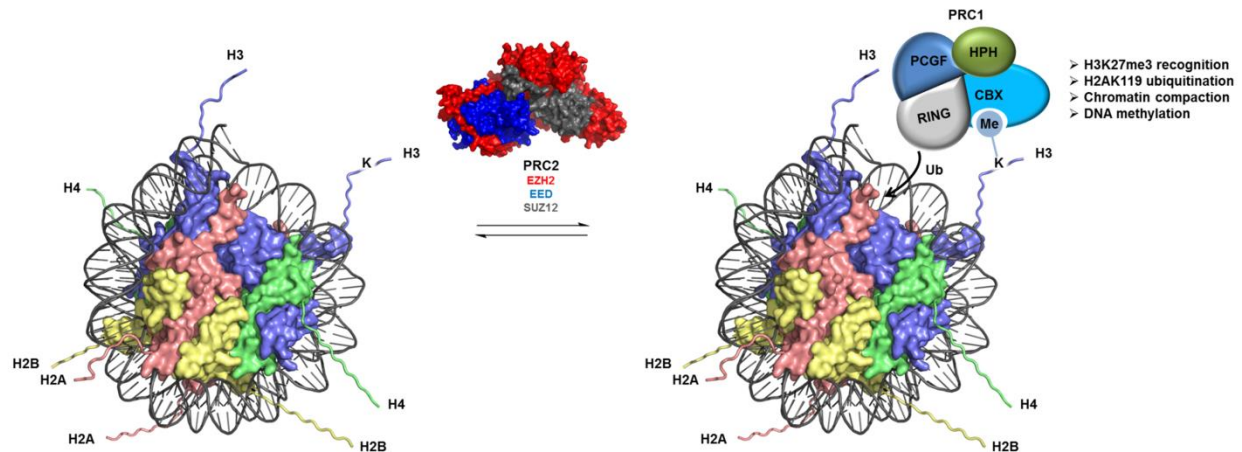
**Figure 1.7.** Domain maps of Pc CBX proteins. Each Pc CBX protein contains an N-terminal chromodomain for binding to H3 and a C-terminal Pc Box for interaction with PRC1. Only domains annotated at [www.uniprot.org](http://www.uniprot.org) are shown.

In addition to their chromodomains, each Pc CBX protein also possesses a C-terminal domain known as a Pc box (**Fig. 1.7**) that mediates their incorporation into Polycomb Repressive Complex 1 (PRC1). This complex is essential for regulating cellular growth, proliferation and stem cell pluripotency<sup>19, 45, 46, 81, 94</sup>. Canonical PRC1 is built around 4 main subunits, a RING finger protein, a PCGF family member, one CBX component and an HPH protein family member. The existence of multiple redundant family members for each of these subunits creates the potential for at least 180 permutations of canonical PRC1<sup>92</sup> (**Fig. 1.8**).



**Figure 1.8.** Composition of canonical PRC1. Boxes list each of the different orthologues for each core component.

Under the canonical polycomb signaling model, PRC1 is hypothesized to function downstream of PRC2, a methyltransferase complex that is minimally composed of one of two methyltransferases, EZH1 or EZH2, and the scaffolding proteins SUZ12 and EED, a known H3K27me3 reader. PRC2 is responsible for deposition of H3K27me3 on nucleosomes which is then thought to recruit PRC1 to specific genomic loci through methyl-lysine binding by the CBX component of PRC1. The RING finger component of PRC1 then monoubiquitinates H2AK119. The presence of H2AK119ub on nucleosomes is associated with compacted chromatin and DNA methylation, resulting in stable repression of the underlying gene<sup>91, 92, 95</sup> (**Fig. 1.9**).



**Figure 1.9.** Canonical Polycomb Signaling pathway. H3K27me3 is deposited by PRC2 (PDB 5HYN), minimally composed of EZH2 (red), EED (blue) and SUZ12 (gray). H3K27me3 then recruits PRC1 through the Pc CBX component of the complex, allowing H2AK119 ubiquitination by the RING finger component. Recruitment of PRC1 to nucleosomes is associated with chromatin compaction, DNA methylation and stable genomic repression. The nucleosome image was adapted from PDB 5B0Z.

This simplistic model was derived from studies in *Drosophila* and provided a useful initial framework in which to begin exploring the biology of PRC1 in mammals. However, there is a growing body of literature that challenges the relevance of the canonical polycomb signaling pathway in mammals. For example, the composition of PRC1 is highly variable. In fact, there are now known to be a subset of PRC1 complexes in cells that differ from canonical PRC1 in that they lack a CBX component entirely. The lack of a known H3K27me3 reader in these non-canonical complexes creates an entire new branch of PRC1 biology that is PRC2-independent<sup>14, 89</sup>.

Additional studies have continued to raise considerable doubt as to the validity of the canonical polycomb signaling model in mammals. One study found that mutations and deletions of the CBX chromodomains that abolish H3K27me3 binding do not significantly alter the association of PRC1 with chromatin or with sites of H3K27me3<sup>96</sup>. One possible explanation for this result could be explained by a recent study that found that EED is a component of both PRC1 and PRC2. The authors of this study postulated that EED orchestrates polycomb signaling by first localizing PRC2 to nucleosomes after which, PRC1 is localized to nucleosomes by binding to EED and competing away the other components of

PRC2<sup>57</sup>. This however is refuted by the observation that even in EED null cells; the association of PRC1 components is unchanged compared to WT<sup>96</sup>.

The canonical role of both the CBX component and deposition of H2AK119ub in PRC1-mediated genomic silencing have also been challenged by studies showing that the ubiquitination module of PRC1 does not require a CBX protein nor H3K27me3 to interact with nucleosomes. Further, the presence of RING1B on chromatin, and not its catalytic activity, is responsible for PRC1-mediated transcriptional silencing and chromatin compaction<sup>12</sup>. Despite these challenges, there is still considerable evidence supporting the canonical signaling model<sup>18, 19, 93</sup>, indicating that signaling through PRC1 and PRC2 is extraordinarily complicated and that these two complexes are likely to be involved in multiple signaling pathways and feedback loops, making general statements about their physiological roles challenging.

## **THE BIOLOGICAL FUNCTION OF PC CBX PROTEINS**

While an in depth understanding of the polycomb signaling pathway(s) remains elusive, it is clear that Pc CBX proteins are critical for normal cellular function. For example, CBX7 is critical for repression of the inactive X chromosome<sup>94</sup>. Additionally, CBX2, -4 and -7 function within a tightly regulated feedback loop that controls cellular differentiation<sup>19, 45, 46, 81, 88, 89, 96</sup>. In embryonic stem cells, the predominant Pc CBX protein is CBX7, where it functions to help the cells maintain a pluripotent state. Interestingly, it does so at least partially through repression of CBX2 and -4. During differentiation, CBX7 is downregulated through expression of certain microRNAs, allowing CBX2 and -4 levels to increase. These two CBXs in turn repress expression of CBX7 in order to help maintain a differentiated state<sup>46, 88</sup>. A similar mechanism was observed for the regulation of hematopoietic stem cell differentiation where CBX7 expression is associated with pluripotency and is down regulated during differentiation when CBX2, -4 and -8 levels begin to increase<sup>19</sup>. The association of CBX7 with pluripotency is also correlated with stem cell self-renewal and proliferation, while expression of the remaining Pc CBXs during differentiation correlates with reduced proliferation and eventually senescence<sup>19, 46, 88</sup>.



## POLYCOMB CHROMODOMAINS IN CANCER

Given their role in regulating differentiation and proliferation, it is not surprising that Pc CBX proteins have been implicated in the development and progression of multiple cancers<sup>15-20, 41-44, 47-49, 62, 80, 82-85, 87, 90</sup>. Despite its shortcoming as general model for signaling through PRC1, there is evidence that the canonical polycomb signaling pathway plays a part in the development and progression of some cancers<sup>18, 19, 93</sup>. Therefore, it is encouraging that EZH2 antagonists have shown promise as therapeutic agents in laboratory models of multiple cancer types and are currently being evaluated in multiple clinical trials<sup>21, 23-25, 32, 36</sup>. While EZH2 antagonists are still many years from being validated as safe and effective therapeutic agents, their initial success is highly suggestive that antagonists targeting other components of polycomb signaling pathways are a viable therapeutic strategy.

Regardless of their biological role with respect to EZH2, Pc CBX proteins present a unique and interesting target class for potential oncology therapeutics. To date, CBX7 has been the most extensively characterized Pc CBX protein and regulation of its expression has been found to contribute to a variety of cancer types<sup>15-20, 42, 43, 48, 62, 80, 82-85, 87</sup>. Interest in the role of CBX7 in cancer began when it was identified in a genetic screen for cDNAs capable of bypassing replicative senescence in normal human prostate epithelial cells. CBX7 was found to dramatically increase cellular lifespan in a manner similar to c-Myc and the ability of CBX7 to confer a proliferative advantage was shown to be dependent on both its chromodomain and Pc box. Therefore, this activity requires both its ability to bind to methyl-lysine and to be incorporated into PRC1. These studies concluded that increased proliferative capacity bestowed by CBX7 overexpression is a conserved mechanism across different cell-types and is driven by repression of the *INK4a/ARF* locus by CBX7<sup>15</sup>. These conclusions were further supported by a follow-up study focusing on the role of CBX7 in multiple prostate cancer cell lines. This study also suggested that CBX7 is important in the transition of prostate tumors to androgen independence<sup>16</sup>.

The notion that increased CBX7 overexpression is a general mechanism for increasing proliferative capacity was soon challenged by additional studies<sup>80, 82-85, 87</sup>. Studies in patient-derived thyroid cancer cell lines indicated that CBX7 functions as a tumor suppressor in thyroid tissues and found that loss of CBX7 expression correlates with increasing malignancy in these tumors. Further complicating matters, this study concluded that CBX7 functions to repress the *INK4a/ARF* locus in thyroid tissue as

well, indicating that the role of CBX7 in regulating cell proliferation is more complicated than simple regulation of the *INK4a/ARF* locus<sup>80</sup>. Continued studies have demonstrated the ability of CBX7 to function as a tumor suppressor in multiple cell lines, including lung, colon, breast and pancreatic cancers<sup>82-85, 87</sup>. One study even suggested that the more generalizable function of CBX7 is to serve as a tumor suppressor based on a CBX7-knockout mouse model which found that *CBX7*-KO mice were especially susceptible to developing adenomas and carcinomas of both the liver and lungs in adulthood. Studies with these mice also suggested that CBX7 positive regulates expression of the *INK4a/ARF* locus<sup>87</sup>, in direct contrast to previous studies<sup>15, 16, 80</sup>. However, despite the evidence for CBX7 serving as a tumor suppressor<sup>80, 82-85, 87</sup>, additional studies in prostate<sup>18, 43</sup>, leukemia<sup>19</sup>, lymphoma<sup>62</sup>, gastric<sup>17</sup>, and ovarian<sup>48</sup> cancers have supported the notion that CBX7 overexpression is generally oncogenic. Interestingly, these studies disagree on the general role of CBX7 in regulating the *INK4a/ARF* locus and how this relates to cell proliferation.

Although less explored than CBX7, a similar conflict exists concerning the roles of CBX2, -4 and -8. In the context of stem cells, CBX2, -4 and -8 expression is associated with reduced proliferative capacity<sup>19, 46, 88</sup>. However, each of these proteins has also been shown to enhance proliferative capacity in at least one cancer type<sup>41, 44, 47, 49, 90, 93</sup>. CBX2 for instance has been found to be upregulated in tumors of the pancreas, central nervous system, colon, endometrium, ovaries, breast and lung. Additionally, analysis of the Oncomine database found that CBX2 mRNA levels are frequently increased in metastatic tumors in comparison to primary tumors. This same study did not uncover a single instance of underexpression of CBX2 in the Oncomine database<sup>90</sup>. Another study uncovered upregulation of CBX2 in neuroendocrine prostate cancer, a highly aggressive subtype of prostate cancer with a median survival of less than a year. This study revealed that, in addition to CBX2, EZH2 is also upregulated in this cancer subtype, suggesting that CBX2 contributes to the disease through the canonical polycomb signaling pathway<sup>93</sup>.

Similar to CBX7, CBX4 and -8 are each capable of bypassing cellular senescence<sup>41, 45, 49</sup>. In human epidermal stem cells, the chromodomain of CBX4 is responsible for repressing senescence, independent of its incorporation into PRC1<sup>45</sup>. Additionally, CBX4 has been found to be upregulated in multiple hepatocellular carcinoma cell lines and clinical samples. Encouragingly, RNAi-mediated

knockdown of CBX4 was found to inhibit cell proliferation and to delay cell cycle progression in these cell lines<sup>47</sup>. Similar to CBX4, CBX8 is capable of bypassing senescence in a PRC1-independent fashion<sup>41, 44, 49</sup>. CBX8 is known to be critical for the development of MLL-AF9 leukemias, again in a PRC1-independent manner, and its removal in mice does not affect normal hematopoietic development nor is it associated with any aberrant phenotype<sup>44</sup>.

In comparison to the other Pc CBX proteins, CBX6 has been severely understudied to date. Little is known about the biological role of CBX6 except that it appears to exist only in a limited subset of PRC1 complexes and preferentially associates with other proteins outside of the context of PRC1<sup>86</sup>. Additionally, only one study to date has evaluated the role of CBX6 in any cancer type, the findings of which suggest that CBX6 functions as a tumor suppressor in glioblastoma multiforme cell lines<sup>20</sup>. Little more can be said of the disease relevance of CBX6 at this time.

## SUMMARY

Readers, writers and erasers of the histone code play a critical role in regulating genomic expression in mammalian organisms. Overexpression and mutation of these proteins contributes to a variety of diseases, especially cancer. Encouragingly, many of these proteins have been shown to be pharmacologically tractable and antagonists of some of these proteins are showing promise in preclinical studies and clinical trials. However, many of these proteins are still unexplored in terms of small-molecule discovery efforts, particularly readers of lysine methylation.

The Pc CBX family of Kme3 readers is an important family of proteins that regulates cellular proliferation and differentiation. Further, these proteins have been linked to numerous cancer types. These proteins are highly similar in their chromodomains and bind to both H3K9me3 and H3K27me3 *in vitro* through a surface-groove recognition mechanism. The binding surface of these chromodomains is more lipophilic and enclosed than the structurally similar HP1 chromodomains, suggesting that they may be more tractable for targeting with small molecules.

Genetic manipulation of individual Pc CBX chromodomains has indicated that the contribution of these proteins in oncogenic signaling pathways is context dependent. The uncertainty surrounding the functions of these proteins in cancer, especially CBX7, is a clear indication that improved tools, especially

chemical probes, are needed to more clearly elucidate the function and therapeutic relevance of these proteins. Further, the roles of Pc CBX proteins in regulating cellular differentiation suggests that chemical probes for these proteins will also be valuable tools for exploring and understanding stem cell biology.

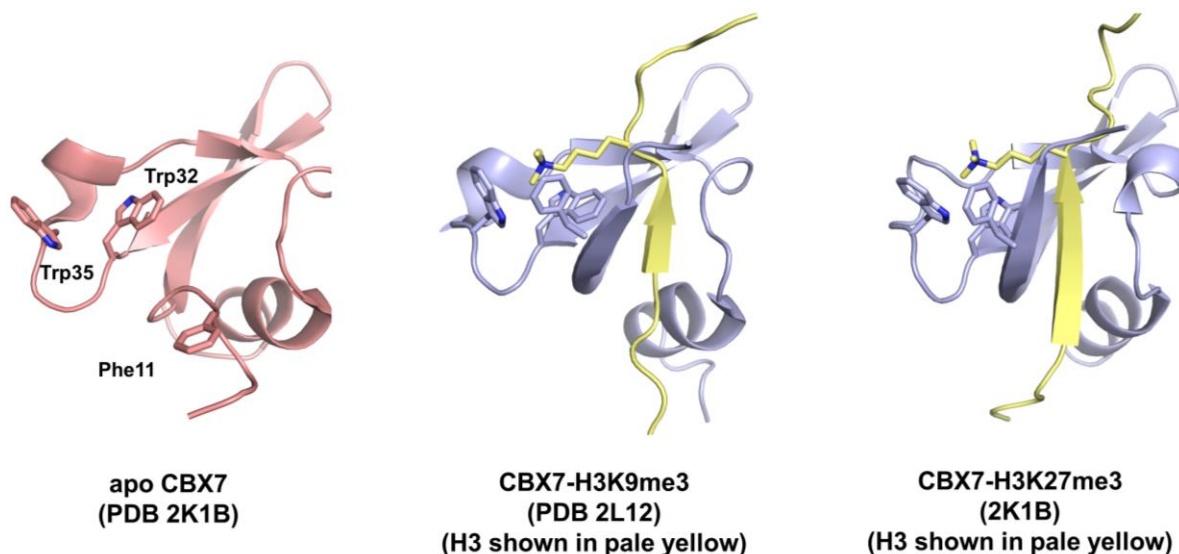
## CHAPTER II: DEVELOPMENT OF HIGH AFFINITY LIGANDS FOR THE CBX7 CHROMODOMAIN

### INTRODUCTION

The pressing biological questions and lack of chemical probes available for Pc CBX proteins prompted us to undertake a discovery effort targeting this family of proteins. We opted to target the chromodomain of CBX7 initially due to its extensive biological characterization<sup>15-19, 42, 43, 46, 48, 62, 80, 82-89, 94</sup>, existing structural information<sup>68</sup> and availability of an established AlphaScreen® competition assay developed in our lab<sup>66, 69, 74, 75</sup>. We set out to produce a validated, selective chemical probe for the chromodomain CBX7 with the understanding that monospecific chemical probes are often not readily feasible within highly homologous protein families, such the Pc CBX chromodomains. Therefore, we understood that any antagonists of the CBX7 chromodomain were likely to exhibit limited selectivity within the Pc CBX family of chromodomains. This is common among first-in-class chemical probes. Nonetheless, the utility of these molecules can still be profound. For example, the first in class bromodomain antagonist, JQ1, was not selective within the BET family of bromodomains but has greatly improved our knowledge of the role of BET bromodomains in numerous disease states<sup>8, 29</sup>. Further, the success of JQ1 in modulating many of these disease states sparked intense interest across multiple pharmaceutical and academic labs across the world, eventually leading to the identification of more potent and selective antagonists of BET bromodomains<sup>29</sup>. It is likely that efforts targeting Pc CBX chromodomains will ultimately follow a similar course. The homology of these proteins not only in terms of their chromodomain as a whole, but especially in the regions within their chromodomains that directly interact with H3 suggests a low probability of discovering molecules that are highly specific to individual chromodomains<sup>31</sup>. Nonetheless, potent, well-characterized molecules targeting these chromodomains are still likely to serve as valuable tools for elucidating the biology of these proteins and, similar to JQ1, may eventually lead to more selective and potent antagonists.

Our decision to focus on CBX7 was due in large part to the previous structural characterization of CBX7, which included NMR structures of the chromodomain in the apo form (PDB 2K1B), bound to

H3K9me3 (PDB 2L12) and bound to H3K27me3 (PDB 2L1B). These studies revealed that the CBX7 aromatic cage (consisting of Phe11, Trp32 and Trp35) is disorganized in the apo form of the chromodomain. Recognition of H3K9me3 and H3K27me3 involves rearrangement of the chromodomain around the histone to form a functional aromatic cage and recognition of both histones occurs via an anti-parallel  $\beta$ -strand interaction (**Fig. 2.1**). In conjunction with these structural studies, mutagenesis studies were also done with both H3K9me3 and H3K27me3 peptides to better understand the binding preferences of CBX7. Encouragingly, these studies revealed that, while CBX7 is selective for ARKme3S motifs, it is not specific for this motif. In fact, these studies uncovered that the arginine of this consensus sequence is amenable to mutation to various hydrophobic residues to produce more potent peptidic ligands for CBX7. These extensive studies revealed that the overall consensus binding motif for CBX7 is actually A(R/I/L/F/Y/V)Kme3(S/T). The discovery of this binding motif led to the identification of a series of peptide sequences from the methyltransferase, SETDB1, which contain one of these consensus binding sequences. Each of these sequences binds to the chromodomain of CBX7 even more tightly than either H3K9me3 or H3K27me3<sup>68</sup>. The success of these studies in finding more potent peptides that interact with CBX7 than either H3 substrate, in addition to the “druggability score of CBX7”<sup>28</sup>, suggested that CBX7 is a tractable target for discovery efforts.



**Figure 2.1.** Comparison of the apo, H3K9me3-bound and H3K27me3 bound structure of CBX7.

Herein, we describe our approach toward the development of potent antagonists of the chromodomain of CBX7. Our initial efforts sought to identify a small molecule based lysine mimetic targeting the CBX7 aromatic cage that proved unsuccessful. This led us to pursue optimization of peptidic antagonists derived from one of the SETDB1 peptide sequences found to bind CBX7 more potently than both H3K9me3 and H3K27me3. Our first focus was to identify the shortest sequence capable of binding to CBX7. Next, we sought to replace the trimethyl ammonium moiety of this sequence. In order to increase the potency of our peptidic ligands, we utilized molecular dynamics simulations to guide our discovery efforts, culminating in the development of UNC3866, the most potent CBX7 antagonist published to date<sup>31, 38-40, 97</sup>. Finally, UNC3866 was derivatized to produce two additional tool compounds. The first compound, UNC4219, is a negative control compound that differs only by a single methyl group and displays no interaction with CBX7 at concentrations up to 100  $\mu$ M. The second tool compound, UNC4195, is a biotinylated analog of UNC3866 that retains high affinity for CBX7 that was synthesized to serve as an affinity reagent for use in biochemical and cell-based experiments.

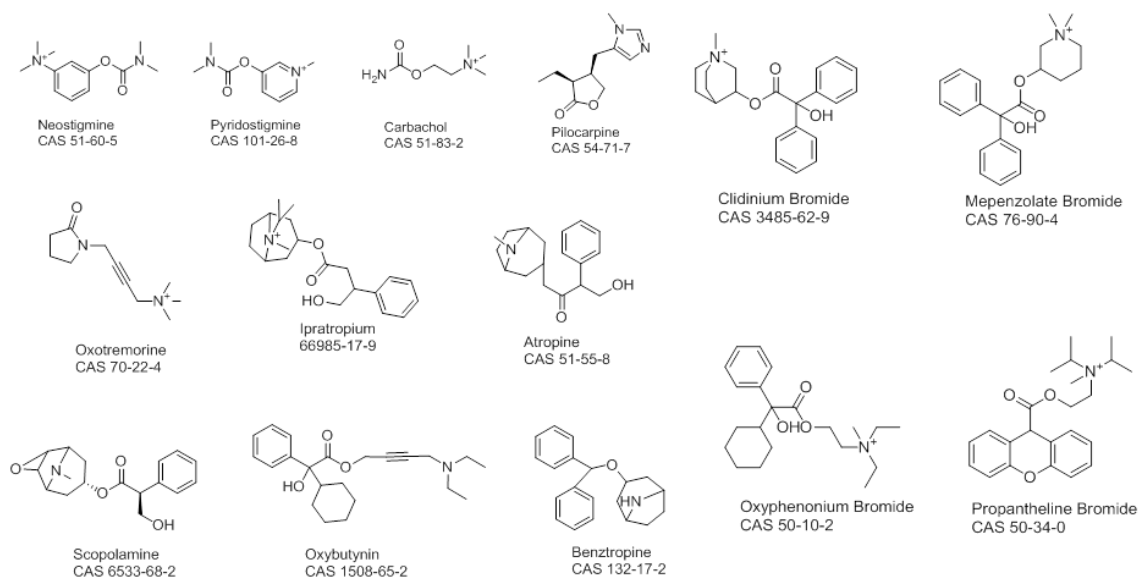
At the onset of the research described in this chapter, there were no reports of antagonists targeting the chromodomain of CBX7. During the course of the work described in this chapter, two studies were published from different laboratories detailing their efforts toward the development of CBX7 antagonists<sup>38, 40</sup>. The first study described a peptide SAR approach, similar to the one described in this chapter, focused around the ALKme3S consensus sequence that was identified from SETDB1<sup>38, 68</sup>. This study identified a peptide with a  $K_d$  of ~200 nM for the chromodomain of CBX7. No studies were reported that extensively detailed the selectivity of this peptide, nor its effects in cell-based assays<sup>38</sup>. A second study was published claiming to identify a weak, small molecule antagonist of the chromodomain of CBX7. SAR studies around this molecule proved ineffective for improving its potency and while cell-based experiments are described, these results are of questionable statistical significance<sup>40</sup>. Our own ITC studies with this molecule and a subsequent small molecule from the same group also reported to interact with CBX7<sup>97</sup> failed to confirm the ability of either molecule to bind to the chromodomain CBX7 *in vitro*.

## RESULTS AND DISCUSSION

### *Screening for small molecule antagonists of CBX7*

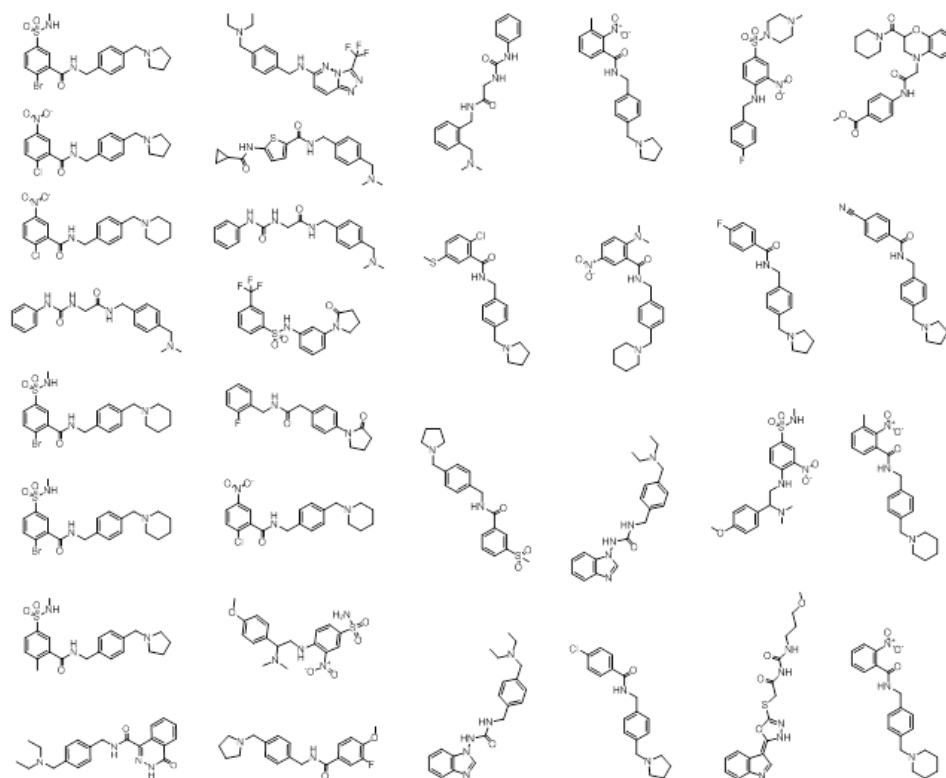
Previously, we screened a library of ~2000 unique small molecules that were synthesized in-house against the chromodomain of CBX7. These compounds had been designed to target multiple protein families and several hundred of these compounds were antagonists of the methyl-lysine reader domains of L3MBTL1, L3MBTL3 and/or 53BP1. Somewhat surprisingly, none of these molecules were inhibitory of the CBX7-H3 interaction when tested at concentrations up to 30  $\mu$ M. We decided to pursue identification of small molecule CBX7 antagonists instead of exploring peptidic inhibitors initially because of the known challenges of developing peptides that antagonize intracellular targets. All peptides that had been shown to interact with CBX7 were minimally 9 amino acids long, putting their molecular weight at ~1000 Da or more<sup>68, 94</sup>. In general, there is a negative correlation between molecular weight and passive membrane permeability. Further, each amide bond within a peptide endows the molecule with two heteroatoms, a carbonyl as a hydrogen bond acceptor and an amide nitrogen, which acts as a hydrogen bond donor. This greatly increases the polar character of the molecules, further reducing their permeability<sup>98</sup>. Additionally, to achieve single-digit micromolar affinity for CBX7, initial peptides identified in the literature required the presence of a trimethylated-lysine residue<sup>68, 94</sup>. The presence of a trimethylammonium is likely to be problematic in biological systems because the permanent cationic nature of the molecule makes it susceptible to trapping in both lysosomes and endosomes while also presenting a metabolic liability in that it can be inactivated by lysine demethylases<sup>99-101</sup>. Therefore, we elected to first pursue approaches focused on identifying small-molecules capable of antagonizing CBX7.





**Figure 2.2.** Structures of commercially available antagonists for quaternary ammonium binding proteins tested against CBX7. The  $IC_{50}$  for all of these compounds was determined to be  $>30\ \mu M$ .

We reasoned that, because CBX7 is a reader of trimethylation with limited-to-no affinity for mono- and di-methylation, our in-house library of Kme antagonists were not active because their reader domain targets were all readers of mono- and dimethylation. To address this, we selected a set of compounds known to antagonize either muscarinic or nicotinic receptors, both of which recognize quaternary amines, with the hope that they may have off-target activity for the Kme3 binding region of CBX7. Disappointingly, none of these compounds displayed antagonism of the CBX7-H3 interaction in AlphaScreen® (**Fig. 2.2**). We simultaneously performed computational docking studies of the 100,000 compound UNC CIBDD diversity library with CBX7 in an effort to find a useful small molecule starting point. The top scorers were narrowed to a list of 32 compounds for screening against CBX7. Despite the fact that many of these compounds contained known methyl-lysine mimetics, none of these showed evidence of CBX7 antagonism at concentrations up to  $30\ \mu M$  (**Fig. 2.3**).



**Figure 2.3.** Structures of top compounds from docking studies with CBX7 and the CICBDD diversity library. IC<sub>50</sub> values for all compounds were >30  $\mu$ M.

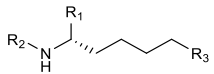
Our lack of success in identifying any small molecules capable of interacting with CBX7 prompted us to revisit a peptide-based approach. As discussed previously, there are many drawbacks to pursuing peptidic antagonists; however, peptides are the only class of molecules known to interact potently ( $K_d \sim 5$   $\mu$ M) with the chromodomain of CBX7. Therefore, in spite of their drawbacks, we performed a series of SAR studies targeting CBX7 using peptidic ligands. We elected to perform these studies around the RGFALKme3STHG peptide sequence identified from SETDB1. We chose this peptide because its published potency ( $K_d = 5$   $\mu$ M) was considerably better than both H3K9me3 ( $K_d = 55$   $\mu$ M) and H3K27me3 ( $K_d = 110$   $\mu$ M) peptides<sup>68</sup>. Additionally, its lack of an arginine at the (-1) position from the Kme3 and the presence of a UV chromophore from the phenylalanine made it synthetically more attractive than other peptide sequences. Based on the aforementioned problems with peptidic antagonists, we focused first on determining the shortest possible peptide capable of antagonizing the CBX7-H3 interaction in an effort to increase the passive permeability of any derivative molecules by decreasing the size of the peptide and


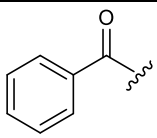
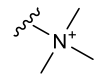

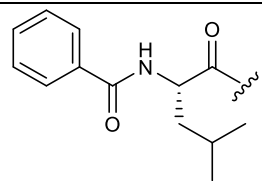
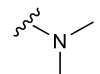

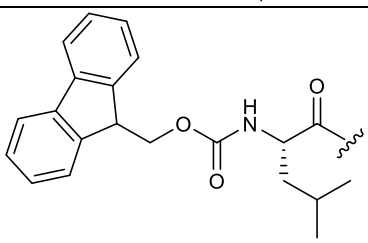
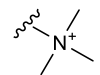

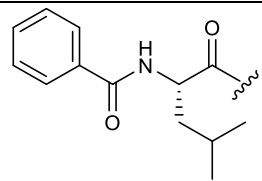
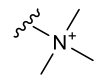

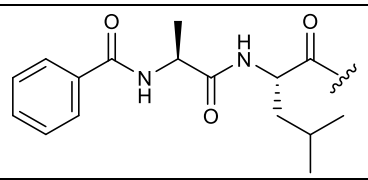
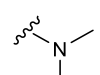

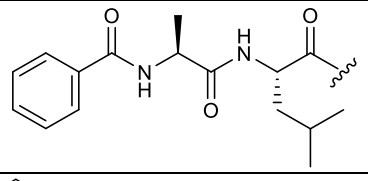
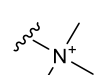

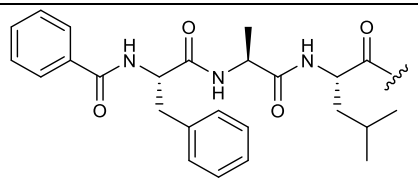
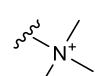

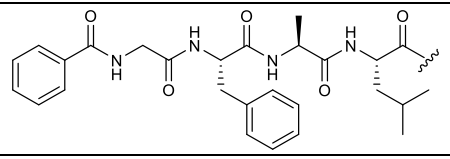
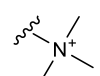
reducing the number of hydrogen bond donors and acceptors<sup>98</sup>. Our second goal for this first SAR study was to determine whether any of the backbone amides of these peptides were dispensable for interaction with CBX7. Our final goal for this first study was to replace the Kme3 of our peptides with a non-quaternary amine functionality.

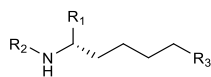
#### *Identification of the minimal binding sequence required for CBX7 antagonism*

In order to determine the minimal sequence necessary for binding to the chromodomain of CBX7. We focused on whether, in the context of shorter peptides, the entire consensus motif (A(R/I/L/F/Y/V)Kme3(S/T)) is needed for binding. All molecules synthesized were capped at the N-terminus with a UV chromophore to aid in purification (**Table 2.1**, compounds **1-18**; for compounds **3** and **10**, the Fmoc protecting group was used as the chromophore to determine if it could pick up any additional interactions to aid in binding). As **Table 2.1** shows, antagonism of the CBX7-H3 interaction requires the entire ALKmeS motif plus an additional residue at the N-terminus; in this context, Phe. Compounds **1-14** were designed to simultaneously interrogate the requirement of the C-terminal serine and the lysine C-terminus. The lack of activity for any of these compounds suggests that additional interactions C-terminal to the Kme3 side chain are required for CBX7 interaction. Compounds **15-18** were synthesized to capture the full ALKme3S motif as well as up to 2 additional residues at the N-terminus as previous work suggested that these residues may be important for binding<sup>68</sup>. Only in the case of 5-residue peptide or longer was antagonism of the CBX7-H3 interaction observed (compounds **17-18**).

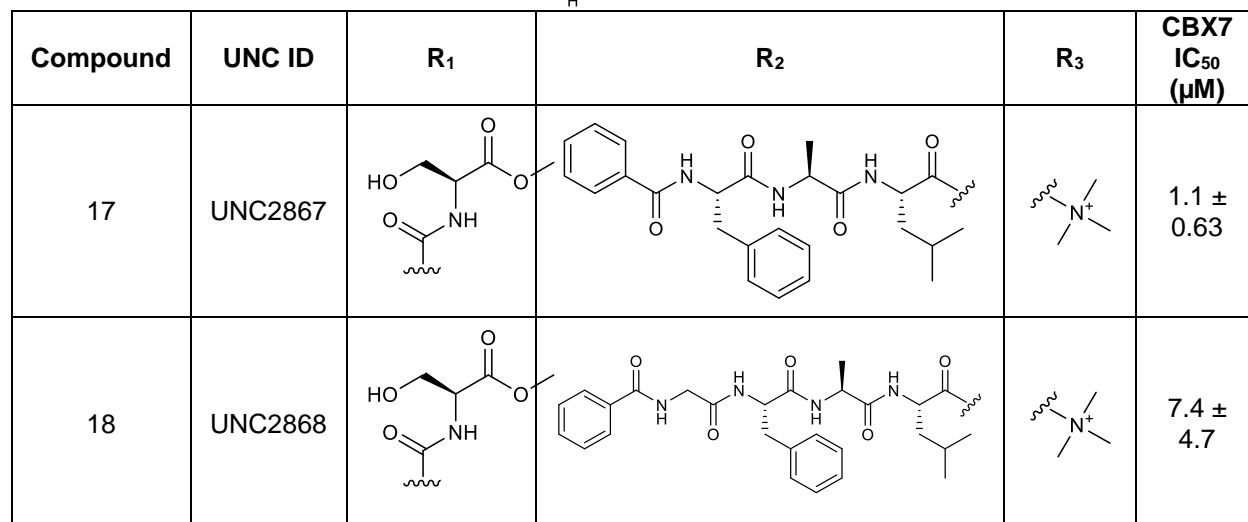
**Table 2.1** Exploration of sequence and length requirements for interaction with the CBX7 chromodomain



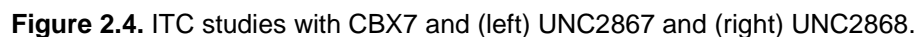
Compound	UNC ID	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CBX7 IC <sub>50</sub> (μM)
1	UNC2404				>30
2	UNC2403				>30
3	UNC2339				>30
4	UNC2401				>30
5	UNC2720				>30
6	UNC2588				>30
7	UNC2934				>30
8	UNC2994				>30



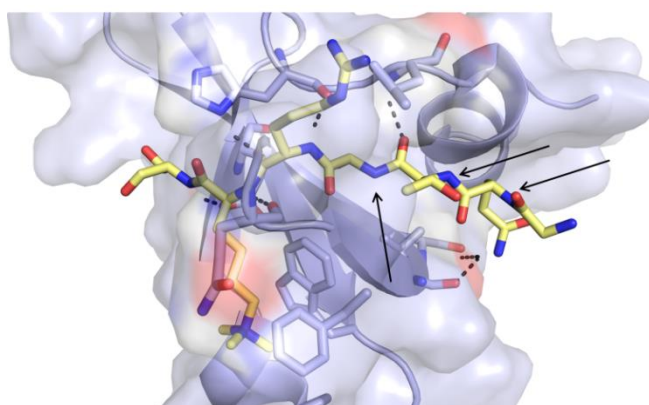
Compound	UNC ID	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CBX7 IC <sub>50</sub> (μM)
9	UNC2746				>30
10	UNC2338				>30
11	UNC2402				>30
12	UNC2745				>30
13	UNC2747				>30
14	UNC2597				>30
15	UNC2596				>30
16	UNC2864				>30



Compounds **17** and **18** were evaluated by ITC for biophysical confirmation of their binding and characterization of the thermodynamic parameters of the interaction (**Fig. 2.4**). While binding was observed for both compounds, only UNC2868 (compound **18**) displayed saturable binding to CBX7 that could be reliably quantified,  $K_d = 10.3 \pm 0.9 \mu\text{M}$ . While it is unknown why UNC2868 displays better binding by ITC than UNC2867 despite appearing weaker in AlphaScreen®, it was nonetheless encouraging that we could validate these molecules as CBX7 antagonists in ITC. We elected to utilize UNC2868 as a scaffold for future SAR studies because of its robust ITC results.

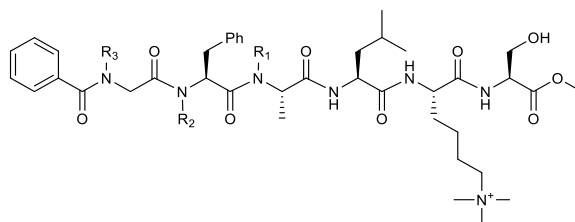


We next explored the importance of the backbone of UNC2868 for interaction with CBX7. To do this, we chose to *N*-methylate select amides of UNC2868 based on the CBX7-H3K9me3 NMR structure (**Fig. 2.5**, PDB 2L12). Based on this structure, we reasoned that the N-H's of the Ser, Kme3 and Leu residues were essential for binding. Therefore, we explored the effect of N-methylating the Ala, Phe and Gly or UNC2868 (**Table 2.2**). As can be seen in **Table 2.2**, methylation of both the Ala and Phe residues causes a loss of detectable binding at concentrations up to 30  $\mu\text{M}$ . On the other hand, N-methylation of the glycine residue had no significant effect on binding when compared to the unmethylated analog, suggesting that this amide was dispensable for binding.



**Figure. 2.5.** Hydrogen bonding between CBX7 and H3K9me3. Hydrogen bonds are shown as black dashes. The arrows indicate which corresponding amides selected for methylation on UNC2868. PDB 2L12

**Table 2.2** Effect of select backbone methylations of UNC2868 on binding to CBX7



Compound	UNC ID	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CBX7 ITC <i>K<sub>d</sub></i> ( $\mu\text{M}$ )
19	UNC3261	Me	H	H	>30
20	UNC3375	H	Me	H	>30
21	UNC3403	H	H	Me	11 $\pm$ 3.0

Data are reported as the average  $\pm$  the standard deviation of the mean ( $n \geq 2$ ).

In parallel with the studies in **Table 2.2**, we performed an SAR study around UNC2868 to find a suitable replacement for the Kme3 residue. Quaternary amines are problematic because they are susceptible to endosomal and lysosomal trapping<sup>99, 100</sup>, and can serve as substrates for demethylase enzymes<sup>101</sup>. We therefore explored the tolerance of the aromatic cage of CBX7 for various non-Kme3 containing peptides in the context of UNC2868. We explored secondary, tertiary, non-basic and di-basic compounds to help identify non-quaternary ammonium species that can antagonize the chromodomain of CBX7 (**Table 2.3**).

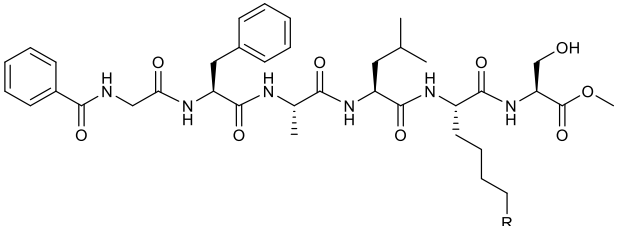
We first synthesized compounds **22-25** as controls for our AlphaScreen® assay. These compounds behaved as expected with unmethylated (**22**), acetylated (**23**) and benzoylated (**24**) analogs displaying no evidence of interaction in the conditions of the assay and reduction of the methylation state to dimethyl (**25**) resulting in loss of potency. We next explored a series of bulkier secondary amines (**26-31**), including one di-basic compound (**31**). In general, none of these modifications were well tolerated by CBX7. Simple isopropylation (**26**) showed weak antagonism, but increasing the size of the alkyl substituent proved detrimental for binding (**27** and **28**). Interestingly, fusion of the cyclopentyl and cyclohexyl groups of these two compounds into a norbornyl compound (**29**) restored activity of the compound to levels similar to the isopropyl derivative (**26**). Further increasing the lipophilicity of the ring system to an adamantyl derivative (**30**) abolished activity in our assay.




The highly lipophilic nature of the norbornyl moiety of **29** led us to question the importance of the cation- $\pi$  interaction between this compound and the aromatic cage of CBX7. If the norbornyl moiety is responsible for the interaction with the aromatic cage residues, then there would be little to no cation- $\pi$  interactions contributing to binding. To investigate this, we synthesized the di-basic **31** in an attempt to recover any lost cation- $\pi$  interactions. While this compound contains an additional bridge carbon, we hypothesized that the addition of a second basic amine could compensate for the increased steric bulk of the compound. Surprisingly, **31** displayed no activity in our assay at concentrations up to 100  $\mu$ M. The lack of activity for this compound prompted us to remove the basicity of the amine through simple acetylation (**32**). Surprisingly, this compound displayed slightly improved affinity for CBX7 compared to **29**. Unfortunately, **32** has limited aqueous solubility, precluding validation of its binding to CBX7 by ITC. Nonetheless, taken together with compounds **24-25**, this data suggests that the presence of basic

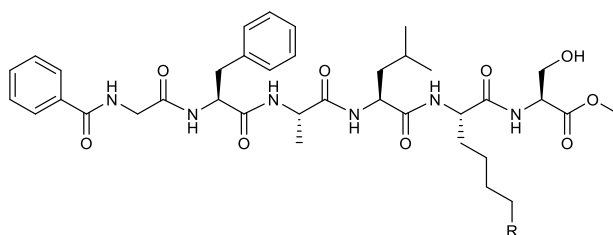


nitrogen is dispensable for interaction with the aromatic cage of CBX7 if sufficient lipophilic interactions are present on the molecule to compensate for the loss of the cation- $\pi$  interaction. However, compound **33** suggests that the presence of a basic nitrogen is preferred as methylation of **29**, rather than acetylation (**32**) resulted in a compound approximately equipotent to the trimethyl-lysine of UNC2868. Encouraged by **33**, we explored other tertiary amines at this position (**34-38**). Interestingly, while considerably less lipophilic than the methyl, norbornyl-lysine of **33**, the diethyl-lysine of **35** was found to be equipotent for CBX7. Restriction of the terminal methyls of this compound in the form of the pyrrolidinyl-lysine of **36** reduced binding approximately 3-fold. Affinity was only recovered when the ring size was increased to 7-atoms (**38**). Finally, we synthesized the more strongly basic **39** and di-basic compound **40** to test whether these could form improved and/or additional cation- $\pi$  interactions with the aromatic cage of CBX7. While the activity of these compounds was detectable in our assay, it was considerably weaker than the Kme3 of UNC2868. Taken together, this data indicate that presence of a Kme3 is not required for interaction with CBX7 and that the loss of the quaternary ammonium of this residue can be compensated for by tertiary amines with certain lipophilic functionalities, but not through additional or improved cation- $\pi$  interactions.

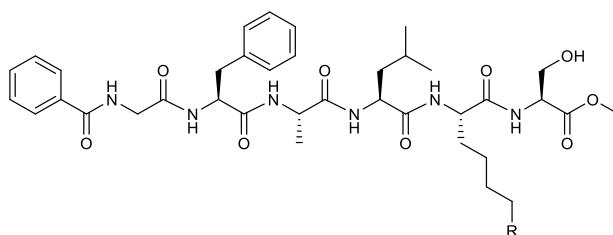
**Table 2.3** Exploration of Kme3 replacements for UNC2868



Compound	UNC ID	R	CBX7 IC <sub>50</sub> (μM)
18	UNC2868		7.4 ± 4.7
22	UNC3189		>100
23	UNC3267		>100



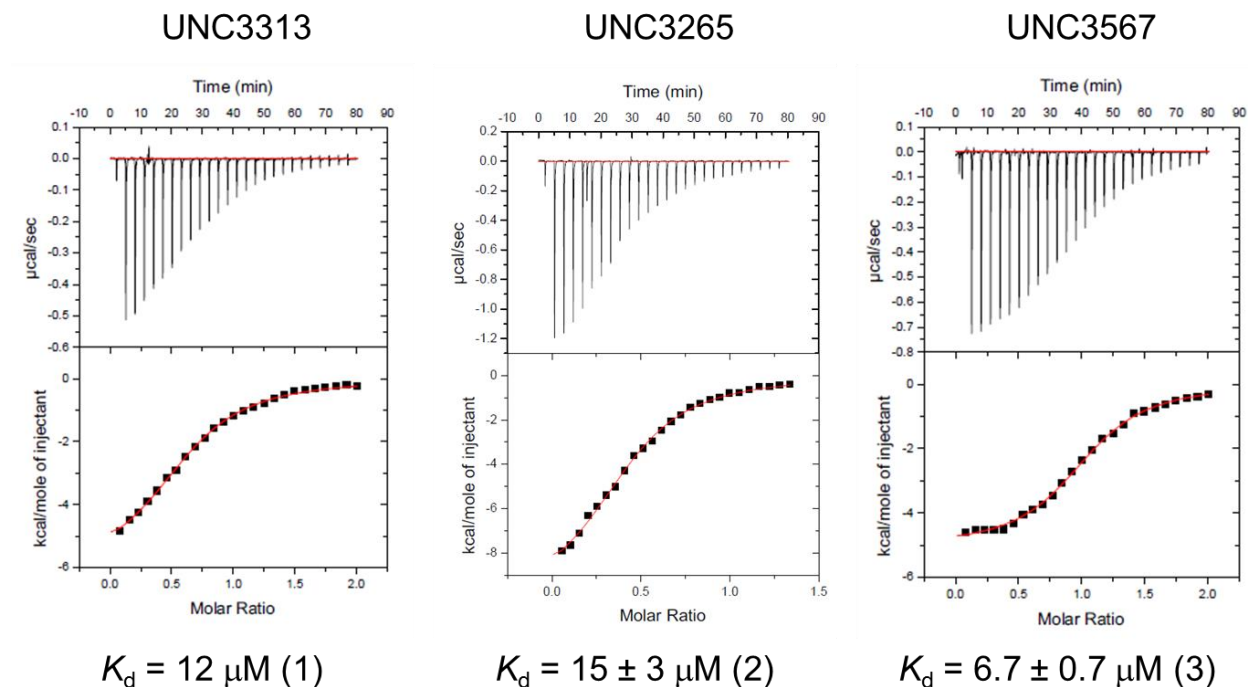
Compound	UNC ID	R	CBX7 IC <sub>50</sub> (μM)
24	UNC3546		>100
25	UNC3188		24 ± 0.53
26	UNC3563		34 ± 1.2
27	UNC3315		51 ± 3.8
28	UNC3314		>100
29	UNC3265		37 ± 2.5
30	UNC3264		>100
31	UNC3266		>100
32	UNC3649		17 ± 8
33	UNC3313		3.9 ± 0.3
34	UNC3260		>100



Compound	UNC ID	R	CBX7 IC <sub>50</sub> (μM)
35	UNC3567		4.0 ± 0.6
36	UNC3580		13 ± 2
37	UNC3581		14 ± 2
38	UNC3648		4.0 ± 0.4
39	UNC3564		41 ± 8
40	UNC3566		52 ± 4.2

Data are reported as the average ± the standard deviation of the mean (n ≥ 2).

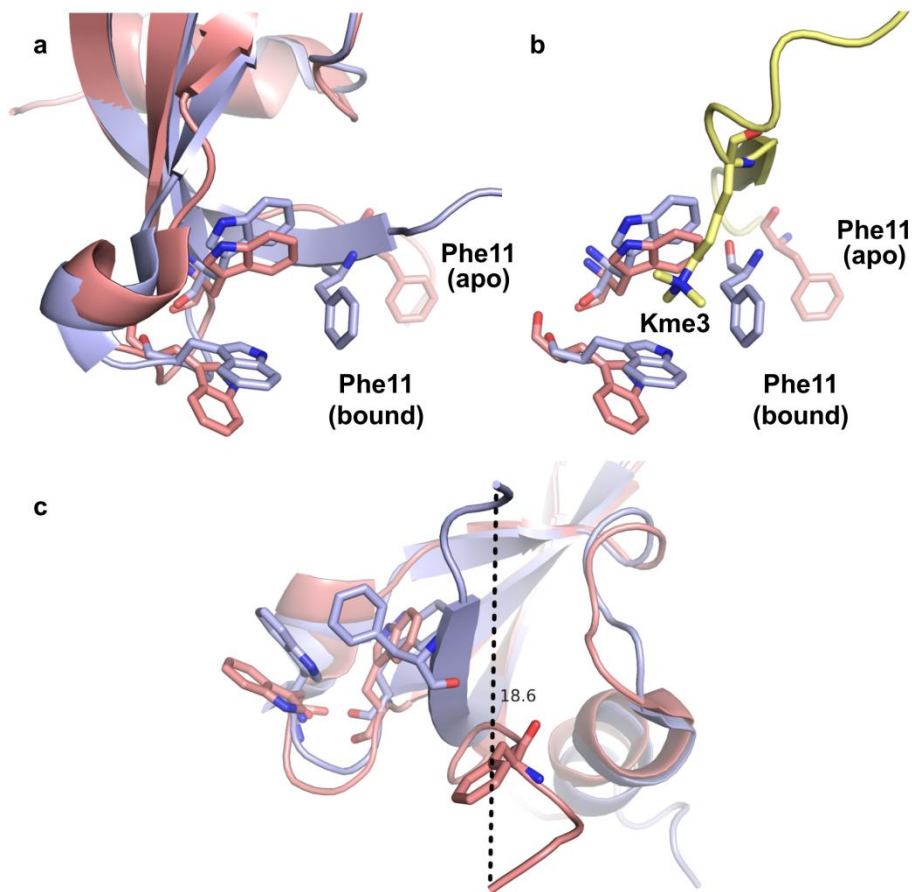
Our studies in **Table 2.3** presented us with 3 suitable replacements for the Kme3 of UNC2868; methyl, norbornyl (**33**, UNC3313), diethyl (**35**, UNC3567), and azepane (**38**, UNC3648). The synthesis of UNC3648 was considerably less efficient than both UNC3313 and UNC3567. Therefore, we elected to forgo this modification in favor of the other two for follow up SAR studies. Binding of **33** (UNC3313); the unmethylated analog of **33**, **29** (UNC3265); and **35** (UNC3567) was confirmed by ITC (**Fig. 2.6**). Interestingly, despite the different activities of UNC3313 and UNC3265 in AlphaScreen®, these compounds were approximately equipotent by ITC (12 μM vs 15 μM, respectively). UNC3567 showed the most potent binding by ITC, though only by approximately 2-fold over UNC3265 and UNC3313. The diethyl-lysine of UNC3567 and the norbornyl of UNC3265 were chosen as methyl-lysine replacements for subsequent analogs because of their validated, similar binding affinities and synthetic ease.



**Figure. 2.6.** ITC studies between CBX7 and peptides with various Kme3 replacements. Numbers in parentheses indicate the number of replicates for each experiment. Data are reported and the average  $\pm$  the standard deviation of the mean.

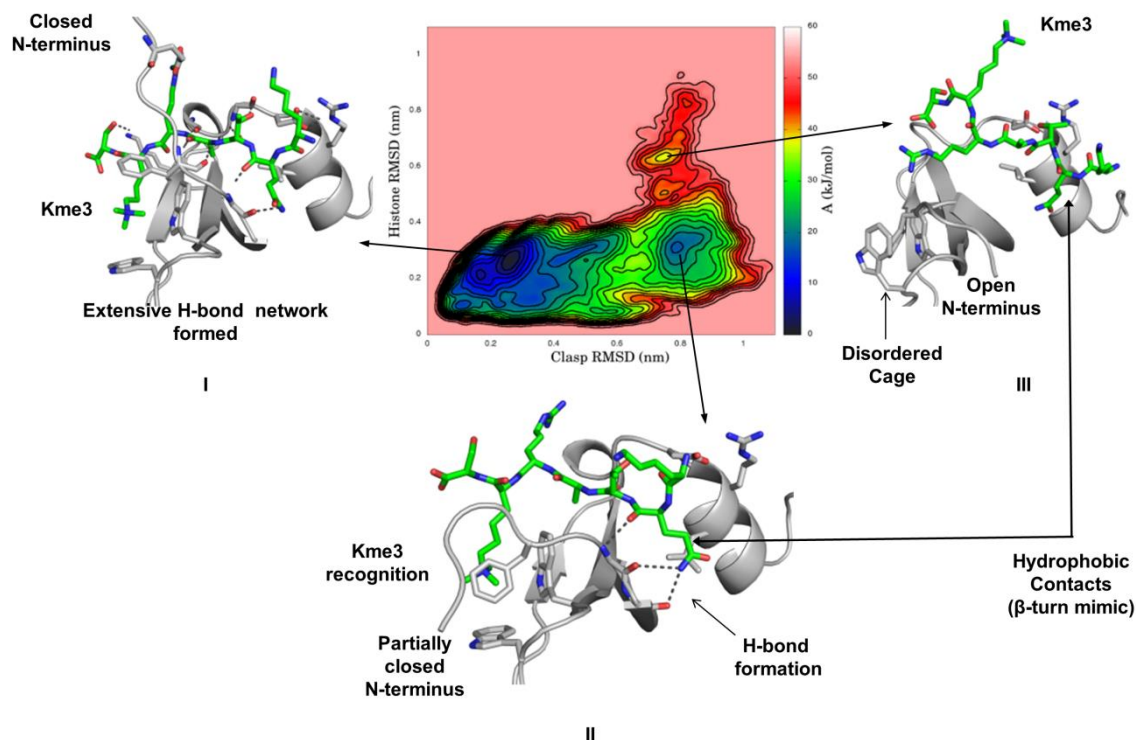
While we were encouraged by our ability to replace the Kme3 of UNC2868, these efforts did not afford any significant improvements in binding affinity for CBX7. We therefore pursued a series of SAR studies focused on improving the potency of our target compounds. Initially, we were intrigued by the ability of the norbornyl functionality of UNC3265 to bind in the absence of any cation- $\pi$  interactions, therefore, we elected to use this moiety as our initial Kme3 replacement. We first started by attempting to reduce the peptidic nature and size of our compounds. Our backbone methylation studies in **Table 2.2** indicated that the N-terminal amide of our antagonists was dispensable for binding. Compound **41** (**Table 2.4**), which does not contain this amide, was therefore synthesized and found to show stronger activity than its parent compound (UNC3265, **29**). While this was an improvement, we were still unable to accomplish submicromolar potencies for our compounds. The extensive amino acid SAR that had been done previously<sup>68</sup> suggested that the core amino acid sequence (FALKme3S) was largely optimized, prompting us to seek out an alternative approach for accessing higher potency compounds.

We were intrigued by the conformational rearrangement that CBX7 has to undergo to organize its aromatic cage in order to allow for Kme3 recognition (**Figures 2.1** and **2.7**). CBX7 has to rearrange in such a way that the N-terminus of the chromodomain is displaced approximately 18.6 Å (**Fig. 2.6c**) in order to position Phe11 in close enough proximity to Trp32 and Trp35 to form a functional aromatic cage. We reasoned that a physical understanding of this process would aid in the design of more potent compounds.



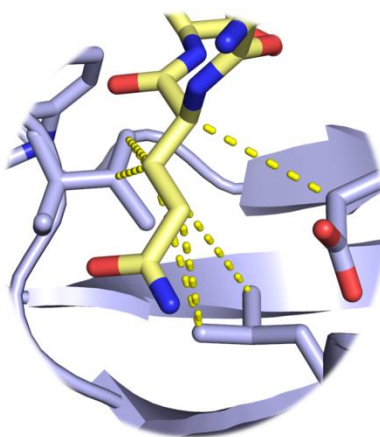
**Figure 2.7.** Comparison of CBX7 in the bound and apo conformations. **(a)** Comparison of the backbone conformations and aromatic cage of CBX7 in the apo (salmon) and H3-bound (light blue) conformations, histone not shown. **(b)** Comparison of the CBX7 aromatic cage in the apo (salmon) and bound (light blue) conformations. H3K9me3 is shown in pale yellow. **(c)** Measurement of the displacement of the N-terminus of CBX7 during binding to H3K9me3. (PDB 2L12)

We therefore turned to adaptively biased molecular dynamics simulations using the previously published NMR structure of CBX7 bound to H3K9me3 (PDB 2L12). We computed possible histone exit paths from this structure and the pathway shown in **Fig. 2.8** was favored when we considered a truncated form of H3K9me3 (residues 4-10), which was most analogous to our peptidic antagonists. Starting from the bound state as shown in **Fig. 2.8**, Frame I, the N-terminus of CBX7 opens by losing partial contact with H3 (transition from Frame I to II). Subsequently, the Kme3 loses contact with the aromatic cage (transition from Frame II to III), leaving H3 engaged with CBX7 primarily through hydrophobic contacts made with V13, D50 and L53 (Frame III). Complete dissociation of H3 and CBX7 occurs following the loss of these contacts.



**Figure 2.8.** Molecular dynamics simulation of the CBX7-H3K9me3 interaction. Free energy plot depicting the calculated binding pathway of CBX7-H3K9me3. Arrows are drawn from potential energy wells to representative structures associated with that well. Frames I-III represent one possible dissociative path of H3 from CBX7, and the reverse of these frames represents the corresponding associative path of CBX7 and H3. The N-terminus label is referring to the N-terminus of CBX7.

Association of CBX7 and H3 is simply the microscopic reversal of this exit path and is analogous to  $\beta$ -hairpin folding between two anti-parallel  $\beta$ -strands. In this case, one  $\beta$ -strand is formed by the histone; the other by residues 8-13 of CBX7. Therefore, establishing the appropriate contacts between the N-terminal portion of the H3 peptide (residue Q5, **Fig. 2.8**, Frame III and **Fig. 2.9**) and CBX7 (namely residues V13, D50 and L53) during association is analogous to forming the critical  $\beta$ -turn during hairpin folding. Once the turn is formed, backbone hydrogen bonds 'zip' the  $\beta$ -hairpin into the fully folded state. In the case of CBX7, in addition to the formation of an extensive hydrogen bond network, this also orients Phe11 of CBX7 to contact the Kme3 and facilitates formation of the aromatic cage. We utilized this pathway for H3 binding as a model for the association of CBX7 with our antagonists and hypothesized that optimizing the N-terminal interactions of these compounds with CBX7 would facilitate the necessary hydrogen bonding for closure of CBX7 around the molecules, thereby increasing ligand affinity.

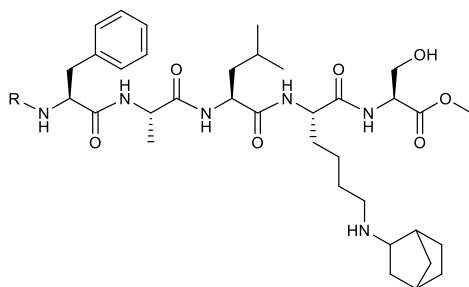


**Figure 2.9** Hydrophobic contacts made by H3Q5 during association of CBX7 and H3K9me3 peptides

Compounds **42-44** were synthesized to begin interrogating our binding hypothesis from the MD simulations. Compounds **42** and **43** were synthesized to mimic the hydrogen bond donation functionality of Gln5 of H3. Neither of these compounds showed significantly stronger interaction with the chromodomain of CBX7 than **41**. However, replacement of the phenyl ring of **41** with a naphthyl moiety (**44**) led to a sub-micromolar antagonist in AlphaScreen®. Unfortunately, **44** exhibits minimal solubility in aqueous buffers, precluding characterization of its interaction with CBX7 in ITC. Nonetheless, these results supported our binding hypothesis derived from the MD simulations wherein hydrophobic interactions at this position are important for nucleating clasp closure and entrance into the bound state.

Further, this data indicates that the hydrophobic contacts are amenable to optimization in order to approve the affinity of our antagonists for CBX7.

**Table 2.4.** Initial N-terminal modifications of a norbornyl-lysine containing CBX7 antagonist



Compound	UNC ID	R	CBX7 IC <sub>50</sub> (μM)
41	UNC3486		10 ± 3
42	UNC3562		5.4 ± 2
43	UNC3565		9.8 ± 0.6
44	UNC3556		0.75 ± 0.01

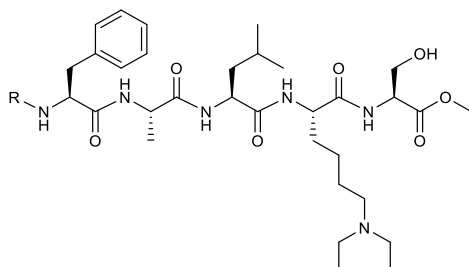
Data are reported as the average ± the standard deviation of the mean (n ≥ 2).

The insoluble nature of **44** prompted us to reconsider the use of norbornyl-lysine as our Kme3 replacement. The need for N-terminal hydrophobic functionalities as suggested by our data in **Table 2.4** indicated that use of a more soluble modified lysine derivative would be important for allowing characterization of our compounds with CBX7 by ITC. Therefore, we chose to use diethyl-lysine as our Kme3 mimetic for future analogs. Compound **45**, which is the diethyl analog of **44** was synthesized and found to be approximately equipotent to **44** and still insoluble, again preventing its characterization by ITC. Compounds **46** and **47**, which each contain an additional heteroatom, were synthesized with the intention of increasing the solubility of this series of compounds. While the indole of **46** was equipotent to naphthyl (**45**), insertion of the additional methylene group in **47** reduced binding approximately 2-3 fold.

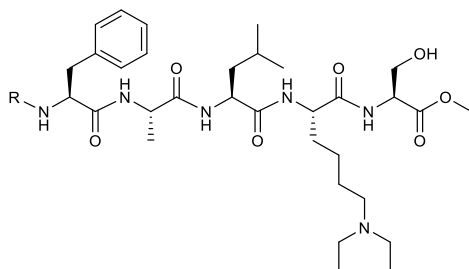


While we do not expect the presence of the primary amine to have any effect on binding based on our studies in **Table 2.2**, we cannot completely rule out this possibility.

**Table 2.5.** N-terminal modifications of a diethyl-lysine containing CBX7 antagonist



Compound	UNC ID	R	CBX7 IC <sub>50</sub> (μM)	CBX7 K <sub>d</sub> (ITC, μM)
45	UNC3772		0.56 ± 0.3	N/D
46	UNC3771		0.56 ± 0.2	N/D
47	UNC3773		1.4 ± 0.4	3.4 (1)
48	UNC4973		N/D	5.2 ± 1
49	UNC4974		N/D	0.33 ± 0.005
50	UNC3867		0.62 ± 0.6	0.22 ± 0.03
51	UNC3769		0.086 ± 0.02	0.14 ± 0.02
52	UNC3865		0.13 ± 0.07	0.16 ± 0.008



53	UNC3866		$0.064 \pm 0.01$	$0.097 \pm 0.002$
54	UNC4975		N/D	$0.24 \pm 0.01$
55	UNC3768		$0.11 \pm 0.09$	N/D
56	UNC3770		$0.22 \pm 0.1$	N/D

Data are reported as the average  $\pm$  the standard deviation of the mean ( $n \geq 2$  except where noted in parentheses).

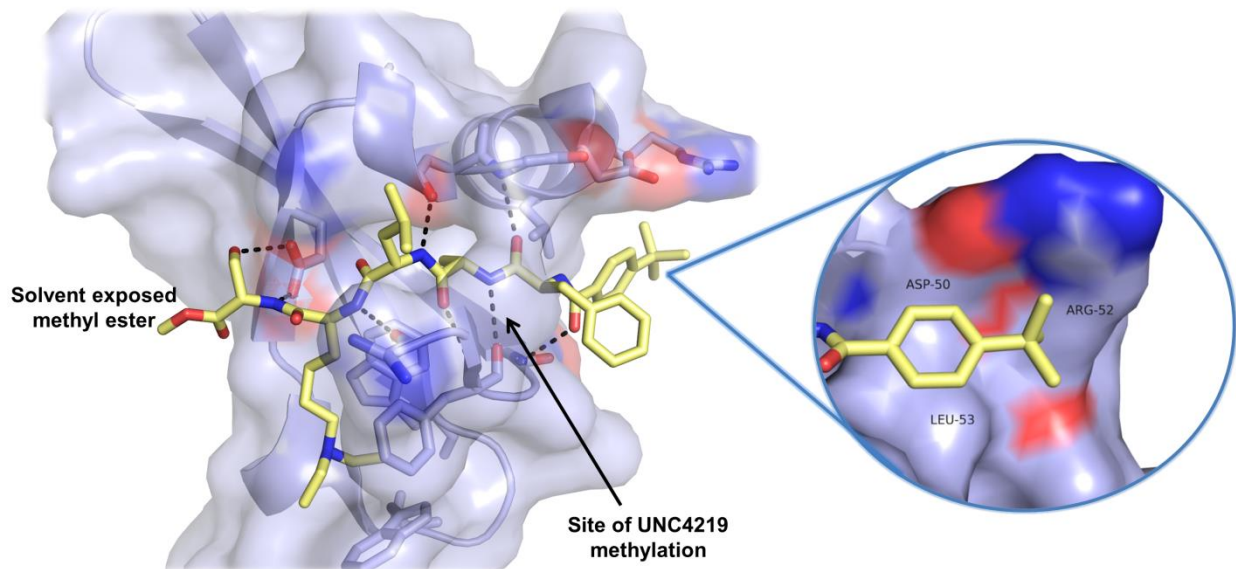
The apparently detrimental effect of increasing the spacing between the N-terminal amide of these molecules and the aromatic ring system prompted us to pursue a series of compounds which eliminated any methylene spacers. Further, the insolubility of the bicyclic ring systems prompted us to pursue alternative hydrophobic functionalities. Our studies in **Table 2.1** indicated that a simple benzoyl cap (**17**) is insufficient for sub-micromolar antagonism of CBX7. Therefore, we performed a short SAR study in which we added a single methyl group at the *ortho*-, *meta*- or *para*- positions to test the effect on binding to CBX7. These compounds proved soluble enough for characterization by ITC. Therefore, we assessed their binding to CBX7 directly using this technique. Placing the methyl group at the *ortho*-position of the benzoyl cap (**48**) showed low micromolar affinity for CBX7. Shifting this methyl group to the *meta*-position (**49**) bestowed sub-micromolar activity to this compound. Shifting to the *para*-position proved to be the optimal position for this alkyl substituent. Consistent with our MD hypothesis, increasing

the lipophilicity of the alkylation at this position (**50-53**) further improved the potency with *tert*-butyl substitution at the *para*-position (**53**, **UNC3866**) providing our most potent antagonist with a  $K_d$  of 97 nM. Encouragingly, we saw consistent trends in both AlphaScreen® and ITC for compounds **50-53**. While the data for compounds **48-53** was consistent with our hypothesis that binding at this portion of the molecule is driven primarily by van der Waals' interactions, we sought to further investigate this by tuning the electronics of the benzoyl cap (compounds **54-56**). The *tert*-butyl of UNC3866 is weakly electron-donating, therefore, we replaced this functionality with a more strongly electron-donating methoxy-functional group. This resulted in an approximately 2.5-fold reduction in binding affinity as determined by ITC. The electron-withdrawing functionalities of **55** and **56** were also detrimental to binding. Particularly interesting is that the trifluoro-methyl group of **56** reduced binding when compared to the isopropyl of **52**. This was surprising because trifluoro-methyl and isopropyl groups are approximately the same size and hydrophobicity. Taken together, this data suggests that the *tert*-butyl of UNC3866 is optimal for interaction with CBX7 in terms of the degree and positioning of alkylation as well as stereoelectronics of the aromatic ring.

In collaboration with Jinrong Min at the Structural Genomics Consortium in Toronto, we solved the X-ray co-crystal structure of UNC3866 bound to CBX7 (PDB 5EPJ) at a resolution of 1.6 Å to improve our understanding for the potency of UNC3866 for CBX7 (**Fig 2.10**). Consistent with our data in **Table 2.1** and **Table 2.2**, each backbone amide of UNC3866 participates in at least one hydrogen bond with the backbone of CBX7. The induced fit binding mechanism coupled with the extensive network of hydrogen bonds likely accounts for the length requirements of our peptidic antagonists as well as our lack of success in discovering small molecule antagonists of the CBX7 chromodomain. The N-terminal rearrangement of CBX7 to form the aromatic cage is likely paid for energetically by the formation of these hydrogen bonds. Without these, the aromatic cage is too disorganized to form significant interactions with shorter peptides and small molecules.

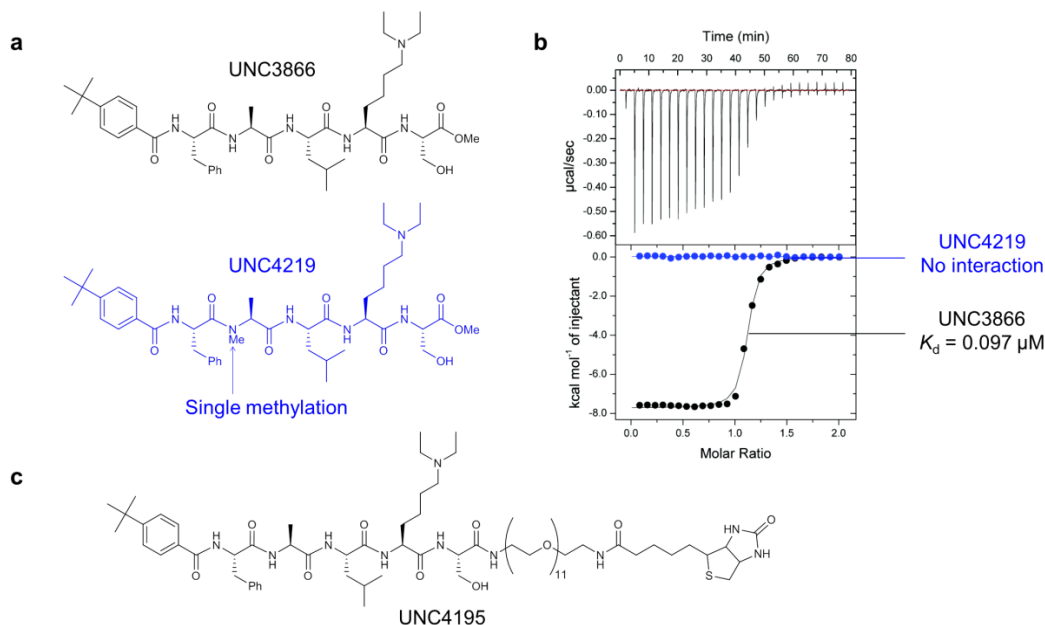
The N-terminal *tert*-butyl benzoyl cap of UNC3866 primarily contacts the side chains of D50, R52, and L53 (**Fig. 3c**). The guanidinium moiety of R52 forms a salt bridge with D50 and this rigidified R52 side-chain forms a hydrophobic groove with L53 within which the *tert*-butyl phenyl group of UNC3866 is embedded. Based on our MDS studies, we suspect that the 70-fold affinity increase observed upon

substitution of our initial benzoyl-glycyl cap, present on UNC3567, for the *tert*-butyl benzoyl cap of UNC3866 is the result of an improvement in the ability of this moiety to nucleate closure of CBX7 around the molecule through interactions with D50, R52 and L53.



**Figure. 2.10.** X-ray co-crystal structure of UNC3866 (pale yellow) bound to CBX7 (light blue). PDB code 5EPJ.

We utilized this crystal structure to design additional analogs of UNC3866 to serve as tool compounds for studying CBX Pc proteins and characterizing UNC3866 (**Fig. 2.11**). We first synthesized a negative control analog, UNC4219, which differs from UNC3866 only by *N*<sub>α</sub>-methylation of the Ala residue of UNC3866 and is inactive at concentrations up to 100 μM as assessed by ITC (**Fig. 2.11b**). This modification was chosen based on both the UNC3866-CBX7 co-crystal structure (**Fig. 2.10**) and our SAR studies in **Table 2.2**. Additionally, the crystal structure indicated that the methyl ester of UNC3866 is solvent exposed and not making any significant interactions with the chromodomain. We therefore generated a biotinylated analog of UNC3866, UNC4195, by hydrolyzing the methyl ester and attaching biotin through an amine terminated PEG-linker. This reagent maintains high affinity for CBX7 as determined by ITC ( $K_d = 220 \pm 22$  nM).



**Figure 2.11.** CBX7 chemical tool kit. (a) Structures of UNC3866 and its corresponding negative control, UNC4219. (b) Representative ITC curves for CBX7 with UNC3866 (black) and UNC4219 (blue). (c) Structure of biotinylated affinity reagent analog of UNC3866, UNC4195.

## CONCLUSIONS

These studies have provided valuable insight into the tractability of targeting CBX7. Our studies suggest that small molecules are unlikely to be capable of potently engaging the chromodomain of CBX7. However, it is possible that our studies aimed at identifying traditional small molecule antagonists of CBX7 were simply outside of the appropriate chemical space. However, the induced-fit mechanism of binding to CBX7 described here suggests that peptidic antagonists are required for formation of and interaction with the aromatic cage of CBX7. An alternative approach could be to search for small molecules capable of binding to the apo conformation of CBX7, although currently, there are no known endogenous ligands capable of binding to this conformation, giving little guidance as to which areas of chemical space would be worth exploring to identify such molecules.

While peptides suffer from a number of drawbacks as discussed, our studies have proven that this class of molecules can be modified to produce potent CBX7 antagonists. These studies have shed considerable light on the binding preferences of CBX7 and have identified the key region of CBX7 to

target for improving affinity of for this chromodomain. While UNC3866 is not a drug-like molecule, the purpose of chemical probe discovery efforts is to produce biologically useful molecules. The utility of chemical probes is based primarily on their selectivity and demonstrated ability to modulate their given target within a biological system. Just as first-in-class chemical probes may exhibit limited selectivity, they also can suffer from poor drug-like properties. Nevertheless, these molecules still have the potential for answering important biological question and their development provides the initial chemical scaffold and SAR necessary for ultimately further improving both their selectivity and drug-like properties. Our studies involving further biophysical and selectivity characterization of UNC3866 are discussed in chapter III and our efforts to characterize UNC3866 in cellular and *in vivo* settings are discussed in chapter IV.

## METHODS

### *Adaptively Biased Molecular Dynamics Simulations*

Adaptive biasing potential<sup>102</sup> was implemented in GROMACS 4.5.5<sup>103</sup>. The bias parameters were  $b=0.8$ ,  $c=0.0005/\delta t$ , where  $\delta t=2$  fs is the molecular dynamics time step. The Gaussian width was 1/4 Angstrom. This choice of bias parameters floods stable states extremely slowly, to minimize adaptation of the bias near transition states. Thus, one may argue for approximate state-to-state dynamics as has recently been done for metadynamics<sup>104</sup> following the principles of hyperdynamics<sup>105</sup>. Using the first frame of the CBX7+H3K9me3 NMR structure, two collective variables were defined: RMSD of clasp, and RMSD of peptide. We used the Kabsch algorithm to align the reference and trajectory during the simulation. All simulations were performed in a cubic box with a minimum of 12 Angstroms between the protein or peptide and the nearest cube face, resulting in 23710 atoms after adding water and salt. System construction included 5 nsec of isothermal–isobaric equilibration at 1 bar. Production simulations were carried out in the canonical ensemble with stochastic velocity rescaling<sup>106</sup>. Van der Waals and direct electrostatic interactions used a 10 Å cutoff. Long-range electrostatics were treated with the particle mesh Ewald approach with a grid spacing of 1.6 Å. The neighbor list was updated every five steps. We ran five 128 ns simulations to get some intuition that could inform ligand design. All images from the simulations were rendered using the PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

### *ITC Experiments*

All ITC measurements were recorded at 25 °C with an AutoITC200 microcalorimeter (MicroCal Inc., MA). All protein and compound stock samples were in the target buffer (25 mM Tris-HCl, pH 8, 150 mM NaCl, and 2 mM  $\beta$ -mercaptoethanol), and then diluted in the same buffer to achieve the desired concentrations: 50  $\mu$ M protein and 0.5 mM compound. The concentration of the protein stock solution was established using the Edelhoch method, whereas compound stock solutions were prepared based on mass. A typical experiment included a single 0.2  $\mu$ l compound injection into a 200  $\mu$ l cell filled with protein, followed by 25 subsequent 1.5  $\mu$ l injections of compound. Injections were performed with a spacing of 180 seconds and a reference power of 8  $\mu$ cal/sec. Control experiments were performed titrating each compound into buffer under identical conditions to determine the heat signals, if any, that arise from diluting the compound. If applicable, the heats of dilution generated were then subtracted from the protein-compound binding curves. The initial data point was routinely deleted. The titration data was analyzed using Origin Software (MicroCal Inc., USA) by non-linear least squares, fitting the heats of binding as a function of the compound:protein ratio to a one site binding model.

### *Kme Reader AlphaScreen® Assays*

The AlphaScreen® assay (Perkin Elmer) was generally performed as previously described.<sup>69</sup> In brief, compound plates (1  $\mu$ L at 10 mM highest concentration; 3-fold, 10-point dilutions in DMSO) were diluted in 1X assay buffer (20 mM TRIS pH 8.0, 25 mM NaCl, 2 mM DTT and 0.05% Tween-20) to 1 mM using a Multimek robotic pipettor (Nanoscreen) and 1  $\mu$ L was spotted into the wells of 384-well low-volume Proxiplates (Perkin Elmer). To these plates 9  $\mu$ L of protein-peptide mix in 1X assay buffer was added by Multidrop (Thermo) to bring the final compound concentration to 100  $\mu$ M and incubated for 30 min at room temperature. Next, 2  $\mu$ L of a 1:1 mixture of streptavidin-conjugate donor and nickel-chelate or  $\alpha$ -GST acceptor beads (45  $\mu$ g/mL in 1X assay buffer) were added and the plates were allowed to incubate for an additional 30 min in the dark at room temperature. After incubation, the plates were read on an EnVision multi-label reader equipped with an HTS AlphaScreen laser (Perkin Elmer). The expression and purification of the constructs used in this assay was described previously<sup>76</sup>.

**Table 2.6** CBX7 peptide substrate used in AlphaScreen®

Protein	Peptide	Peptide sequence
CBX7	H3K9Me3	ARTKQTARK(Me3)STGGKAPRKQL-K(Biotin)-NH <sub>2</sub>

#### *Protein Expression and Purification*

All expression constructs were transformed into Rosetta BL21(DE3)pLysS competent cells (Novagen, EMD Chemicals, San Diego, CA). Protein expression was induced by growing cells at 37°C with shaking until the OD<sub>600</sub> reached ~0.6-0.8 at which time the temperature was lowered to 18°C and expression was induced by adding 0.5mM IPTG and continuing shaking overnight. Cells were harvested by centrifugation and pellets were stored at -80°C.

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



pooled and concentrated using Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

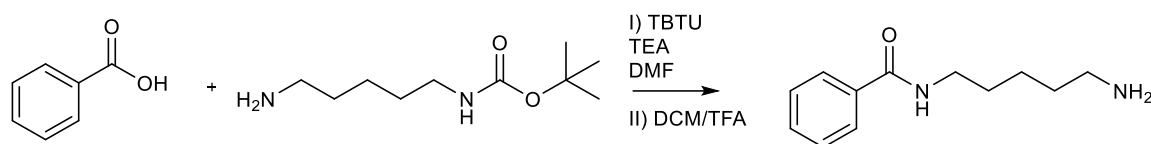
GST-tagged proteins were purified by re-suspending thawed cell pellets in 30ml of lysis buffer (1xPBS, 5mM DTT, 1X EDTA free protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)) per liter of culture. Cells were lysed on ice by sonication as described for His-tagged proteins. Clarified cell lysate was loaded onto a GSTrap FF column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 10 column volumes of binding buffer (1xPBS, 5mM DTT) using a AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was washed with 10 column volumes of binding buffer and protein was eluted in 100% elution buffer (50mM Tris pH 7.5, 150mM NaCl, 10mM reduced glutathione) over 10 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2ml in Amicon Ultra-15 concentrators, 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare, Piscataway, NJ) that had been preequilibrated with 1.2 column volumes of sizing buffer (25mM Tris pH 7.5, 250mM NaCl, 2mM DTT, 5% glycerol) using an ATKA FPLC (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2ml/min collecting 3ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

#### *General chemistry procedures*

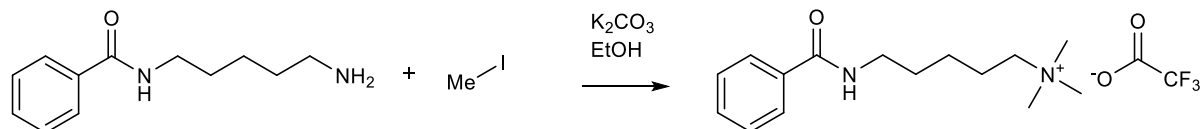
Analytical LCMS data for all compounds were acquired using an Agilent 6110 Series system with the UV detector set to 220 nm. Samples were injected (<10  $\mu$ L) onto an Agilent Eclipse Plus 4.6  $\times$  50 mm, 1.8  $\mu$ m, C18 column at room temperature. Mobile phases A (H<sub>2</sub>O + 0.1% acetic acid) and B (MeOH + 0.1% acetic acid) were used with a linear gradient from 10% to 100% B in 5.0 min, followed by a flush at 100% B for another 2 minutes with a flow rate of 1.0 mL/min. Mass spectra (MS) data were acquired in positive ion mode using an Agilent 6110 single quadrupole mass spectrometer with an electrospray ionization (ESI) source. Reverse phase column chromatography was performed with a Teledyne Isco CombiFlash®R<sub>f</sub> 200 using C18 RediSep®R<sub>f</sub> Gold columns with the UV detector set to 220 nm and 254

nm. Mobile phases of A (H<sub>2</sub>O + 0.1% TFA) and B (MeOH or MeCN) were used with default column gradients. Preparative HPLC was performed using an Agilent Prep 1200 series with the UV detector set to 220 nm and 254 nm. Samples were injected onto a Phenomenex Luna 250 × 30 mm, 5 μm, C18 column at room temperature. Mobile phases of A (H<sub>2</sub>O + 0.1% TFA) and B (MeOH or MeCN) were used with a flow rate of 40 mL/min. A general gradient of 0-15 minutes increasing from 10 to 100% B, followed by a 100% B flush for another 5 minutes. Small variations in this purification method were made as needed to achieve ideal separation for each compound. Analytical LCMS (at 220 nm) was used to establish the purity of targeted compounds. All compounds that were evaluated in biochemical and biophysical assays had >95% purity as determined by LCMS.

#### Synthesis of UNC2404 (1)



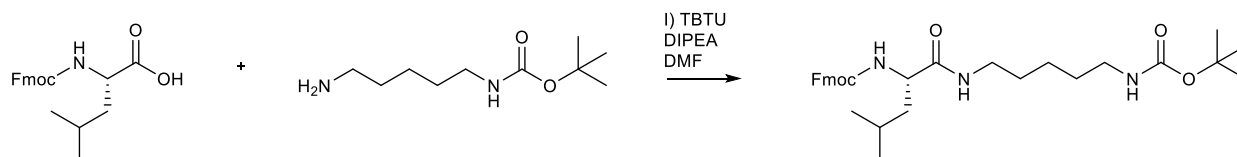
Benzoic acid (200 mg, 1.6 mmol, 1.0 eq) and TBTU (680 mg, 2.1 mmol, 1.3 eq.) were dissolved in 6 mL of DMF. N-boc-cadaverine (1.2, 2.0 mmol, 410 μL) was added to the solution followed by TEA (690 μL, 4.9 mmol, 3.0 eq.). The solution was stirred overnight at room temperature. The mixture was diluted with brine (100 mL) and extracted 3X with DCM (50 mL). The organic extracts were combined and dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated by rotary evaporation. The residue was taken up in 1.5 mL DCM. TFA (1 mL, excess) was added to the solution and the mixture was stirred at room temperature until complete. The reaction was concentrated by rotary evaporation and by vacuum-dried overnight to yield 303 mg (90%) of crude product. The crude product was carried forward without further purification as a TFA salt. MSI (ESI): 207 [M+H]<sup>+</sup>



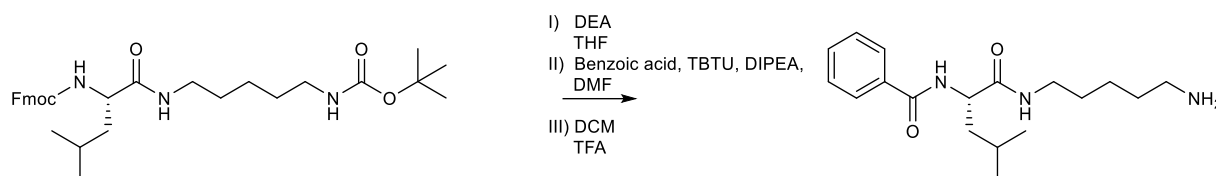
N-(5-aminopentyl)benzamide 2,2,2-trifluoroacetate (30 mg, 1.5 mmol, 1.0 eq.) and K<sub>2</sub>CO<sub>3</sub> (1.2 g, 8.8 mmol, 6.0 eq.) were dissolved in 1.5 mL of EtOH. Iodomethane was added (920 μL, 15 mmol, 10 eq.) and

the reaction was stirred at room temperature overnight. Iodomethane was added (400  $\mu$ L, 6.4 mmol, 5.0 eq.) and the mixture was stirred until complete. The reaction was concentrated by rotary evaporation, taken up in 1600 mL of methanol and purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 212 mg (58%) of the title compound as a trifluoroacetate salt. MSI (ESI): 249.25 [M]<sup>+</sup>

### Synthesis of UNC2403 (2)

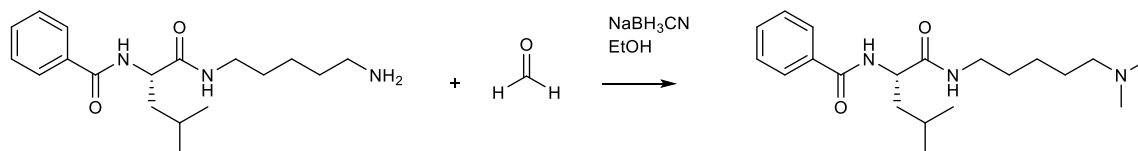


Fmoc-Leu-OH (500 mg, 1.4 mmol, 1.0 eq) and TBTU (590 mg, 3.1 mmol, 1.3 eq.) were dissolved in 5 mL of DMF. *N*-boc-cadaverine (1.2, 1.7 mmol, 350  $\mu$ L) was added to the solution followed by DIPEA (540  $\mu$ L, 1.8 mmol, 2.2 eq.). The solution was stirred for 2 hours at room temperature. The mixture was diluted with brine (100 mL) and extracted 3X with DCM (50 mL). The organic extracts were combined and dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated by rotary evaporation. The residue was purified by normal phase column chromatography (hexanes/EtOAc) to yield 747 mg (98%) of the desired product. MSI (ESI): 538 [M+H]<sup>+</sup>



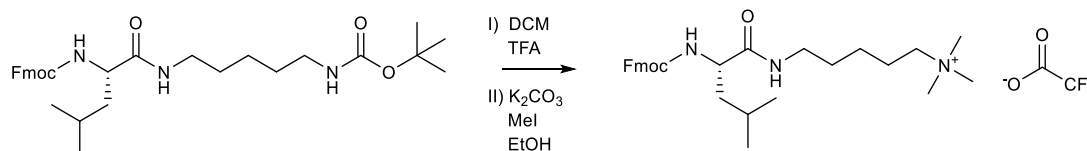
*tert*-butyl (S)-(5-(2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-methylpentanamido)pentyl) carbamate (100 mg, 0.19 mmol, 1.0 eq) was dissolved in 1.8 mL of THF. Diethylamine (190  $\mu$ L, 1.9 mmol, 10 eq.) was added and the mixture was stirred at room temperature overnight. The mixture was concentrated by rotary evaporation and dried under high vacuum at room temperature overnight. The crude product was dissolved in 1 mL of DMF. Benzoic acid (27 mg, 0.22 mmol, 1.2 eq.), TBTU (78 mg, 0.24 mmol, 1.3 eq.) and TEA (78  $\mu$ L, 0.56 mmol, 3.0 eq.) were added and the mixture was stirred at room temperature overnight. The solution was diluted in EtOAc (50 mL) and washed 3X with brine (50 mL). The organic extract was dried over Na<sub>2</sub>SO<sub>4</sub>, filtered, concentrated by rotary evaporation and purified by normal phase column chromatography (hexanes/EtOAc) to yield 64.8 mg (83%) of *tert*-butyl (S)-(5-(2-benzamido-4-

methylpentanamido)pentyl)carbamate. MSI (ESI): 420 [M+H]<sup>+</sup>. *tert*-butyl (S)-(5-(2-benzamido-4-methylpentanamido)pentyl)carbamate was dissolved in 1.5 mL DCM. TFA (1 mL) was added and the reaction was stirred until completion. The crude product was concentrated by rotary evaporation and carried through without further purification. MSI (ESI): 320 [M+H]<sup>+</sup>



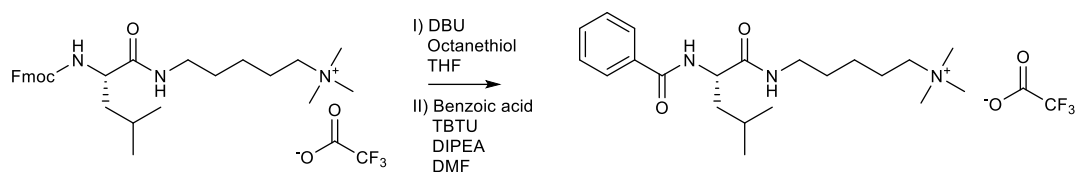
(S)-*N*-(1-((5-aminopentyl)amino)-4-methyl-1-oxopentan-2-yl)benzamide TFA salt (49 mg, 0.15 mmol, 1.0 eq.) was dissolved in EtOH (1.5 mL). Sodium cyanoborohydride (29 mg, 0.46 mmol, 3.0 eq.) and 37% aqueous formaldehyde (50  $\mu$ L, 0.68 mmol, 4.4 eq.) mmol, were added and the mixture was stirred until completion. The mixture was concentrated by rotary evaporation, taken up in 1600 mL of MeOH and purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 21.5 mg (30%) of UNC2403 as a TFA salt. MSI (ESI): 348.3 [M+H]<sup>+</sup>

### Synthesis of UNC2339 (3)



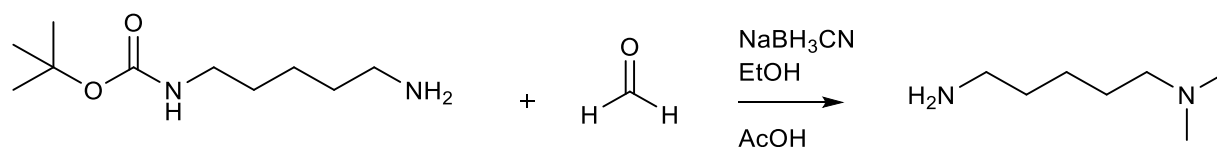
*tert*-butyl (S)-(5-(2-(((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-methylpentanamido)pentyl) carbamate (250 mg, 0.46 mmol, 1.0 eq.) was dissolved in 4 mL of DCM. TFA (3 mL, excess) was added and the mixture was stirred at room temperature for 1 hour. The solution was concentrated by rotary evaporation. The crude residue was taken up in 1 mL EtOH. Iodomethane (290  $\mu$ L, 4.6 mmol, 10 eq) and K<sub>2</sub>CO<sub>3</sub> (190 mg, 1.4 mmol, 3.0 eq.) were added and the mixture was stirred overnight at room temperature. Iodomethane (290  $\mu$ L, 4.6 mmol, 10 eq) and K<sub>2</sub>CO<sub>3</sub> (190 mg, 1.4 mmol, 3.0 eq.) were added and the mixture was stirred overnight. The solution was concentrated by rotary evaporation and purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 226 mg (83%) of the title compound as a trifluoroacetate salt. MSI (ESI): 480.4 [M]<sup>+</sup>

### Synthesis of UNC2401 (4)



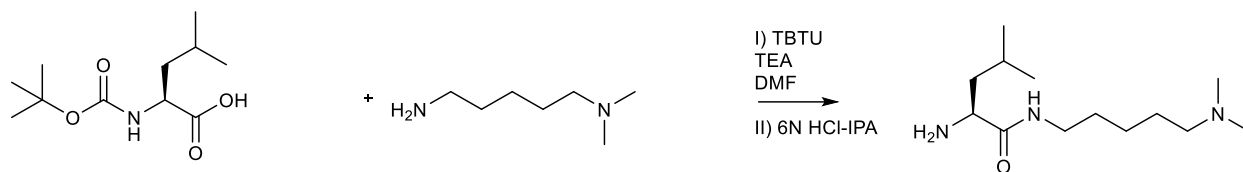
(*S*)-5-(2-((((9*H*-fluoren-9-yl)methoxy)carbonyl)amino)-4-methylpentanamido)-*N,N,N*-trimethylpentan-1-aminium 2,2,2-trifluoroacetate (170 mg, 0.29 mmol, 1.0 eq.) was dissolved in THF (3 mL). Octanethiol (500  $\mu$ L, 2.9 mmol, 10 eq.) was added followed by dropwise addition of DBU (1.3  $\mu$ L, .009 mmol, 0.3 eq.). The solution was stirred at room temperature. After one hour, 0.5 eq. of DBU was added. Upon completion, the solution was concentrated by rotary evaporation. The resulting residue was triturated with 2 mL of ice cold ether. The ether was decanted followed by three additional ether triturations. The residue was placed on under high vacuum overnight to yield 92 mg (86%) of (*S*)-5-(2-amino-4-methylpentanamido)-*N,N,N*-trimethylpentan-1-aminium. (*S*)-5-(2-amino-4-methylpentanamido)-*N,N,N*-trimethylpentan-1-aminium was dissolved in 1 mL of DMF. Benzoic acid (36 mg, 0.30 mmol, 1.2 eq.), TBTU (100 mg, 0.32 mmol, 1.3 eq.) and DIPEA (95  $\mu$ L, 0.54 mmol, 2.2 eq.) were added and the mixture was stirred at room temperature for 2 hours. The solution was diluted w/ DCM (50 mL) and brine (100 mL). The aqueous layer was extracted 3X with 50 mL of DCM. Product was retained in the aqueous layer, which was concentrated and resuspended in 1600  $\mu$ L MeOH. The suspension was then filtered and purified by HPLC (( $\text{H}_2\text{O}$  + 0.1% TFA)/MeOH) to yield 41 mg (35%) of UNC2401 as a trifluoroacetate salt. MSI (ESI): 362.3  $[\text{M}]^+$

### Synthesis of UNC2720 (5)

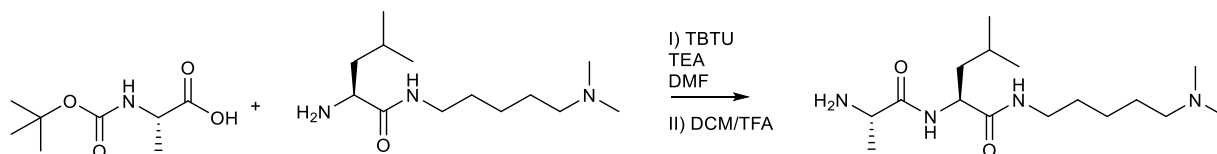


Boc-cadaverine (100  $\mu$ L, 0.49 mmol, 1.0 eq.) was dissolved in 5 mL of EtOH. Formaldehyde (37% aq., 160  $\mu$ L, 2.2 mmol, 4.4 eq.) and sodium cyanoborohydride (190 mg, 3.0 mmol, 6.0 eq.) were added and the mixture was stirred at room temperature for 15 minutes. Acetic acid (140  $\mu$ L, 2.5 mmol, 5eq.) was added and the mixture was stirred at room temperature until complete. The solution was concentrated by

rotary evaporation and taken up in 5 mL of a 6N solution of HCl in isopropyl alcohol. Upon completion, the mixture was concentrated under rotary evaporation and carried forward without further purification. MSI (ESI): 131 [M+H]<sup>+</sup>

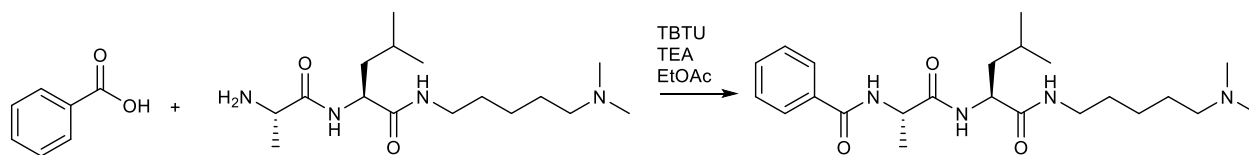


Boc-Leu-OH monohydrate (100 mg, 0.41 mmol, 1.0 eq.), *N,N*-dimethylpentane-1,5-diamine·2TFA salt (64 mg, 0.49 mmol, 1.2 eq.) and TBTU (170 mg, 0.54 mmol, 1.3 eq.) were dissolved in DMF (5 mL). TEA (340  $\mu$ L, 3.0 mmol, 6.0 eq.) was added and the reaction was stirred at room temperature overnight. The mixture was diluted with EtOAc (50 mL) and washed 3X with saturated bicarbonate (50 mL). The aqueous washes were combined and extracted 3X with 50 mL EtOAc. The organic extracts were combined, dried over Na<sub>2</sub>SO<sub>4</sub>, concentrated by rotary evaporation and dried overnight under high vacuum. The residue was taken up 5 mL of a 6N solution of HCl in isopropyl alcohol. The solution was stirred at room temperature until the reaction was complete. The solution was concentrated by rotary evaporation and dried overnight under reduced pressure to afford the product as a dihydrochloride salt in quantitative yield. MSI (ESI): 122 [M+2H]<sup>2+</sup>, 244 [M+H]<sup>+</sup>. The product was carried forward without further purification.



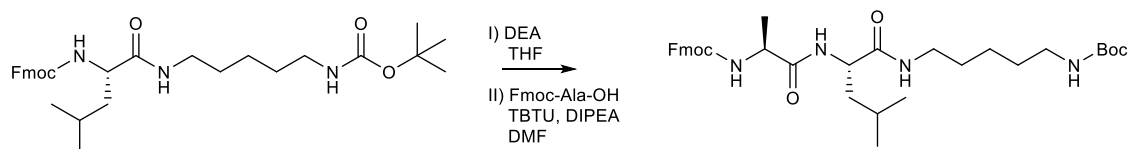
Boc-Ala-OH (91 mg, 0.48 mmol, 1.0 eq.), *(S)*-2-amino-*N*-(5-(dimethylamino)pentyl)-4-methylpentanamide·2HCl (140 mg, 0.58 mmol, 1.2 eq.), and TBTU (200 mg, 0.63 mmol, 1.3 eq.) were dissolved in DMF (5 mL). TEA (400  $\mu$ L, 2.9 mmol, 6 eq.) was added and the reaction was stirred at room temperature for 2 hours. The solution was diluted with 50 mL of EtOAc and washed 3X with saturated bicarbonate (50 mL). The aqueous washes were combined and extracted 3X with EtOAc (50 mL). The organic extracts were combined and dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated by rotary evaporation. The residue was taken up in 5 mL of DCM and stirred at room temperature until complete. The solution was

concentrated by rotary evaporation and dried under reduced pressure overnight to yield the product in nearly quantitative yield as a TFA salt. MSI (ESI): 157 [M+2H]<sup>2+</sup>, 314 [M+H]<sup>+</sup>

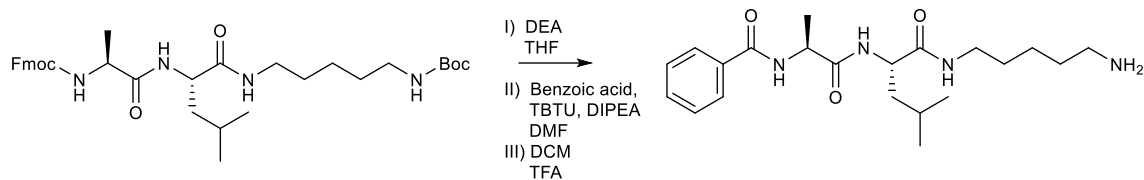


Benzoic acid (49 mg, 0.40 mmol, 1.0 eq.), (S)-2-(((S)-2-aminopropanamido)-N-(5-(dimethylamino)pentyl)-4-methylpentanamide (150 mg, 0.48 mmol, 1.2 eq.) and TBTU (170 mg, 0.52 mmol, 1.3 eq.) were dissolved in EtOAc (4 mL). TEA (340  $\mu$ L, 2.4 mmol, 6 eq.) was added and the mixture was stirred at room temperature until complete. The solution was concentrated and taken up in 1600  $\mu$ L of MeOH, filtered and purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 51 mg (30% of UNC2720 as a TFA salt. MSI (ESI): 419.3 [M+H]<sup>+</sup>

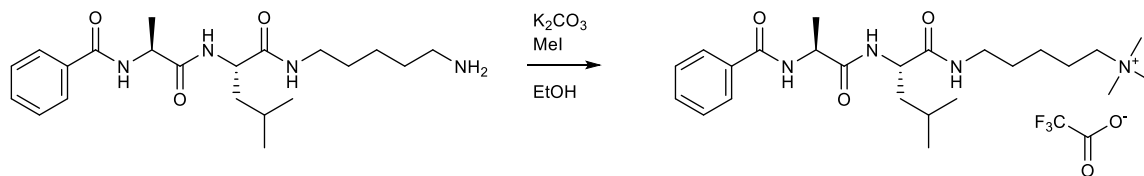
#### Synthesis of UNC2588 (**6**)



*tert*-butyl (S)-5-(2-((((9H-fluoren-9-yl)methoxy)carbonyl)amino)-4-methylpentanamido)pentyl carbamate (300 mg, 0.56 mmol, 1.0 eq.) was dissolved in THF (5.5 mL). DEA (580  $\mu$ L, 5.6 mmol, 10 eq.) was added and the solution was stirred overnight at room temperature. The mixture was concentrated by rotary evaporation and dried under reduced pressure overnight. The residue was taken up in DMF (2 mL) and stirred at room temperature. Fmoc-Ala-OH (210 mg, 0.67 mmol, 1.2 eq.), TBTU (230 mg, 0.73 mmol, 1.3 eq.) and DIPEA (210  $\mu$ L, 1.2 mmol, 2.2 eq.) were added and the reaction was stirred at room temperature until complete. The solution was diluted with EtOAc (50 mL), washed 2X with brine (50 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated by rotary evaporation to give 122 mg (36%) of the desired product. MSI (ESI): 609 [M+H]<sup>+</sup>. The crude product was carried forward without further purification.



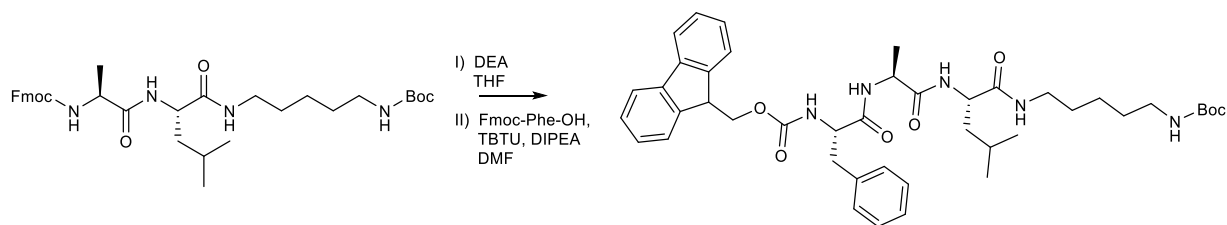
*tert*-butyl ((5*S*,8*S*)-1-(9*H*-fluoren-9-yl)-8-isobutyl-5-methyl-3,6,9-trioxo-2-oxa-4,7,10-triazapentadecan-15-yl)carbamate was dissolved in THF (2 mL) and stirred at room temperature. Diethylamine (200  $\mu$ L, excess) was added and the reaction was stirred at room temperature until complete. The solution was concentrated by rotary evaporation and dried under reduced pressure overnight. The crude product was taken up in DMF (1 mL) and stirred at room temperature. Benzoic acid (17 mg, 0.14 mmol, 1.0 eq.), TBTU (57 mg, 0.18 mmol, 1.3 eq.) and TEA (57  $\mu$ L, 0.41 mmol, 3.0 eq.) were added and the solution was stirred overnight. The mixture was diluted with EtOAc (50 mL) and washed 3X with brine (50 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated by rotary evaporation. The residue was taken up in 1.4 mL of DCM and stirred at room temperature. TFA (1 mL) was added and the reaction was stirred at room temperature until completion. The solution was concentrated by rotary evaporation and carried forward without further purification. MSI (ESI): 391 [M+H]<sup>+</sup>



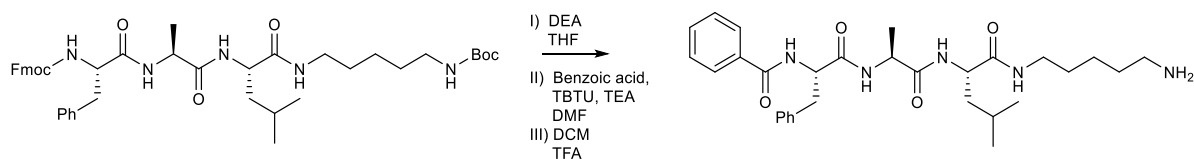
*N*-((*S*)-1-(((*S*)-1-((5-aminopentyl)amino)-4-methyl-1-oxopentan-2-yl)amino)-1-oxopropan-2-yl)benzamide (54 mg, 0.14 mmol, 1.0 eq.), K<sub>2</sub>CO<sub>3</sub> (110 mg, 0.82 mmol, 6 eq.), iodomethane (86  $\mu$ L, 1.4 mmol, 10 eq.) were taken up in EtOH (1.4 mL). The solution was stirred at room temperature. The mixture was concentrated by rotary evaporation and taken up in 1600  $\mu$ L MeOH and filtered. The filtrate was purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 18 mg (24%) of UNC2588 as a trifluoroacetate salt. MSI (ESI): 433.4 [M]<sup>+</sup>



## Synthesis of UNC2934 (7)

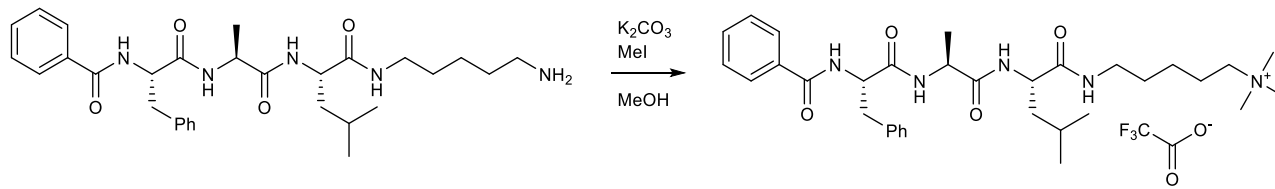


*tert*-butyl ((5*S*,8*S*)-1-(9*H*-fluoren-9-yl)-8-isobutyl-5-methyl-3,6,9-trioxo-2-oxa-4,7,10-triazapentadecan-15-yl)carbamate (300 mg, 0.50 mmol, 1.0 eq.) was taken up in 2.5 mL THF and stirred at room temperature. Diethylamine (2.5 mL, excess) was added and the reaction was stirred at room temperature until complete. Once complete, the solution was concentrated by rotary evaporation and carried forward without further purification. MSI (ESI): 387 [M+H]<sup>+</sup>. The crude residue was taken up in DMF (5 mL) and stirred at room temperature. Fmoc-Phe-OH (230 mg, 0.60 mmol, 1.2 eq.), TBTU (210 mg, 0.65 mmol, 1.3 eq.), and DIPEA (110  $\mu$ L, 0.65 mmol, 1.3 eq.) were added and the reaction was stirred at room temperature until complete. The solution was diluted with EtOAc (50 mL), and washed 3X with brine (75 mL). Precipitate was observed in the organic layer. The organic layer suspension was collected and concentrated by rotary evaporation. The residue was purified by normal phase column chromatography (DCM/20% MeOH in DCM). Product-containing fraction were combined, concentrated by rotary evaporation and purified by normal phase column chromatography to yield 260 mg (68%) of the desired product. MSI (ESI): 756 [M+H]<sup>+</sup>



*tert*-butyl ((5*S*,8*S*,11*S*)-5-benzyl-1-(9*H*-fluoren-9-yl)-11-isobutyl-8-methyl-3,6,9,12-tetraoxo-2-oxa-4,7,10,13-tetraazaoctadecan-18-yl)carbamate (100 mg, 0.13 mmol, 1.0 eq.) was dissolved in DMF (1 mL) and stirred at room temperature. Diethylamine (1 mL, excess) was added and the reaction was stirred at room temperature until complete. The solution was concentrated by rotary evaporation and dried under reduced pressure overnight. The crude product was taken up in DMF (1 mL) and stirred at room temperature. Benzoic acid (19 mg, 0.13 mmol, 1.2 eq.), TBTU (55 mg, 0.17 mmol, 1.3 eq.) and TEA (55  $\mu$ L, 0.40 mmol, 3.0 eq.) were added and the solution was stirred until complete. The mixture was diluted

with EtOAc (50 mL) and washed 3X with brine (50 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated by rotary evaporation. The residue was purified by normal phase column chromatography (Hexanes/EtOAc). Product-containing fractions were combined, concentrated by rotary evaporation and taken up in 0.5 mL of DCM and stirred at room temperature. TFA (0.5 mL) was added and the reaction was stirred at room temperature until completion. The solution was concentrated by rotary evaporation and carried forward without further purification. MSI (ESI): 538 [M+H]<sup>+</sup>



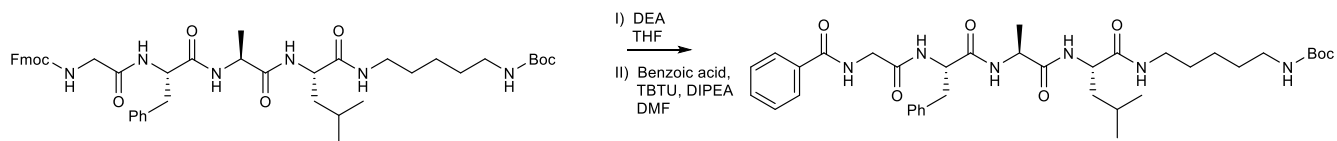
*N*-((*S*)-1-(((*S*)-1-(((*S*)-1-((5-aminopentyl)amino)-4-methyl-1-oxopentan-2-yl)amino)-1-oxopropan-2-yl)amino)-1-oxo-3-phenylpropan-2-yl)benzamide (24 mg, 0.045 mmol, 1.0 eq.), K<sub>2</sub>CO<sub>3</sub> (37 mg, 0.27 mmol, 6 eq.), iodomethane (28  $\mu$ L, 0.45 mmol, 10 eq.) were taken up in MeOH (0.5 mL). The was heated to 110°C by microwave irradiation and stirred vigorously for 10 minutes. The solution was diluted with 1 mL of MeOH and filtered. The filtrate was purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 18 mg (56%) of UNC2934 as a trifluoroacetate salt. MSI (ESI): 580.4 [M]<sup>+</sup>

#### Synthesis of UNC2994 (8)

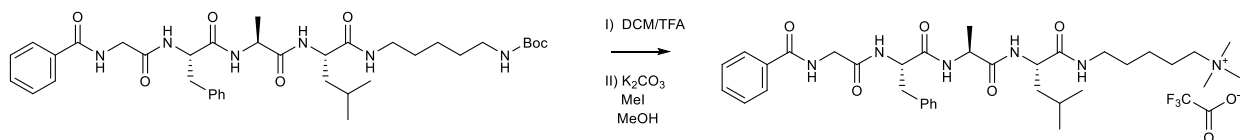


*tert*-butyl ((5*S*,8*S*,11*S*)-5-benzyl-1-(9*H*-fluoren-9-yl)-11-isobutyl-8-methyl-3,6,9,12-tetraoxo-2-oxa-4,7,10,13-tetraazaoctadecan-18-yl)carbamate (160 mg, 0.21 mmol, 1.0 eq.) was dissolved in DMF (1 mL) and stirred at room temperature. Diethylamine (1 mL, excess) was added and the reaction was stirred at room temperature until complete. The solution was concentrated by rotary evaporation and dried under reduced pressure overnight. The crude product was taken up in DMF (1 mL) and stirred at room temperature. Fmoc-Gly-OH (74 mg, 0.25 mmol, 1.2 eq.), TBTU (87 mg, 0.27 mmol, 1.3 eq.) and DIPEA (47  $\mu$ L, 0.27 mmol, 1.3 eq.) were added and the solution was stirred until complete. The mixture was diluted with EtOAc (50 mL) and washed 3X with brine (50 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub>

and concentrated by rotary evaporation. The residue was purified by normal phase column chromatography (Hexanes/EtOAc) to yield 130 mg (76%) of the desired product. MSI (ESI): 813 [M+H]<sup>+</sup>

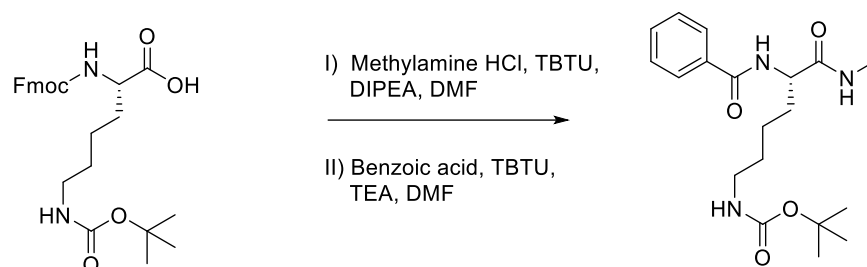


(9*H*-fluoren-9-yl)methyl *tert*-butyl ((4*S*,7*S*,10*S*)-4-benzyl-10-isobutyl-7-methyl-2,5,8,11-tetraoxo-3,6,9,12-tetraazaheptadecane-1,17-diyl)dicarbamate carbamate (130 mg, 0.16 mmol, 1.0 eq.) was dissolved in DMF (0.75 mL) and stirred at room temperature. Diethylamine (0.75 mL, excess) was added and the reaction was stirred at room temperature until complete. The solution was concentrated by rotary evaporation and dried under reduced pressure overnight. The crude product was taken up in DMF (1.5 mL) and stirred at room temperature. Benzoic acid (23 mg, 0.19 mmol, 1.2 eq.), TBTU (66 mg, 0.20 mmol, 1.3 eq.) and DIPEA (55  $\mu$ L, 0.31 mmol, 2.0 eq.) were added and the solution was stirred until complete. The mixture was diluted with EtOAc (50 mL) and washed 3X with brine (50 mL). The organic layer was dried over Na<sub>2</sub>SO<sub>4</sub> and concentrated by rotary evaporation. The residue was purified by normal phase column chromatography (Hexanes/EtOAc) to yield 91 mg (83%) of the desired product. MSI (ESI): 695 [M+H]<sup>+</sup>

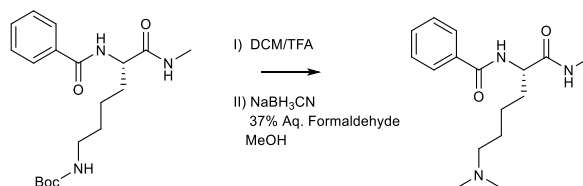


*tert*-butyl ((6*S*,9*S*,12*S*)-6-benzyl-12-isobutyl-9-methyl-1,4,7,10,13-pentaoxo-1-phenyl-2,5,8,11,14-pentaazanonadecan-19-yl)carbamate (91 mg, 0.13 mmol, 1.0 eq.) was taken up in 1.3 mL of DCM and the solution was stirred at room temperature. TFA (0.5 mL) was added and the reaction was stirred until completion. The solution was concentrated by rotary evaporation. The residue was taken up in MeOH (2 mL) transferred to a microwave reaction vessel. Iodomethane (82  $\mu$ L, 1.3 mmol, 10 eq.) and K<sub>2</sub>CO<sub>3</sub> (110 mg, 0.79 mmol, 6.0 eq.) was added, the vessel was sealed and placed into a microwave reactor. The solution was heated to 110°C for 10 min with vigorous stirring. The solution was filtered and purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 29.1 mg (30%) of UNC2994 as a trifluoroacetate salt. MSI (ESI): 637 [M]<sup>+</sup>

## Synthesis of UNC2746 (9)

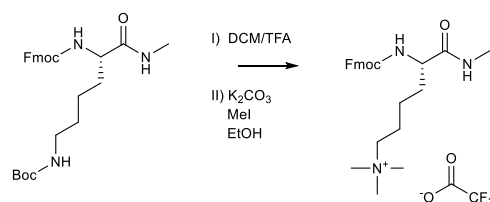


Fmoc-Lys-OH (150 mg, 0.32 mmol, 1.0 eq.), Methylamine hydrochloride (26 mg, 0.38 mmol, 1.2 eq.), TBTU (130 mg, 0.42 mmol, 1.3 eq.) and DIPEA (120  $\mu$ L, 0.70 mmol, 2.2 eq.) were dissolved in DMF (3 mL) and stirred at room temperature until the reaction was complete. The solution was diluted with EtOAc (50 mL), washed 3X with brine (50 mL), dried over  $\text{Na}_2\text{SO}_4$  and concentrated by rotary evaporation. The crude product was taken up in DMF (2.5 mL) and stirred at room temperature. Benzoic acid (33 mg, 0.27 mmol, 1.0 eq.), TBTU (110 mg, 0.35 mmol, 1.3 eq.) and TEA (110  $\mu$ L, 0.80 mmol, 3.0 eq.) were added and the solution was stirred overnight at room temperature. The solution was diluted with EtOAc (50 mL), washed 3X with brine (50 mL), dried over  $\text{Na}_2\text{SO}_4$  and concentrated by rotary evaporation. The crude product was purified by normal phase column chromatography (hexanes/ EtOAc) to yield 65 mg (67%) of the desired product. MSI (ESI): 264  $[\text{M}+\text{H}]^+$



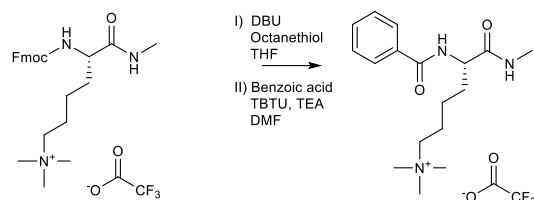
Bz-Lys(Boc)-NMe (65 mg, 0.18 mmol, 1.0 eq.) was dissolved in DCM (2 mL). Two mL of TFA was added and the reaction was stirred at room temperature overnight. The solution was concentrated by rotary evaporation and taken up in MeOH (2 mL). Sodium cyanoborohydride (67 mg, 1.1 mmol, 6.0 eq.), 37% Aq. formaldehyde (59  $\mu$ L, 0.79 mmol, 4.4 eq.) was added and the solution was stirred until the reaction was complete. The solution was filtered and purified by HPLC ( $(\text{H}_2\text{O} + 0.1\% \text{TFA})/\text{MeOH}$ ) to yield 14 mg (19%) of UNC2746 as a TFA salt. MSI (ESI): 292.2  $[\text{M}+\text{H}]^+$

### Synthesis of UNC2338 (10)



Fmoc-Lys(Boc)-NMe (210 mg, 0.43 mmol, 1.0 eq.) was dissolved in DCM (4 mL). TFA (2 mL) was added and the solution was stirred at room temperature until the reaction was complete. The solution was concentrated by rotary evaporation and the residue was taken up in 3.5 mL EtOH. Iodomethane (240  $\mu$ L, 3.8 mmol, 10 eq.) and  $K_2CO_3$  (180 mg, 1.3 mmol, 3.3 eq.) were added and the solution was stirred overnight at room temperature. The solution was concentrated by rotary evaporation, taken up in 1600  $\mu$ L MeOH, filtered, and purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 140 mg (68%) of UNC2338 as a trifluoroacetate salt. MSI (ESI): 424.3 [M]<sup>+</sup>

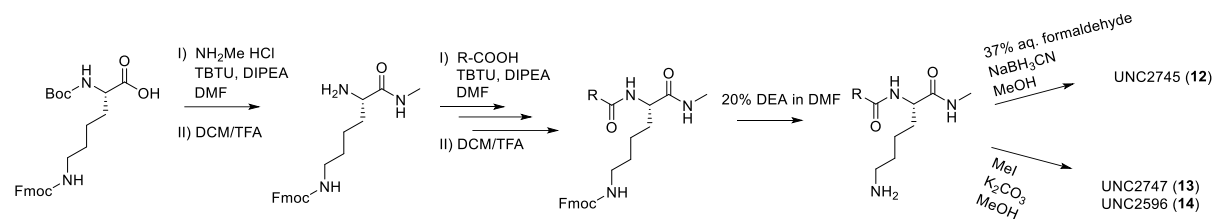
### Synthesis of UNC2402 (11)



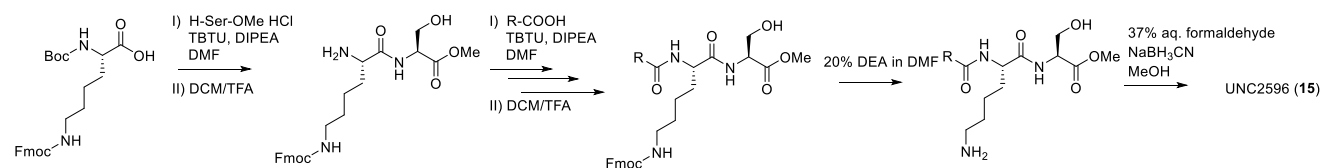
Fmoc-Lys(me3)-NMe (100 mg, 0.19 mmol, 1.0 eq.) was dissolved in THF (2 mL). The solution was stirred at room temperature and octanethiol (330  $\mu$ L, 1.9 mmol, 10 eq.) and DBU (2.8  $\mu$ L, 0.019 mmol, 0.1 eq.) were added. The solution was stirred overnight at room temperature. Ice cold ether was added to triturate the product. Solvent was decanted off and the crude residue was dried under reduced pressure overnight. The residue was taken up in DMF (1 mL). The solution was stirred at room temperature and benzoic acid (28 mg, 0.23 mmol, 1.2 eq.), TBTU (78 mg, 0.24 mmol, 1.3 eq.) and TEA (58  $\mu$ L, 0.41 mmol, 2.2 eq.) were added and the reaction was stirred overnight at room temperature. The solution was diluted with brine (50 mL) and extracted 3X with DCM (50 mL). The aqueous layer was collected, concentrated, taken up in 1600  $\mu$ L MeOH, filtered and purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield 31 mg (39%) of UNC2402 as a trifluoroacetate salt. MSI (ESI): 306.25 [M]<sup>+</sup>

## Synthesis of compounds **12-57**:

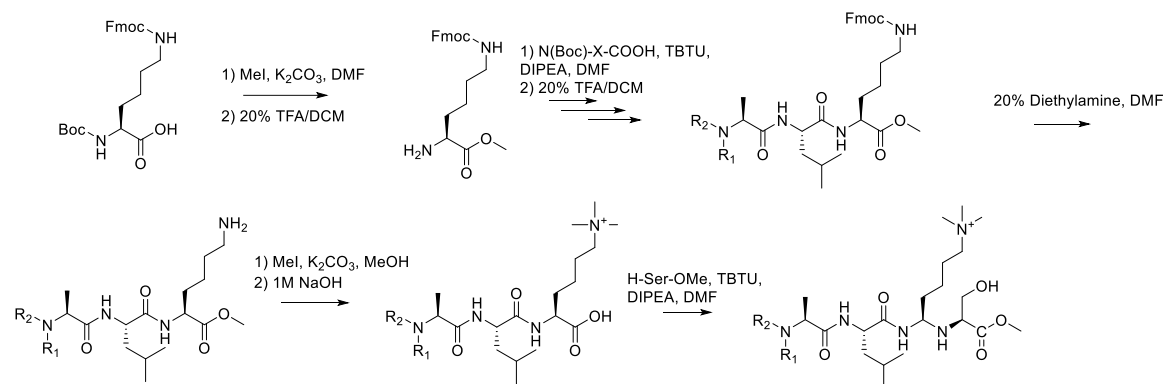
### Synthetic scheme I: **UNC2745 (12)**, **UNC2747(13)** and **UNC2597 (14)**



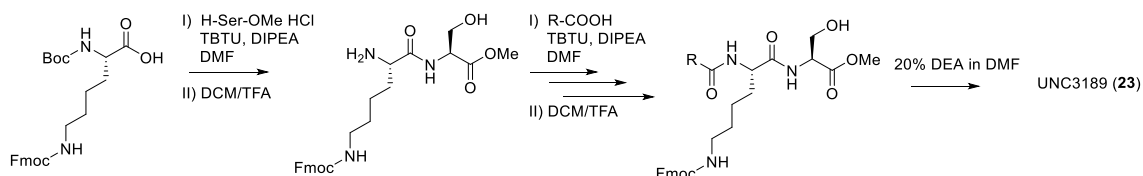
### Synthetic scheme II: **UNC2596 (15)**



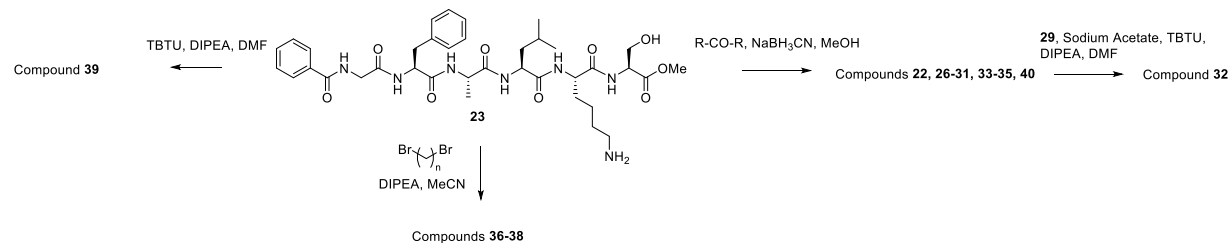
### Synthetic scheme III: compounds **16-21**



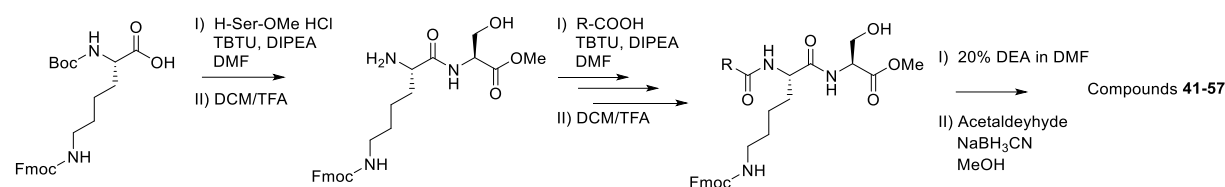
### Synthetic scheme IV: **UNC3189 (23)**



### Synthetic Scheme V: compounds **23-40**



### Synthetic Scheme VI: compounds **41-57**



### General Procedures

#### Amide couplings followed by Boc removal

To a solution of 0.1 M carboxylic acid (1.0 eq) in DMF was added 1.2 equivalents of amine, 1.3 equivalents of TBTU and 2.2 equivalents of DIPEA. The reaction was stirred at room temperature for 30 min to overnight. The mixture was diluted with EtOAc (100 mL) and washed 3X with brine (100 mL). The organic layer was then dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated by rotary evaporation. The resulting residue was dissolved in a mixture of 20% TFA in DCM. The solution was stirred at room temperature for 30 minutes to overnight. The mixture was concentrated by rotary evaporation and purified by reverse-phase column chromatography ((Water + 0.1% TFA)/Methanol) to yield the TFA salt of the target compound. Typical yields ranged from 60-90%.

#### Amide couplings followed by Fmoc-removal

To a solution of 0.1 M carboxylic acid (1.0 eq) in DMF was added 1.2 equivalents of amine, 1.3 equivalents of TBTU and 2.2 equivalents of DIPEA. The reaction was stirred at room temperature for 30 min to overnight. The mixture was diluted with EtOAc (100 mL) and washed 3X with brine (100 mL). The organic layer was then dried over Na<sub>2</sub>SO<sub>4</sub>, filtered and concentrated by rotary evaporation. The resulting residue was dissolved in a mixture of 20% diethylamine in DMF. The solution was stirred at room temperature for 30 minutes to overnight. The mixture was concentrated by rotary evaporation and purified by reverse-phase column chromatography ((Water + 0.1% TFA)/Methanol) to yield the TFA salt of the target compound. Typical yields ranged from 40-70%.

#### *Reductive aminations with aldehydes*

The lysine containing peptide (1.0 eq) was dissolved in methanol at a concentration of 0.1 M. Sodium cyanoborohydride (3.0 eq) and aldehyde (8.0 eq.) were added and the solution was stirred at RT until the amine starting material was consumed. The mixture was then concentrated and purified by reverse phase column chromatography ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield the target compound as a TFA salt. Typical yields ranged from 60-90%.

#### *Reductive aminations with ketones*

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

#### *Cyclic alkylations*

The lysine containing peptide (1.0 eq) was dissolved in MeCN at a final concentration of 0.1 M. Dibromoalkane (1.2 eq) and DIPEA (3.1 eq) were added. The solution was heated at 110°C in a microwave reactor for 1 hour. The solution was then concentrated by rotary evaporation and the residue was purified by HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeOH) to yield the target compound as a TFA salt. Yields ranged from 25-70%.

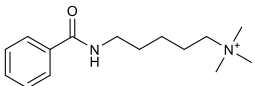
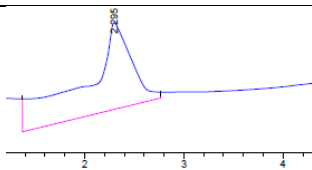
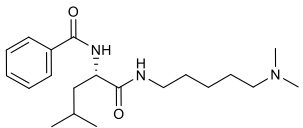
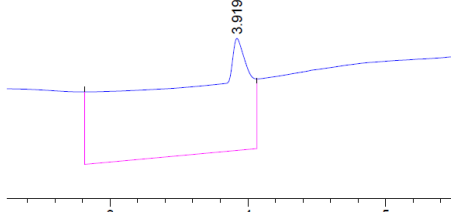
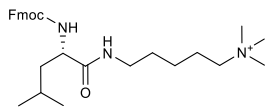
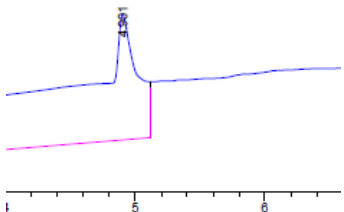
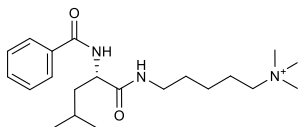
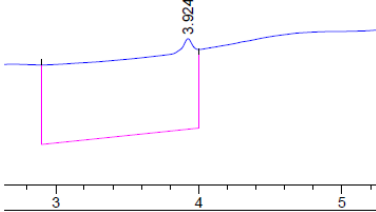
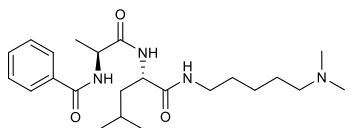
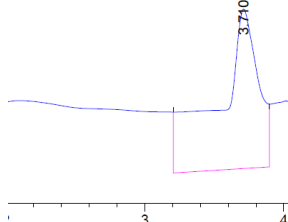
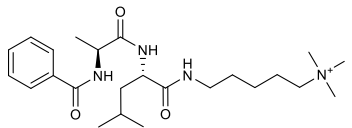
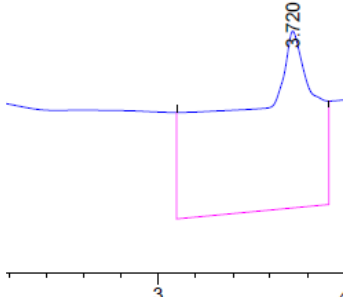
#### *Lysine Trimethylation*

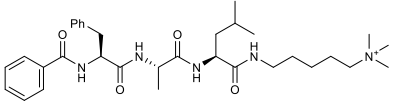
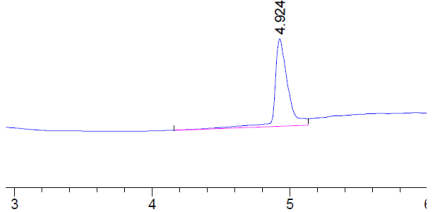
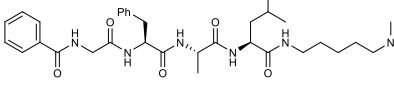
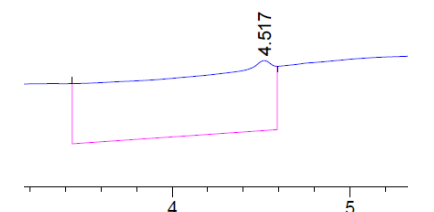
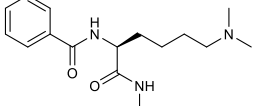
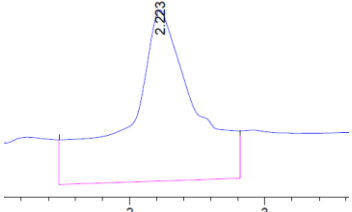
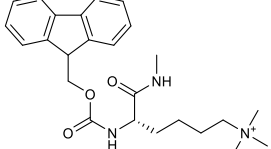
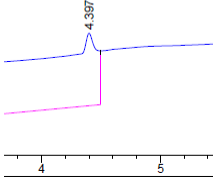
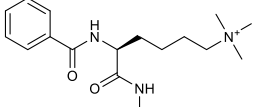
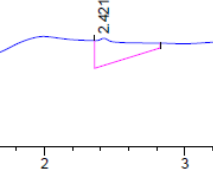
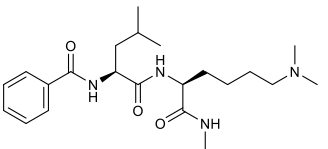
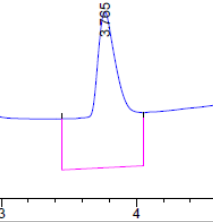
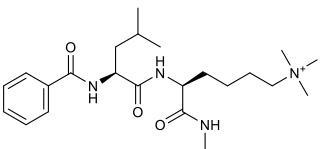
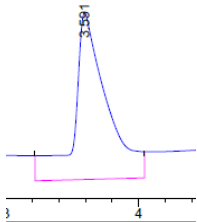
The lysine containing peptide (1.0 eq) was dissolved in MeOH at a final concentration of 0.1 M. Iodomethane (10.0 eq) and K<sub>2</sub>CO<sub>3</sub> (6.0 eq) were added and the mixture was stirred overnight at room temperature. The mixture was concentrated by rotary evaporation and dissolved in a solution of 1M NaOH at a concentration of 0.1 M and stirred for 30 minutes at room temperature. The solution was then brought to pH 7 with 37% aq HCl and concentrated by rotary evaporation. The residue was purified by



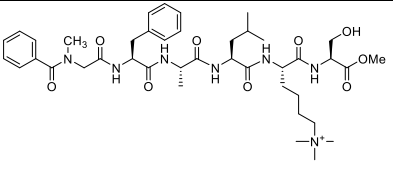
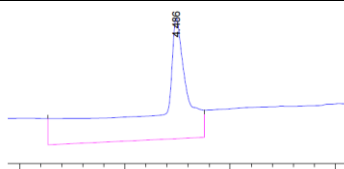
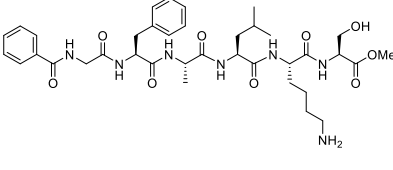
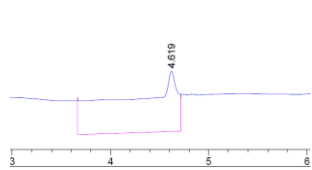
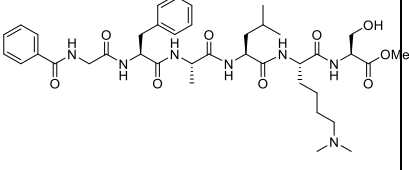
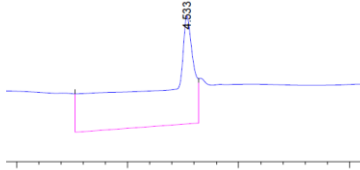
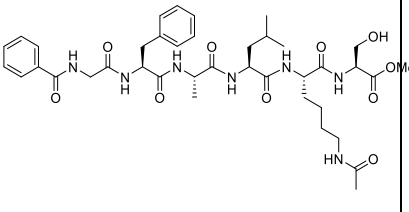
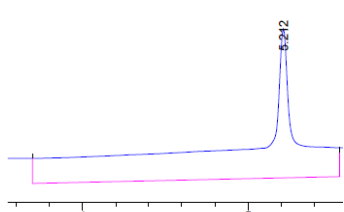
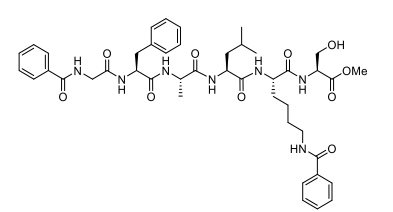
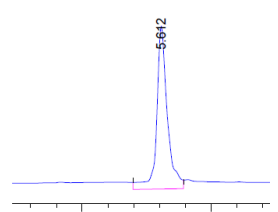
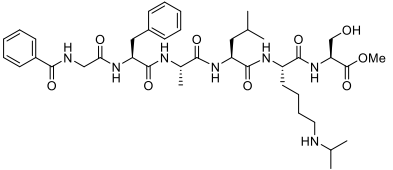
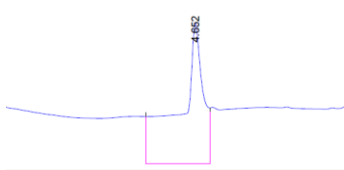
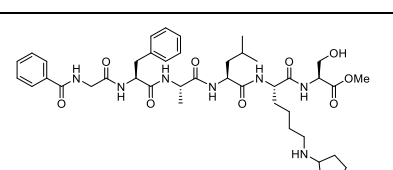
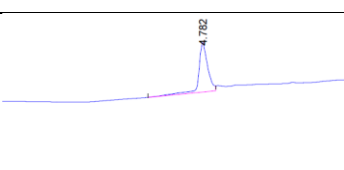
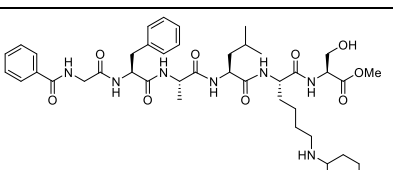
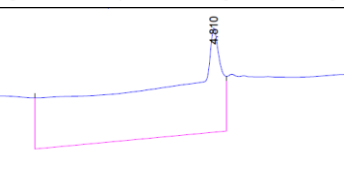
HPLC ((H<sub>2</sub>O + 0.1% TFA)/MeCN) to yield the target compound as a TFA salt. Typical yields ranged from 50-80%.

**Table 2.7** Compound structures and characterization by LCMS

Compound ID	Structure	LCMS trace (220 nM)	Observed Mass
UNC2404 (1)			249.25 [M] <sup>+</sup>
UNC2403 (2)			348.3 [M+H] <sup>+</sup>
UNC2339 (3)			480.4 [M] <sup>+</sup>
UNC2401 (4)			362.3 [M] <sup>+</sup>
UNC2720 (5)			419.3 [M+H] <sup>+</sup>
UNC2588 (6)			433.4 [M] <sup>+</sup>

UNC2934 (7)			580.4 [M] <sup>+</sup>
UNC2994 (8)			637 [M] <sup>+</sup>
UNC2746 (9)			292.2 [M+H] <sup>+</sup>
UNC2338 (10)			424.3 [M] <sup>+</sup>
UNC2402 (11)			306.25 [M] <sup>+</sup>
UNC2745 (12)			405.30 [M+H] <sup>+</sup>
UNC2747 (13)			419.30 [M] <sup>+</sup>

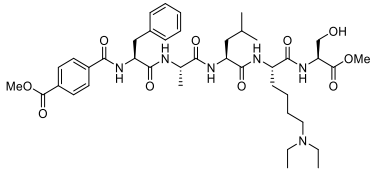
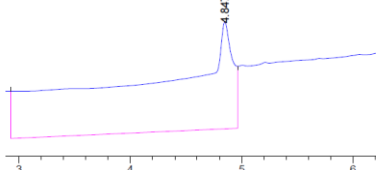
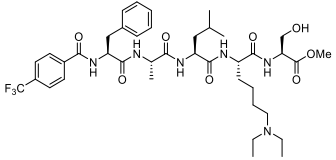
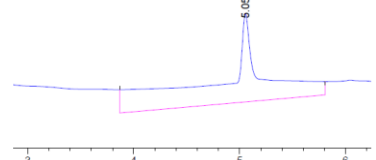
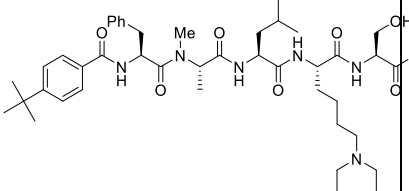
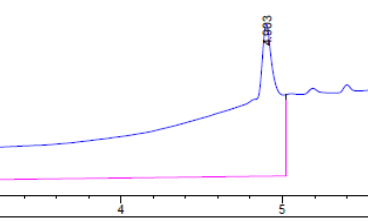
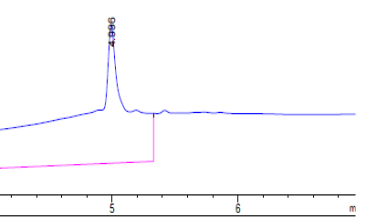
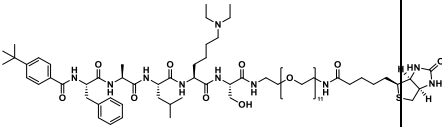
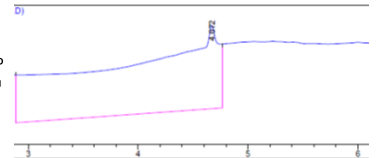
<p>UNC2597 (14)</p>			<p>490.35 [M]<sup>+</sup></p>
<p>UNC2596 (15)</p>			<p>490.40 [M+H]<sup>+</sup></p>
<p>UNC2864 (16)</p>			<p>578.35 [M]<sup>+</sup></p>
<p>UNC2867 (17)</p>			<p>363.30 [M+H]<sup>2+</sup> 725.40 [M]<sup>+</sup></p>
<p>UNC2868 (18)</p>			<p>391.80 [M+H]<sup>2+</sup> 782.45 [M]<sup>+</sup></p>
<p>UNC3261 (19)</p>			<p>796.45 [M]<sup>+</sup></p>
<p>UNC3375 (20)</p>			<p>398.80 [M+H]<sup>2+</sup> 796.40 [M]<sup>+</sup></p>

UNC3403 (21)			398.80 [M+H] <sup>2+</sup> 796.40 [M] <sup>+</sup>
UNC3189 (23)			370.80 [M+2H] <sup>2+</sup> 740.40 [M+H] <sup>+</sup> 762.40 [M+Na] <sup>+</sup>
UNC3188 (22)			384.80 [M+2H] <sup>2+</sup> 768.40 [M+H] <sup>+</sup>
UNC3267 (24)			782.35 [M+H] <sup>+</sup>
UNC3546 (25)			844.50 [M+H] <sup>+</sup>
UNC3563 (26)			402.80 [M+H+Na] <sup>2+</sup> 782.40 [M+H] <sup>+</sup>
UNC3315 (27)			808.45 [M+H] <sup>+</sup>
UNC3314 (24)			411.85 [M+2H] <sup>2+</sup> 422.80 [M+H+Na] <sup>2+</sup> 822.45 [M+H] <sup>+</sup>

UNC3265 (29)			417.85 [M+2H] <sup>2+</sup> 834.50 [M+H] <sup>+</sup>
UNC3264 (30)			448.85 [M+H+Na] <sup>2+</sup> 874.50 [M+H] <sup>+</sup>
UNC3266 (31)			425.30 [M+2H] <sup>2+</sup> 849.45 [M+H] <sup>+</sup>
UNC3649 (32)			876.00 [M+H] <sup>+</sup> 898.00 [M+Na] <sup>+</sup>
UNC3313 (33)			435.80 [M+Na] <sup>+</sup> 848.50 [M+H] <sup>+</sup>
UNC3260 (34)			424.90 [M+2H] <sup>2+</sup> 435.85 [M+H+Na] <sup>2+</sup> 848.50 [M+H] <sup>+</sup>
UNC3567 (35)			398.80 [M+2H] <sup>2+</sup> 409.80 [M+H+Na] <sup>2+</sup> 796.45 [M+H] <sup>+</sup>
UNC3580 (36)			397.75 [M+2H] <sup>2+</sup> 408.80 [M+H+Na] <sup>2+</sup> 794.40 [M+H] <sup>+</sup>
UNC3581 (37)			404.75 [M+2H] <sup>2+</sup> 415.75 [M+H+Na] <sup>2+</sup> 808.40 [M+H] <sup>+</sup>
UNC3648 (38)			411.80 [M+2H] <sup>2+</sup> 422.80 [M+H+Na] <sup>2+</sup> 822.50 [M+H] <sup>+</sup>

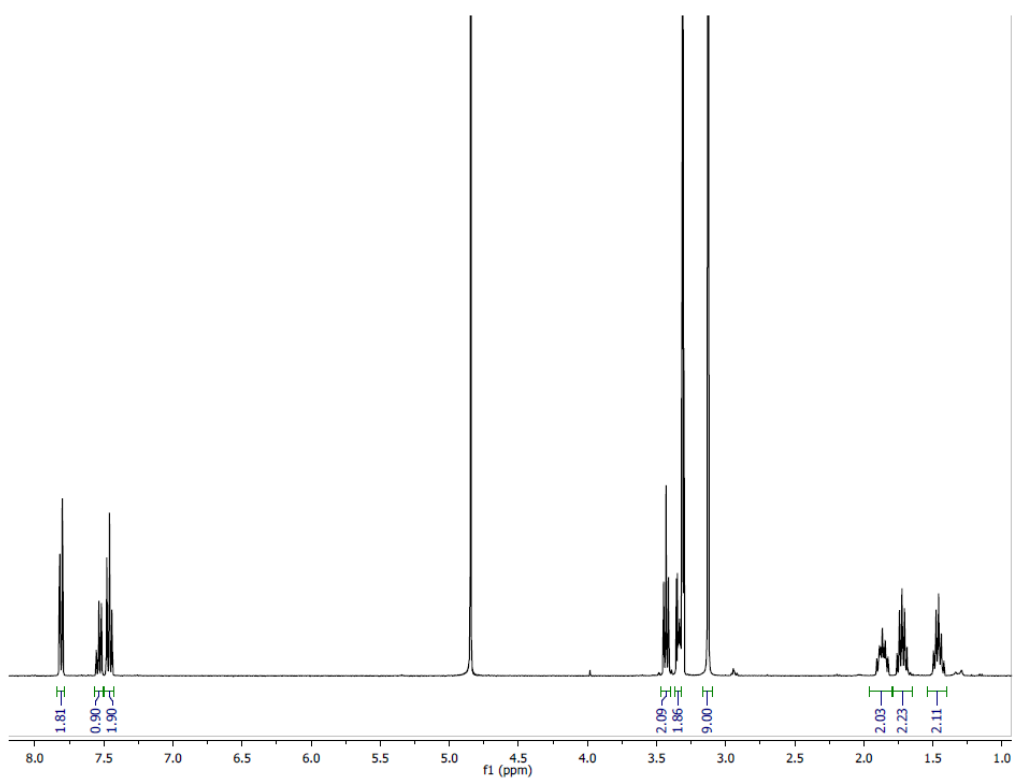
UNC3564 (39)			419.80 [M+2H] <sup>2+</sup> 430.85 [M+H+Na] <sup>2+</sup> 838.45 [M+H] <sup>+</sup>
UNC3566 (40)			405.35 [M+2H] <sup>2+</sup> 416.25 [M+H+Na] <sup>2+</sup> 809.50 [M+H] <sup>+</sup>
UNC3486 (41)			791.45 [M+H] <sup>+</sup>
UNC3562 (42)			806.50 [M+H] <sup>+</sup>
UNC3565 (43)			806.45 [M+H] <sup>+</sup>
UNC3556 (44)			841.50 [M+H] <sup>+</sup>
UNC3772 (45)			803.50 [M+H] <sup>+</sup>
UNC3771 (46)			407.80 [M+H+Na] <sup>2+</sup> 792.45 [M+H] <sup>+</sup>
UNC3773 (47)			416.80 [M+H+Na] <sup>2+</sup> 832.50 [M+H] <sup>+</sup>

UNC4973 (48)			753.45 [M+H] <sup>+</sup>
UNC4974 (49)			377.30 [M+2H] <sup>2+</sup> 754.45 [M+H] <sup>+</sup>
UNC3867 (50)			388.10 [M+H+Na] <sup>2+</sup> 753.10 [M+H] <sup>+</sup>
UNC3769 (51)			767.50 [M+H] <sup>+</sup>
UNC3865 (52)			402.30 [M+H+Na] <sup>2+</sup> 781.50 [M+H] <sup>+</sup>
UNC3866 (53)			409.35 [M+H+Na] <sup>2+</sup> 795.50 [M+H] <sup>+</sup>
UNC4975 (54)		<p>220 nm</p> <p>254 nm</p>	385.25 [M+2H] <sup>2+</sup> 769.45 [M+H] <sup>+</sup>

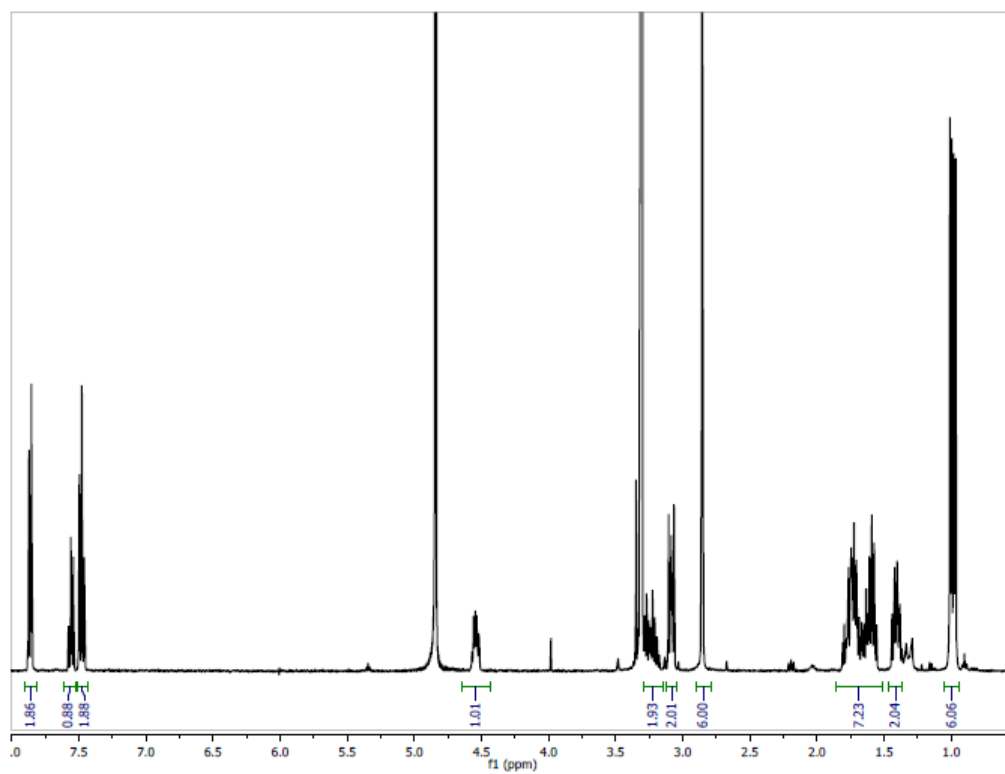
<p>UNC3768 (55)</p>			<p>797.40 [M+H]<sup>+</sup></p>
<p>UNC3770 (56)</p>			<p>807.40 [M+H]<sup>+</sup></p>
<p>UNC4219 (57)</p>		<p><i>Rotamer a</i></p>  <p><i>Rotamer b</i></p> 	<p>809 [M+H]<sup>+</sup></p>
<p>UNC4195 (58)</p>			<p>511 [M+3H]<sup>3+</sup> 767 [M+2H]<sup>2+</sup></p>



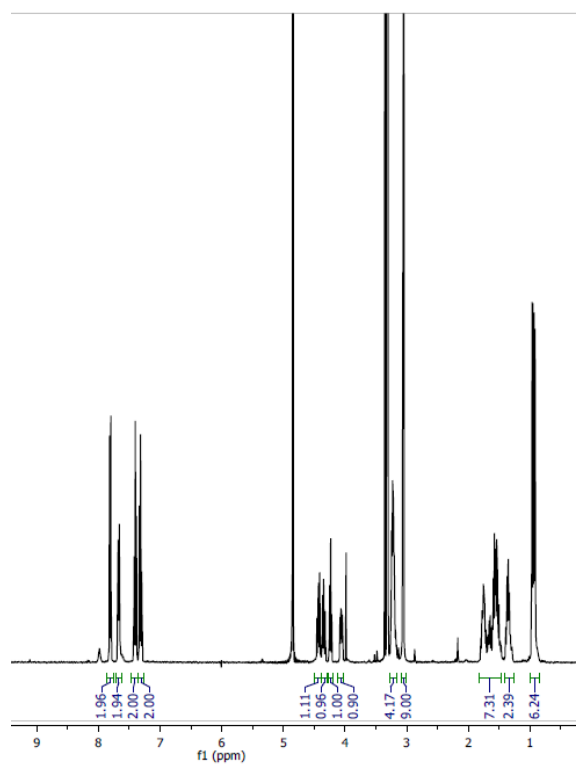
<sup>1</sup>H NMR spectrum of UNC2404 as a trifluoroacetate salt



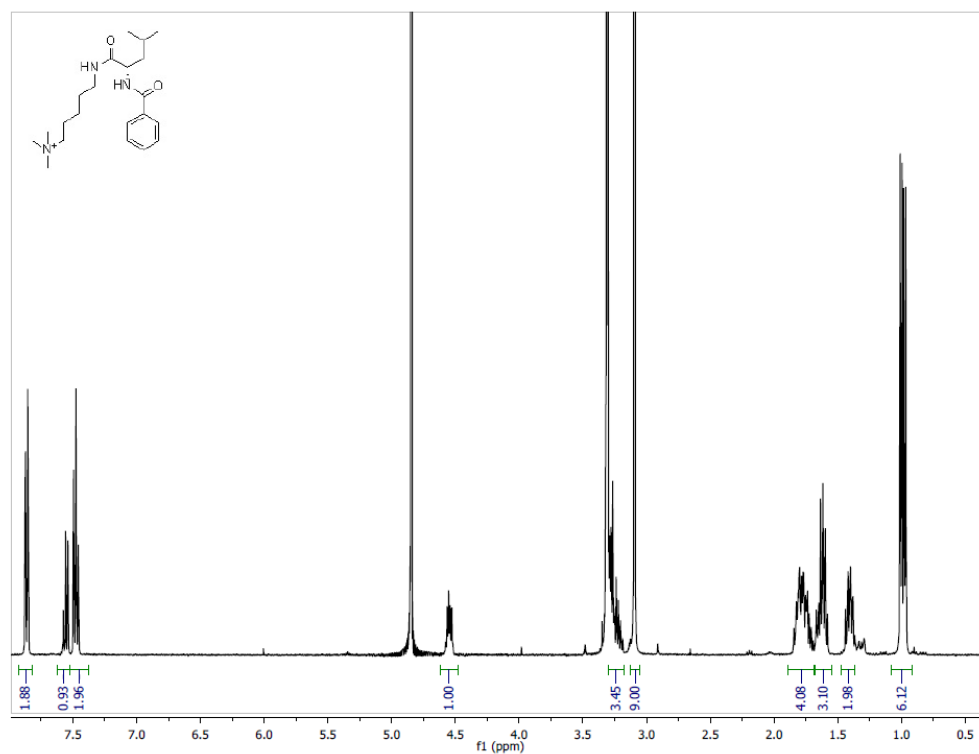
<sup>1</sup>H NMR spectrum of UNC2403 as a TFA salt



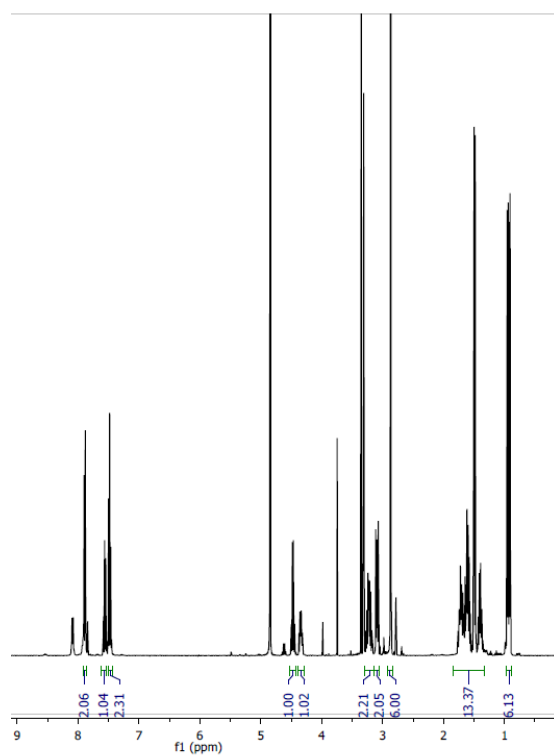
<sup>1</sup>H NMR spectrum of UNC2339 as a TFA salt



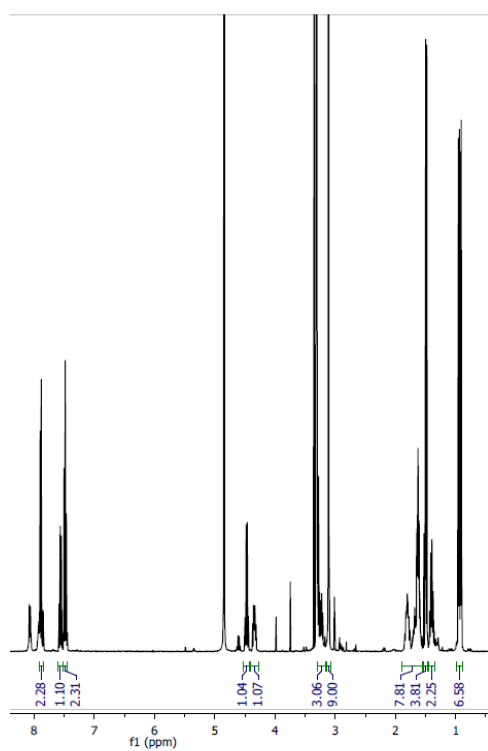
<sup>1</sup>H NMR spectrum of UNC2339 as a trifluoroacetate salt



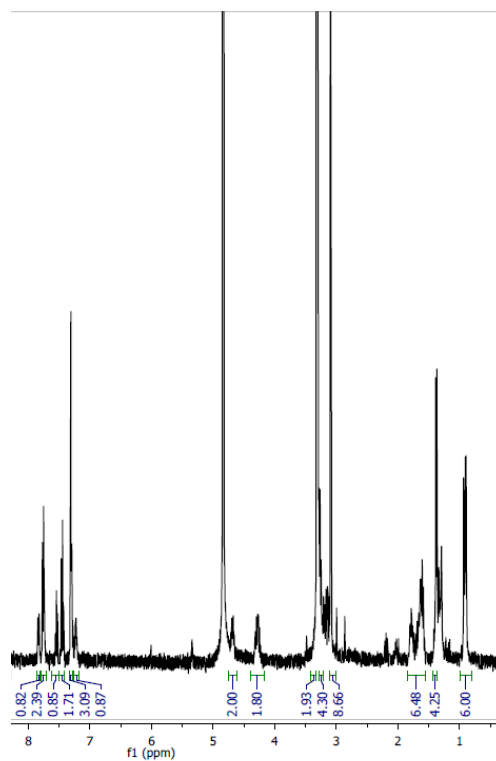
<sup>1</sup>H NMR spectrum of UNC2720 as a TFA salt



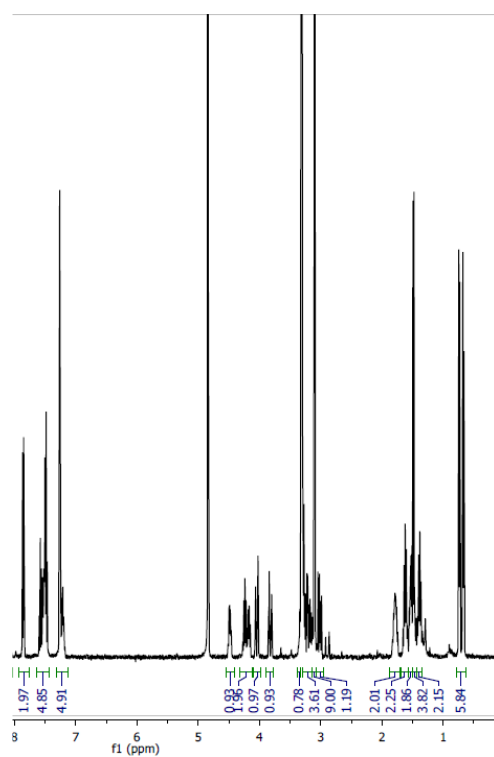
$^1\text{H}$  NMR spectrum of UNC2588 as a trifluoroacetate salt



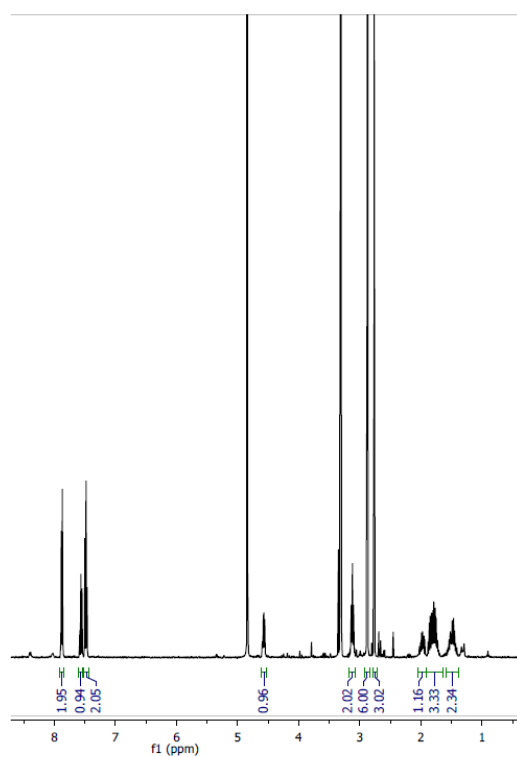
$^1\text{H}$  NMR spectrum of UNC2934 as a trifluoroacetate salt



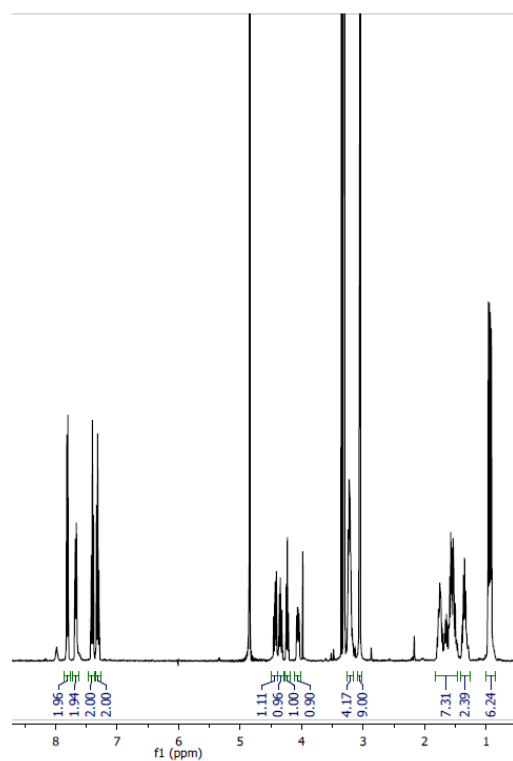
<sup>1</sup>H NMR spectrum of UNC2934 as a trifluoroacetate salt



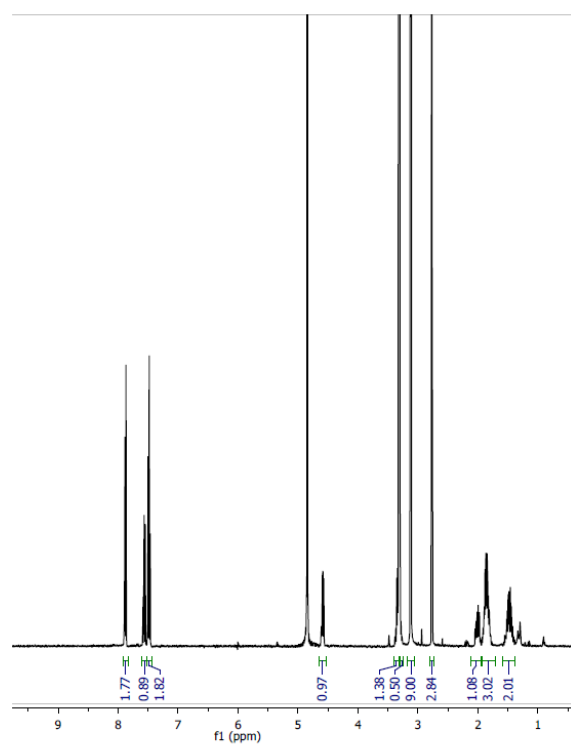
<sup>1</sup>H NMR spectrum of UNC2746 as a TFA salt



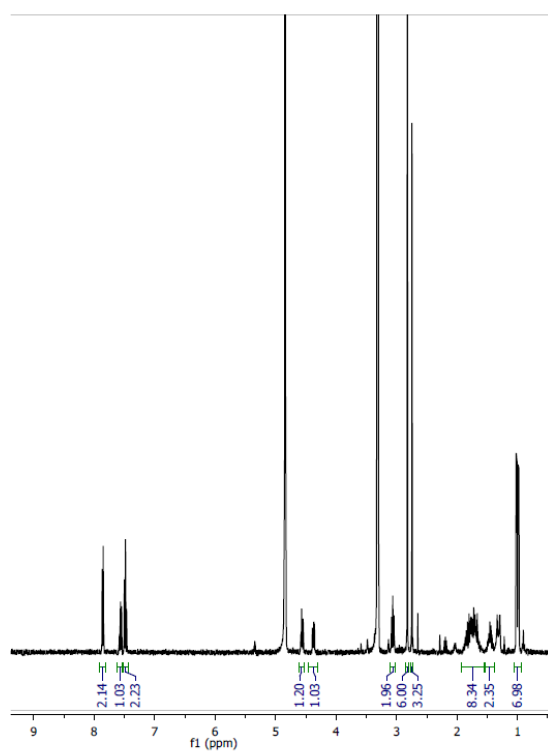
<sup>1</sup>H NMR spectrum of UNC2338 as a trifluoroacetate salt



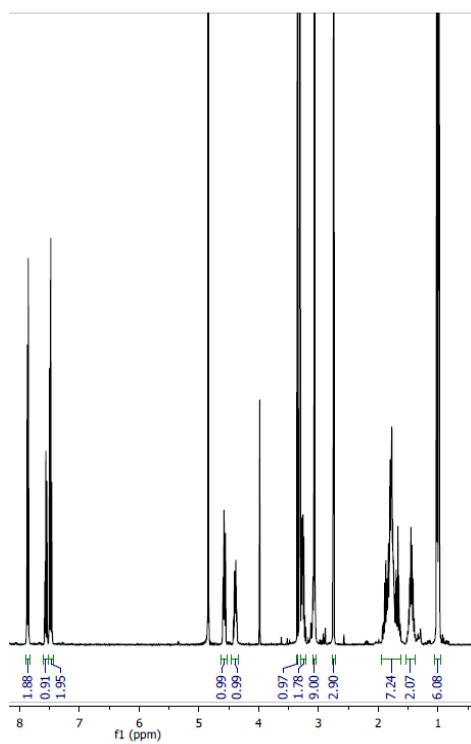
<sup>1</sup>H NMR spectrum of UNC2402 as a trifluoroacetate salt



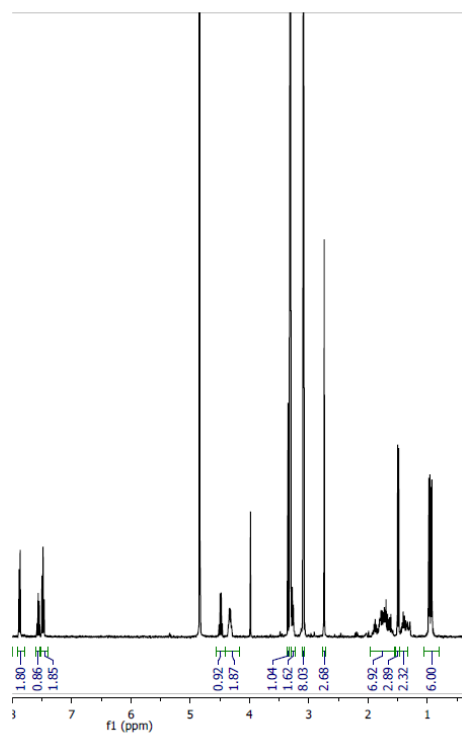
<sup>1</sup>H NMR spectrum of UNC2745 as a TFA salt



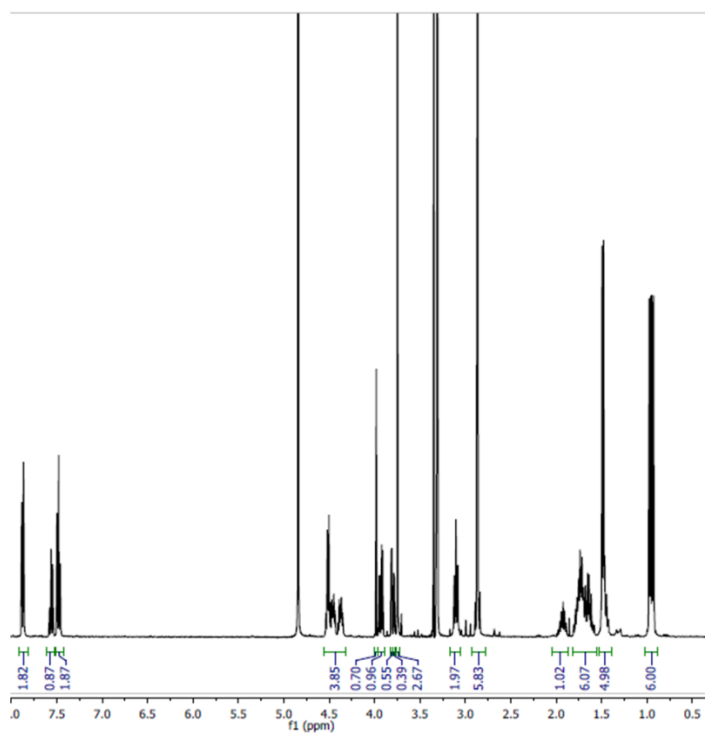
<sup>1</sup>H NMR spectrum of UNC2747 as a trifluoroacetate salt



<sup>1</sup>H NMR spectrum of UNC2597 as a trifluoroacetate salt

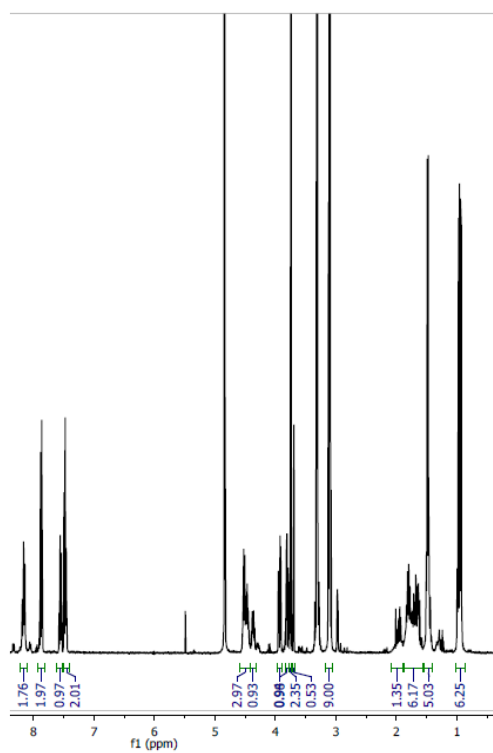


<sup>1</sup>H NMR spectrum of UNC2596 as a TFA salt

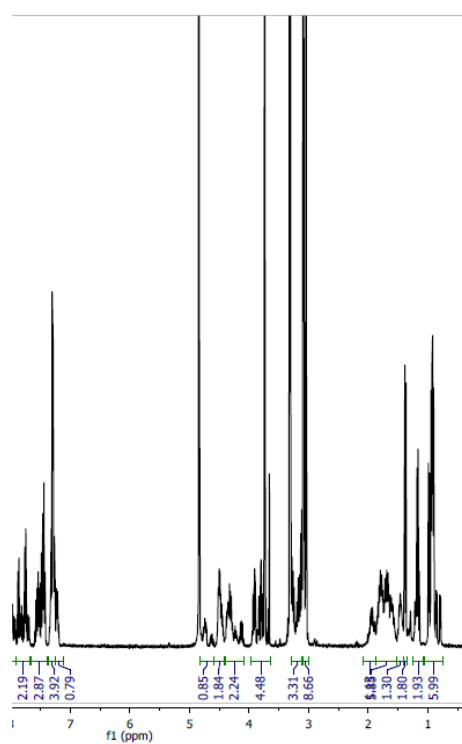




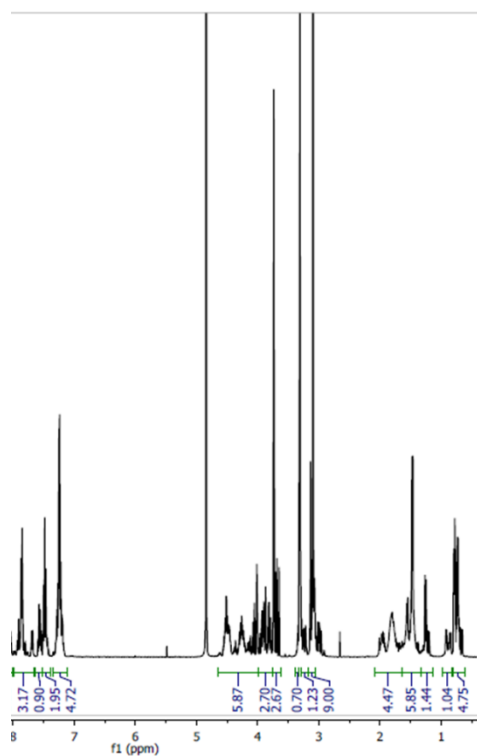
<sup>1</sup>H NMR spectrum of UNC2864 as a trifluoroacetate salt



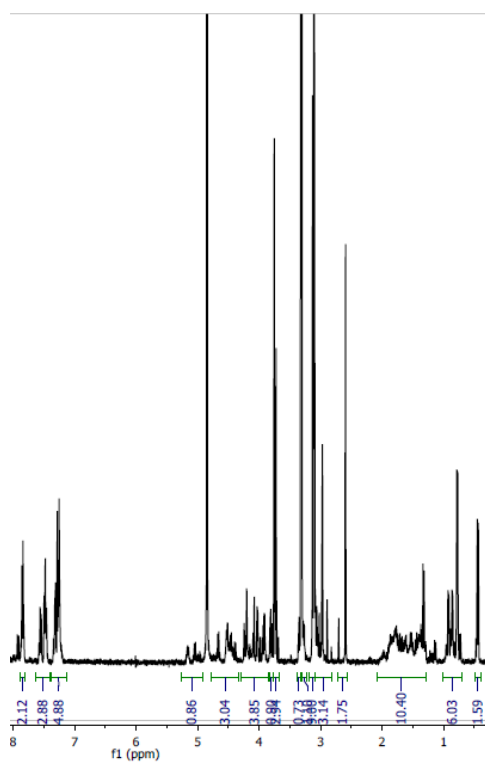
<sup>1</sup>H NMR spectrum of UNC2867 as a trifluoroacetate salt



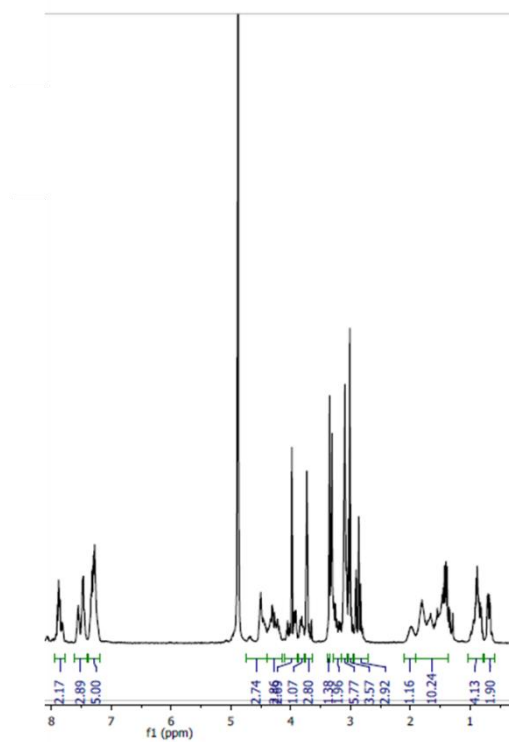
<sup>1</sup>H NMR spectrum of UNC2868 as a trifluoroacetate salt



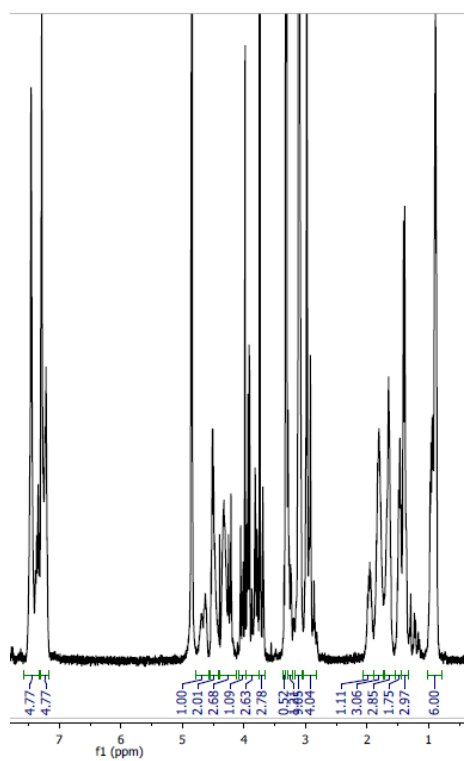
<sup>1</sup>H NMR spectrum of UNC3261 as a trifluoroacetate salt



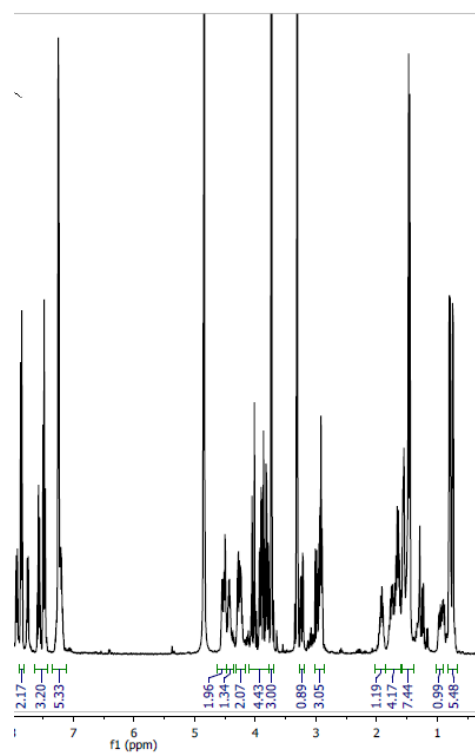
$^1\text{H}$  NMR spectrum of UNC3375 as a trifluoroacetate salt



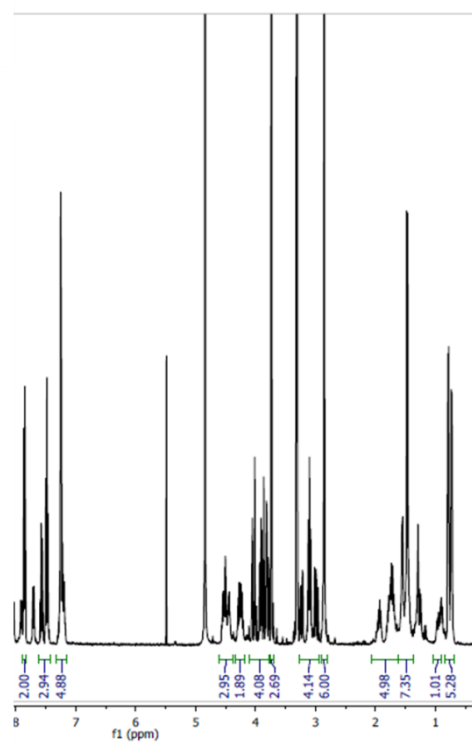
$^1\text{H}$  NMR spectrum of UNC3403 as a trifluoroacetate salt



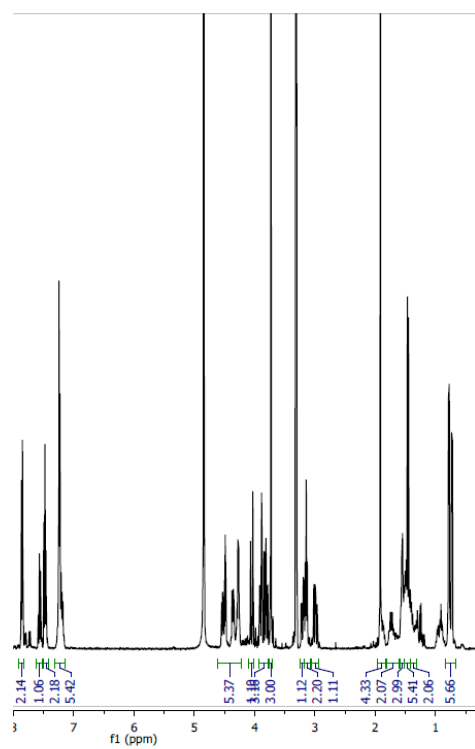
<sup>1</sup>H NMR spectrum of UNC3189 as a TFA salt



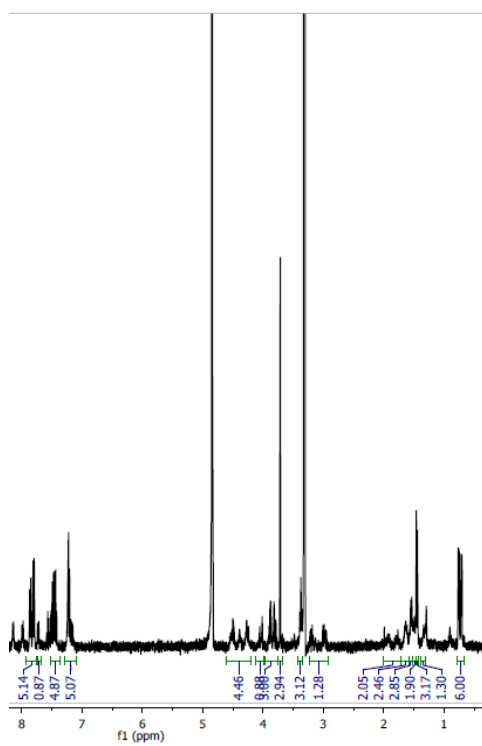
<sup>1</sup>H NMR spectrum of UNC3188 as a TFA salt



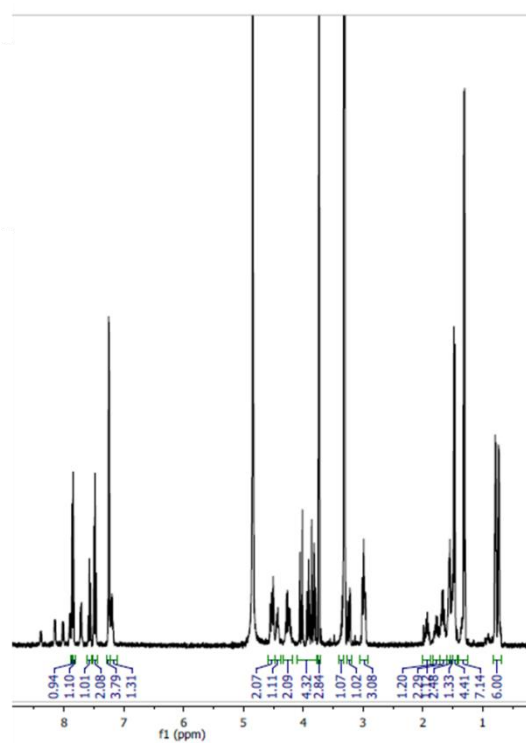
<sup>1</sup>H NMR spectrum of UNC3267



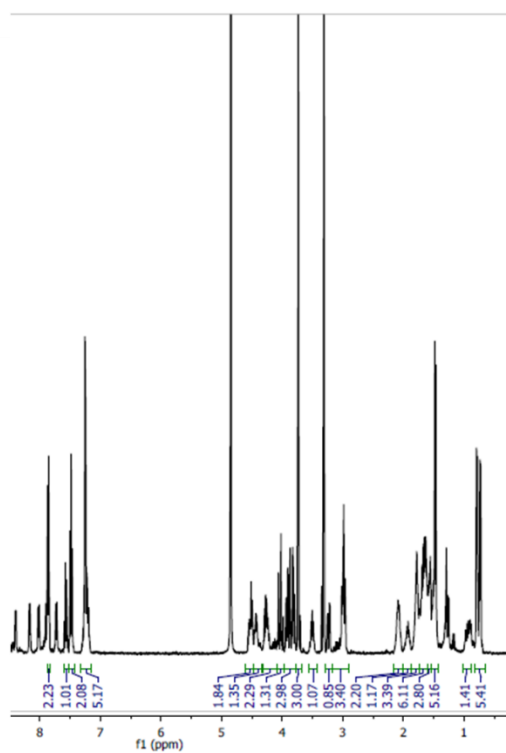
<sup>1</sup>H NMR spectrum of UNC3546



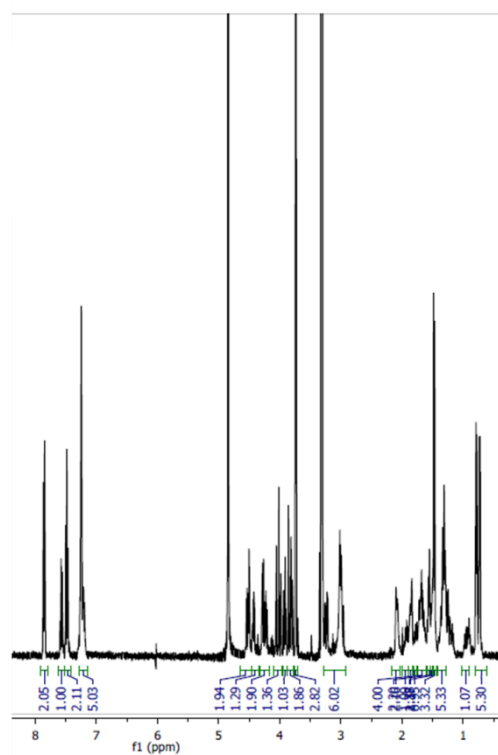
<sup>1</sup>H NMR spectrum of UNC3563 as a TFA salt



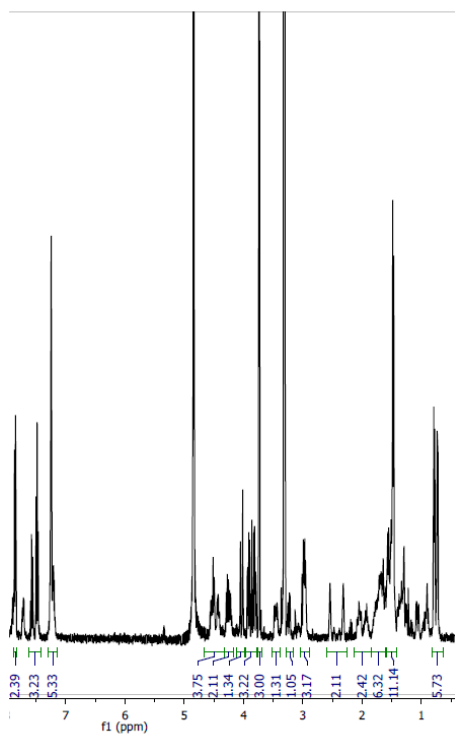
<sup>1</sup>H NMR spectrum of UNC3315 as a TFA salt



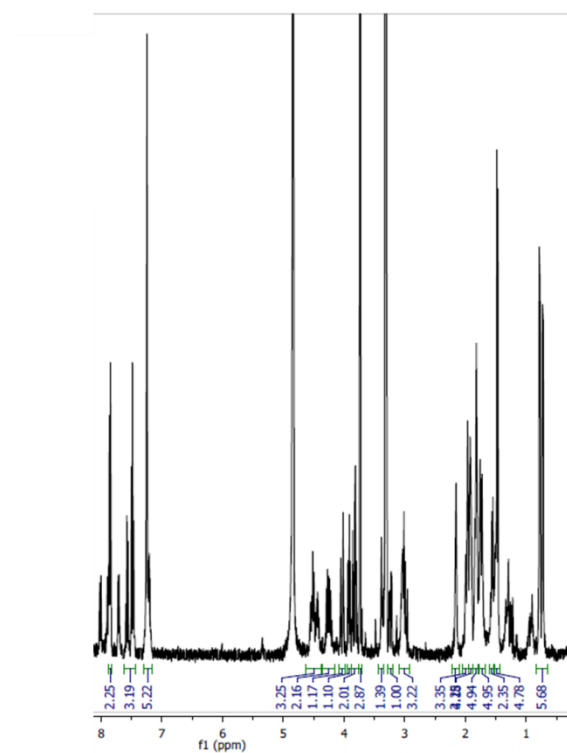
<sup>1</sup>H NMR spectrum of UNC3314 as a TFA salt



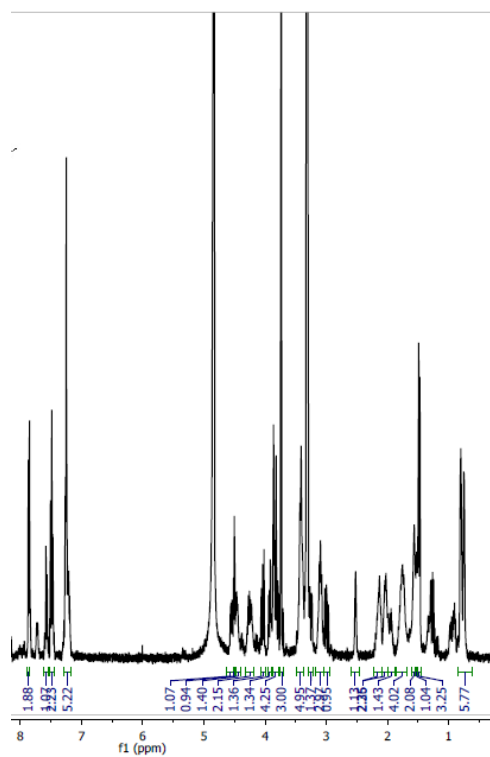
<sup>1</sup>H NMR spectrum of UNC3265 as a TFA salt



$^1\text{H}$  NMR spectrum of UNC3264 as a TFA salt

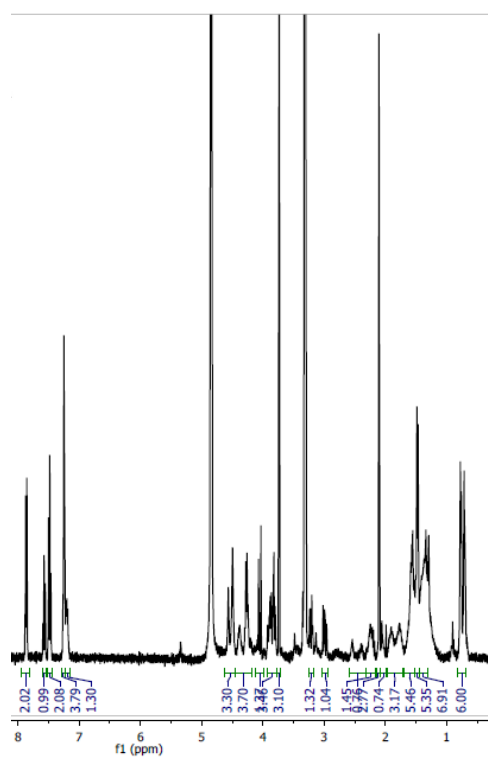


$^1\text{H}$  NMR spectrum of UNC3266 as a TFA salt

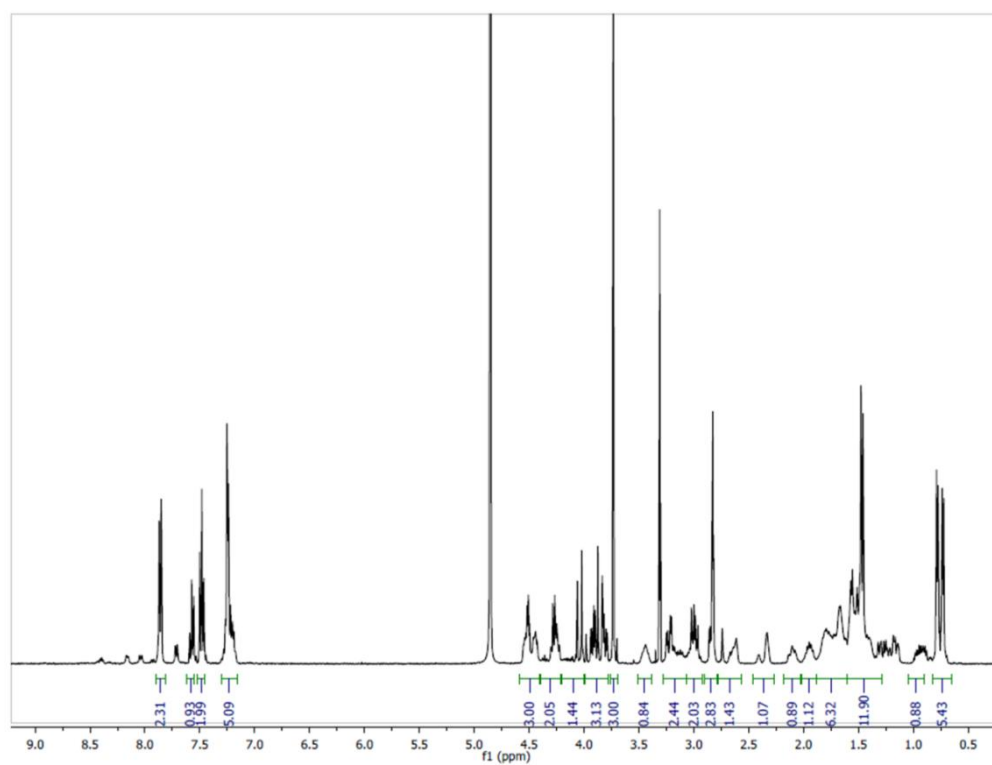




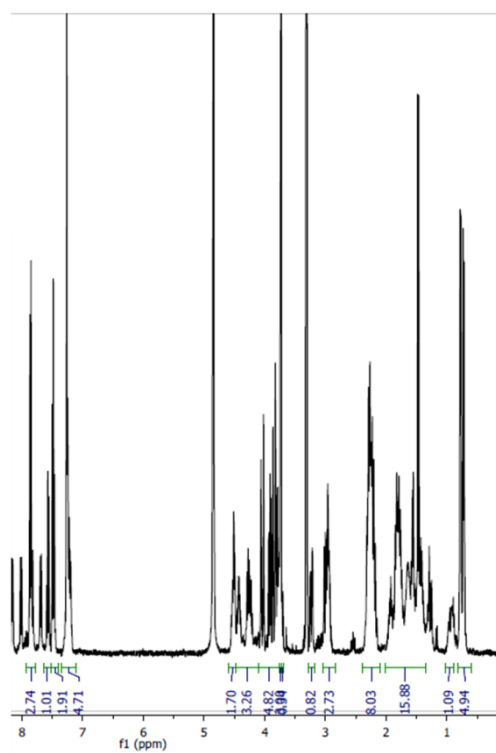
$^1\text{H}$  NMR spectrum of UNC3649



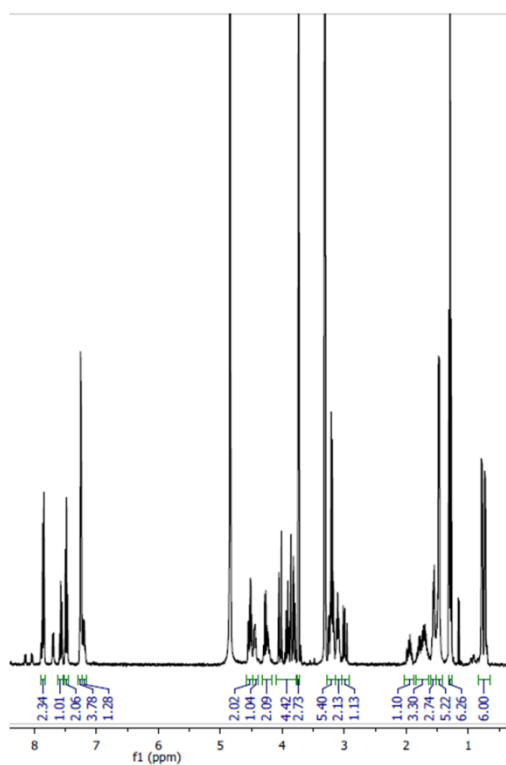
$^1\text{H}$  NMR spectrum of UNC3313 as a TFA salt



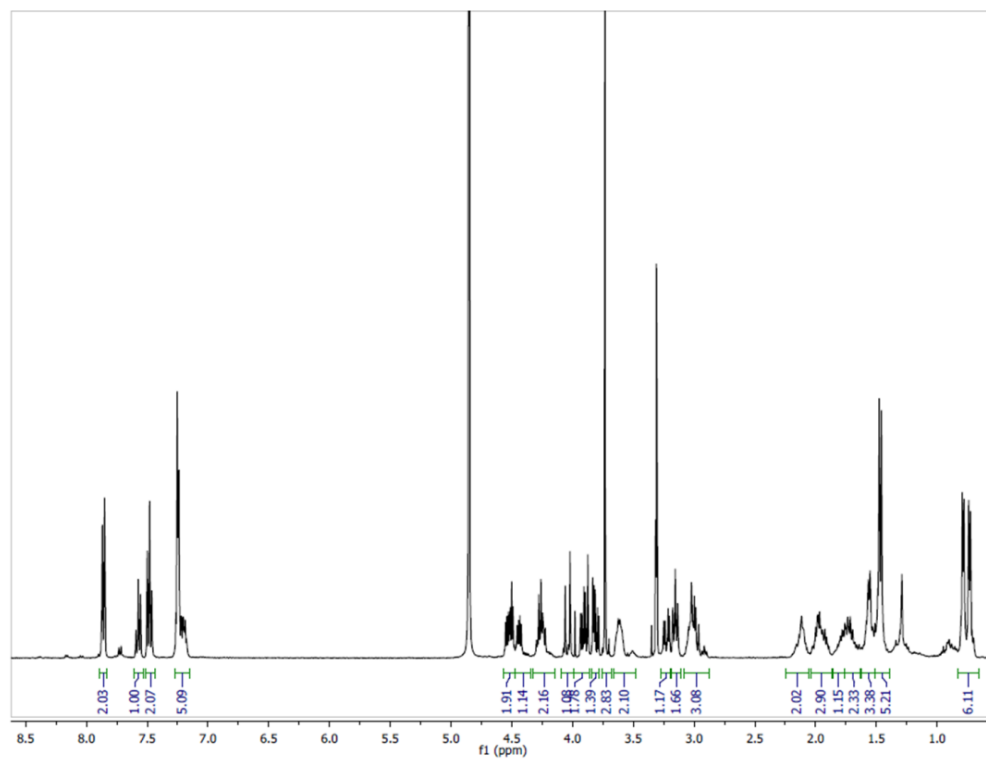
$^1\text{H}$  NMR spectrum of UNC3260 as a TFA salt



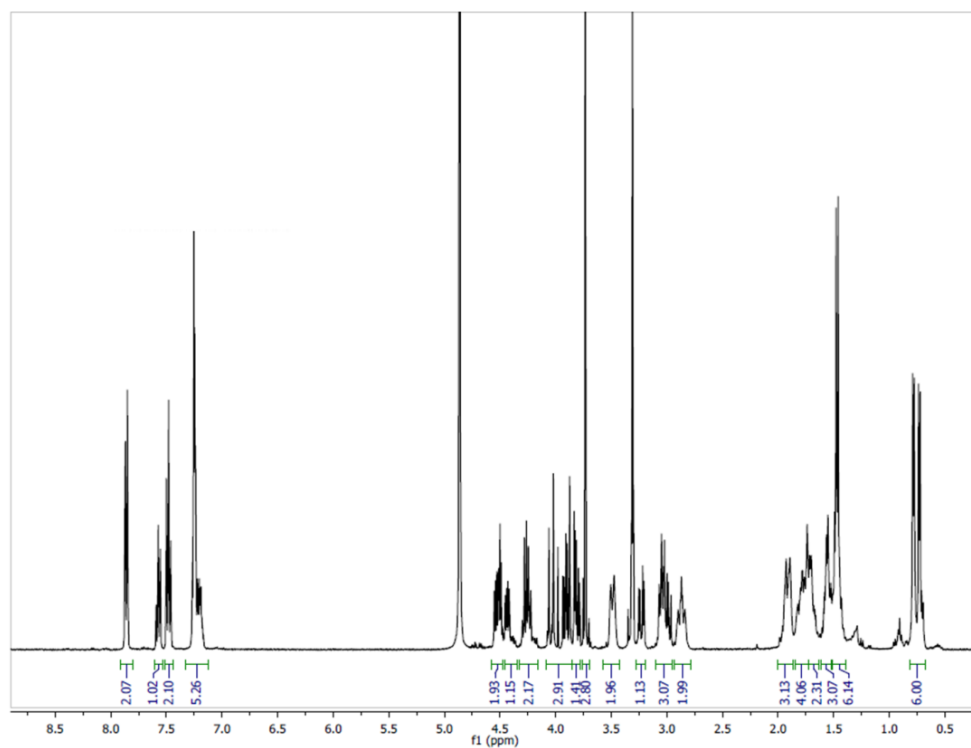
$^1\text{H}$  NMR spectrum of UNC3567 as a TFA salt



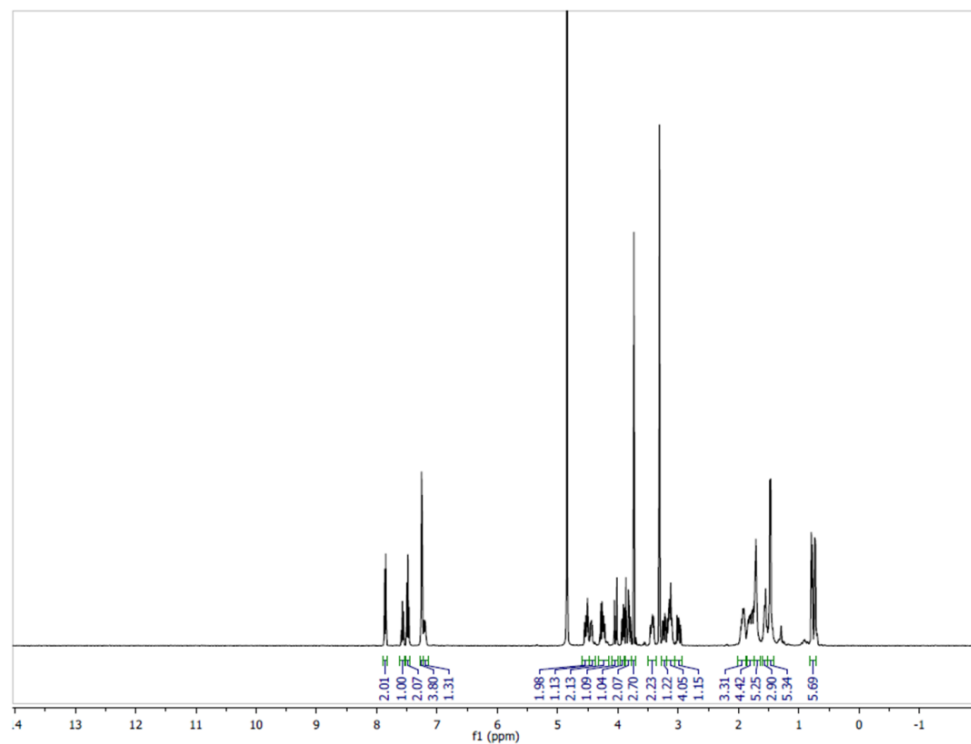
$^1\text{H}$  NMR spectrum of UNC3580 as a TFA salt



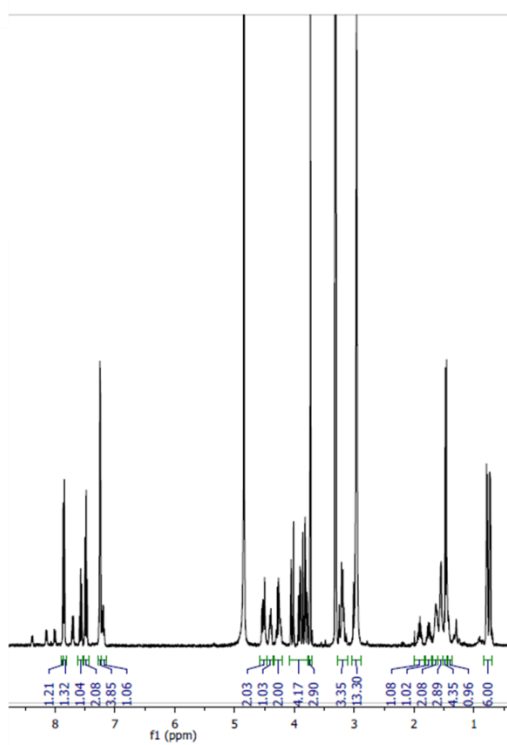
$^1\text{H}$  NMR spectrum of UNC3581 as a TFA salt



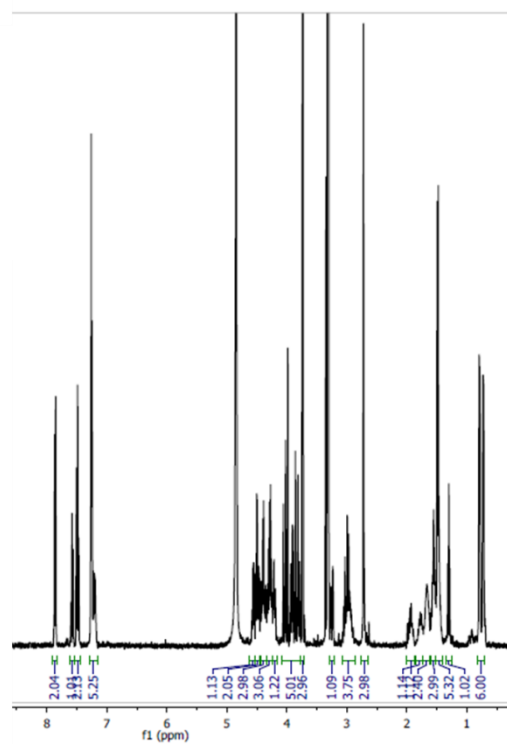
<sup>1</sup>H NMR spectrum of UNC3648 as a TFA salt



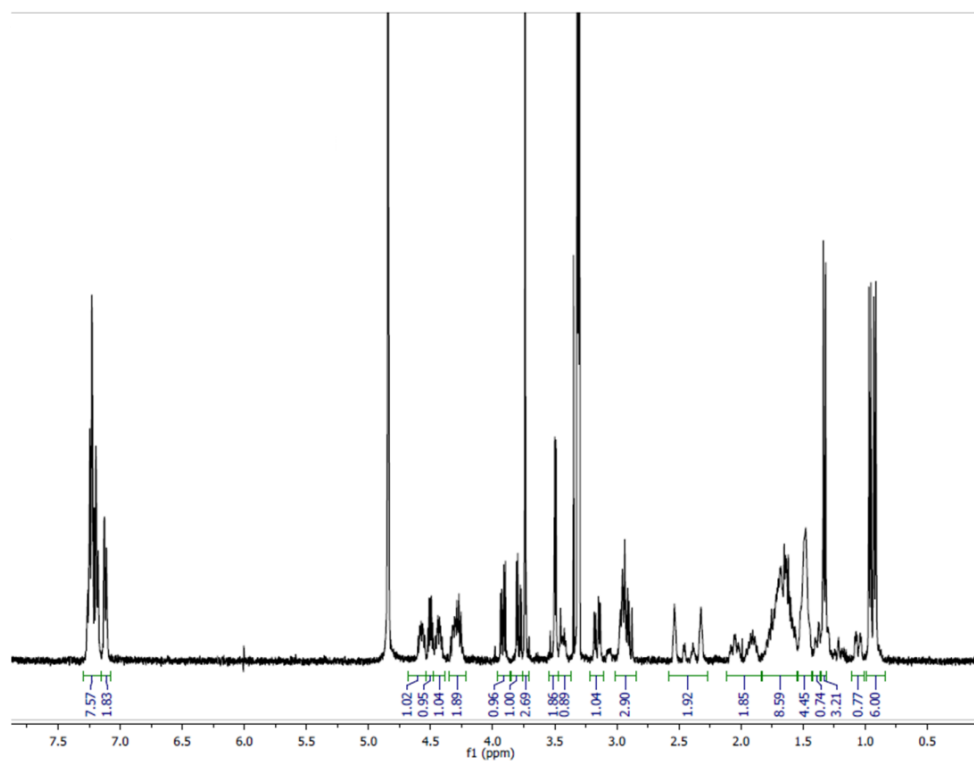
<sup>1</sup>H NMR spectrum of UNC3564 as a TFA salt



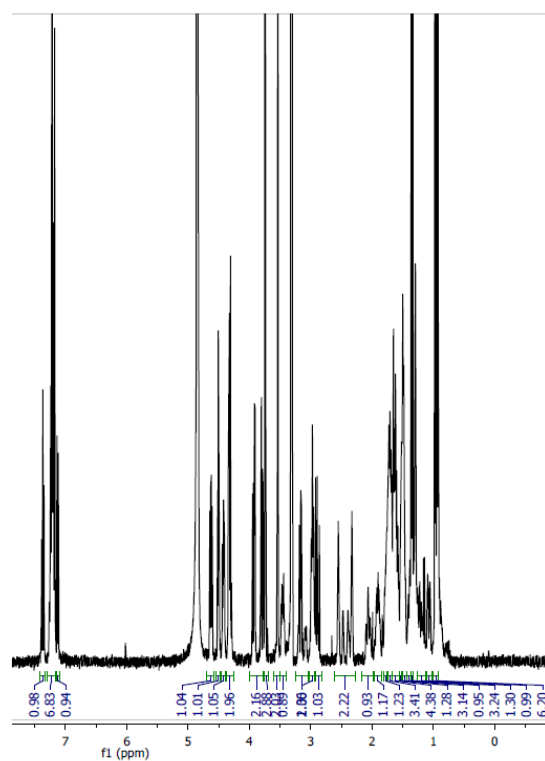
$^1\text{H}$  NMR spectrum of UNC3566 as a TFA salt



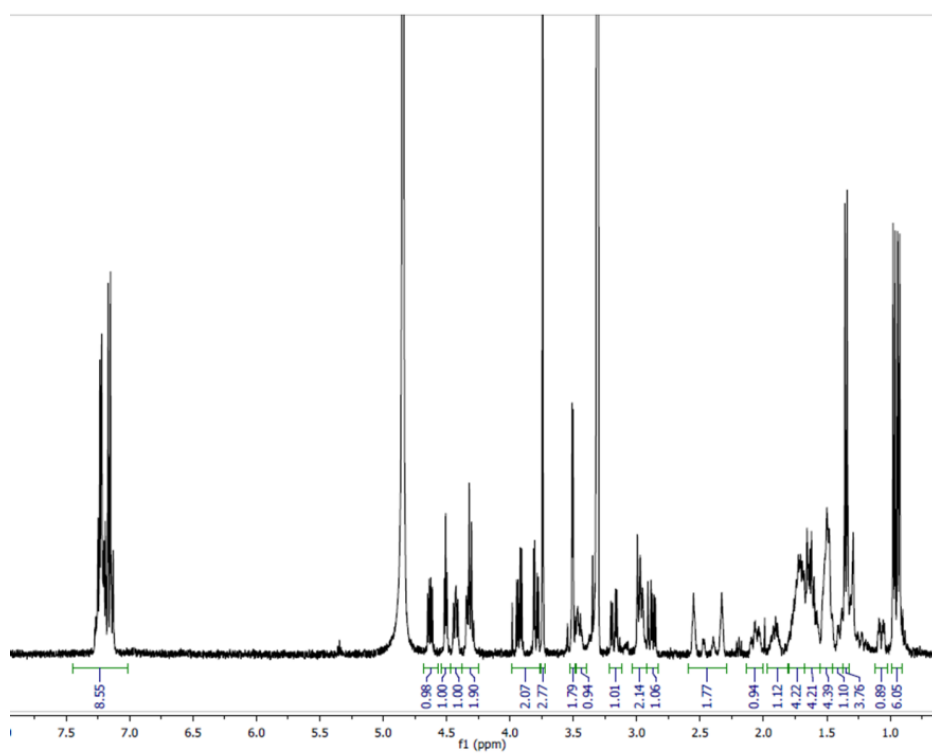
$^1\text{H}$  NMR spectrum of UNC3486 as a TFA salt



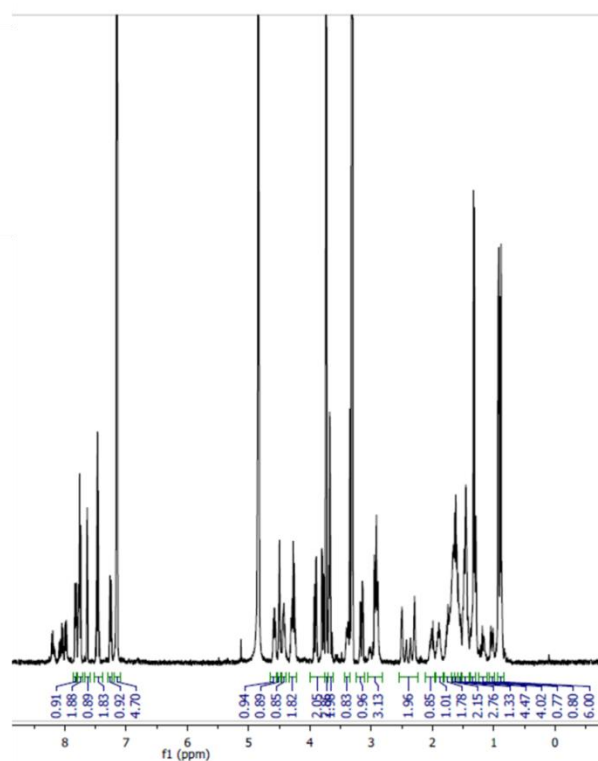
*<sup>1</sup>H NMR spectrum of UNC3562 as a TFA salt*



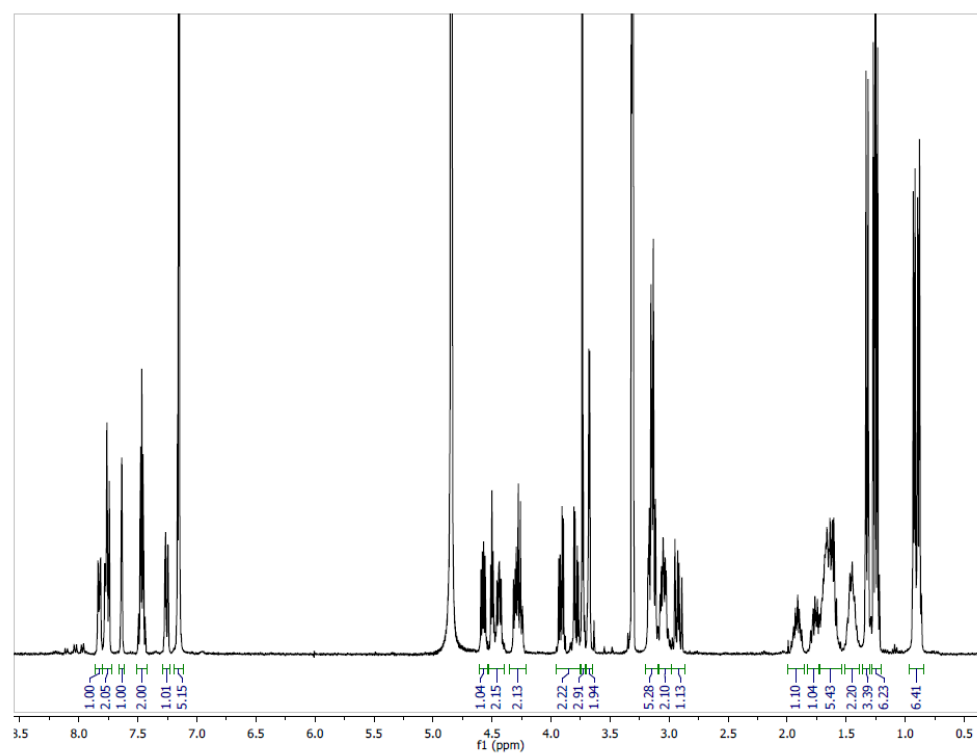
*<sup>1</sup>H NMR spectrum of UNC3565 as a TFA salt*



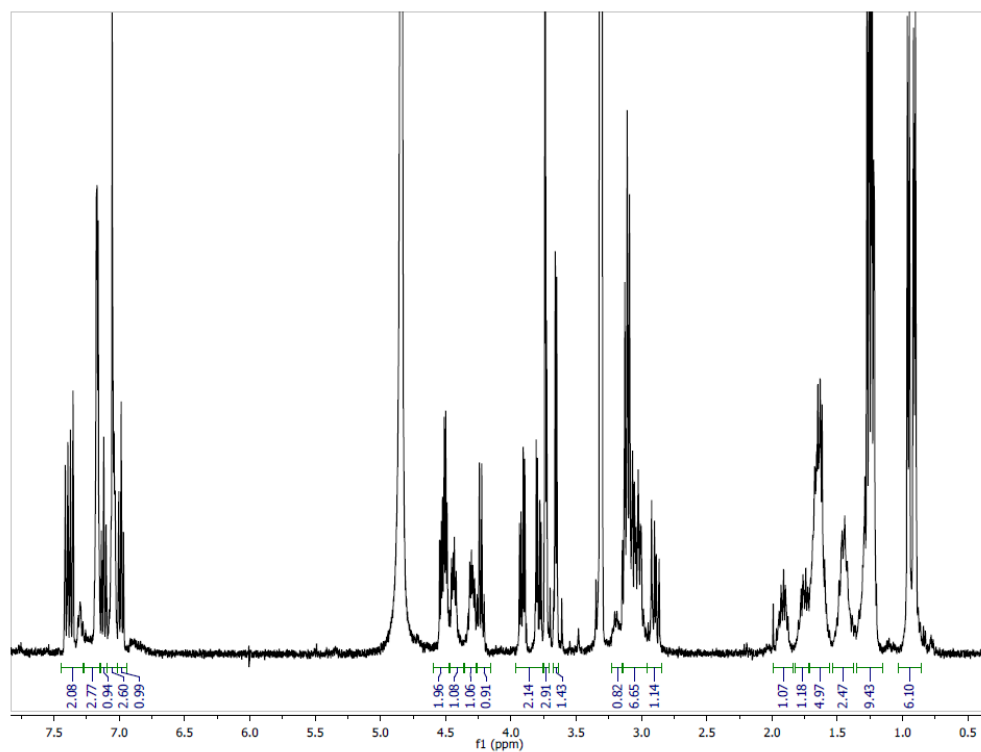
<sup>1</sup>H NMR spectrum of UNC3556 as a TFA salt



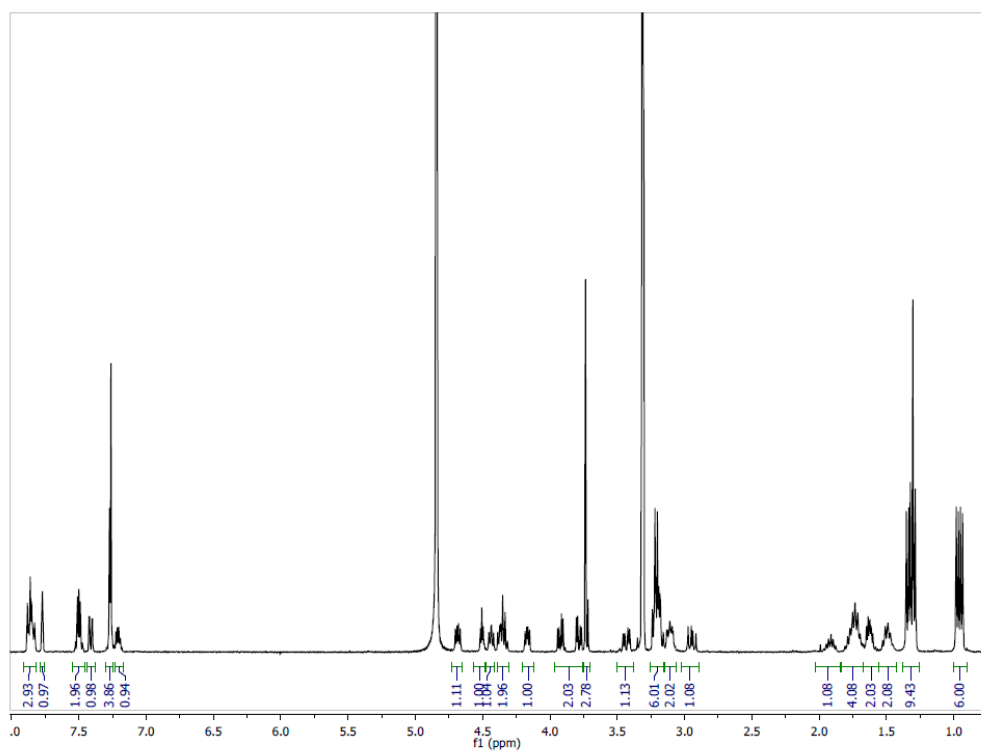
<sup>1</sup>H NMR spectrum of UNC3772 as a TFA salt



<sup>1</sup>H NMR spectrum of UNC3771 as a TFA salt

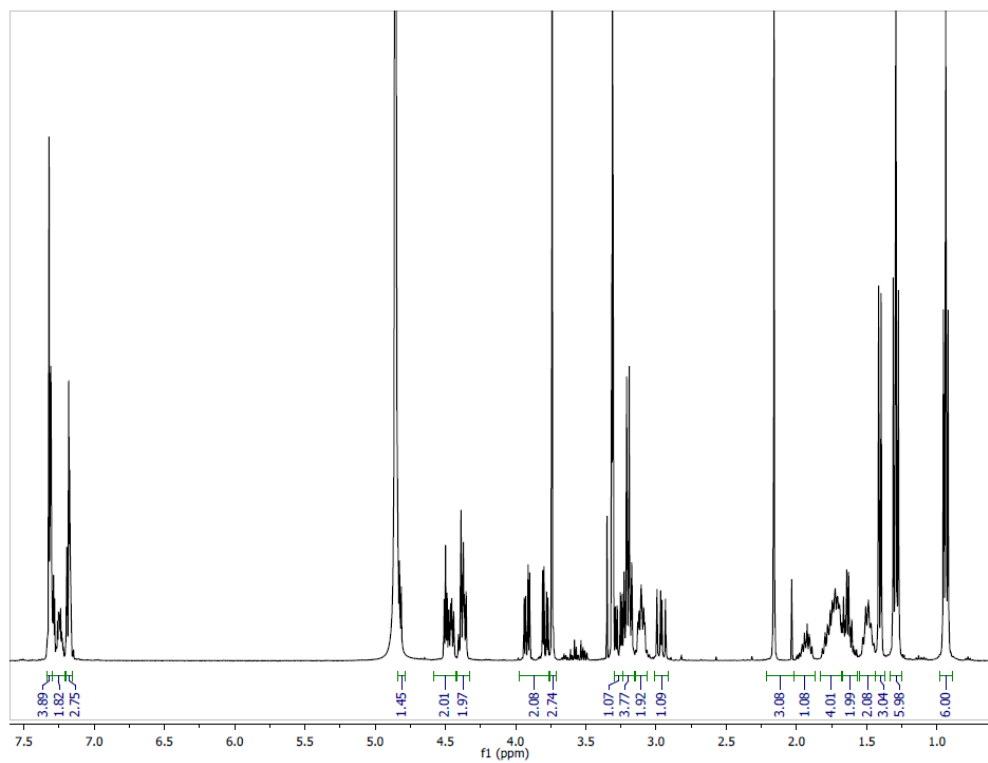


<sup>1</sup>H NMR spectrum of UNC3773 as a TFA salt

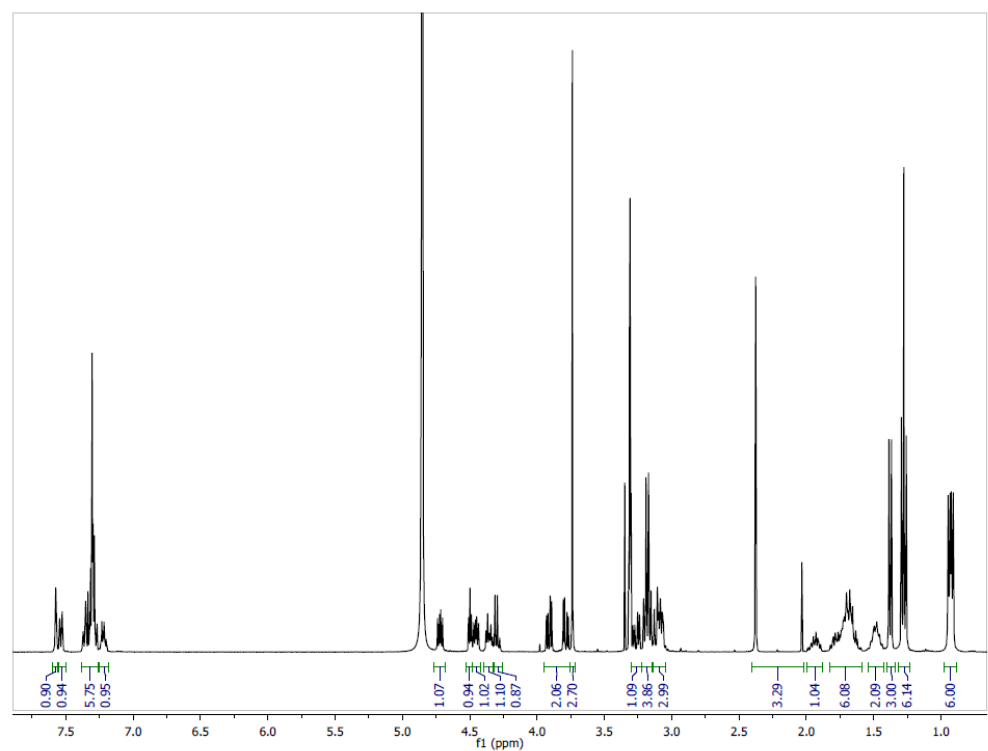




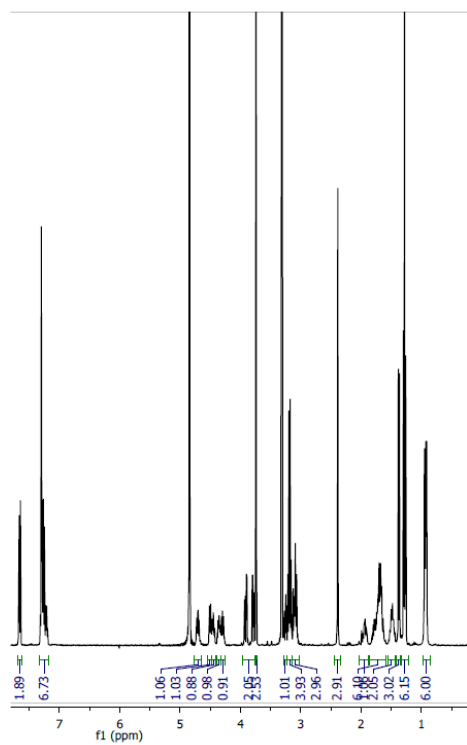
$^1\text{H}$  NMR spectrum of UNC4973 as a TFA salt



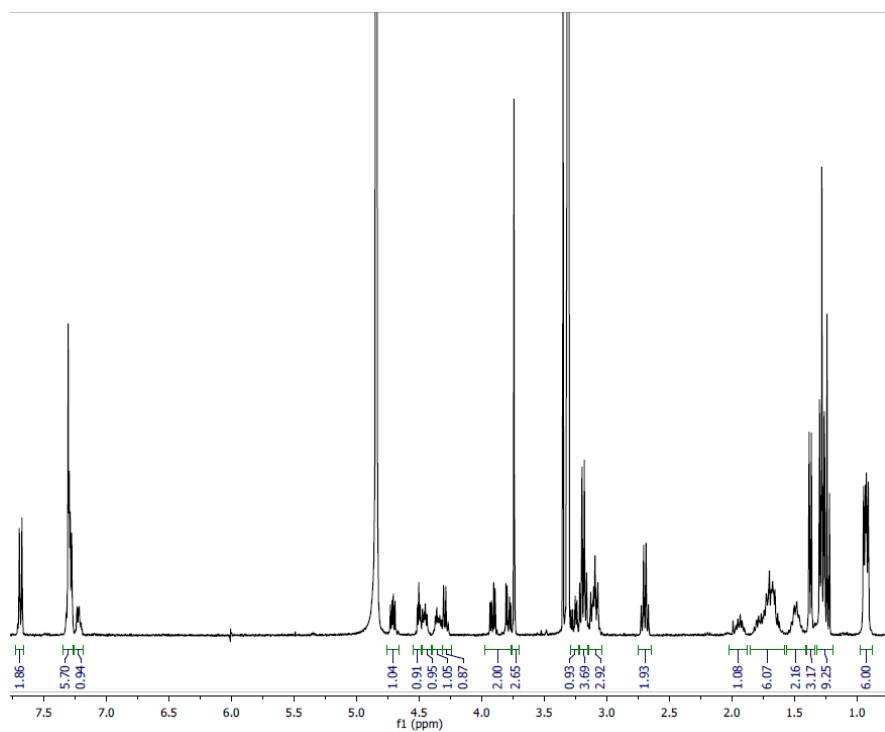
$^1\text{H}$  NMR spectrum of UNC4974 as a TFA salt



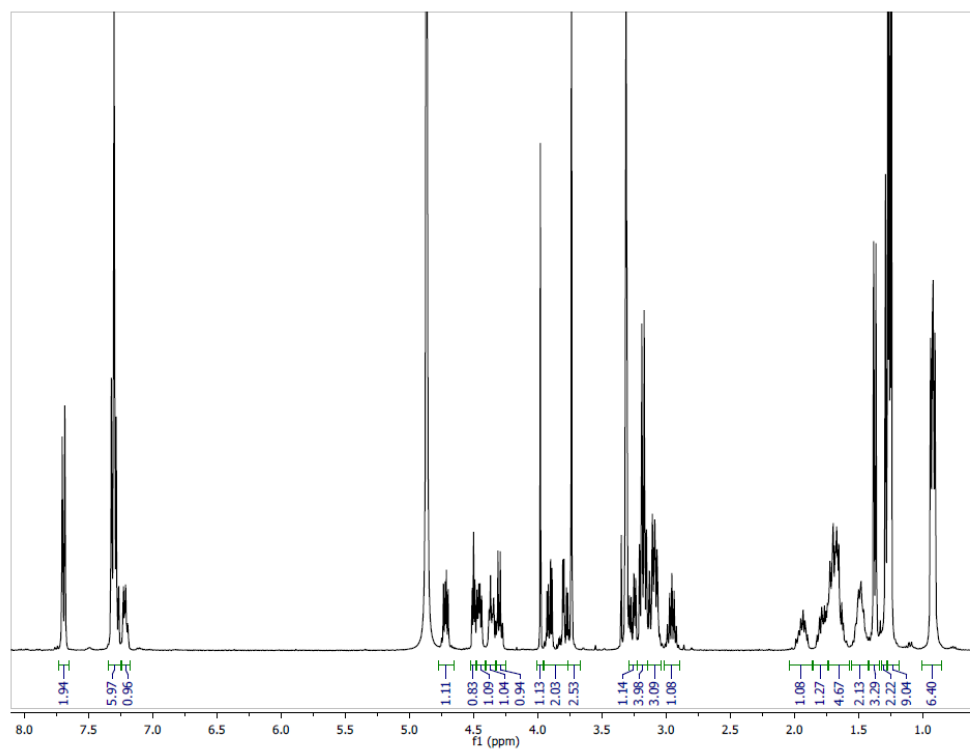
<sup>1</sup>H NMR spectrum of UNC3867 as a TFA salt



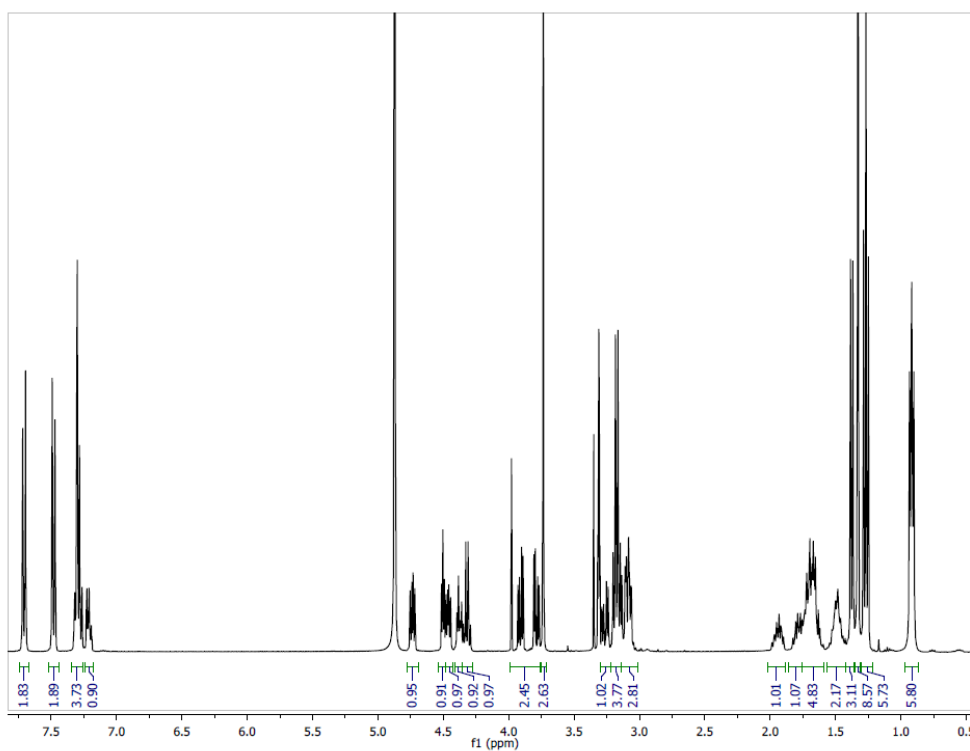
<sup>1</sup>H NMR spectrum of UNC3769 as a TFA salt



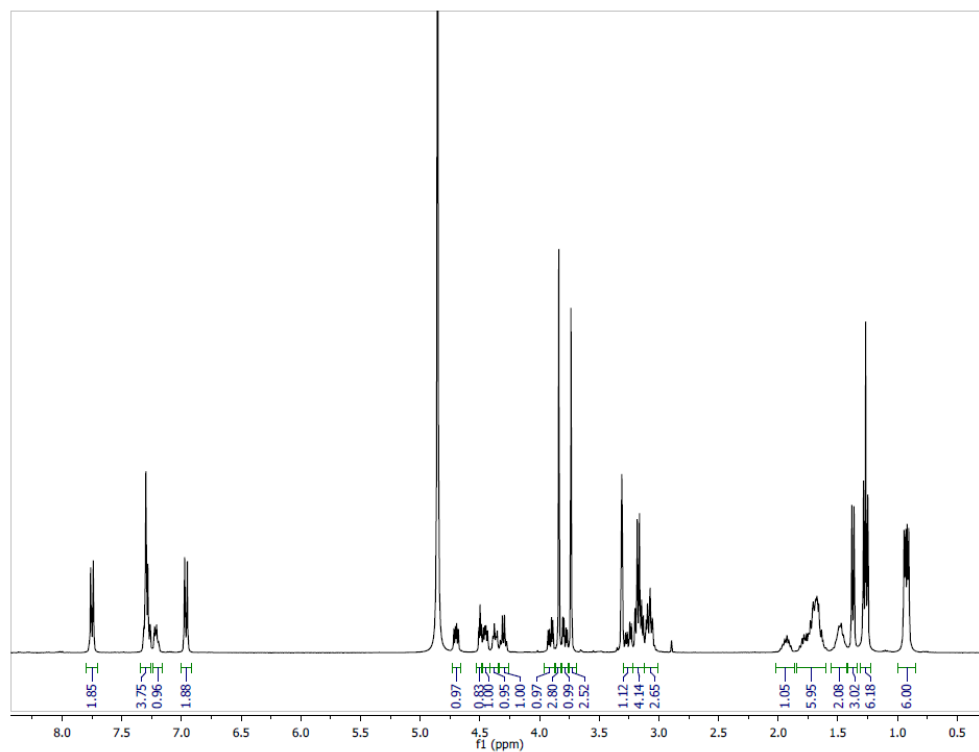
<sup>1</sup>H NMR spectrum of UNC3865 as a TFA salt



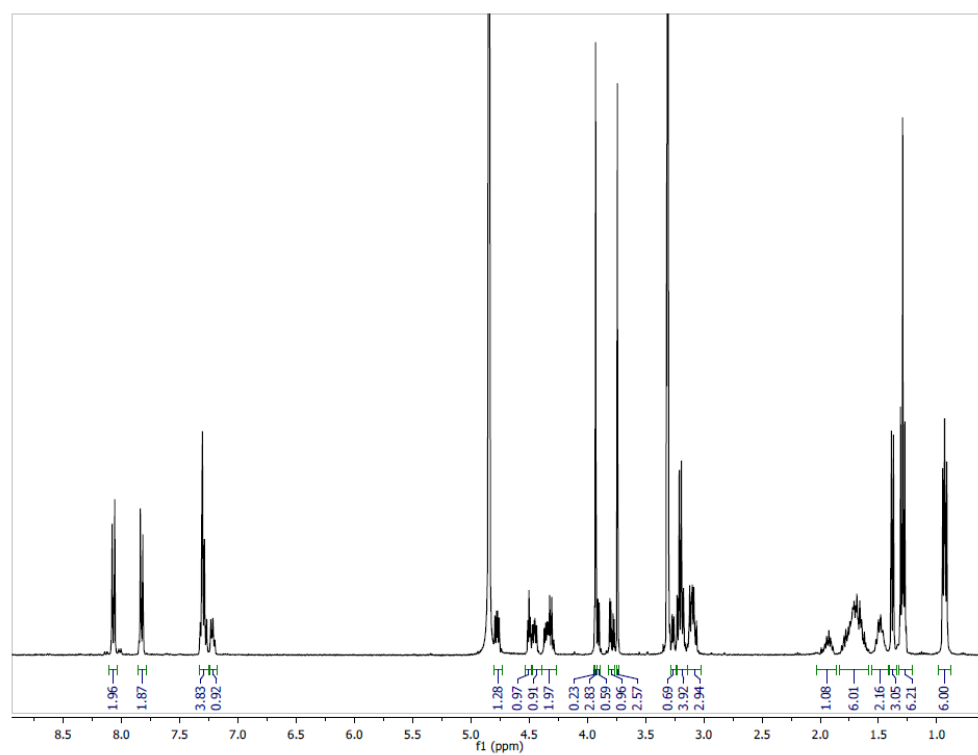
<sup>1</sup>H NMR spectrum of UNC3866 as a TFA salt



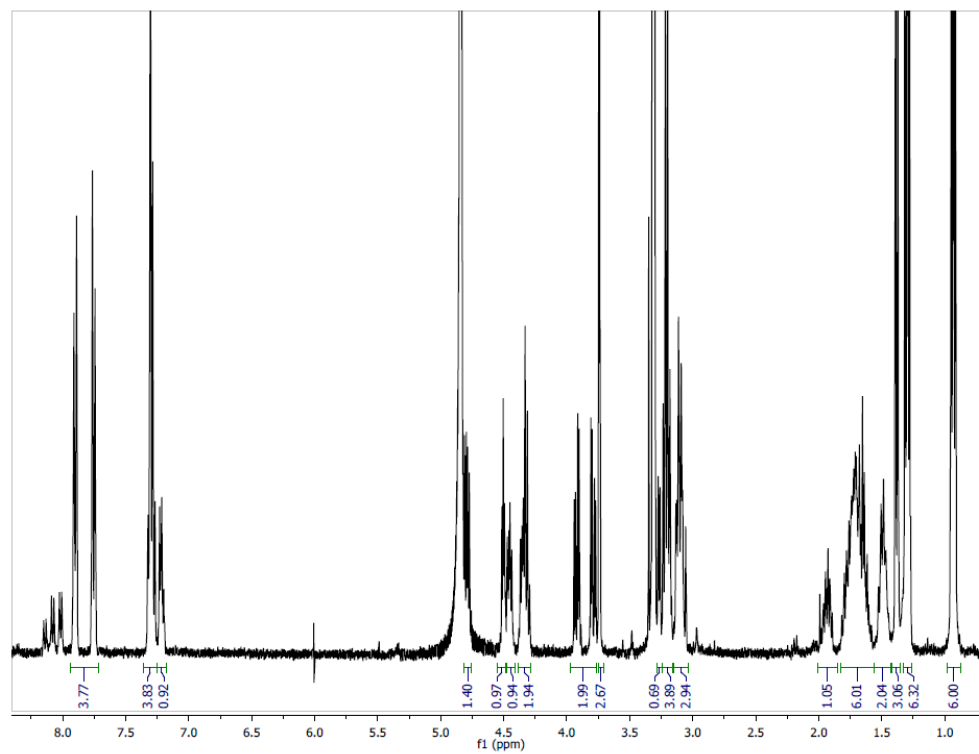
<sup>1</sup>H NMR spectrum of UNC4975 as a TFA salt



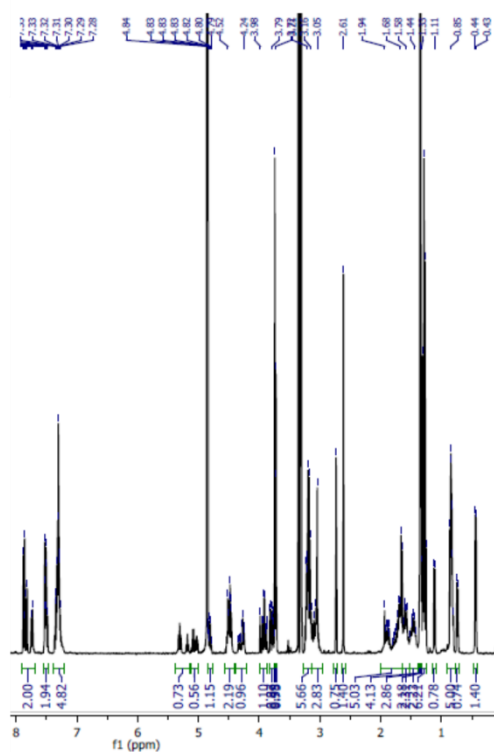
<sup>1</sup>H NMR spectrum of UNC3768 as a TFA salt



$^1\text{H}$  NMR spectrum of UNC3770 as a TFA salt



$^1\text{H}$  NMR spectrum of UNC4219 as a TFA salt



## CHAPTER III: BIOCHEMICAL AND BIOPHYSICAL CHARACTERIZATION OF CBX ANTAGONISTS

### INTRODUCTION

The development of UNC3866 was an important step toward a biologically useful CBX7 antagonist. However, while achieving high affinity for the target of interest is a critical, potency is only one of several parameters that define a high quality chemical probe. Qualification of a molecule for use as a chemical probe requires that the molecule be thoroughly characterized to the extent that any phenotypic effects produced by the molecule can be reliably attributed to modulation of its intended target. Efforts to characterize molecules for use as a chemical probe should be focused on the five core principles that undergird any quality chemical probe discovery effort<sup>58</sup>:

- (1) Molecular profiling
- (2) Mechanism of action
- (3) Identity of the active species
- (4) Proven utility as a probe
- (5) Availability

Molecular profiling involves providing “sufficient *in vitro* potency and selectivity data to associate the *in vitro* profile of a molecule with its cellular or *in vivo* profile”<sup>58</sup>. Our studies in Chapter II laid the initial foundations for UNC3866 with respect to this principle. These studies demonstrated on-target potency of UNC3866 as an H3-competitive antagonist of CBX7 (IC<sub>50</sub> ~60 nM). We further validated this interaction by ITC ( $K_d$  ~97 nM) and X-ray crystallography. Our studies in this chapter were focused on complete satisfaction of this principle through:

- (1) Characterization of the *in vitro* selectivity of the UNC3866 versus the CBX7 orthologues, CBX2, -4, -6 and -8, as well as against a broader panel of molecular targets
- (2) Kinetic characterization of the interaction between CBX7 and UNC3866.

Principles 2-5 as they relate to UNC3866 will be addressed in Chapter IV.

Key to our initial decision to focus on CBX7 was the availability of a high-throughput screening assay already developed in our lab<sup>66, 69, 74, 75</sup>. We had previously been unsuccessful in developing high throughput assays for other CBX chromodomains with H3 substrates, precluding rapid, cost-effective selectivity profiling of our compounds against these orthologous chromodomains. The development of UNC3866 as a potent CBX7 antagonist made overcoming this challenge a necessity that we successfully addressed, as discussed in this chapter, in order to cost-effectively evaluate and understand the selectivity profile for UNC3866 amongst the Pc CBX chromodomains.

Our purpose for characterizing the kinetics of the UNC3866-CBX7 interaction was two-fold. First, it is becoming increasingly apparent that the cellular and *in vivo* efficacy of many compounds is best correlated with the kinetics of their interaction with the intended molecular target (specifically with the dissociative kinetics of the interaction)<sup>107</sup>. Therefore, knowledge of this binding parameter is critical for the development of high quality chemical probes. Secondly, we sought to use our kinetic characterization of the interaction of UNC3866 with CBX7 to enable further interrogation of the binding mechanism proposed by our molecular dynamics simulations described in Chapter II that were critical for the development of UNC3866. An understanding of the kinetics of the CBX7-UNC3866 interaction is important for further validation of this hypothesis in order to continue its application in the design of improved CBX antagonists.

Herein we describe our efforts to more fully characterize UNC3866 to enable its use as a chemical probe for the chromodomain of CBX7. We profiled UNC3866 against a panel of >250 potential molecular targets, and found that it is highly selective for two families of chromodomains. We developed high throughput assays for these families of chromodomains, providing important insight into the selectivity profile of UNC3866 and its related analogs. In addition to our screening efforts, we also solved co-crystal structures of UNC3866 with multiple chromodomains as a means to further understand the selectivity determinants for these proteins. Finally, we studied the binding kinetics of a subset of our CBX7 antagonists by surface plasmon resonance experiments (SPR). These experiments provided critical evidence supporting our induced fit binding mechanism that enabled the design of a second-generation CBX7 antagonist, UNC4976, which displays dramatically increased target residence time in SPR experiments compared to UNC3866.

## RESULTS AND DISCUSSION

### *UNC3866 selectivity*

Knowledge of the activity profile of a molecule is essential in order to associate its cellular effects with modulation of a specific molecular target(s) of interest. Thus, the use and classification of a molecule as a chemical probe is dependent upon an in-depth understanding of its activity against a broad panel of potential molecular targets<sup>58</sup>. Accordingly, we profiled the selectivity of UNC3866 against a wide range of epigenetic and non-epigenetic targets. First, we focused on the selectivity of UNC3866 against our in-house panel of Kme reader domains. This panel consists of a small set of Kme readers encompassing proteins from the MBT, PHD and Tudor families for which we have established AlphaScreen® and/or ITC assays (**Table 3.1**). Encouragingly, UNC3866 is more than 100-fold selective for CBX7 over the nine members of this Kme reader panel.

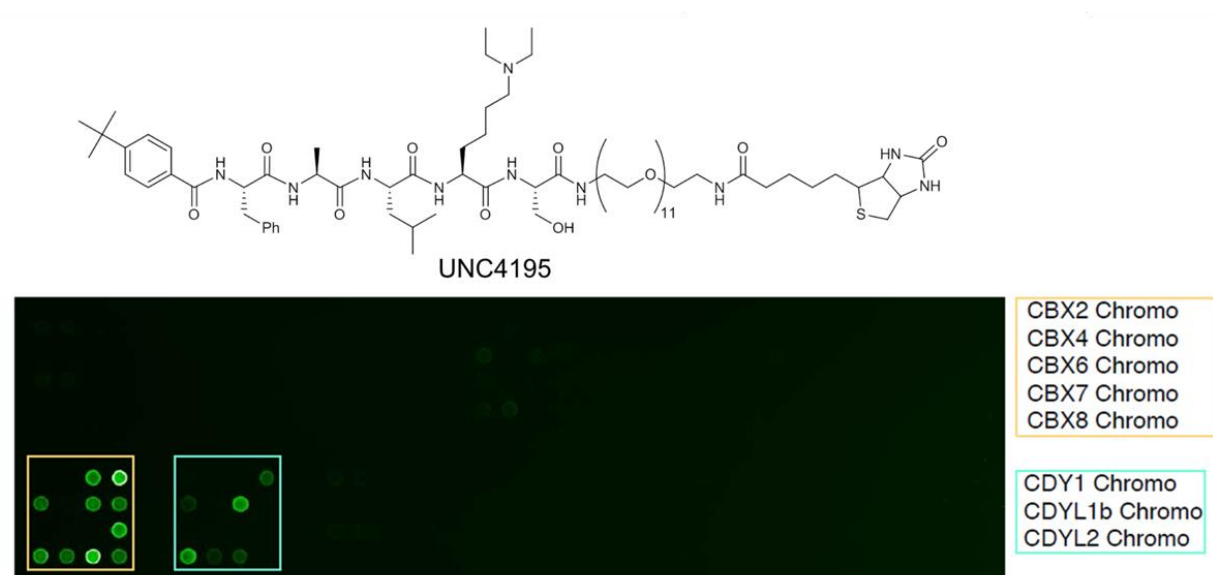
**Table 3.1.** Characterization of UNC3866 selectivity by AlphaScreen® or ITC against a panel of Kme reader domains.

ID	AlphaScreen® IC <sub>50</sub> (nM)								ITC K <sub>d</sub> (nM)	
	CBX7	53BP1	JARID1A	PHF1	PHF19	PHF23	UHRF1	MBTD1	L3MBTL1	L3MBTL1
UNC3866	64 ± 1.0	>10,000	>10,000	>10,000	>10,000	>10,000	>10,000	>10,000	>10,000	>10,000

\*IC<sub>50</sub> represents the average of at least 2 replicates ± the standard deviation.

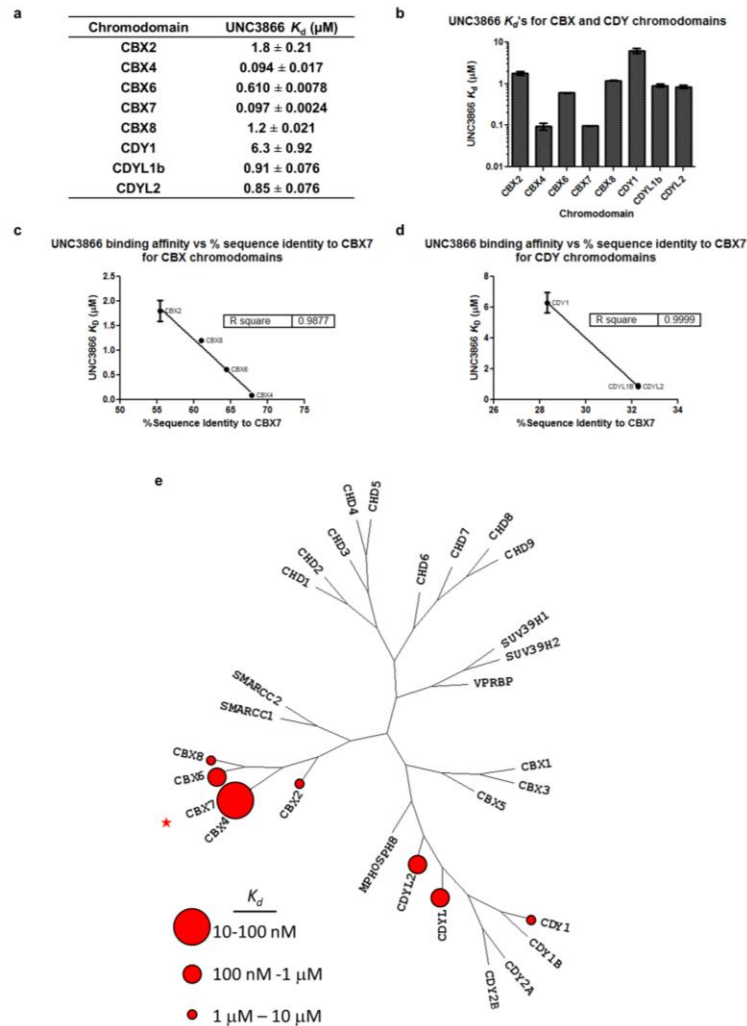
We next utilized UNC4195, our biotinylated analog of UNC3866, in order to facilitate further selectivity studies. In collaboration with Mark Bedford at MD Anderson, we used this derivative to evaluate binding to 96 purified chromatin-associated effector proteins, including at least 19 known H3K9me3 or H3K27me3 binding proteins and 28 distinct chromodomains, spotted in duplicate onto a nitrocellulose coated membrane (**Fig. 3.1**) Binding interactions were visualized with fluorescently tagged streptavidin, revealing that UNC4195 binds to two families of chromodomains. As expected, a positive binding interaction was observed with CBX7 and the other PRC1 orthologues, CBX2, -4, -6 and -8. We did not detect binding to CBX1, -3 and -5 (also known as HP1β, -γ and -α, respectively) which are highly homologous to CBX7 in their chromodomains. The array also revealed binding to three members of the CDY family of chromodomains: CDY1, CDYL1b, and CDYL2.





**Figure 3.1.** Characterization of UNC3866 selectivity with UNC4195. UNC4195 selectively interacts with 2 subfamilies of chromodomains on a microarray containing 96 epigenetic reader proteins across multiple domain types (Bedford lab, MD Anderson).

We expressed and purified all proteins that demonstrated positive binding to UNC4195 on the array and subsequently quantified the affinity for UNC3866 by ITC (**Fig. 3.2**). In agreement with the microarray data, we detected binding to each of these chromodomains. The affinity of UNC3866 for CBX2, -4, -6 and -8 is also surprisingly well correlated with the percent sequence identity of each chromodomain relative to that of CBX7 (**Fig. 3.2c**). UNC3866 is equipotent for CBX4, which is most similar to CBX7, while it is 18-, 6- and 12-fold selective for CBX4/7 over CBX2, -6 and -8, respectively. Additionally, UNC3866 is 65-fold selective for CBX4/7 over CDY1 and 9-fold selective for CBX4/7 over CDYL1b and CDYL2. While some work has been done to explore the biochemistry of CDY proteins<sup>67</sup>, their biological function is less well characterized than Polycomb CBX proteins which makes assessing the potential consequences of antagonizing these proteins in a cellular context challenging, making this is a “known unknown” in the use of UNC3866.



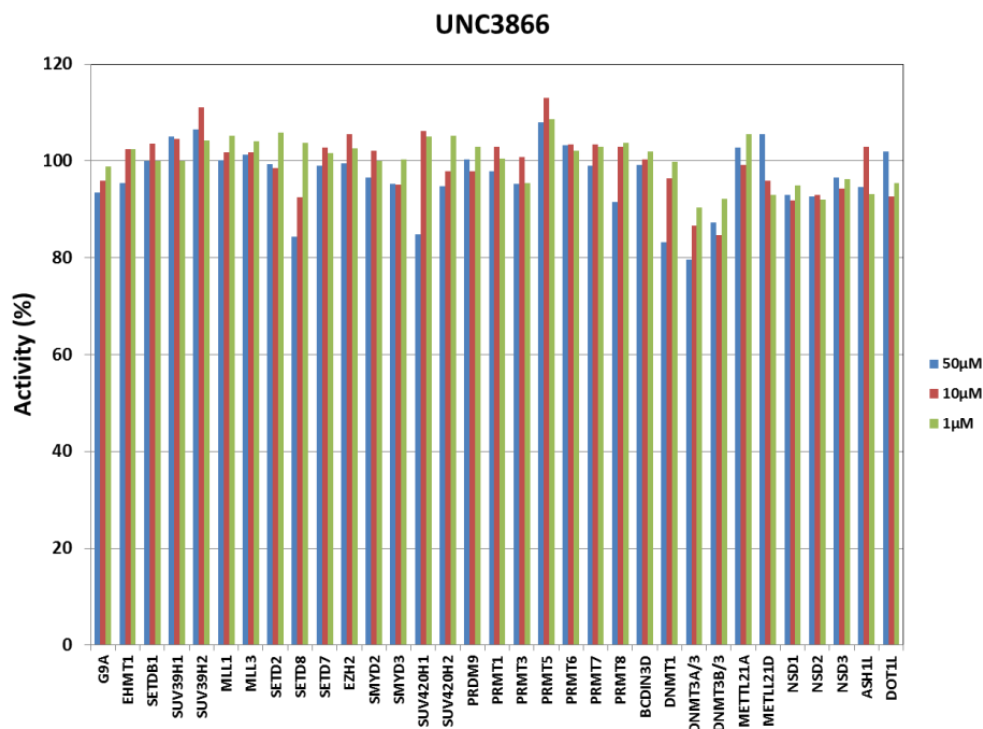
**Figure 3.2.** Selectivity studies with UNC3866 and its chromodomain targets. **(a)** Binding constants of UNC3866 between each of its chromodomain targets identified on the microarray (**Fig. 3.1**). Data are reported as the mean  $\pm$  the standard deviation of at least two replicates. **(b)** Graphical illustration of the UNC3866 binding constants for its chromodomain targets. **(c)** Comparison of the UNC3866 binding constant for each CBX chromodomain bound by UNC3866 versus that chromodomains % sequence identity to CBX7. **(d)** Comparison of the UNC3866 binding constant for each CDY chromodomain bound by UNC3866 versus that chromodomain's % sequence identity to CBX7. **(e)** Chromodomain phylogenetic tree and the relative affinity of UNC3866 for its target chromodomains. The phylogenetic tree was generated using:

[http://apps.thesgc.org/resources/phylogenetic\\_trees/index.php?treeGroup=epig&tree\\_type=domain\\_base\\_d&domain=CHROMO](http://apps.thesgc.org/resources/phylogenetic_trees/index.php?treeGroup=epig&tree_type=domain_base_d&domain=CHROMO)

UNC3866 was also evaluated against a panel of 48 bromodomains, 33 protein methyltransferases, and 7 lysine demethylases and found to be inactive in each case (**Table 3.1**, **Fig. 3.3**, and **Table 3.2**). UNC3866 was tested against a general pharmacology panel (Cerep) consisting of 49 G protein-coupled receptors (GPCRs), 5 ion channels and 3 transporter proteins (**Table 3.3**), as well as the National Institute of Mental Health's Psychoactive Drug Screening Program (NIMH PDSP) panel which includes 22 additional GPCRs. Follow-up functional assays<sup>108, 109</sup> for targets displaying >50% inhibition of control radioligand binding at 10  $\mu$ M UNC3866 were also performed with both UNC3866 and UNC4219 (negative control) (**Figures 3.4-3.7**). Both UNC3866 and UNC4219 showed nearly identical profiles in these follow-up assays with both compounds showing very weak antagonism for the NK2 and V<sub>1a</sub> receptors (**Fig. 3.5** and **Fig. 3.7a**). The similar profiles of UNC3866 and UNC4219 indicate that UNC4219 will be a valuable negative control in determining the biological consequences specific to chromodomain antagonism.

**Table 3.2.** Activity of UNC3866 against a panel of bromodomains as measured by  $\Delta T_m$ . UNC3866 is shows no activity for the bromodomains in this panel at 25  $\mu$ M ( $\Delta T_m < 1^\circ\text{C}$ ).

UNC3866-Bromodomain profile					
Target	UNC3866 $T_m$ shift	Target	UNC3866 $T_m$ shift	Target	UNC3866 $T_m$ shift
ASH1L	-0.39	BRDT(2)	0.21	PB1(4)	0.56
ATAD2	-0.62	BRPF1A	-0.04	PB1(5)	-0.15
BAZ1A	-1.44	BRPF1B	0.47	PB1(6)	0.28
BAZ1B	0.37	BRPF3	0	PCAF	-0.07
BAZ2A	-0.23	BRWD3(2)	-0.57	PHIP(2)	0.64
BAZ2B	0.17	CECR2	0.58	SMARCA2	0.14
BRD1	0.49	CREBBP	0.35	SMARCA4	0.24
BRD2(1)	0.19	EP300	0.7	SP140	0.07
BRD2(2)	0.04	FALZ	0	TAF1(1)	0.44
BRD3(1)	0.53	GCN5L2	0	TAF1(2)	0.62
BRD3(2)	0.33	ATAD2B	0	TAF1L(1)	0.45
BRD4(1)	0.88	SP140L	0.53	TAF1L(2)	0.72
BRD4(2)	0.2	MLL	0.04	TIF1-bromo	0.27
BRD7	0.08	PB1(1)	-0.03	TIF1-phd-bromo	0.49
BRD9	0.23	PB1(2)	0.24	TRIM28	0.47
BRDT(1)	-0.19	PB1(3)	0.33	WDR9(2)	0.95



**Figure 3.3.** Activity of UNC3866 against a panel of methyltransferase enzymes. Methyltransferase inhibition is evaluated by monitoring the transfer of a radioactive methyl group from SAM to lysine or arginine. No significant inhibition was observed in the presence of up to 50  $\mu$ M UNC3866.

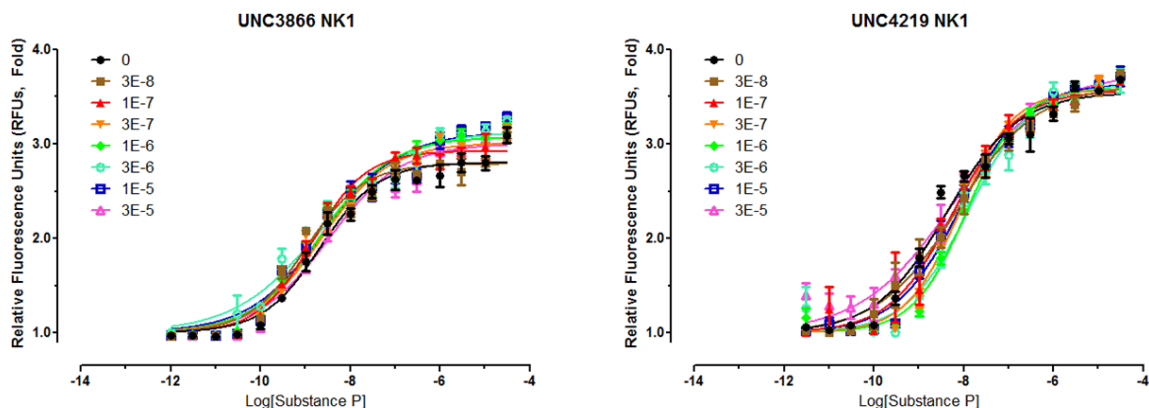
**Table 3.3.** Activity of UNC3866 against a set of demethylase enzymes as determined by AlphaScreen®. Treatment of 100  $\mu$ M UNC3866 resulted in <50% inhibition of all targets binding to their respective histone substrates.

Target	% inhibition at 100 $\mu$ M
JARID1A	18.7
JARID1B	17.8
JARID1C	17.8
JARID1D	-4.7
JMJD1A	15.0
JMJD2C	-6.5
JMJD3	47.9

**Table 3.4.** Cerep selectivity profile of UNC3866. Highlighted in yellow are targets which resulted in >50% inhibition of binding of the control radioligand upon incubation with UNC3866 (10  $\mu$ M). Follow-up dose-response assays for targets displaying >50% inhibition and closely related targets are shown in Supplementary Figures 5-8. Information pertaining to these assays can be found at [www.cerep.fr](http://www.cerep.fr)

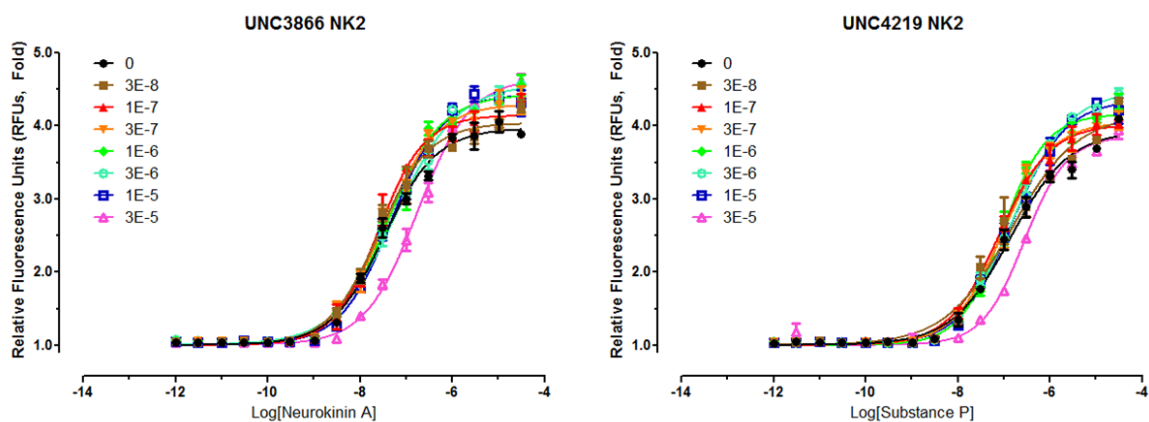
Assay	% Inhibition of Control Specific Binding	Average % of Control Specific Binding (N=2)	Reference Compound
A1 (h) (antagonist radioligand)	4	96.3	DPCPX
A2A (h) (agonist radioligand)	-2	102.3	NECA
A3 (h) (agonist radioligand)	3	96.9	IB-MECA
alpha 1 (non-selective) (antagonist radioligand)	-11	111.4	prazosin
alpha 2 (non-selective) (antagonist radioligand)	-7	107.3	yohimbine
beta 2 (h) (agonist radioligand)	-1	100.8	ICI 118551
AT1 (h) (antagonist radioligand)	-4	104.1	saralasin
BZD (central) (agonist radioligand)	-17	117.4	diazepam
B2 (h) (agonist radioligand)	-7	106.9	NPC 567
CB1 (h) (agonist radioligand)	-13	113.2	CP 55940
CCK1 (CCKA) (h) (agonist radioligand)	33	67.2	CCK-8s
D1 (h) (antagonist radioligand)	4	95.7	SCH 23390
D2S (h) (antagonist radioligand)	40	60.5	(+)butaclamol
ETA (h) (agonist radioligand)	-12	112.5	endothelin-1
GABA (non-selective) (agonist radioligand)	1	99.2	GABA
GAL2 (h) (agonist radioligand)	-21	121.4	galanin
CXCR2 (IL-8B) (h) (agonist radioligand)	-12	111.7	IL-8
CCR1 (h) (agonist radioligand)	7	93.4	MIP-1alpha
H1 (h) (antagonist radioligand)	-6	106.3	pyrilamine
H2 (h) (antagonist radioligand)	1	98.6	cimetidine
MC4 (h) (agonist radioligand)	16	84.0	NDP-alpha -MSH
MT1 (ML1A) (h) (agonist radioligand)	8	92.1	melatonin
M1 (h) (antagonist radioligand)	1	99.0	pirenzepine
M2 (h) (antagonist radioligand)	-6	106.1	methoctramine
M3 (h) (antagonist radioligand)	7	93.2	4-DAMP
NK2 (h) (agonist radioligand)	75	24.6	[Nleu10]-NKA (4-10)

Assay	% Inhibition of Control Specific Binding	Average % of Control Specific Binding (N=2)	Reference Compound
NK3 (h) (antagonist radioligand)	8	92.4	SB 222200
Y1 (h) (agonist radioligand)	-20	119.9	NPY
Y2 (h) (agonist radioligand)	10	90.1	NPY
NTS1 (NT1) (h) (agonist radioligand)	-10	109.8	neurotensin
delta 2 (DOP) (h) (agonist radioligand)	29	71.4	DPDPE
kappa (KOP) (agonist radioligand)	91	9.5	U 50488
mu (MOP) (h) (agonist radioligand)	51	48.6	DAMGO
NOP (ORL1) (h) (agonist radioligand)	23	76.8	nociceptin
EP4 (h) (agonist radioligand)	5	95.2	PGE2
5-HT1A (h) (agonist radioligand)	19	81.3	8-OH-DPAT
5-HT1B (antagonist radioligand)	-11	110.9	serotonin
5-HT2A (h) (antagonist radioligand)	39	61.2	ketanserin
5-HT2B (h) (agonist radioligand)	7	93.0	(±)DOI
5-HT3 (h) (antagonist radioligand)	-21	120.8	MDL 72222
5-HT5a (h) (agonist radioligand)	7	93.1	serotonin
5-HT6 (h) (agonist radioligand)	-5	104.7	serotonin
5-HT7 (h) (agonist radioligand)	8	92.2	serotonin
sst (non-selective) (agonist radioligand)	17	83.1	somatostatin-14
VPAC1 (VIP1) (h) (agonist radioligand)	-9	109.5	VIP
V1a (h) (agonist radioligand)	85	14.6	[d(CH2)51,Tyr(Me)2]-AVP
Ca2+ channel (L, verapamil site) (phenylalkylamine) (antagonist radioligand)	16	84.2	D 600
KV channel (antagonist radioligand)	0	99.8	alpha -dendrotoxin
SKCa channel (antagonist radioligand)	-8	107.9	apamin
Na+ channel (site 2) (antagonist radioligand)	35	64.5	veratridine
Cl- channel (GABA-gated) (antagonist radioligand)	7	93.3	picrotoxinin
norepinephrine transporter (h) (antagonist radioligand)	-13	113.4	protriptyline
dopamine transporter (h) (antagonist radioligand)	7	92.7	BTCP
5-HT transporter (h) (antagonist radioligand)	5	94.9	imipramine

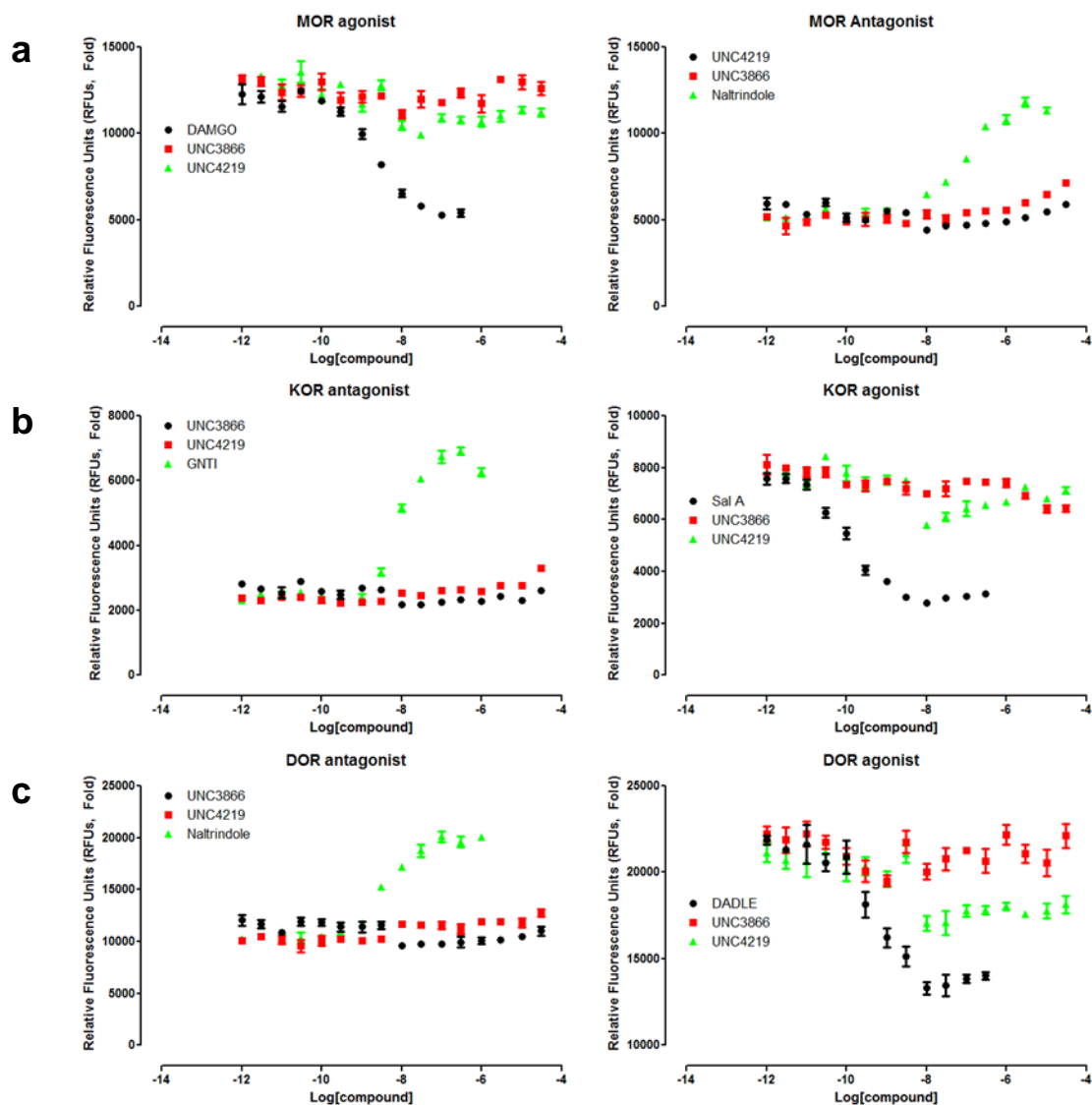


**Figure 3.4.** NK1 receptor dose-response functional assays. (a) Both UNC3866 and UNC4219 display no agonist activity for the NK1 receptor. Assays (calcium mobilization) were performed with HEK293 cells stably expressing the NK1 receptor using FLIPRTETRA (Molecular Devices). Detailed assay protocols and data analysis are available from the PDSP assay protocol book:

<https://pdspdb.unc.edu/pdspWeb/?site=functional>.

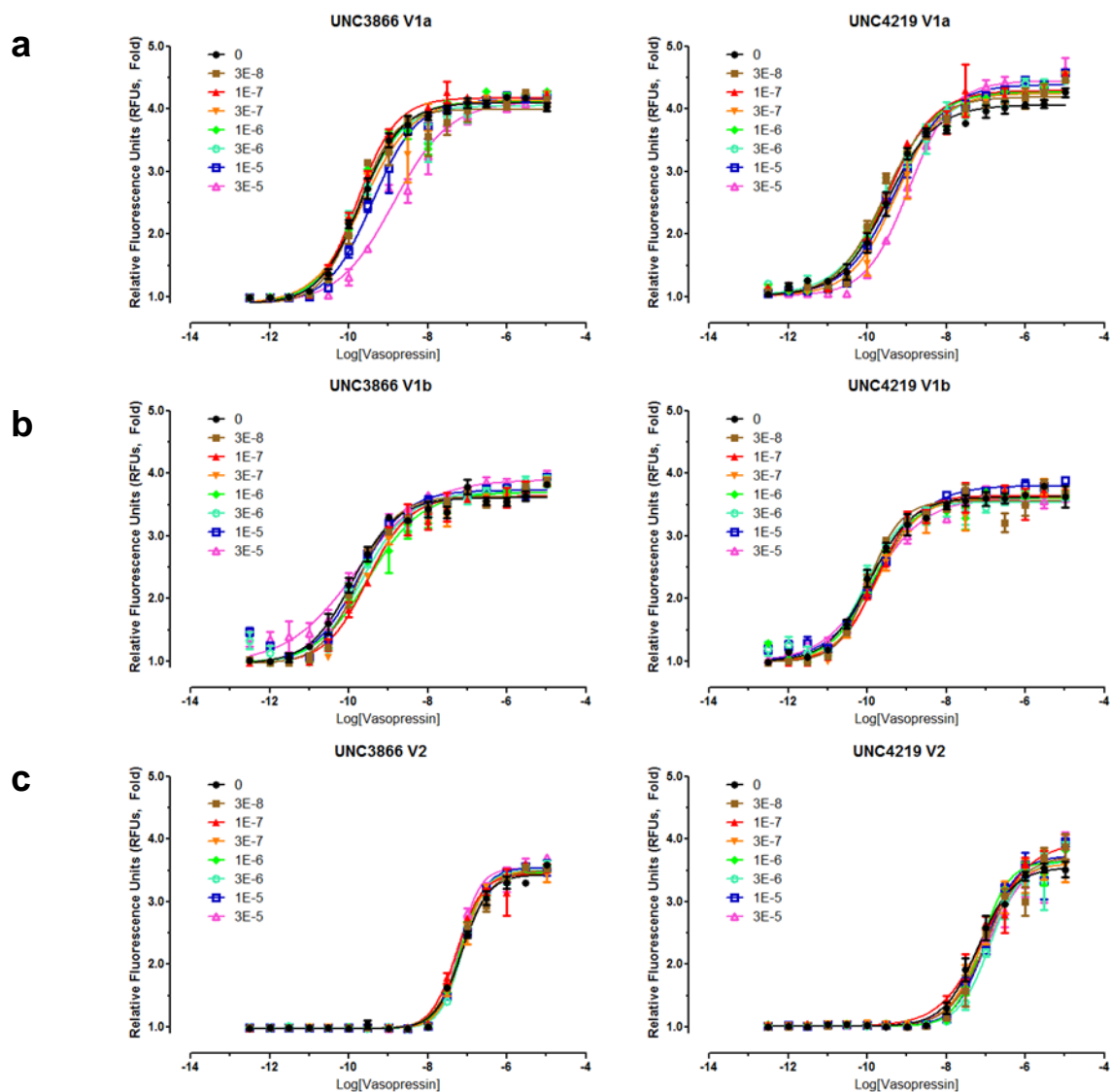


**Figure 3.5.** NK2 receptor dose-response functional assays. Both UNC3866 and UNC4219 display weak antagonist activity for the NK2 receptor against either Neurokinin A or Substance P. Significant antagonism was primarily observed at the highest concentration (30  $\mu$ M). Dose-responses were determined in the absence and presence of a graded concentration of UNC3866. Assays (calcium mobilization) were performed with HEK293 cells stably expressing the NK2 receptor using FLIPRTETRA (Molecular Devices). Detailed assay protocols and data analysis are available from the PDSP assay protocol book: <https://pdspdb.unc.edu/pdspWeb/?site=functional>.



**Figure 3.6.** Opioid receptor dose-response functional assays. (a) Neither UNC3866 nor UNC4219 display agonist or antagonist activity against the  $\mu$ -opioid receptor. DAMGO and Naltrexone are MOR controls. (b) Neither UNC3866 nor UNC4219 display agonist or antagonist activity against the  $\kappa$ -opioid receptor. Sal A and GNTI are KOR controls. (c) Neither UNC3866 nor UNC4219 display agonist or antagonist activity against the  $\delta$ -opioid receptor. DADLE and Naltrindole are DOR controls. Assays were carried out with transiently transfected HEK293 T cells using a split luciferase reporter (GloSensor cAMP, Promega). Detailed assay protocols and data analysis are available from the PDSP assay protocol book <https://pdspdb.unc.edu/pdspWeb/?site=functional>. To measure antagonist activity, reference agonist was used at 50 nM in each assay (DADLE for DOR; Salvinorin A for KOR; DAMGO for MOR).





**Figure 3.7.** Vasopressin receptor dose-response functional assays. (a) Both UNC3866 and UNC4219 display weak antagonist activity for the  $V_{1A}$  receptor. (b) Neither UNC3866 nor UNC4219 displays antagonist activity against the  $V_{1B}$  or (c)  $V_2$  receptors. Dose-responses were determined in the absence and presence of a graded concentration of UNC3866. Vasopressin serves as a control in each case. Assays (calcium mobilization) were carried out with CHO cells stably expressing  $V_{1A}$ ,  $V_{1B}$ , or  $V_2$  receptors using FLIPRTETRA (Molecular Devices). Detailed assay protocols and data analysis are available from the PDSP assay protocol book <https://pdspdb.unc.edu/pdspWeb/?site=functional>.

### SAR studies with Pc CBX, HP1 CBX and CDY chromodomains

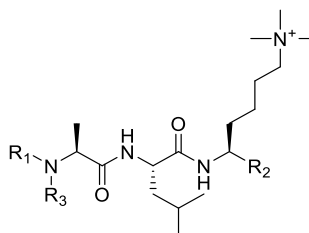
In order to better understand the chromodomain selectivity profile of UNC3866, we developed AlphaScreen<sup>®69</sup> assays for three representative members of the Pc CBX chromodomains (CBX4, -7 and -8), with CBX4 and -7 being the most potent targets of UNC3866, and CBX8 representing the remaining Pc CBX chromodomains (CBX2, -6, and -8), all of which display somewhat weaker affinity for UNC3866<sup>31</sup>. We also developed assays for one member of the CDY family of chromodomains, CDYL2, and a representative of the HP1 family of chromodomains, CBX5, in order to better understand the molecular determinants for selectivity between these three structurally related families of chromodomains. Each of these assays use His-tagged recombinant protein and UNC4195, our biotinylated analog of UNC3866, as the bait ligand. It should be noted that our previously reported AlphaScreen<sup>®</sup> studies with UNC3866 were performed utilizing a similar assay with a biotinylated H3K9me3 bait ligand and CBX7. The improved potency of UNC4195 for our target chromodomains over the H3K9me3 peptide allowed us to reduce the amount of all reagents used in the assay and also reduce the tight binding limit. For CBX5, which does not appreciably bind to UNC4195, we again utilized a biotinylated version of the endogenous substrate, H3K9me3<sup>68</sup>. We tested many of the compounds described in Chapter II against these proteins as well as several new compounds (**Tables 3.5-3.8**).

We selected a subset of our series of peptides designed to interrogate the length requirements for binding to CBX7 based on the ALKme3S motif and found that, minimally, the 5 residue UNC2867 (**17**) was the shortest peptide capable of not only CBX7 antagonism, but also antagonism all three of the other CBX chromodomains in our panel (**Table 3.5**, compounds **7**, **14**, **16-18**). UNC2867 displayed modest affinity for CBX4 and -7, weak affinity for CBX8 and CBX5, and submicromolar potency for CDYL2. In contrast to the Pc CBX chromodomains, the chromodomain of CDYL2 maintains weak affinity for the shorter peptides, UNC2934 (**7**), UNC2597 (**14**), and UNC2864 (**16**). UNC2868 (**18**) displays somewhat reduced affinity for each panel member relative to UNC2867 (**17**), suggesting that there are subtle differences in the way these different chromodomains recognize the N-terminus of these molecules. UNC2934 (**7**) binding affinities were consistent with previous observations that binding of all CBX chromodomains, not just CBX7 requires the presence of a serine C-terminal to the methyl-lysine<sup>38, 68</sup>. Interestingly, this was not the case for CDYL2. The tolerance of CDYL2 for shorter peptide antagonists

relative to CBX chromodomains suggests that targeting this chromodomain through a more traditional small molecule approach may be a viable strategy.

We also evaluated the effect of our *N*-methylated backbone analogs of UNC2868 (**18**) on chromodomain binding (peptides **19-21**). Methylation of the alanine amide nitrogen UNC3261 (**19**) resulted in loss of binding to all chromodomains tested. Methylation of the phenylalanine amide nitrogen UNC3375 (**20**) selectively bound to CDYL2 while methylation of the glycine amide in compound UNC3403 (**21**) had little effect on binding to any of our panel members when compared to the corresponding unmethylated analog, UNC2868 (**18**). UNC3375 (**20**) indicates that CDY-selective antagonists will be accessible through the use of *N*-methylation at the phenylalanine position of UNC3866.

**Table 3.5.** Determination of peptide length requirements for chromodomain antagonism of ALKme3S peptides



				IC <sub>50</sub> (μM) <sup>a</sup>				
				Polycomb			CDY	HP1
Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	CBX4	CBX7	CBX8	CDYL2	CBX5 (HP1α)
UNC2934 (7)	Bz-Phe	-H	-H	>60	>60	>60	37 ± 4.2	>60
UNC2597 (14)	Bz	-CONMe	-H	>60	>60	>60	54 ± 16	>60
UNC2864 (16)	Bz	-CO-Ser-OMe	-H	>60	>60	>60	16 ± 1.9	>60
UNC2867 (17)	Bz-Phe	-CO-Ser-OMe	-H	6.9 ± 1.3	2.6 ± 0.39	33 ± 7.1	0.57 ± 0.042	49 ± 4.1
UNC2868 (18)	Bz-Gly-Phe	-CO-Ser-OMe	-H	58 ± 5.3	18 ± 4.5	>60	1.0 ± 0.23	>60
UNC3261 (19)	Bz-Gly-Phe	-CO-Ser-OMe	-Me	>60	>60	>60	>60	>60
UNC3375 (20)	Bz-Gly-N(me)Phe	-CO-Ser-OMe	-H	>60	>60	>60	15 ± 6.4	>60
UNC3403 (21)	Bz-N(me)Gly-Phe	-CO-Ser-OMe	-H	57 ± 6.0	26 ± 8.3	>60	0.47 ± 0.14	>60

<sup>a</sup>IC<sub>50</sub> values are reported as the mean ± the standard deviation of at least two biological replicates each performed in duplicate.

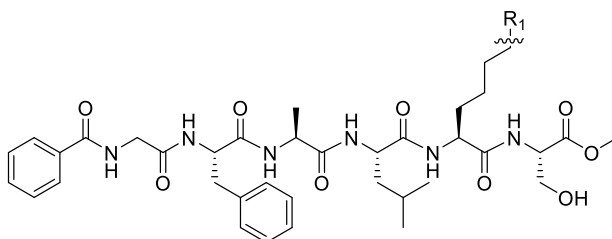
We screened our library of Kme3 replacements against our chromodomain panel to determine the tolerance of the various aromatic cages on these proteins to non-natural lysine analogs. (Table 3.6) Consistent with previous studies, the non-methylated lysine analog of UNC2868 (**18**), UNC3189 (**22**), displayed no antagonism of any of the chromodomains in the panel<sup>68</sup>. Replacement of the quaternary amine with bulky secondary amines (compounds **26-30**) was not well tolerated by all CBX chromodomains, while the CDYL2 chromodomain is able to bind these ligands with modest affinity. For example, the isopropyl-lysine of UNC3563 (**26**) showed no affinity for CBX chromodomains but single-digit micromolar affinity for CDYL2. Incorporation of a cyclopentyl- or cyclohexyl-lysine (UNC3315, **27** and UNC3314, **28**, respectively) did not recover CBX activity. Norbornyl-lysine (UNC3265, **29**) showed weak CBX7 antagonism, but no antagonism for CBX4 or -8 while being equipotent for CDYL2 when compared to cyclopentyl alone (UNC3315, **27**). Adamantyl substitution (UNC3264, **30**) resulted in no measurable CBX activity and CDYL2 activity comparable to the cyclohexyl compound (UNC3314, **28**), indicating that increasing the size of the fused ring system beyond norbornyl (UNC3265, **29**) is not beneficial for binding to any of the chromodomain families in our panel. Taken together, this data suggests that CDYL2 is more accommodating to various secondary amine Kme3 replacements than the CBX chromodomains.

We next focused on our tertiary substitutions at the  $\epsilon$ -nitrogen of the lysine residue. Unsurprisingly<sup>67, 68</sup>, we found that dimethylation (UNC3188, **25**) reduced the activity for both CBX7 and CDYL2 about 3-fold relative to the analogous Kme3 compound (UNC2868, **18**). However, the diethyl-lysine of UNC3567 (**35**) resulted in similar affinities for all chromodomains as UNC2868 (**18**), albeit with slight improved affinity for CBX8. Our pyrrolidinyl derivative, UNC3580 (**36**), resulted in a reduction in potency across the panel, while increasing the ring size to a piperidinyl group (UNC3581, **27**) or a 7-membered azepane (UNC3648, **38**) also did not appear to be beneficial. Our bulkier dicyclobutyl-lysine analog (UNC3260, **34**) also did not display any significant enhancements in potency for any chromodomain.

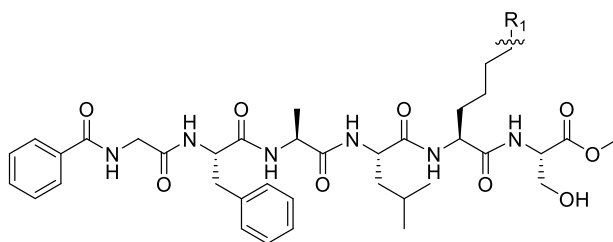
We again tested our acetyl, norbornyl-lysine derivative (UNC3649, **32**) and found that, in contrast to our studies in Chapter II, binding to CBX7 was undetectable in our new assay format. Binding of this compound was also undetectable to all other CBX chromodomains. However, this compound did show affinity for CDYL2 in our assay, albeit with an approximately 7-fold reduction in potency. Changing from

the acetyl, norbornyl to the methyl, norbornyl of UNC3313 (**33**) produced the most potent analog of UNC2868 (**18**) for all members of our chromodomain panel (except CBX5), further emphasizing the importance of a basic amine. However, the more strongly basic tetramethylguanidine derivative (UNC3564, **39**) and dibasic compounds UNC3261 (**31**) and UNC3566 (**40**) showed no activity for any of the CBX chromodomains and weakened affinity for CDYL2, suggesting that enhanced cation- $\pi$  interactions are not beneficial for binding to our chromodomain panel. Taken together, the SAR data resulting from modification of the lysine indicates that the presence of a quaternary amine is not required for engagement of our chromodomain panel; however, both the diethyl-lysine containing compound (**18**) and methyl, norbornyl-lysine analog (UNC3313, **33**) presented themselves as viable Kme3 replacements for Pc CBX and CDY chromodomains. Our data indicates that substantially improving both the selectivity and potency of this series of compounds through modification of the lysine amine presents a significant challenge for CBX chromodomains, but may be a more viable strategy for CDY chromodomain ligand optimization.

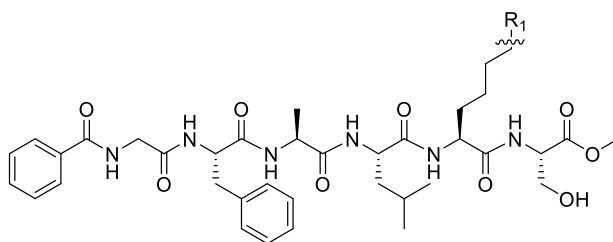
**Table 3.6.** Exploration of secondary and tertiary amine Kme3 replacements.



Compound	R <sub>1</sub>	IC <sub>50</sub> (μM) <sup>a</sup>				
		Polycomb			CDY	HP1
		CBX4	CBX7	CBX8	CDYL2	CBX5
UNC3189 (22)	NH <sub>2</sub>	>60	>60	>60	>60	>60
UNC3563 (26)	HN(CH <sub>3</sub> ) <sub>2</sub>	>60	>60	>60	2.3 ± 0.48	>60
UNC3315 (27)	HN(Cyclopentyl)	>60	>60	>60	3.0 ± 0.28	>60
UNC3314 (28)	HN(Cyclohexyl)	>60	>60	>60	17 ± 4.3	>60



Compound	R <sub>1</sub>	IC <sub>50</sub> (μM) <sup>a</sup>				
		Polycomb			CDY	HP1
		CBX4	CBX7	CBX8	CDYL2	CBX5
UNC3265 (29)		>60	51 ± 10	>60	5.0 ± 0.48	>60
UNC3264 (30)		>60	>60	>60	14 ± 2.0	>60
UNC3188 (25)		>60	51 ± 10	>60	2.8 ± 0.41	>60
UNC3567 (35)		46 ± 19	16 ± 3.7	47 ± 19	0.67 ± 0.071	>60
UNC3580 (36)		>60	53 ± 5.6	>60	1.1 ± 0.25	>60
UNC3581 (37)		>60	41 ± 6.6	>60	0.54 ± 0.039	>60
UNC3648 (38)		>60	28 ± 3.6	31 ± 7.8	0.59 ± 0.080	>60
UNC3260 (34)		>60	32 ± 8.9	34 ± 27	1.8 ± 0.56	>60
UNC3313 (33)		25 ± 8.9	7.3 ± 1.3	31 ± 22	0.30 ± 0.042	>60
UNC3649 (32)		>60	>60	>60	38 ± 13	>60



Compound	R <sub>1</sub>	IC <sub>50</sub> (μM) <sup>a</sup>				
		Polycomb			CDY	HP1
		CBX4	CBX7	CBX8	CDYL2	CBX5
UNC3564 (39)		>60	>60	>60	2.5 ± 0.48	>60
UNC3266 (31)		>60	>60	>60	24 ± 9.7	>60
UNC3566 (40)		>60	>60	>60	8.6 ± 0.62	>60

<sup>a</sup>IC<sub>50</sub> values are reported as the mean ± the standard deviation of at least two biological replicates each performed in duplicate.

We next tested our variations to the N-terminus of our molecules in the context of diethyl-lysine (**Table 3.7**). We synthesized additional molecules from those in Chapter II to more fully understand the SAR at this position (UNC4940, **59** and UNC4972, **60**). UNC4940 (**59**) which is the diethyl-lysine analog of UNC2867 (**17**), showed similar or improved activities toward each of the chromodomains bound by UNC2867 (**17**). Interestingly, UNC4940 (**59**) demonstrates significantly more affinity for CBX5 in contrast to UNC2867 (**17**), providing our first evidence that the aromatic cage of CBX5 can also accommodate non-natural lysine analogs with improved affinity compared to the natural Kme3.

We next explored our series of analogs methylated at the *ortho*-, *meta*- and *para*-positions of the aromatic cap (**48-50**). Methylation at the *ortho*-position (UNC4973, **48**) of the ring dramatically reduced potency for all chromodomains except for CDYL2. Affinity for all chromodomains except CBX5 is recovered to levels comparable to UNC4940 (**59**) when the methyl group is shifted to the *meta*-position UNC4974 (**49**). Installation of *para*-Me (UNC3867, **50**) resulted in further improvements in binding to both CBX7 and CBX4, with submicromolar affinities in each case. Interestingly, CBX8 did not display the same

preference as CBX4 and -7 for UNC3867 (**50**) over UNC4974 (**49**). For CDYL2, the methyl placement had little to no effect on binding.

We next evaluated our compounds exploring the effect of increasing steric bulk at the *para*-position (compounds **51-53**). Overall, increasing lipophilicity had opposing effects on the affinity for Polycomb vs HP1 chromodomains. Introduction of a *para*-ethyl or *para*-isopropyl substituent as in UNC3769 (**51**) and UNC3865 (**52**), respectively, increased the affinity ~2-5 fold for CBX4, -7, and -8, as well as CDYL2, relative to UNC3867 (**50**), while the affinity for CBX5 was largely unchanged or slightly weakened. *t*-Butyl substitution (UNC3866, **53**) further improved potency for CBX4 and -7 while having no significant effect on the affinity for CBX8 and causing a loss of detectable CBX5 affinity. Taken together, these data indicate that potency and selectivity can be tuned by systematically adjusting both the position of alkylation and degree of lipophilicity on the benzoyl cap at the N-terminus.

We also evaluated our compounds incorporating electron-withdrawing groups at the *para*-position of the N-terminal phenyl cap (UNC3768, **55**, and UNC3770, **56**). The methyl ester at the *para*-position of UNC3768 (**55**) resulted in an approximately 15-fold loss in potency for both CBX4 and -7 relative to compound UNC3866 (**53**), but only an approximately 3-fold loss in potency for CBX8, further suggesting that CBX4 and -7 depend more strongly on the presence of bulky alkyl substituents at this position. The strongly electron-withdrawing trifluoromethyl substituent of UNC3770 (**56**) also weakened binding to the Pc chromodomains, although to a lesser extent than UNC3768 (**55**). CDYL2 affinity is largely unaffected by replacement of *para*-alkyl substituents with electron-withdrawing groups, consistent with the generally less stringent binding preferences of CDYL2. Inclusion of an electron-donating group as in UNC4975 (**54**) also did not further improve binding to Pc CBX chromodomains, further indicating that alkyl substituents at this position are preferred. In contrast, these results suggest that CBX5 prefers less sterically demanding functional groups at the N-terminus.

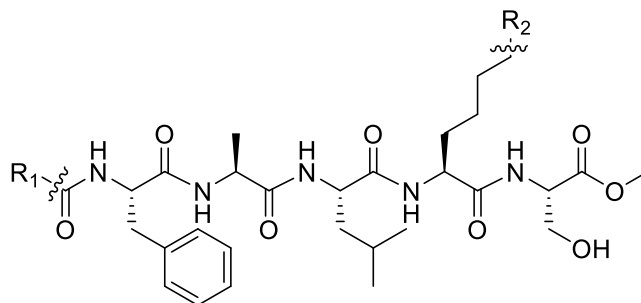
We next investigated the importance of the spacing between the N-terminal hydrophobic group and the rest of the peptide. As can be seen with UNC4972 (**60**), addition of a single methylene group relative to UNC3866 (**53**) resulted in an approximately 100-fold loss of potency for CBX4 and -7 and almost a 20-fold loss for CBX8, while CDYL2 was largely unable to distinguish between the *t*-butyl phenyl and *t*-butyl benzyl moieties. Additionally, our larger bicyclic aromatic capping groups with a methylene



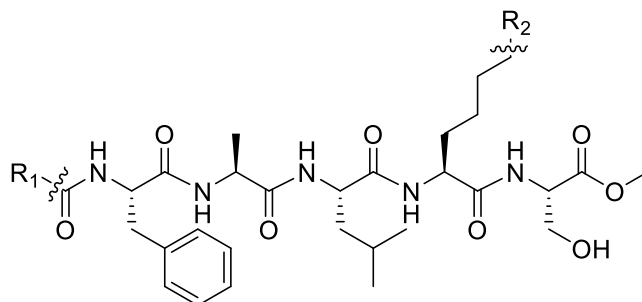
spacer (UNC3771, **45** and UNC3772, **46**) were also less well tolerated by CBX4, -7 and -8. Consistent with our prior results, the increased steric demand of these functionalities was detrimental to CBX5 activity.

Our studies in **Table 3.6** indicated that the tolerance of Pc CBX chromodomains to unnatural Kme analogs is fairly limited, yet we were interested in exploring a small number of lysine modifications, as well as the native Kme3 residue, in the context of our more potent peptide scaffold. We generated the Kme3 analog of UNC3866 (**53**), UNC4938 (**61**), and three asymmetric tertiary amine derivatives, UNC4941 (**62**) UNC4971 (**63**), and UNC4976 (**64**). The isopropyl group of UNC4941 (**62**) and UNC4971 (**63**) was hypothesized to make additional contacts with the aromatic cage of the Pc CBX proteins and to more completely fill the cage based on the crystal structures of UNC3866 (**53**) bound to these chromodomains (**Fig. 3.8a**). The methyl, norbornyl derivative was synthesized because of its improved potency in **Table 3.6** over diethyl-lysine. Mirroring prior results, the Kme3 containing compound, UNC4938 (**61**) was about equipotent to UNC3866 (**53**), while the bulkier lysine mimetics of UNC4941 (**62**), UNC4971 (**63**) and UNC4976 (**64**) did not appear to pick up any additional interactions to further improve potency for Pc CBX chromodomains. However, UNC4976 (**64**) did show improved affinity for both CDYL2 and CBX5.

**Table 3.7.** Effect of N-terminal variations and select lysine modifications.

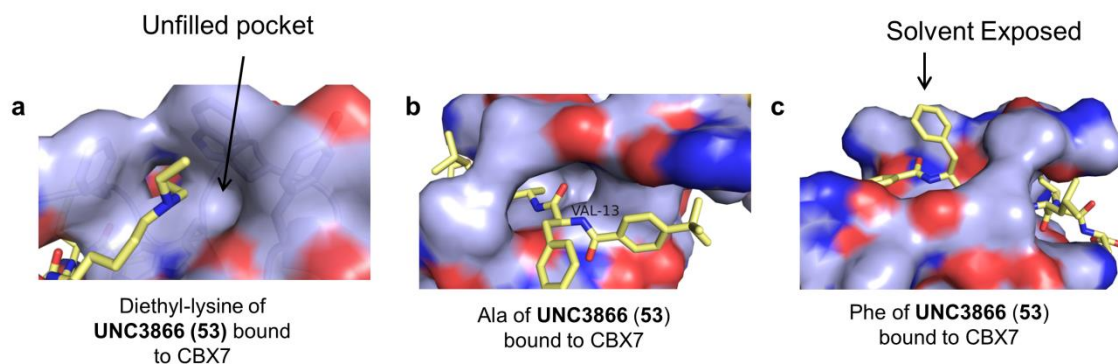


Compound	R <sub>1</sub>	R <sub>2</sub>	IC <sub>50</sub> (μM) <sup>a</sup>				
			Polycomb			CDY	HP1
			CBX4	CBX7	CBX8	CDYL2	CBX5
UNC4940 (59)			3.0 ± 0.82	1.2 ± 0.69	4.8 ± 1.4	0.34 ± 0.021	6.9 ± 0.52
UNC4973 (48)			58 ± 5.4	25 ± 2.5	>60	0.56 ± 0.054	>60
UNC4974 (49)			2.7 ± 1.0	0.74 ± 0.11	5.1 ± 1.4	0.34 ± 0.12	>60
UNC3867 (50)			0.77 ± 0.030	0.22 ± 0.049	4.8 ± 0.39	0.28 ± 0.16	21 ± 4.3
UNC3769 (51)			0.19 ± 0.074	0.081 ± 0.020	1.5 ± 0.19	0.10 ± 0.025	14 ± 2.4
UNC3865 (52)			0.16 ± 0.021	0.061 ± 0.016	1.9 ± 0.46	0.10 ± 0.026	38 ± 6.3
UNC3866 (53)			0.083 ± 0.028	0.037 ± 0.017	1.3 ± 0.48	0.14 ± 0.043	>60
UNC3768 (55)			1.2 ± 0.45	0.57 ± 0.087	4.8 ± 0.87	0.11 ± 0.026	26 ± 2.7
UNC3770 (56)			0.48 ± 0.11	0.23 ± 0.041	3.0 ± 0.23	0.13 ± 0.019	33 ± 8.6



Compound	R <sub>1</sub>	R <sub>2</sub>	IC <sub>50</sub> (μM) <sup>a</sup>				
			Polycomb			CDY	HP1
			CBX4	CBX7	CBX8	CDYL2	CBX5
UNC4975 (54)			0.75 ± 0.11	0.25 ± 0.026	3.3 ± 0.44	0.26 ± 0.018	12 ± 2.3
UNC4972 (60)			9.7 ± 4.4	4.4 ± 0.67	23 ± 2.3	0.20 ± 0.025	>60
UNC3771 (46)			3.4 ± 0.66	1.7 ± 0.61	15 ± 2.3	0.11 ± 0.031	>60
UNC3772 (45)			5.2 ± 0.72	1.1 ± 0.23	24 ± 4.2	0.084 ± 0.028	>60
UNC4938 (61)			0.075 ± 0.021	0.035 ± 0.0064	1.5 ± 0.52	0.11 ± 0.034	>60
UNC4941 (62)			0.075 ± 0.014	0.035 ± 0.0084	1.0 ± 0.21	0.13 ± 0.010	>60
UNC4971 (63)			0.062 ± 0.016	0.032 ± 0.0078	1.3 ± 0.34	0.11 ± 0.022	>60
UNC4976 (64)			0.058 ± 0.0065	0.020 ± 0.0017	0.89 ± 0.17	0.044 ± 0.0014	37 ± 1.9.9

<sup>a</sup>IC<sub>50</sub> values are reported as the mean ± the standard deviation of at least two biological replicates each performed in duplicate.



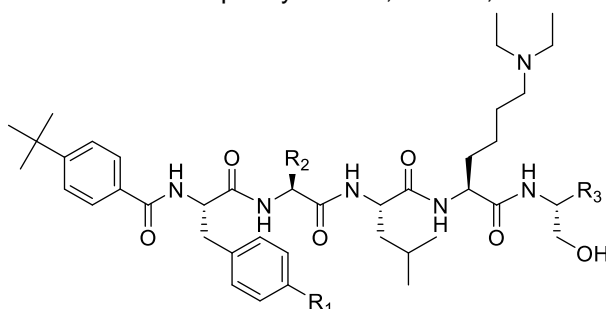
**Figure 3.8.** UNC3866 bound to CBX7 (PDB 5EPJ). (a) Interaction of the CBX7 aromatic cage with diethyl-lysine. (b) Binding pocket of CBX7 occupied by the Ala side chain of UNC3866. (c) Position of Phe of UNC3866 when bound to CBX7.

Previous studies from the Hof group found that the residue at the (-2) position from the modified lysine can have profound effects on the selectivity and potency of peptide antagonists for Pc CBX chromodomains<sup>39</sup>. We explored this in the context of our antagonists by testing the effect of increasing the size of the alkyl side chain at this position (**Table 3.8**, UNC4939, **65**, and UNC4030, **66**). We found that replacing the methyl side chain of alanine with an ethyl or isopropyl group reduced affinity for CBX4, -7, and CDYL2 while slightly increasing the affinity for CBX8. This was not surprising as previously solved co-crystal structures with UNC3866 (**53**) revealed that the pocket occupied by this alanine side chain is larger in CBX8 than in CBX4 and -7 owing to substitution of Val13 in CBX4/7 with Ala13 in CBX8 (**Fig. 3.8b**)<sup>31</sup>. The increased potency and modest selectivity of UNC4030 (**66**) for CBX8 may make it a useful starting point for the development of more potent and selective CBX8 compounds.

Modifications to the C-terminus of UNC3866 (**53**) (compounds **67-69**) were reasonably well tolerated while the methyl ester of UNC3866 (**53**) appears to be the preferred functionality at this position. Replacement with the corresponding acid (UNC4007, **67**), dimethyl amide (UNC4087, **68**), or removal of the C-terminal carbonyl entirely (UNC4051, **69**), resulted in a mere 2- to 4-fold reduction in potency for Pc CBX chromodomains. Interestingly, UNC4007 (**67**) resulted in a slightly larger reduction in potency for CDYL2 (5-fold), suggesting that additional exploration of the C-terminus of these molecules may be useful for further tuning selectivity.

The co-crystal structure of CBX7 with UNC3866 (**53**) revealed that the phenylalanine of UNC3866 (**53**) is largely solvent exposed (**Fig. 3.8c**); therefore, we hypothesized that this was an additional position we could use for tagging to access additional tool compounds. An amine functional group was installed on the phenylalanine side chain, and as expected from the crystallographic information, UNC4946 (**70**) displayed similar affinities to UNC3866 (**53**) for each of the chromodomains in our panel. Further, because the boc group of UNC4961 (**71**) does not interfere with binding, this position presents a suitable location to place larger chemical moieties in future studies.

**Table 3.8.** Exploration of modifications to the phenylalanine, alanine, and C-terminus of UNC3866.



				IC <sub>50</sub> (μM) <sup>a</sup>				
Compound	R <sub>1</sub>	R <sub>2</sub>	R <sub>3</sub>	Polycomb			CDY	HP1
				CBX4	CBX7	CBX8	CDYL2	CBX5
UNC4939 ( <b>65</b> )	-H	-Et	COOMe	0.64 ± 0.19	0.24 ± 0.017	0.98 ± 0.22	0.50 ± 0.071	>60
UNC4030 ( <b>66</b> )	-H	-iPr	COOMe	7.9 ± 1.0	3.1 ± 0.87	0.70 ± 0.15	0.75 ± 0.026	>60
UNC4007 ( <b>67</b> )	-H	-Me	COOH	0.24 ± 0.027	0.13 ± 0.021	2.0 ± 0.92	0.74 ± 0.12	>60
UNC4087 ( <b>68</b> )	-H	-Me	CON(Me) <sub>2</sub>	0.14 ± 0.038	0.047 ± 0.012	2.1 ± 0.50	0.36 ± 0.055	>60
UNC4051 ( <b>69</b> )	-H	-Me	-H	0.20 ± 0.031	0.075 ± 0.012	3.5 ± 0.68	0.41 ± 0.061	>60
UNC4946 ( <b>70</b> )	-CH <sub>2</sub> NH <sub>2</sub>	-Me	COOMe	0.17 ± 0.046	0.063 ± 0.028	2.4 ± 0.61	0.075 ± 0.019	53 ± 4.7
UNC4961 ( <b>71</b> )	-CH <sub>2</sub> NHBoc	-Me	COOMe	0.098 ± 0.011	0.056 ± 0.016	1.7 ± 0.32	0.16 ± 0.012	44 ± 11

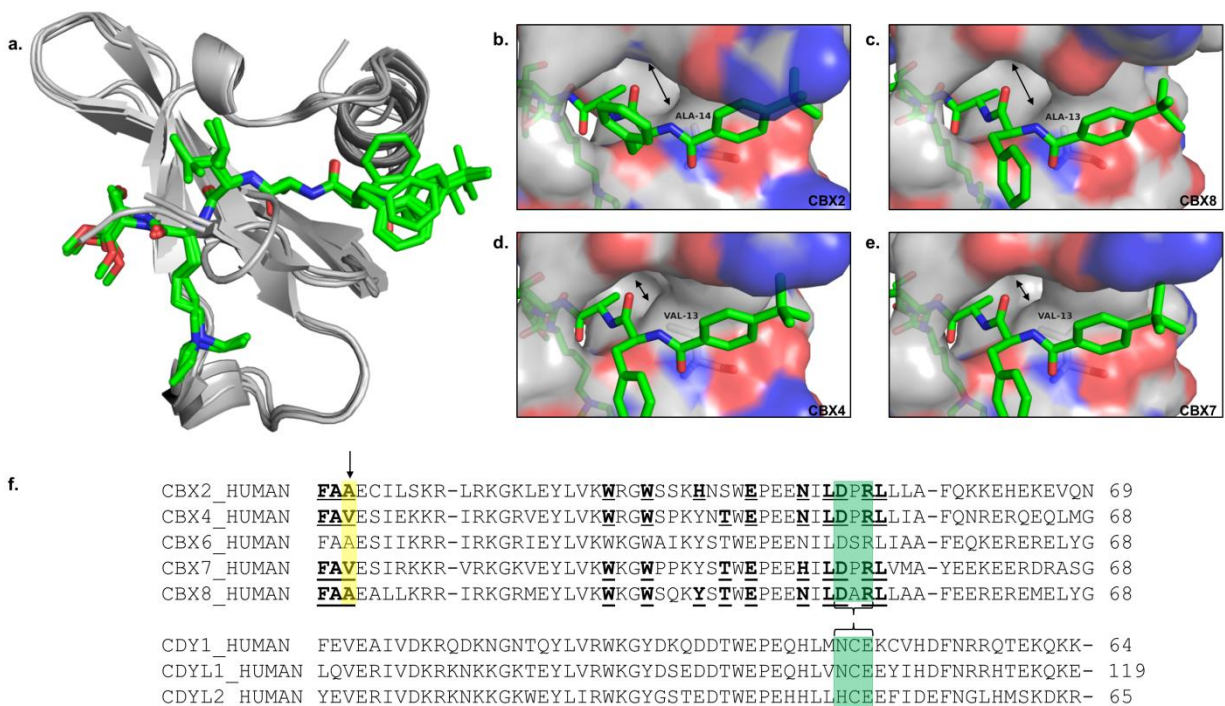
<sup>a</sup>IC<sub>50</sub> values are reported as the mean ± the standard deviation of at least two biological replicates each performed in duplicate.

#### UNC3866 co-crystallization studies

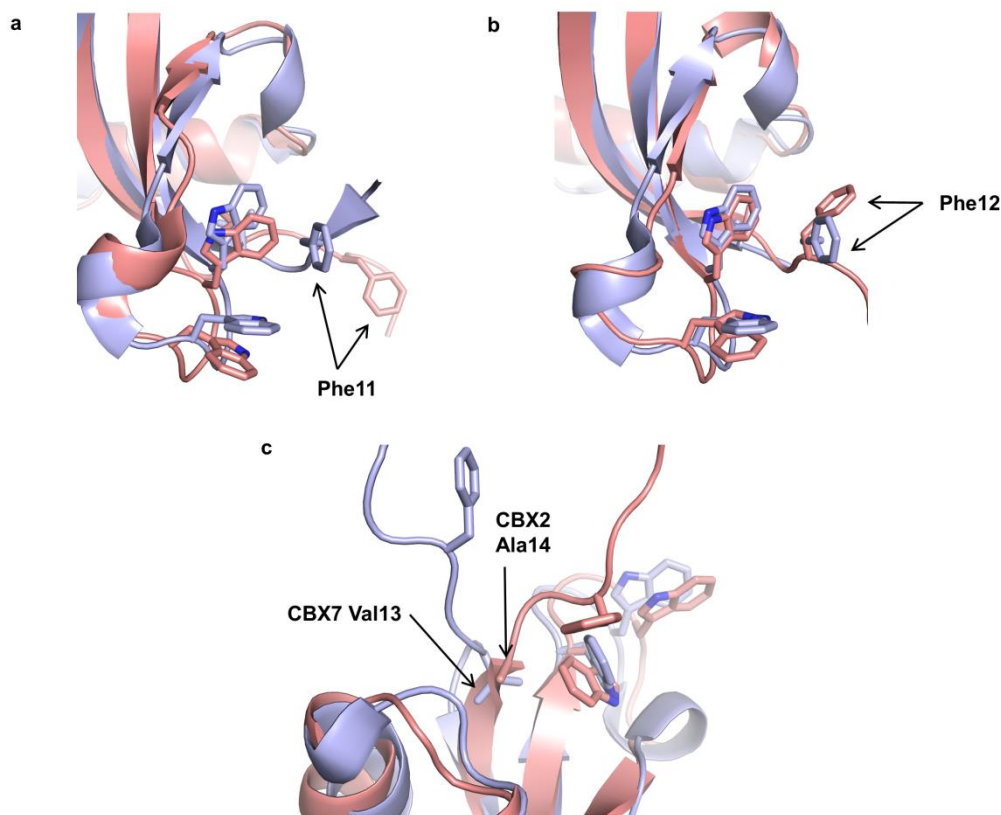
In addition to the co-crystal structure of UNC3866 bound to CBX7, our collaborators at the SGC solved X-ray crystal structures of UNC3866 bound to CBX2 (PDB code 5EPK), CBX4 (PDB code 5EPL), and CBX8 (PDB code 5EQ0) in order to shed further light on the selectivity profile of UNC3866 (**Fig. 3.9**). Alignment of these structures clearly illustrates that UNC3866 binds in a similar fashion to each chromodomain (**Fig. 3.9a**). These structures are consistent with our observations in **Table 3.8** where larger side chains at the (-2) position from the modified lysine are preferred for CBX8. The UNC3866

alanine binding pocket in CBX2 and -8 is much more open in these chromodomains when compared to this pocket in CBX4 and -7 (**Fig. 3.9b-f**). The smaller pocket in CBX4 and -7 more tightly molds to the alanine, enhancing the van der Waals' interactions with UNC3866. This single residue is likely the predominant factor that determines the higher affinity of UNC3866 for CBX4 and -7. Structurally, this simple substitution appears to have profound consequences on the conformational preferences of these chromodomains (**Fig. 3.10**).

Presently, the only Pc CBX chromodomains whose structures have been solved in the apo conformation are CBX2 and CBX7. As already described in Chapter II, CBX7 binds to its substrates through an induced fit binding mechanism that is responsible for aromatic cage formation and Kme3 recognition. In the case of CBX2 however, the aromatic cage is largely pre-organized in the absence of a peptide substrate (**Fig. 3.10b**). Alignment of the apo conformations of CBX2/7 reveals that the residue where their structures diverge is at the Ala/Val residues (**Fig. 3.10c**). This divergence likely requires CBX7-UNC3866 binding to undergo a much more complicated binding pathway than CBX2. Our data in **Table 3.7** strongly agree with this notion as binding to CBX4/7 is highly sensitive to changes at this position of the molecule, consistent with this position being important for initiating the induced fit binding mechanism. In contrast, binding to CBX8, which, like CBX2, contains an Ala at this position, is much less effected by variations in N-terminus of our antagonists. Thus, the added affinity for CBX4/7 is likely gained through the ability of UNC3866 to initiate the induced fit binding mechanism whereas, in the case of CBX2/6/8, the bound conformation is largely preformed, providing less of a potency boost from the *t*-butyl group of UNC3866.



**Figure 3.9.** Co-crystal structures of UNC3866 with CBX2 (PDB code 5EPK), CBX4 (PDB code 5EPL), CBX7 (PDB code 5EPJ) and CBX8 (PDB code 5EQ0). (a) Overlay of CBX2, -4, -7 and -8 (gray) co-crystal structures with UNC3866 (green). (b) UNC3866 alanine binding pocket of CBX2. (c) UNC3866 alanine binding pocket of CBX8. (d) UNC3866 alanine binding pocket of CBX4. (e) UNC3866 alanine binding pocket of CBX7. (f) Alignment of the sequences of the chromodomain targets of UNC3866. The bold and underlined amino acids are the hypothesized residues of each chromodomain that directly interact with UNC3866. These residues contain at least 1 atom located within 4 Å of any atom of UNC3866 in the respective crystal structure. Highlighted in yellow in the CBX family is the residue hypothesized to predominantly determine the selectivity of UNC3866 for CBX4 and -7 (Val in CBX4 and -7; Ala in CBX2, -6, and -8). Highlighted in green is the portion of the CBX proteins that interacts with the *tert*-butyl cap of UNC3866 (D-X-R) and the corresponding residues of the CDY chromodomains (N/H-C-E). The sequence diversity at this position may contribute to the selectivity of UNC3866 for CBX4 and -7.

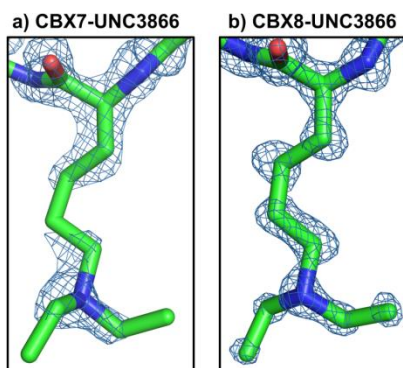


**Figure 3.10.** Comparison of CBX2 and CBX7 in the apo and UNC3866-bound conformations. **(a)** Structural alignment of apo CBX7 (salmon, PDB 2K1B) and CBX7 bound to UNC3866 (light blue, UNC3866 not shown, PDB 5EPJ). **(b)** Structural alignment of apo CBX2 (salmon, PDB 2D9U) and CBX2 bound to UNC3866 (light blue, UNC3866 not shown, PDB 5EPK). **(c)** Structural alignment of apo CBX2 (salmon, PDB 2D9U) and apo CBX7 (light blue, PDB 2K1B).

Aside from understanding the selectivity of UNC3866, we were also interested in using our co-crystal studies to learn more about the interactions between the Polycomb CBX chromodomains and the diethyl-lysine of UNC3866. While the H3K27me3 side-chain has been modeled as adopting a fully extended, all anti-conformation<sup>68</sup>, the lysine side chain of UNC3866 when bound to CBX proteins in our crystal structures is modeled as adopting a gauche conformation about the C<sub>γ</sub>-C<sub>δ</sub> bond in **Fig. 3.11a**. Although the corresponding electron density is comparatively weak for CBX7, we model the side chain in the gauche conformation based on the higher resolution CBX8-UNC3866 co-crystal structure (1.2 Å as compared to 1.6 Å for the CBX7-UNC3866 structure, **Fig. 3.11**). While the electron density of the CBX8-



UNC3866 structure more strongly supports the gauche conformation of this side chain in the bound state (**Fig. 3.11b**), the ambiguity in the position of these atoms may reflect that the side chain in UNC3866 remains somewhat flexible when bound to CBX7. Further, this apparent flexibility in the bound state may explain the ability of CBX7 and the other Polycomb chromodomains to accommodate the diethyl functional group, which is more sterically demanding than the endogenous Kme3 recognized by these domains.



**Figure 3.11** Comparison of the UNC3866 diethyl-lysine side chain electron densities from the CBX7 (PDB code 5EPJ) or CBX8 (PDB code 5EQ0) co-crystal structures. **(a)** The modelled gauche conformation of the lysine side chain (green) in the 1.6 Å CBX7-UNC3866 is only weakly supported by the electron density (blue). **(b)** Compared to the CBX7-UNC3866, the modelled gauche conformation of the lysine side chain (green) in the 1.2 Å CBX8-UNC3866 structure is more strongly supported by the electron density (blue). Taken together, the electron densities for these two structures suggest that the gauche conformation is the favored, but possibly not the exclusive, conformation of the side chain when UNC3866 binds to either of these two chromodomains. Both electron density maps were prepared after deletion of the inhibitor atoms from the model and subsequent torsion-based simulated annealing with PHENIX. The maps are contoured at the 3.0  $\sigma$  level.

#### *Kinetic characterization of CBX7 antagonists by surface plasmon resonance*

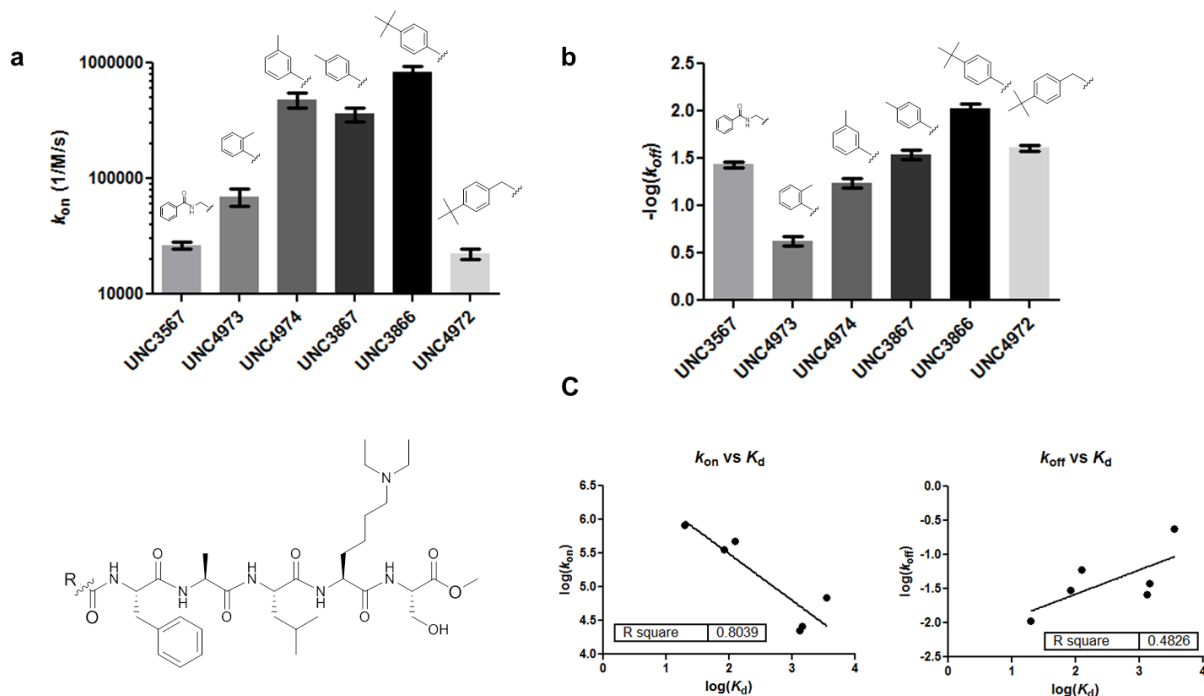
As extensively discussed above, our proposed binding model for our CBX7 antagonists involves an induced-fit mode of recognition in which the N-terminus of these ligands makes the first specific

contact with CBX7 and that initial hydrophobic interaction nucleates a series of backbone hydrogen bonds, allowing CBX7 to close and the aromatic cage to form, generating the stably bound state. While we were able to apply this model to develop more potent chromodomain antagonists, largely through modification of the N-terminal aromatic capping group, as described in Chapter II, we were motivated to further interrogate the validity of the model. We therefore turned to surface-plasmon resonance (SPR) studies with CBX7 and a subset of our compounds that differ at the N-terminus in order to determine if the binding kinetics could provide any evidence to further support this induced-fit hypothesis.

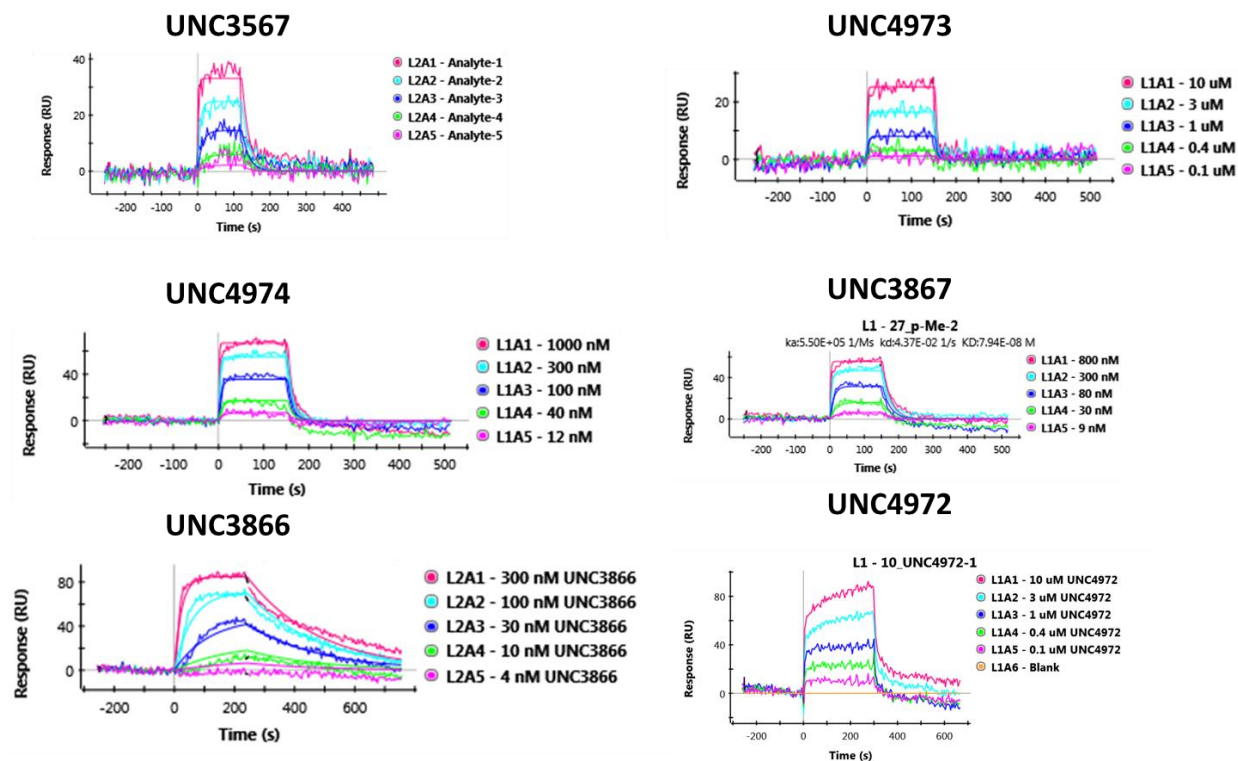
When first comparing the SPR data for UNC3567 (**Fig. 3.12**), which contains the larger benzoyl glycine N-terminus, to the *ortho*-Me phenyl derivative (UNC4973), both of which have similar IC<sub>50</sub>'s for CBX7, we observed that the  $k_{on}$  for UNC3567 was slightly faster but that the  $k_{off}$  increased in a proportional manner, resulting in comparable affinities. When the methyl group of compound UNC4973 is shifted to the *meta*-position (UNC4974), the higher potency observed for this compound is largely driven by a significant increase in  $k_{on}$ . Moving the methyl substituent to the *para*-position (UNC3867) did not substantially change the  $k_{on}$  relative to compound UNC4974 but did slightly improve  $k_{off}$ , resulting in the 3-fold improvement in CBX7 potency for UNC3867 over UNC4974 observed in AlphaScreen®. Inclusion of a *para*-*t*-butyl moiety in UNC3866 further improves both  $k_{on}$  and  $k_{off}$ , resulting in one of our most potent CBX7 inhibitors to date.

We were intrigued by the fact that the majority of the potency improvements resulting from modification of the N-terminus were driven primarily by enhanced association rates instead of decreased dissociation rates<sup>107</sup>. Based on molecular dynamics simulations, we believe that N-terminal portion of these molecules forms the first stable interaction with CBX7, which is consistent with modifications at this position primarily affecting the association kinetics. More specifically, the improvement in association kinetics may result from a reduction in the conformational flexibility at the N-terminus, thereby increasing the likelihood of a productive binding collision between CBX7 and the antagonist. SPR results with CBX7 and UNC4972, which is approximately 100-fold less potent than UNC3866 for CBX7, demonstrate that insertion of a flexible methylene linker primarily reduced  $k_{on}$  while  $k_{off}$  was minimally affected. Finally, analysis of the  $k_{on}$  and  $k_{off}$  rates of these select compounds with respect to their  $K_d$ 's (**Fig. 3.12c**) revealed

that the binding is better correlated with  $k_{on}$  as opposed to  $k_{off}$ , providing even stronger support for this hypothesis.

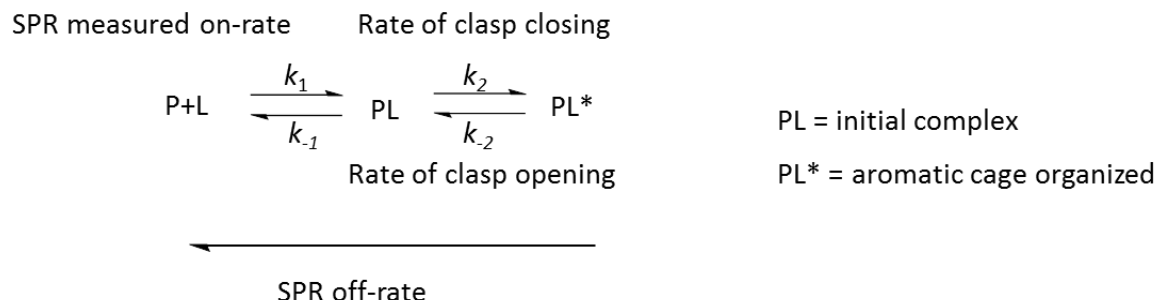


**Figure 3.12.** (a) Comparison of compound-CBX7 associates rates. (b) Comparison of compound-CBX7 dissociation rates. Negative log values of the dissociation rates (1/s) are shown. Data are reported as the mean  $\pm$  the standard error of the mean ( $n \geq 5$ ). (c) CBX7 binding affinities of compounds with N-terminal variations are better correlated with the association rates as compared to the dissociation rates. The log transformations of  $k_{on}$  (1/M/s),  $k_{off}$  (1/s) and  $K_d$  (nM) are plotted.



**Figure 3.13.** Example SPR sensorgrams of CBX7 with each respective compound.

Having demonstrated even stronger support for our binding hypothesis between CBX7 and our antagonists, we were encouraged to further apply this binding pathway for the development of improved CBX7 antagonists. One potential pitfall for the use of UNC3866 in biological settings that our SPR studies revealed is the relatively short target residence time (~100 seconds). A short target residence time indicates that, at equilibrium *in vitro*, there is a rapid, continuous cycle of binding, unbinding and rebinding occurring between the target and its antagonist. This can be problematic in the context of a biological setting. Once a compound has dissociated from its target in a cellular context, it can diffuse away from the target. Diffusion away from the target allows the compound to be acted on by metabolic enzymes, be bound by off target proteins or be effluxed out of the cell. Increasing the residence time creates a long lived antagonist-target complex that prevents diffusion away from the target, producing more efficient target modulation<sup>107</sup>.



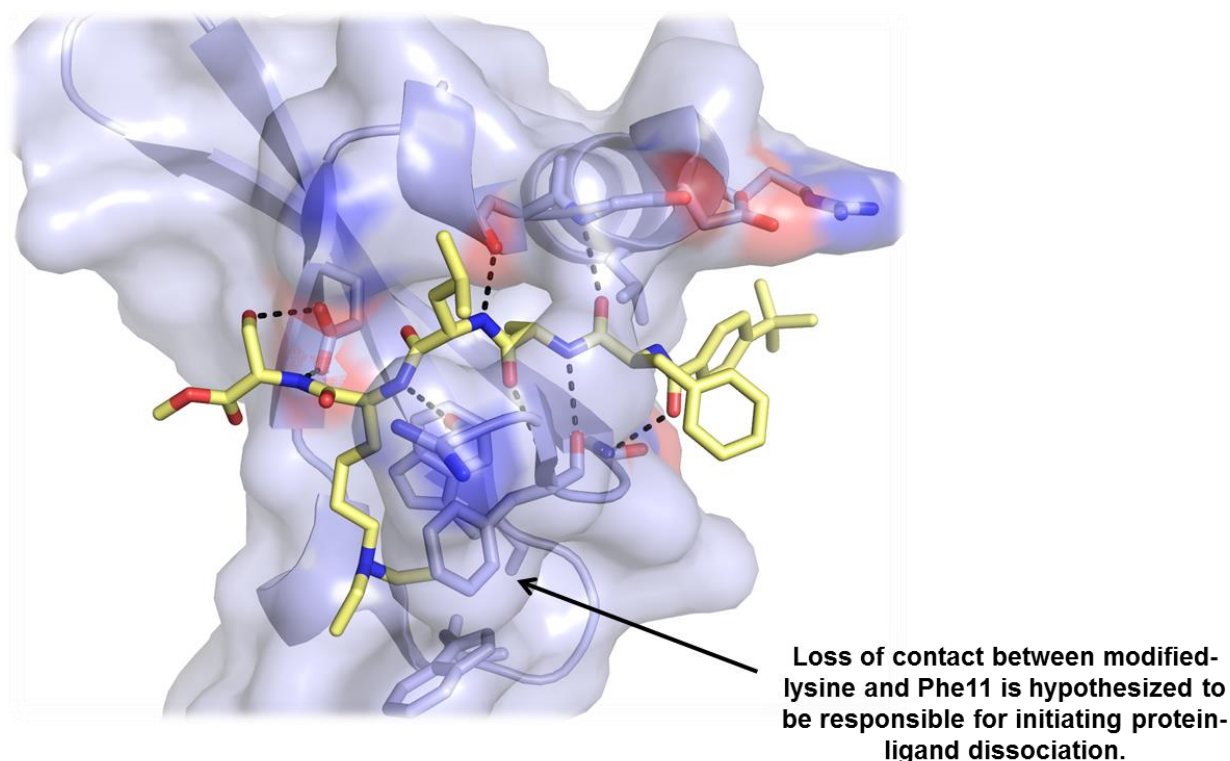
**Figure 3.14.** Induced fit binding equilibrium between CBX7 and peptidic antagonists.

In order to rationally design antagonists with improved off-rates, we again turned to our induced fit binding hypothesis. We approximated our hypothesized binding pathway as a two-step binding equilibrium based on a theoretical treatment of the binding kinetics of induced-fit binding pathways set forth by Weikl and von Deuster<sup>110</sup> (**Fig. 3.14**). The first step involves association of CBX7 (**P**) with the N-terminus of the ligand (**L**), leading to the initial formation of the antagonist-target complex (**PL**). This equilibrium will be dictated by its own association ( $k_1$ ) and dissociation ( $k_{-1}$ ) rate constants. The transition from **PL** to the stably bound state (**PL\***), wherein the N-terminus of CBX7 wraps around the ligand to form a functional aromatic cage, is then governed by its own set of association ( $k_2$ ) and dissociation ( $k_{-2}$ ) rate constants.

In the theoretical model set forth by Weikl and von Deuster, for their treatment of induced fit binding kinetics, the transition into the ground state is assumed to be much faster than the binding and unbinding processes, that is,  $k_2 \gg k_1$ . Therefore, the on-rate of an induced fit system is approximately  $k_1$ . Since the stably bound state (**PL\***) is the ground state conformation, then  $(k_2/k_{-2})$  must be  $\gg 1$ , leading them to assume that  $k_{-2}$  is the dominate rate during dissociation. Encouragingly, deductions from our SPR experiments lead to the same conclusions. The nature of SPR makes it incapable of distinguishing between **PL** and **PL\***, thus the SPR measured association rate will best represent  $k_1$ , the rate limiting step by Weikl and von Deuster. Rigidifying and optimizing the N-terminus of our antagonists primarily increased the SPR association rate ( $k_1$ ) while having a limited effect on the SPR dissociation rate. Because SPR cannot distinguish between **PL** and **PL\***, the measured SPR dissociation rate must be the combined rate of both ( $k_{-1}$ ) and ( $k_{-2}$ ). However, because our modifications to the N-terminus of our

antagonists had minimal effects on the SPR measured off-rate (~3-fold vs ~70-fold improvements in the association rate);  $k_{-2}$  is likely the rate-limiting step along our proposed binding pathway, in perfect agreement with the assumptions of Weikl and von Deuster.

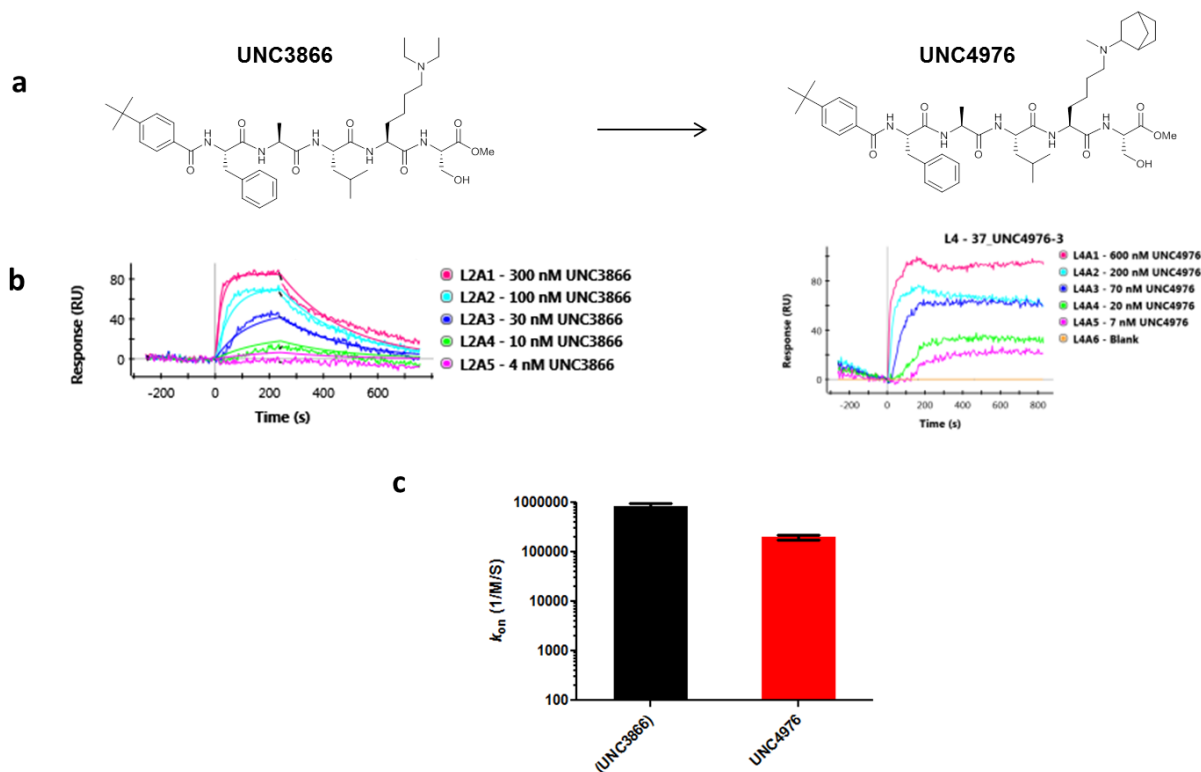
Therefore, we hypothesized that modifications to our antagonists that can decrease  $k_{-2}$  will likely improve the residence time of these molecules. Our proposed binding pathway hypothesizes that dissociation is initiated when Phe11 of the aromatic cage loses contact with the modified-lysine residue (**Fig. 3.15**), allowing the N-terminus of CBX7 to open up ( $k_{-2}$ ) followed by complete protein-ligand dissociation ( $k_{-1}$ ). Therefore, modifications to the antagonist which prevent opening of Phe11 and the CBX7 N-terminus should decrease  $k_{off}$ , increasing the residence time of the target complex.



**Figure 3.15.** X-ray co-crystal structure of UNC3866 bound to CBX7 (PDB code 5EPJ).

While efforts to improve the off-rates of our CBX7 antagonists is still an on-going area of optimization, our initial approach to test this hypothesis was to test the effect of replacing the diethyl-lysine of UNC3866 with one of our other possible Kme3 replacements. The methyl, norbornyl-lysine derivative UNC4976 (**Fig. 3.16a**) has a CBX7  $K_d$  in ITC of  $53 \pm 25$  nM and is approximately equipotent in

AlphaScreen IC<sub>50</sub> to UNC3866. However, the complex formed between CBX7 and UNC4976 is exceptionally long-lived with little-to-no dissociation observed over the course of SPR experiments. Further, this compound possesses a similar association rate with CBX7 as UNC3866 (**Fig. 3.16c**). This results further supports our binding hypothesis and lays the basis for future design seeking to further improve the kinetics of the interaction between CBX7 and our antagonists.



**Figure 3.16.** (a) Structures of UNC3866 and UNC4976. (b) SPR sensorgrams for UNC3866 (left) and UNC4976 (right) with CBX7. (c) comparison of UNC3866 and UNC4976 association rates.

## CONCLUSIONS

We extensively profiled the selectivity of UNC3866 and found it to be a highly selective CBX4/7 antagonist. Further, we found that our negative control compound, UNC4219 has a similar off-target profile as UNC3866, making these two molecules a powerful toolset for interrogating CBX biology. We developed AlphaScreen® assays CDYL2 and for multiple CBX chromodomains, enabling us to better understanding the selectivity determinants for each of these chromodomains. These studies, coupled with our X-ray crystallographic studies will aid in the future development of antagonists with varying

chromodomain selectivity profiles. Further, our studies suggest CBX2, -6 and -8 may in fact be tractable targets for more traditional small molecule discovery efforts. In the meantime, the selectivity and potency of UNC3866 makes it a potentially powerful chemical tool for studying CBX4/7 in cellular systems. Our studies characterizing its effects in cellular systems are described in Chapter IV.

Our SPR studies with CBX7 provided powerful support for our hypothesized induced fit binding mechanism. These studies revealed what is to our knowledge, the first example of a protein-ligand system in which the association kinetics can be rationally tuned. The identification of the slow-dissociating CBX7 antagonist, UNC4976, further demonstrated the utility of our molecular dynamics based binding model and will serve as the basis for future SAR efforts seeking to optimize the dissociative kinetics of our antagonists with CBX7, paving the way for a second-generation CBX7 chemical probe.

## METHODS

### *Kme Reader panel AlphaScreen® assays*

The AlphaScreen® assay (Perkin Elmer) was generally performed as previously described.<sup>69</sup> In brief, compound plates (1  $\mu$ L at 10 mM highest concentration; 3-fold, 10-point dilutions in DMSO) were diluted in 1X assay buffer (20 mM TRIS pH 8.0, 25 mM NaCl, 2 mM DTT and 0.05% Tween-20) to 1 mM using a Multimek robotic pipettor (Nanoscreen) and 1  $\mu$ L was spotted into the wells of 384-well low-volume Proxiplates (Perkin Elmer). To these plates 9  $\mu$ L of protein-peptide mix in 1X assay buffer was added by Multidrop (Thermo) to bring the final compound concentration to 100  $\mu$ M and incubated for 30 min at room temperature. Next, 2  $\mu$ L of a 1:1 mixture of streptavidin-conjugate donor and nickel-chelate or  $\alpha$ -GST acceptor beads (45  $\mu$ g/mL in 1X assay buffer) were added and the plates were allowed to incubate for an additional 30 min in the dark at room temperature. After incubation, the plates were read on an EnVision multi-label reader equipped with an HTS AlphaScreen laser (Perkin Elmer). The expression and purification of the constructs used in this assay was described previously<sup>76</sup>.

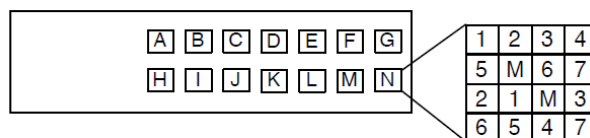


**Table 3.9.** Kme reader AlphaScreen® proteins and their corresponding peptide substrates.

Protein	Peptide	Peptide sequence
53BP1	H4K20Me2	Biotin-AHX-KGGAKRHRK(Me2)VLRDNIQ-OH
L3MBTL1	H4K20Me1	Biotin-AHA-KGGAKRHRK(Me1)VLRDNIQ-OH
L3MBTL3	H4K20Me2	Biotin-AHX-KGGAKRHRK(Me2)VLRDNIQ-OH
MBTD1	H4K20Me1	Biotin-AHA-KGGAKRHRK(Me1)VLRDNIQ-OH
CBX7	H3K9Me3	ARTKQTARK(Me3)STGGKAPRKQL-K(Biotin)-NH2
UHRF1	H3K9Me3	ARTKQTARK(Me3)STGGKAPRKQL-K(Biotin)-NH2
PHF23	H3K4Me3	ARTK(Me3)QTARKSTGGKAPRKQYT-K(Biotin)-NH2
JARID1A	H3K4Me3	ARTK(Me3)QTARKSTGGKAPRKQYT-K(Biotin)-NH2
PHF1	H3K36me3	KSAPSTGGVK(Me3)KPHRYRPGTV-K(biotin)-NH2
PHF19	H3K36me3	KSAPSTGGVK(Me3)KPHRYRPGTV-K(biotin)-NH2

#### Protein microarray layout

#### Biomatik Full Array 3



<b>TUDOR</b> A1) 53BP1(1-2) A2) 53BP1(1-2)* A3) TDRD1-2 A4) TDRD1-3 A5) TDRD1-4 A6) TDRD2 A7) TDRD3	<b>TUDOR</b> B1) TDRD4-1 B2) TDRD4-3 B3) TDRD4-4 B4) TDRD5 B5) TDRD6-5 B6) TDRD6-6 B7) TDRD6-7	<b>TUDOR</b> C1) TDRD8 C2) TDRD9 C3) TDRD10 C4) TDRD11 C5) TDRD12-1 C6) ARID4A C7) ARID4B	<b>TUDOR</b> D1) AKAP1 D2) LBR D3) PHF1 D4) PHF19 D5) PHF20 D6) PHF20-2 D7) PHF20L1	<b>TUDOR</b> E1) SETDB1 E2) SMN E3) SMN* E4) SPF30 E5) UHRF1 E6) ZGPAT E7) JMJD2A(1-2)	<b>TUDOR+PHD</b> F1) JMJD2B F2) JMJD2C F3) MTF2 F4) PHF1* <b>TUDOR</b> F5) Pombe1* F6) Spindlin1* F7) TDRKH*	<b>PHD</b> G1) ING2 G2) TAF3 G3) TRIM24* G4) CHD5(1-2)* <b>CHROMO+PHD</b> G5) CHD3 G6) CHD4 G7) CHD5
<b>CHROMO</b> H1) ARID4A H2) ARID4B H3) CBX2 H4) CBX4 H5) CBX6 H6) CBX7 H7) CBX8	<b>CHROMO</b> I1) HP1α/CBX5 I2) HP1β I3) HP1γ/CBX3 I4) CDY1 I5) CDYL1b I6) CDYL2 I7) MSL3	<b>CHROMO</b> J1) CHD1 J2) CHD2 J3) CHD6 J4) CHD7 J5) CHD8 J6) CHD9 J7) TIP60	<b>CHROMO</b> K1) MPP8 K2) MRG15 K3) MYST1/MOF K4) SMARCC1 K5) SMARCC2 K6) SUV39H1 K7) SUV39H2	<b>AGENET</b> L1) FMR1 L2) FXR1 L3) FXR2 <b>ANK</b> L4) GLP L5) IκB FL <sup>2</sup> <b>BAH</b> L6) ORC1	<b>IDCL</b> M1) PCNA <b>MBT</b> M2) L3MBTL1(1-3) M3) L3MBTL1(1-3)* M4) L3MBTL3 <b>PWWP</b> M5) DNMT3A <sup>2</sup> M6) NSD1 PWWP+PHD* <b>Z1-CW</b> M7) ZCWPW1	<b>BROMO</b> N1) SP140* N2) BRD4(1)* N3) WDR9(1-2)* N4) SMAP* <b>HEAT</b> N5) N-CAPD3(1-4) N6) N-CAPD3(5-8)

**Figure 3.17**

#### ITC studies

All ITC measurements were recorded at 25 °C with an AutoITC200 microcalorimeter (MicroCal Inc., MA). All protein and compound stock samples were in the target buffer (25 mM Tris-HCl, pH 8, 150 mM NaCl, and 2 mM β-mercaptoethanol), and then diluted in the same buffer to achieve the desired concentrations: 50 μM protein and 0.5 mM compound. The concentration of the protein stock solution

was established using the Edelhoch method, whereas compound stock solutions were prepared based on mass. A typical experiment included a single 0.2  $\mu$ l compound injection into a 200  $\mu$ l cell filled with protein, followed by 25 subsequent 1.5  $\mu$ l injections of compound. Injections were performed with a spacing of 180 seconds and a reference power of 8  $\mu$ cal/sec. Control experiments were performed titrating each compound into buffer under identical conditions to determine the heat signals, if any, that arise from diluting the compound. If applicable, the heats of dilution generated were then subtracted from the protein-compound binding curves. The initial data point was routinely deleted. The titration data was analyzed using Origin Software (MicroCal Inc., USA) by non-linear least squares, fitting the heats of binding as a function of the compound:protein ratio to a one site binding model.

#### *Bromodomain Selectivity Assays*

Thermal melting experiments were carried out using an Mx3005p Real Time PCR machine (Stratagene). Proteins were buffered in 10 mM HEPES pH 7.5, 500 mM NaCl and assayed in a 96-well plate at a final concentration of 2  $\mu$ M in 20  $\mu$ L volume. Compounds were added at a final concentration of 25  $\mu$ M. SYPRO Orange (Molecular Probes) was added as a fluorescence probe at a dilution of 1:1000. Excitation and emission filters for the SYPRO-Orange dye were set to 465 and 590 nm, respectively. The temperature was raised with a step of 3  $^{\circ}$ C per minute from 25  $^{\circ}$ C to 96  $^{\circ}$ C and fluorescence readings were taken at each interval.

#### *Methyltransferase Selectivity Assays*

The effect of UNC3866 on the methyltransferase activity of G9a, EHMT1, SUV39H1, SUV39H2, SETDB1, SETD8, SUV420H1, SUV420H2, SETD7, MLL1 trimeric complex, MLL3 tetrameric complex, EZH2 trimeric complex, PRMT1, PRMT3, PRMT5-MEP50 complex, PRMT6, PRMT7, PRMT8, PRDM9, SETD2, SMYD2, SMYD3, BCDIN3D and DNMT1 was assessed by monitoring the incorporation of tritium-labeled methyl group to lysine or arginine residues of peptide substrates using Scintillation Proximity Assay (SPA) as previously described<sup>111</sup>. Assays were performed in a 20  $\mu$ L reaction mixture containing 3H-SAM (Cat. # NET155V250UC; Perkin Elmer; [www.perkinelmer.com](http://www.perkinelmer.com)) at substrate concentrations close to the  $K_m$  values for each enzyme. Three concentrations (1  $\mu$ M, 10  $\mu$ M, and 50  $\mu$ M)

of UNC3866 were used in all selectivity assays. To stop the enzymatic reactions, 7.5 M Guanidine hydrochloride was added, followed by 180 µl of buffer (20 mM Tris, pH 8.0). The reactions were mixed and then transferred to a 96-well FlashPlate (Cat. # SMP103; Perkin Elmer; [www.perkinelmer.com](http://www.perkinelmer.com)). The reaction mixtures in Flash plates were incubated for 1 hour and the CPM were measured using a TopCount plate reader (Perkin Elmer, [www.perkinelmer.com](http://www.perkinelmer.com)). The CPM counts in the absence of compound for each data set were defined as 100% activity. In the absence of the enzyme, the CPM counts in each data set were defined as background (0%).

For DOT1L, NSD1, NSD2, NSD3, ASH1L, METTL21A, METTL21D, DNMT3A/3L, and DNMT3B/3L, a filter-based assay was used. In this assay, 20 µl of reaction mixtures were incubated at 22 °C for 1 hour, 100 µl of 10% trichloroacetic acid (TCA) was added, and the samples were mixed and transferred to filter-plates (Millipore; cat.# MSFBN6B10; [www.millipore.com](http://www.millipore.com)). Plates were centrifuged at 2000 rpm (Allegra X-15R - Beckman Coulter, Inc.) for 2 min, followed by two additional 10% TCA washes, one ethanol wash (180 µl), and centrifugation. Plates were dried and 100 µl of MicroO (MicroScint-O; Cat. # 6013611, Perkin Elmer; [www.perkinelmer.com](http://www.perkinelmer.com)) was added to each well, centrifuged and removed. Another 70 µl of MicroO was added and the CPM was measured using a TopCount plate reader.

#### *Demethylase AlphaScreen® Assays*

The demethylase AlphaScreen® assay was performed in 384-well plate format using white proxiplates (Perkin Elmer) and compound transfers (100 nl) were performed using an ECHO 550 Acoustic Dispenser (Labcyte). All subsequent steps were carried out in assay buffer (50 mM HEPES pH 7.5, 0.1% (w/v) Bovine Serum Albumin and 0.01 % (v/v) Tween-20). In brief, 5 ml of assay buffer containing demethylase enzyme at 2X final assay concentration (see **Table 3.12** for assay specifics) was pre-incubated for 15 minutes with dilutions of compound. The enzyme reaction was initiated by addition of substrate (5 ml) consisting of L-Ascorbic Acid (100 mM), 2-oxoglutarate (2X final assay concentration), Ferrous Ammonium Sulphate (2X final assay concentration) and Histone H3 substrate peptide (2X final assay concentration) and the enzyme reaction was allowed to proceed for the required time. The final assay DMSO concentration was 1%. The enzyme reaction was stopped by addition of 5 ml assay buffer containing EDTA (30 mM) and NaCl (800 mM). Streptavidin Donor beads (0.08 mg/ml) and Protein-A conjugated acceptor beads (0.08 mg/ml) were pre-incubated for 1 hour with anti-methyl mark antibody

(4X final assay concentration) and the presence of Histone H3 product methyl mark was detected by addition of the pre-incubated AlphaScreen® beads (5 ml). Detection was allowed to proceed for 2 hour at room temperature and the assay plates were read in a BMG Pherastar FS plate reader (Excitation 680 nM / Emission 570 nM). Data were normalized to the no enzyme control and the IC<sub>50</sub> determined from the nonlinear regression curve fit using GraphPad Prism 5.

**Table 3.10. Biotinylated peptide substrates.**

Peptide	Supplier	Code
H3K4Me3 (JARIDs)	Anaspec	64192
H3K27Me3 (JMJD3)	Anaspec	64367
H3K9Me3 (JMJD2C)	Anaspec	64360
H3K9Me2 (JMJD1A)	Anaspec	64359

**Table 3.11. Anti-methyl mark antibodies.**

Assay	Supplier	Code
Anti-H3K4Me2 (JARIDs)	Cell Signalling Technology	9725S
Anti-H3K9Me2 (JMJD2C)	Abcam	ab1220
Anti-H3K9Me1 (JMJD1A)	Abcam	ab8896
Anti-H3K27Me2 (JMJD3)	Millipore	07-452

**Table 3.12. Demethylase Alphascreen® assay specifics.**

Assay	Enzyme (nM)	Peptide (nM)	2-OG (μM)	Fe <sup>2+</sup> (μM)	Incubation (min)	Antibody (Conc.)
JARID1A	4	100	5	10	25	1:2500
JARID1B	2	100	5	10	20	1:2500
JARID1C	2	100	5	10	20	1:2500
JARID1D	2	100	5	10	15	1:2500
JMJD2C	1	30	10	1	15	50 mg/ml
JMJD1A	0.5	60	5	10	5	40 mg/ml
JMJD3	0.1	60	10	10	15	40 mg/ml

#### *NIMH PDSP Functional Assays*

The NK1 and NK2 receptors; Δ-, κ-, and μ-opioid receptors; and V<sub>1A</sub>, V<sub>1B</sub> and V<sub>2</sub> receptor functional assays were performed according to the NIMH PDSP Assay Protocol Book Version II found at <https://pdspdb.unc.edu/pdspWeb/?site=binding>

*Conditions for newly developed AlphaScreen assays*

**Table 3.13.** Protein-bait ligand concentrations used for AlphaScreen® assays.

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

*Protein expression and purification*

All expression constructs were transformed into Rosetta BL21(DE3)pLysS competent cells (Novagen, EMD Chemicals, San Diego, CA). Protein expression was induced by growing cells at 37°C with shaking until the OD<sub>600</sub> reached ~0.6-0.8 at which time the temperature was lowered to 18°C and expression was induced by adding 0.5mM IPTG and continuing shaking overnight. Cells were harvested by centrifugation and pellets were stored at -80°C.

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

NaCl, 2mM DTT, 5% glycerol) using an ATKA Purifier (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2ml/min collecting 3ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

GST-tagged proteins were purified by re-suspending thawed cell pellets in 30ml of lysis buffer (1xPBS, 5mM DTT, 1X EDTA free protease inhibitor cocktail (Roche Diagnostics, Indianapolis, IN)) per liter of culture. Cells were lysed on ice by sonication as described for His-tagged proteins. Clarified cell lysate was loaded onto a GSTrap FF column (GE Healthcare, Piscataway, NJ) that had been pre-equilibrated with 10 column volumes of binding buffer (1xPBS, 5mM DTT) using a AKTA FPLC (GE Healthcare, Piscataway, NJ). The column was washed with 10 column volumes of binding buffer and protein was eluted in 100% elution buffer (50mM Tris pH 7.5, 150mM NaCl, 10mM reduced glutathione) over 10 column volumes. Peak fractions containing the desired protein were pooled and concentrated to 2ml in Amicon Ultra-15 concentrators, 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Concentrated protein was loaded onto a HiLoad 26/60 Superdex 200 prep grade column (GE Healthcare, Piscataway, NJ) that had been preequilibrated with 1.2 column volumes of sizing buffer (25mM Tris pH 7.5, 250mM NaCl, 2mM DTT, 5% glycerol) using an ATKA FPLC (GE Healthcare, Piscataway, NJ). Protein was eluted isocratically in sizing buffer over 1.3 column volumes at a flow rate of 2ml/min collecting 3ml fractions. Peak fractions were analyzed for purity by SDS-PAGE and those containing pure protein were pooled and concentrated using Amicon Ultra-15 concentrators 10,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL).

#### *Affinity tag removal*

The N-terminal affinity tag was removed from CBX2, CBX6, CDY1, CDYL1b and CDYL2 by thrombin cleavage according to manufacturer's recommendations (Novagen, EMD Chemicals, San Diego, CA). Briefly, purified protein was incubated with biotinylated thrombin at a final concentration of 1 unit thrombin per milligram tagged protein for 16 hours at 4<sup>0</sup>C. The cleavage reaction was then passed over either a HisTrap FF column (His-tagged proteins) or a GSTrap FF column (GST-tagged proteins) to

remove any protein that still retained the tag. The column flow through was collected and incubated with streptavidin agarose to remove thrombin from the sample. Streptavidin agarose was removed from the sample by centrifugation and the protein was concentrated using Amico Ultra-15 concentrators, 3,000 molecular weight cut-off (Merck Millipore, Carrigtwohill Co. Cork IRL). Protein was exchanged into a buffer containing 25mM Tris pH 7.5, 150mM NaCl, 2mM  $\beta$ -mercaptoethanol prior to use in ITC.

#### *Preparation of proteins for crystallization*

Expression and purification followed a previous description<sup>68</sup>. The coding DNA fragments of single chromodomains of CBX2 (residues 9-62), CBX4 (residues 8-65), CBX7 (residues 8-62) and CBX8 (residues 8-61) were subcloned into modified pET28-MHL or pET28a-Lic vectors to encode N-terminal His-tagged fusion proteins. Recombinant proteins were overexpressed in *E. coli* BL21 (DE3) Codon plus RIL cells (Stratagene) at 15 °C, purified by affinity chromatography on Ni-nitrilotriacetate resin (Qiagen), and treated with TEV or thrombin protease to remove the tag. The proteins were further purified by gel filtration over a Superdex75 column (GE Healthcare, Piscataway, NJ). For crystallization, purified proteins were concentrated to 10 mg/mL in a buffer containing 20 mM Tris, pH 7.5, 150 mM NaCl and 1 mM DTT.

#### *Crystallization*

Purified protein was mixed with compound at a 1:2 molar ratio and crystallized using the sitting drop vapor diffusion method at 18–20°C by mixing 0.5  $\mu$ L of the protein with 0.5  $\mu$ L of the reservoir solution. Reservoir solutions were 20% PEG 3350, 0.2 M KSCN for CBX2, 30% PEG 3350, 0.2 M NaCl, 0.1 M sodium cacodylate, pH 5.5 for CBX4, 20% PEG 3350, 0.2 M ammonium formate for CBX7, and 1.4 M sodium citrate, 0.1 M Hepes, pH 7.5, 5% glycerol for CBX8. Crystals were soaked in a cryoprotectant consisting of 85% reservoir solution and 15% glycerol, and flash-frozen in liquid nitrogen.

#### *Data collection/structure determination and refinement*

Diffraction data for model refinement were collected at a copper rotating anode (CBX2, CBX7) or APS beam line 19ID (CBX4, CBX8; wavelength 0.979 Å). Diffraction intensities were integrated with XDS<sup>112</sup> and merged with AIMLESS<sup>113</sup>. Molecular replacement was performed with the program

PHASER<sup>114</sup>. Geometry restraints for the modified lysyl residue as well as the 4-t-butylbenzoyl protecting group were prepared with eLBOW/Mogul, restraints for selected peptide links with jLigand. Current crystallographic models underwent iterative re-building, refinement, validation in COOT, REFMAC, MOLPROBITY, respectively. Favored regions of the respective model's MOLPROBITY Ramachandran plot included at least 98% of the residues. Anisotropic displacement parameters were analyzed on the PARVATI server. PDB\_EXTRACT as well as PHENIX, IOTBX and CCP4 software were used to compile data for model deposition in the Protein Data Bank and for publication.

Specifically:

CBX2. Coordinates from PDB entry 3H91<sup>68</sup> were used for molecular replacement.

CBX4. The structure of an isomorphous crystal was solved by molecular replacement with CBX2 coordinates. ARP/WARP was used in “map improvement” and “automated model building” modes. The model refinement was further refined against data from the current crystal.

CBX7. The structure was first solved in the lower symmetry P41 setting by molecular replacement with coordinates from PDB entry 4MN3<sup>38</sup>. After merging of data in space group P41212, molecular replacement was performed with preliminarily refined CBX7 coordinates.

CBX8. The structure of an isomorphous crystal was solved by molecular replacement with CBX7 coordinates. ARP/WARP was used in “map improvement” and “automated model building” modes. The model was further refined against data from an additional crystal and, subsequently, the current crystal.



**Table 3.14.** Data collection and refinement statistics (molecular replacement).

Protein (PDB ID)	CBX2 (5EPK)	CBX4 (5EPL)	CBX7 (5EPJ)	CBX8 (5EQ0)
<b>Data collection</b>				
Space group	P6	P6 <sub>1</sub> 22	P4 <sub>1</sub> 2 <sub>1</sub> 2	P2 <sub>1</sub>
Cell dimensions				
<i>a</i> , <i>b</i> , <i>c</i> (Å)	78.89,78.89,30.61	65.18,65.18,133.06	40.29,40.29,82.09	30.57,30.01,36.10
$\alpha$ , $\beta$ , $\gamma$ (°)	90.00,90.00,120.00	90.00,90.00,120.00	90.00,90.00,90.00	90.00,100.17,90.00
Resolution (Å)	34.16-1.80(1.84-1.80)	44.35-1.81(1.84-1.81)	36.17-1.60(1.63-1.60)	35.54-1.18(1.20-1.18)
<i>R</i> <sub>sym</sub> or <i>R</i> <sub>merge</sub>	0.067(0.886)	0.064(0.664)	0.039(0.187)	0.038(0.343)
<i>I</i> / $\sigma$ <i>I</i>	29.2(3.1)	37.4(5.4)	51.8(13.8)	20.5(3.7)
Completeness (%)	100.0(100.0)	99.9(98.7)	99.9(98.2)	99.5(94.5)
Redundancy	10.6(10.3)	20.4(19.6)	12.4(12.3)	3.5(3.3)
<b>Refinement</b>				
Resolution (Å)	34.00-1.80	44.00-1.81	36.17-1.60	35.54-1.18
No. reflections				
work/free	9752/506	14370/1566	8760/671	20239/1069
<i>R</i> <sub>work</sub> / <i>R</i> <sub>free</sub>	0.189/0.212	0.183/0.216	0.237/0.260	0.156/0.186
No. atoms	569	1195	572	630
Protein	460	938	454	511
Ligand	61	131	57	57
Water	40	118	48	50
Other	8	8	13	12
<i>B</i> -factors	31.5	27.1	12.8	11.9
Protein	31.4	26.9	12.1	11.2
Ligand	32.2	23.4	12.0	10.2
Water	31.4	32.6	18.8	20.8
Other	32.6	27.1	19.3	17.0
R.m.s. deviations				
Bond lengths (Å)	0.019	0.018	0.015	0.021
Bond angles (°)	1.9	1.8	1.6	2.0

\*Numbers in parentheses are for highest-resolution shell.

All images were rendered using the PyMOL Molecular Graphics System, Version 1.7.4 Schrödinger, LLC.

#### Surface Plasmon Resonance Experiments

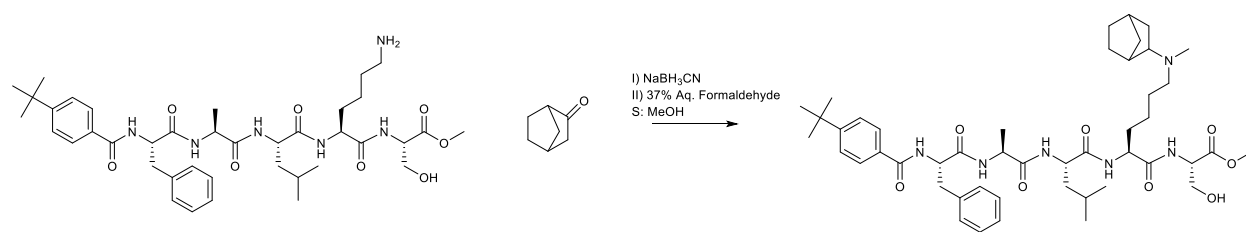
SPR experiments were performed on a ProteOn™ XPR36 Interaction Array System. C-terminally His-tagged CBX7 was immobilized using a ProteOn™ HTE sensor chip. Compounds were flowed at a rate of 50-100  $\mu$ L/minute with a contact time of 150 seconds. Dissociation steps were run for 300 seconds at a

flow rate of 100  $\mu$ L/minute. Double referencing subtraction was done with buffer and protein blank channels to account for nonspecific binding to the sensor chip.

### Compound synthesis and characterization

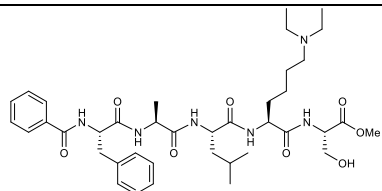
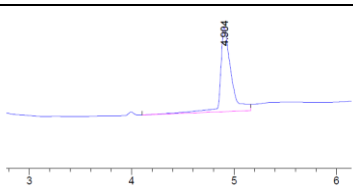
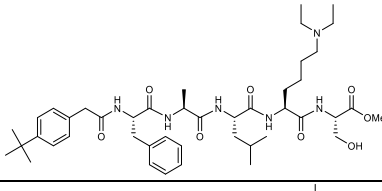
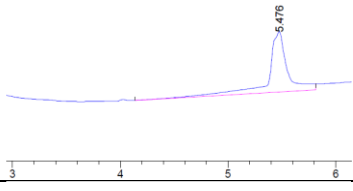
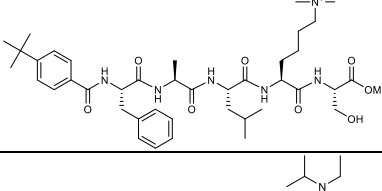
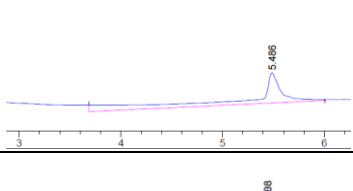
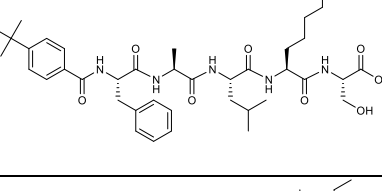
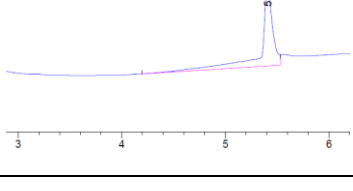
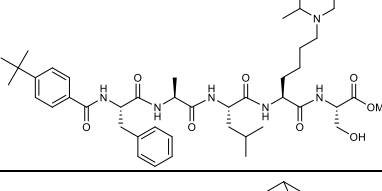
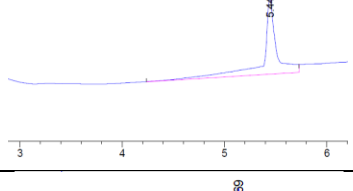
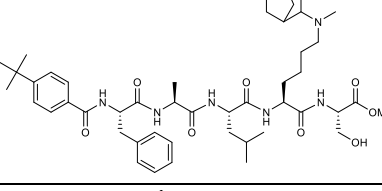
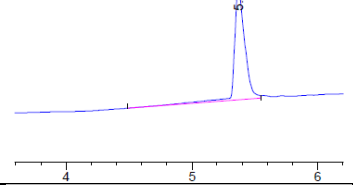
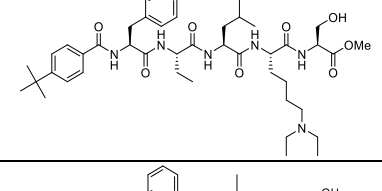
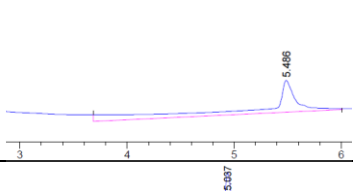
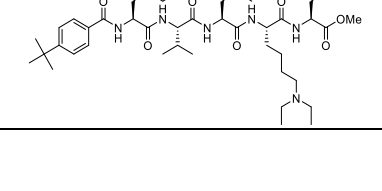
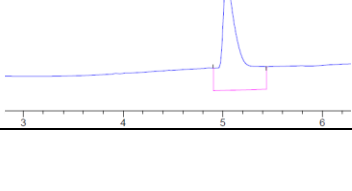
General experimental procedures used for all compounds except UNC4976 are described in depth in the **METHODS** section of **CHAPTER II**.

#### Synthesis of UNC4976 (**64**)



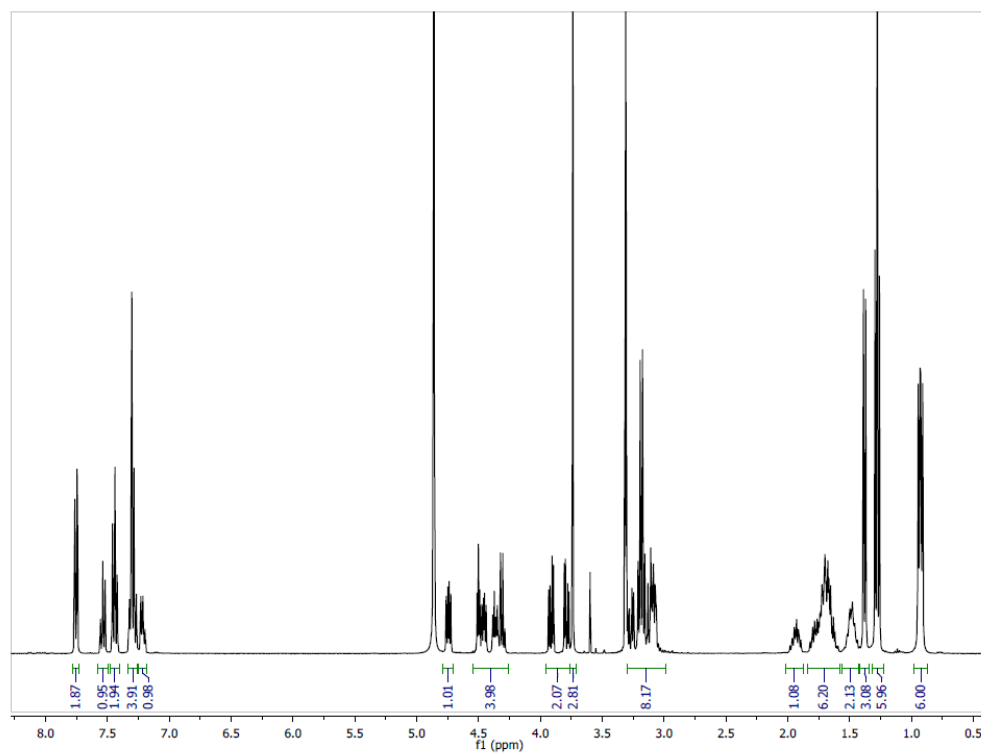
methyl (4-(*tert*-butyl)benzoyl)-L-phenylalanyl-L-alanyl-L-leucyl-L-lysyl-L-serinate (200 mg, 0.23 mmol, 1.0 eq.) was dissolved in 2 mL of MeOH. Norcamphor (210 mg, 1.9 mmol, 8.0 eq.) and sodium cyanoborohydride (44 mg, 0.70 mmol, 3.0 eq.) were added and the solution was stirred at 50°C until complete consumption of methyl (4-(*tert*-butyl)benzoyl)-L-phenylalanyl-L-alanyl-L-leucyl-L-lysyl-L-serinate (~2 hours). The mixture was cooled to room temperature and 50  $\mu$ L of 37% Aq. Formaldehyde (excess) was added and the solution was stirred at room temperature. Upon completion, the solution was concentrated by rotary evaporation and loaded onto silica gel. The crude mixture was purified by normal phase column chromatography with the UV detector to 220 nm and 254 nm. Mobile phases of A (DCM) and B (Methanol) were used with a gradient of 0 to 40% methanol. Product containing fractions were combined, concentrated by rotary evaporation and the resulting residue was purified by reverse phase column chromatography (( $\text{H}_2\text{O}$  + 0.1% TFA)/MeCN) to yield 100.2 mg (45%) of the title compound as a TFA salt. MSI (ESI): 847.50  $[\text{M}+\text{H}]^+$

**Table 3.15.** Compound structures and characterization by LCMS

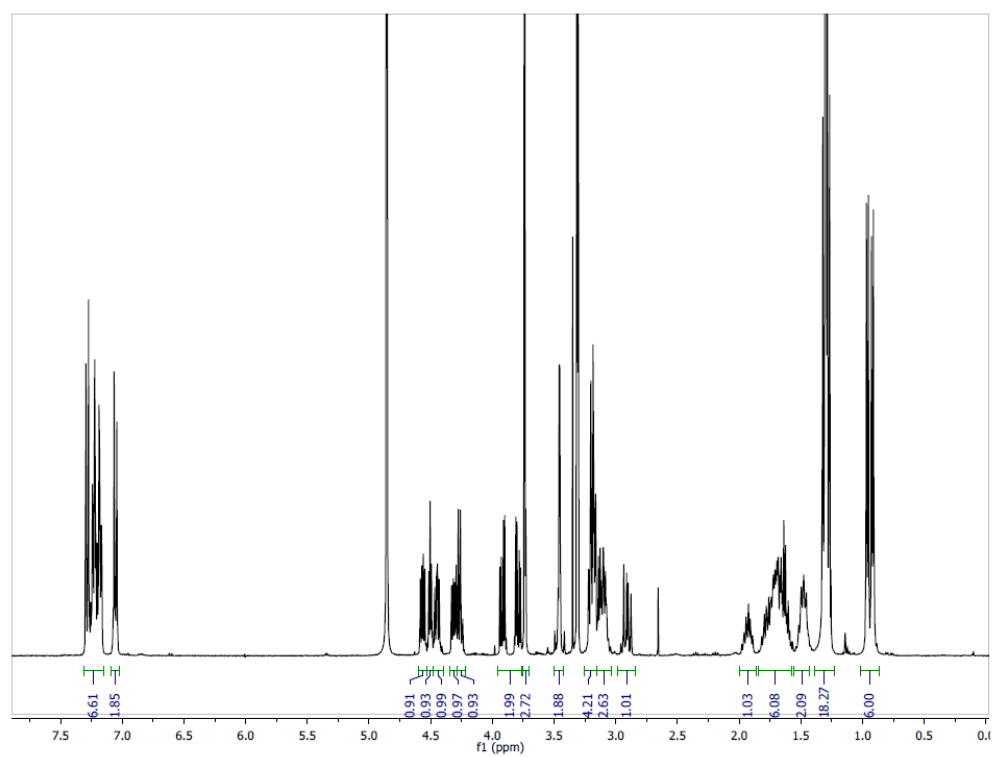
Compound ID	Structure	LCMS trace (220 nm)	Observed Mass
UNC4940 (59)			739.40 [M+H] <sup>+</sup>
UNC4972 (60)			809.50 [M+H] <sup>+</sup>
UNC4938 (61)			391.35 [M+H] <sup>+</sup> 781.45 [M] <sup>+</sup>
UNC4941 (62)			405.35 [M+2H] <sup>2+</sup> 809.50 [M+H] <sup>+</sup>
UNC4971 (63)			823.50 [M+H] <sup>+</sup>
UNC4976 (64)			847.5 [M+H] <sup>+</sup>
UNC4939 (65)			809.50 [M+H] <sup>+</sup>
UNC4030 (66)			823.15 [M+H] <sup>+</sup>

UNC4007 (67)			391.20 [M+2H] <sup>2+</sup> 402.15 [M+H+Na] <sup>2+</sup> 781.10 [M+H] <sup>+</sup>
UNC4087 (68)			404.65 [M+2H] <sup>2+</sup> 808.20 [M+H] <sup>+</sup>
UNC4051 (69)			737.20 [M+H] <sup>+</sup>
UNC4946 (70)		<p>220 nm</p> <p>254 nm</p>	412.85 [M+2H] <sup>2+</sup>
UNC4961 (71)			924.55 [M+H] <sup>+</sup>

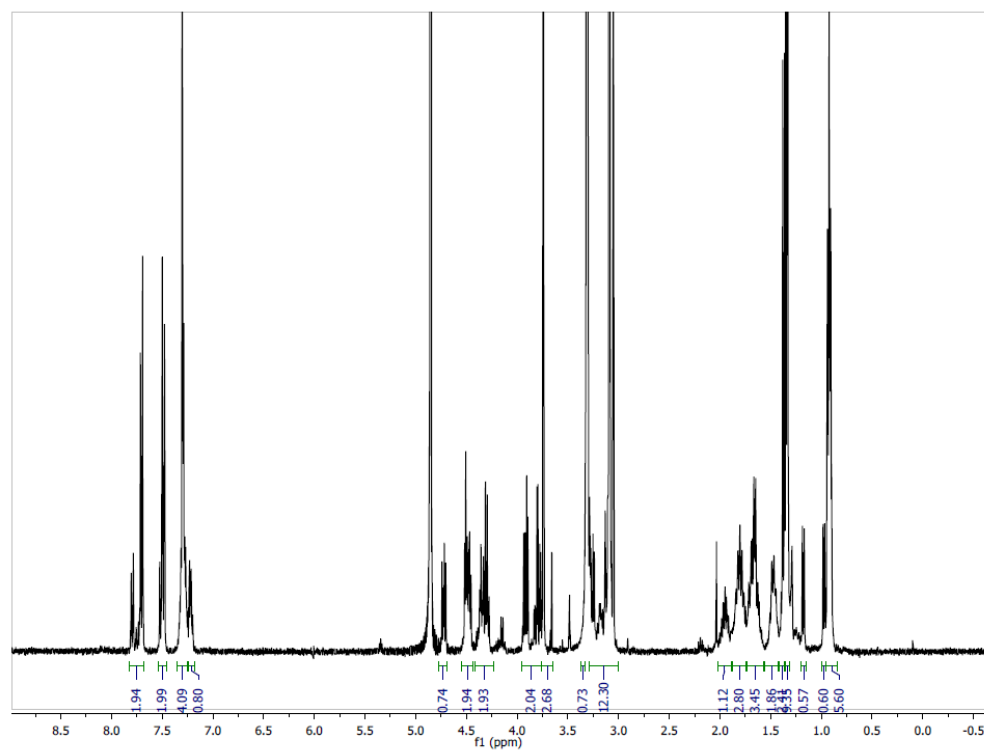
$^1\text{H}$  NMR spectrum of UNC4940 as a TFA salt



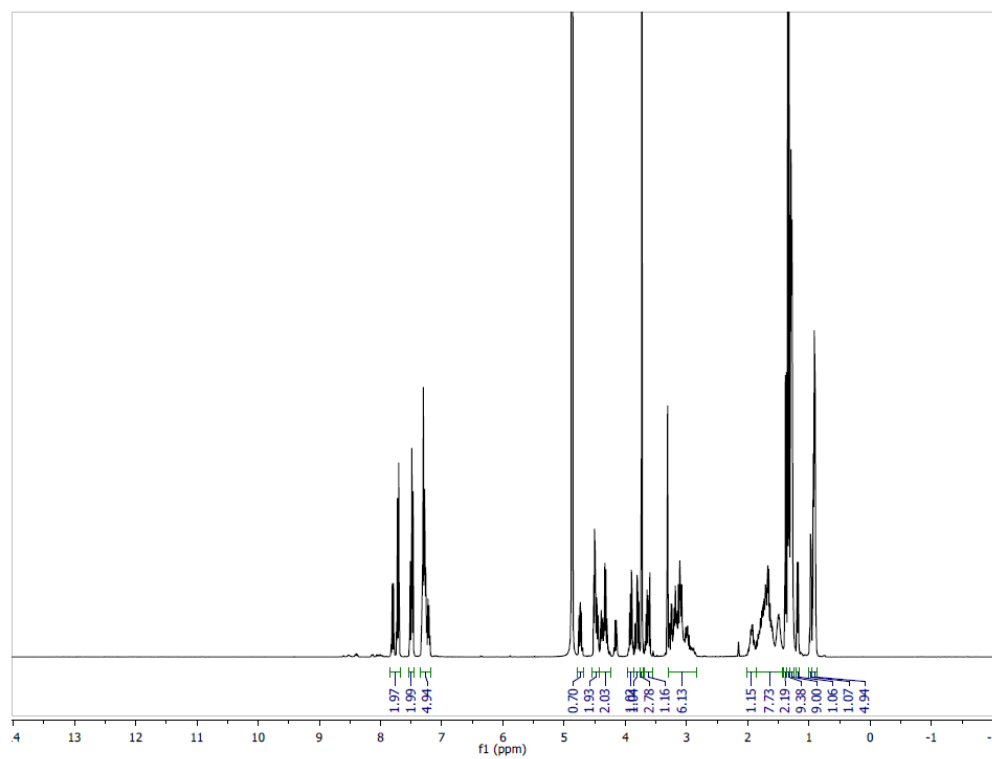
$^1\text{H}$  NMR spectrum of UNC4972 as a TFA salt



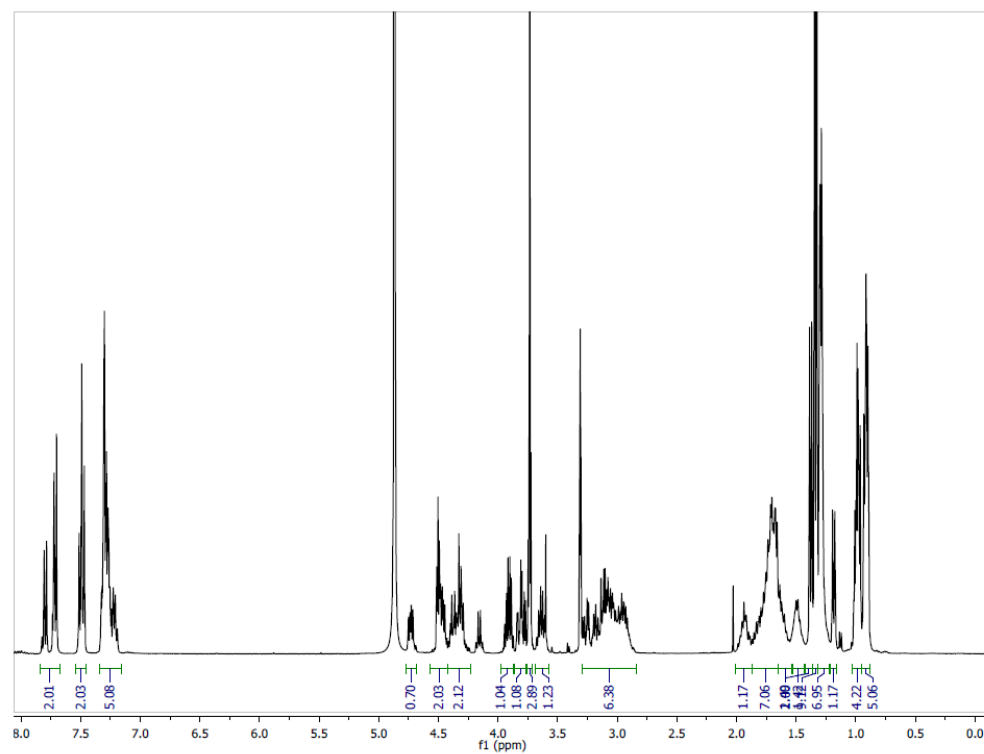
$^1\text{H}$  NMR spectrum of UNC4938 as a trifluoroacetate salt



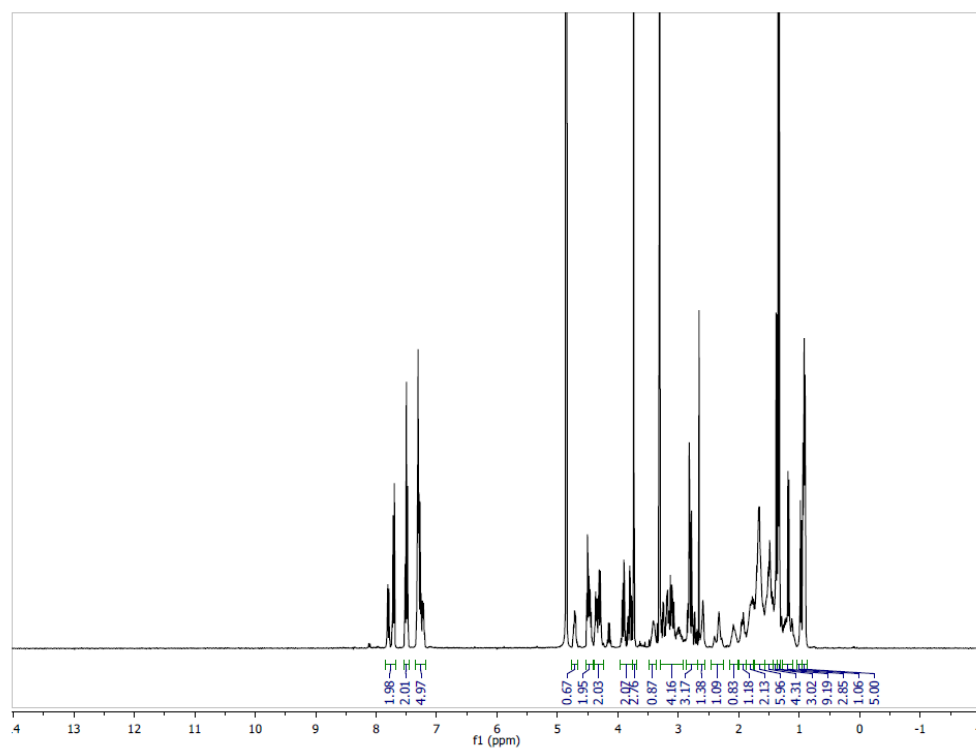
$^1\text{H}$  NMR spectrum of UNC4941 as a TFA salt



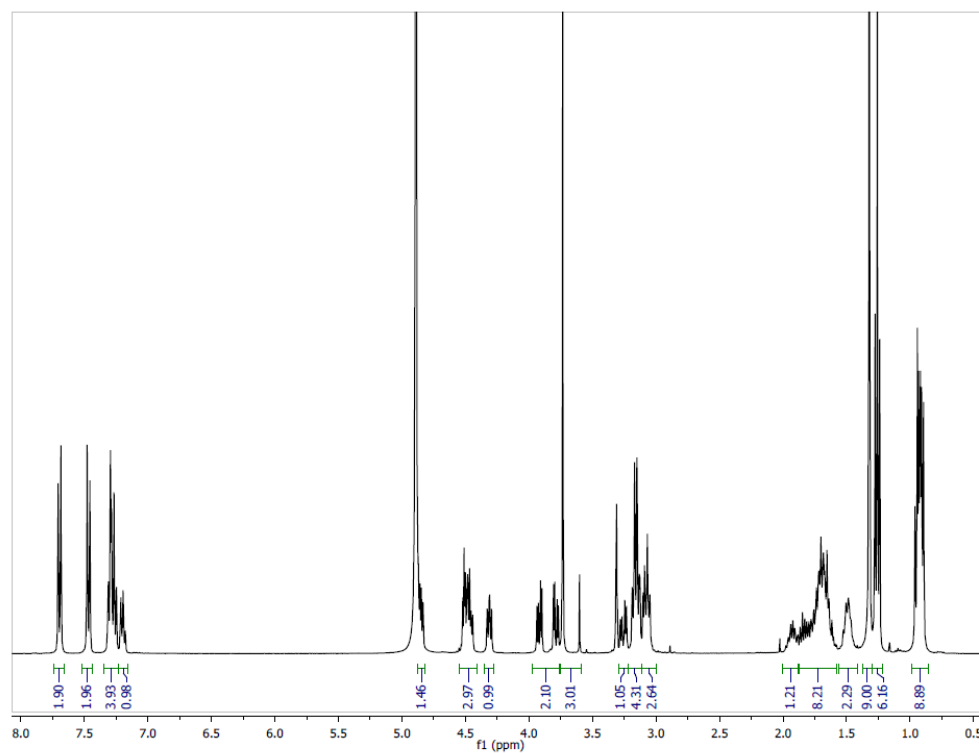
$^1\text{H}$  NMR spectrum of UNC4971 as a TFA salt



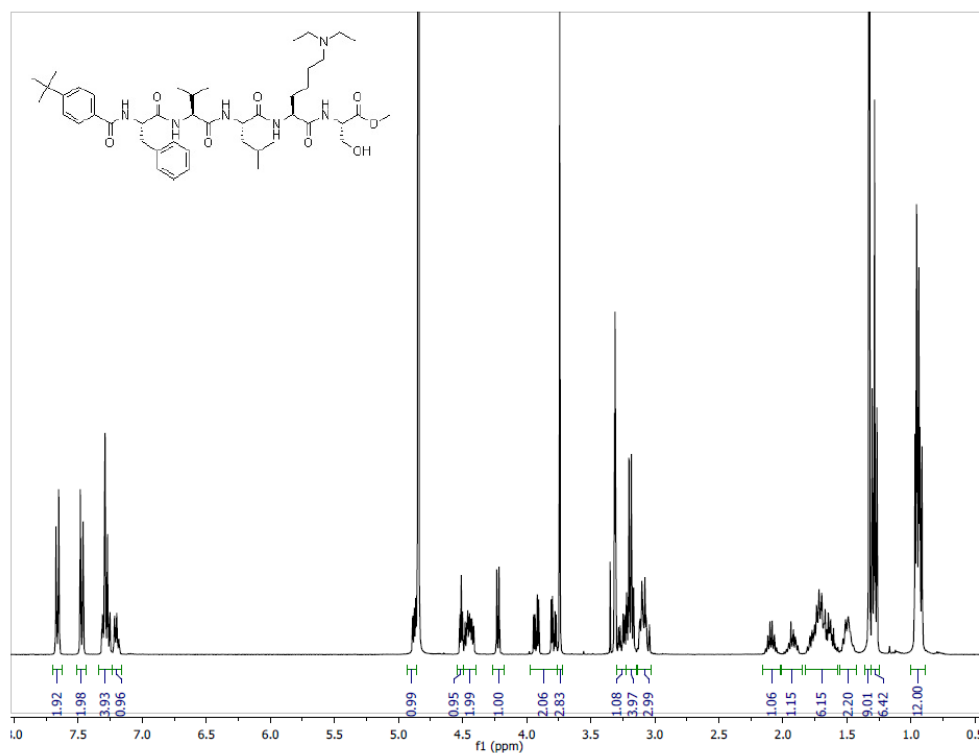
$^1\text{H}$  NMR spectrum of UNC4976 as a TFA salt



$^1\text{H}$  NMR spectrum of UNC4939 as a TFA salt

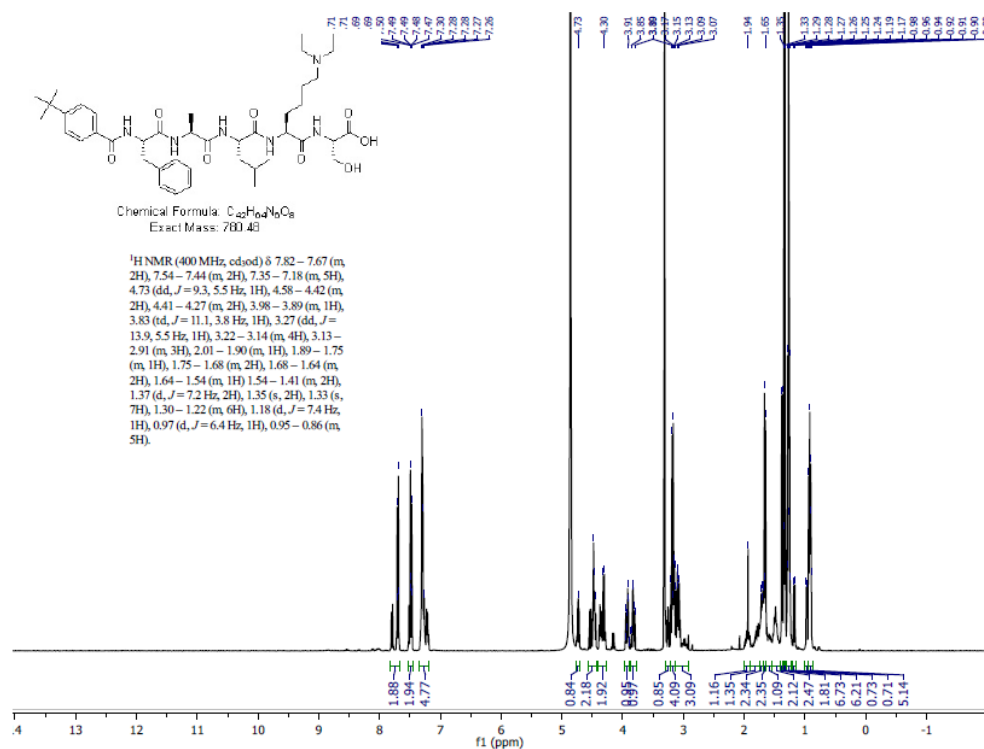


$^1\text{H}$  NMR spectrum of UNC4030 as a TFA salt

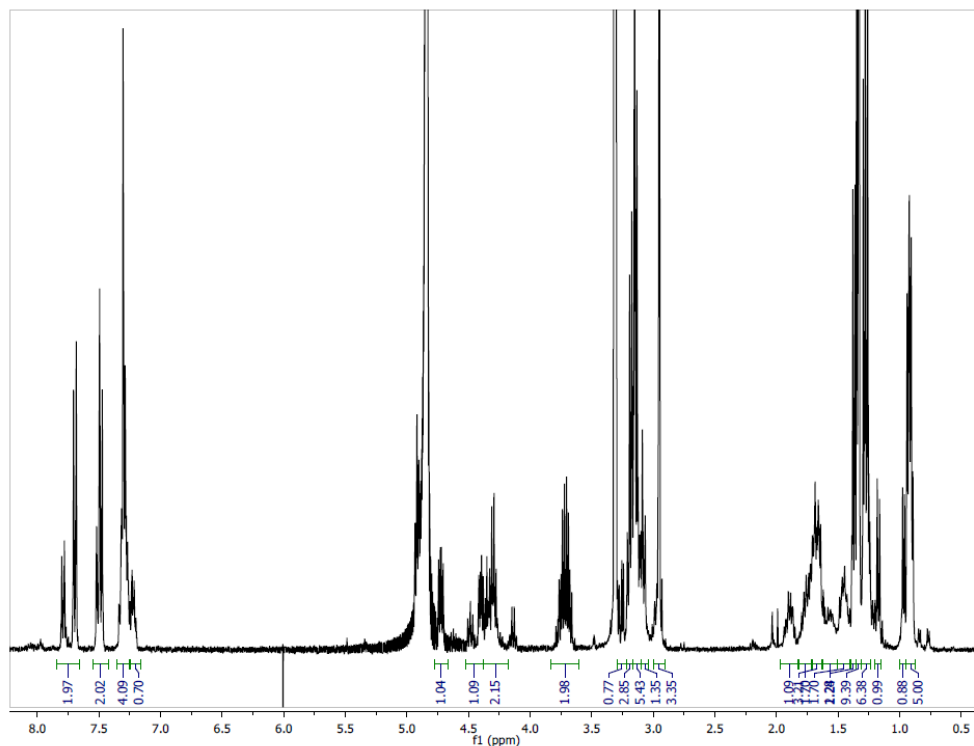




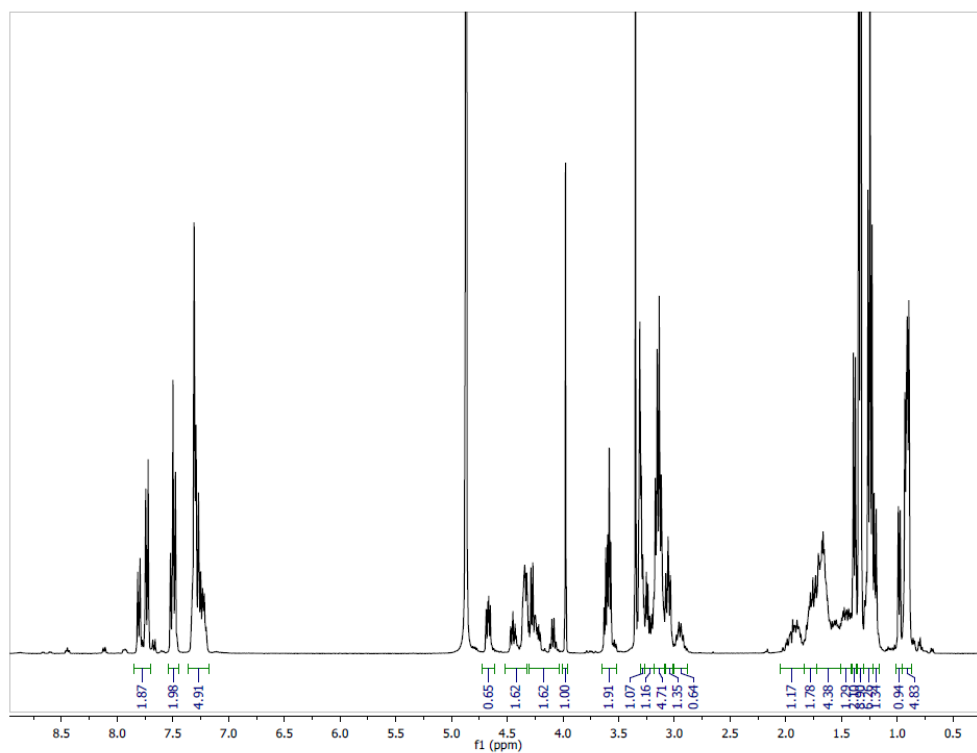
<sup>1</sup>H NMR spectrum of UNC4007 as a TFA salt



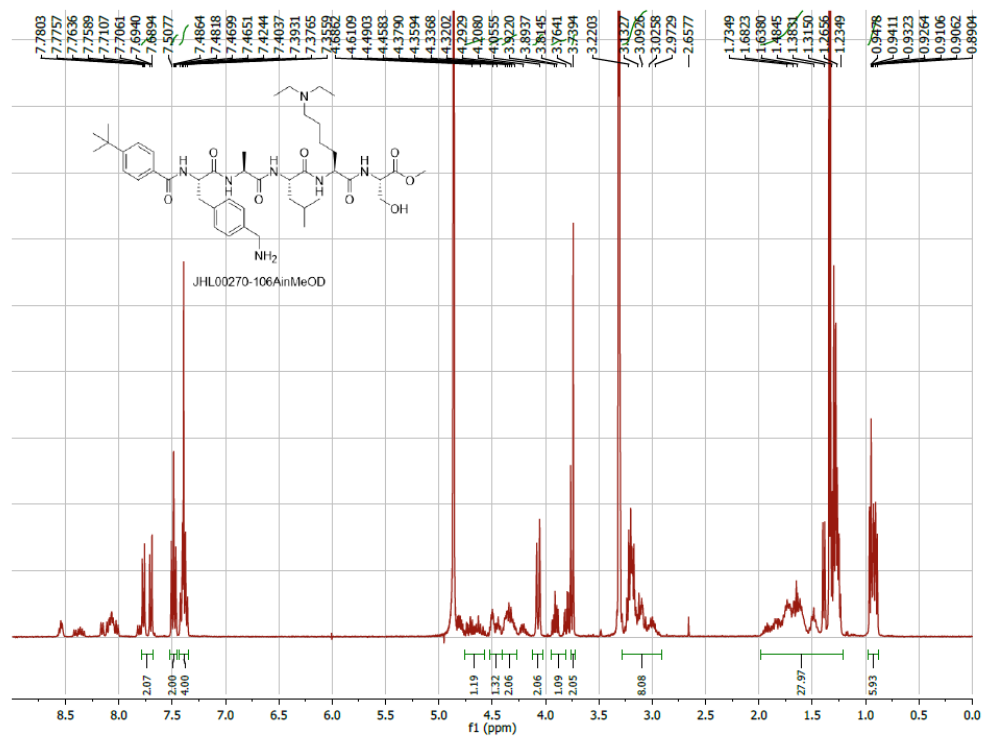
<sup>1</sup>H NMR spectrum of UNC4087 as a TFA salt



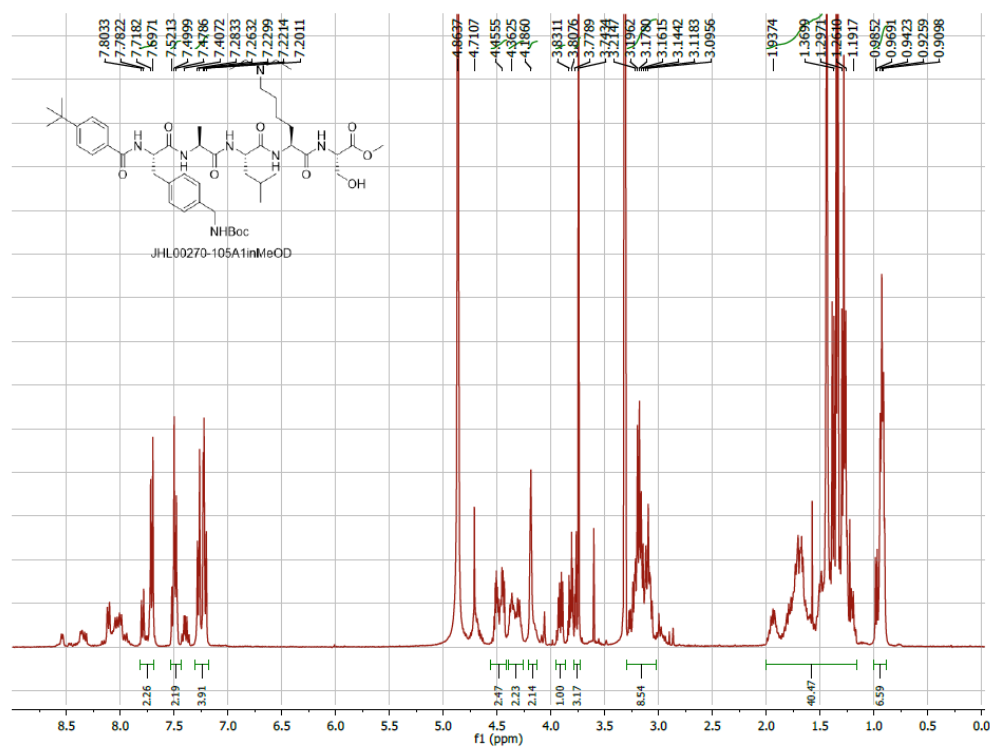
*<sup>1</sup>H NMR spectrum of UNC4051 as a TFA salt*



*<sup>1</sup>H NMR spectrum of UNC4946 as a TFA salt*



$^1\text{H}$  NMR spectrum of UNC4961 as a TFA salt



## CHAPTER IV: CELLULAR CHARACTERIZATION OF UNC3866

### INTRODUCTION

Our selectivity studies in Chapter III uncovered that UNC3866 is highly selective for two families of chromodomains and is overall a selective antagonist for CBX4/7. Further, the limited number of GPCR targets engaged by CBX7 can be controlled for in biological settings through the use of UNC4219 in parallel studies, as the off-target profiles of UNC3866 and UNC4219 are nearly identical. While our studies have demonstrated that UNC3866 has an acceptable molecular profile for use as a chemical probe, four of the core principles of a quality chemical probe remained following our studies in Chapter III<sup>58</sup>:

- (1) Molecular profiling
- (2) Mechanism of action**
- (3) Identity of the active species**
- (4) Proven utility as a probe**
- (5) Availability**

Mechanism of action studies involve providing sufficient data demonstrating modulation of the intended molecular target in a dose-dependent fashion by the probe candidate. In the case of methyl-lysine reader proteins, this particular principle is difficult to satisfy when compared to evaluating antagonists of enzymatic proteins. Typical approaches for demonstrating modulation of reader proteins have focused on techniques such as Fluorescence Recovery after Photobleaching (FRAP) and the Cellular Thermal Shift Assay (CETSA)<sup>8, 75, 115</sup>. Unfortunately, CBX proteins are not amenable to either of these studies, the reasons for which will be discussed in this chapter. This forced us to pursue an alternative approach to demonstrate target engagement by UNC3866. Confidence in mechanism of action studies is inseparably tied to knowledge of the active species. Characterization of both the permeability and metabolic stability of any probe candidate is essential for understanding its mechanism of action and attributing its phenotypic effects to modulation of the target of interest. The utility of a probe candidate (Principle 4) can

be proven through satisfaction of principles 1-3 coupled with carefully controlled cellular experiments demonstrating phenotypic effects through target modulation.

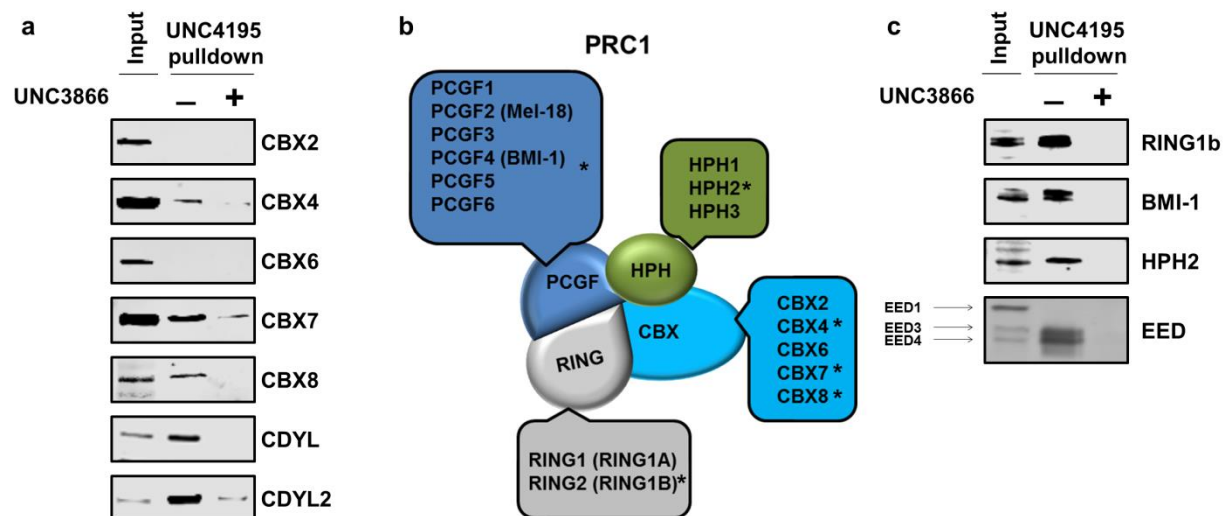
To evaluate the utility of UNC3866 as a CBX7 chemical probe, we focused on well characterized role of CBX7 in prostate tissue. The ability of CBX7 to regulate normal epithelial and cancerous prostate cell proliferation has been well-studied. CBX7 increases cell proliferation in prostate tissue by bypassing replicative senescence, although the precise mechanism by which it does so is unclear<sup>15, 16, 18, 43</sup>. Further, CBX7 may be an important contributor in the conversion of prostate cancer cells to androgen-independence<sup>16</sup>.

Herein we report our efforts to characterize UNC3866 in both cellular and *in vivo* contexts. We primarily focused our studies in the context of PC3 cells, an advanced staged prostate cancer cell model previously shown to have its growth regulated by CBX7<sup>15, 16, 18, 43</sup>. We describe our studies characterizing the cellular effects of UNC3866 and its validation as a cellular chemical probe for the chromodomains of PRC1.

## RESULTS AND DISCUSSION

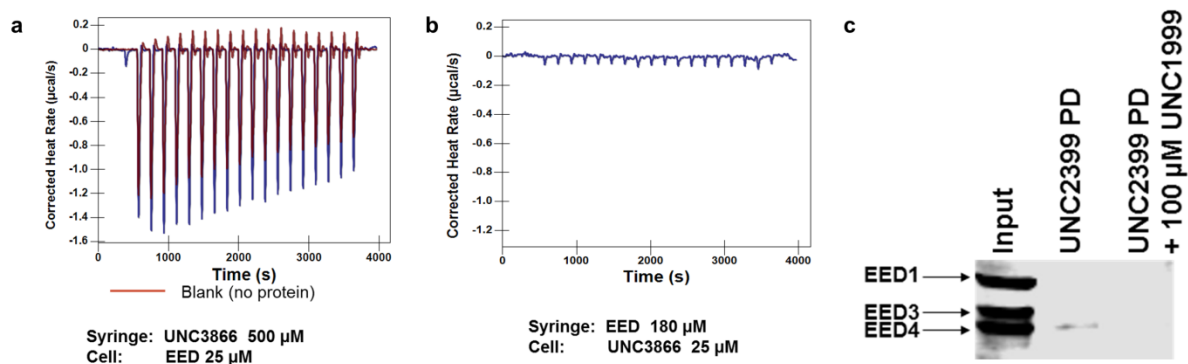
### *UNC3866 engages canonical PRC1*

Our studies described in Chapters II and III utilized recombinant protein constructs consisting only of the chromodomain of each of UNC3866's targets: CBX2/4/6/7/8, CDY1, CDYL1b and CDYL2. To verify that UNC3866 is capable of interacting with the full-length, endogenous forms of these proteins, we utilized our biotinylated derivative, UNC4195 in a series of pull-down studies from PC3 cell lysates (**Fig. 4.1**). We blotted for each of the chromodomain targets of UNC3866 except CDY1, for which we were unable to obtain a reliable antibody. We detected pull-down of all chromodomain targets of UNC3866 except CBX2 and -6 (**Fig. 4.1b**). Furthermore, pull-down is antagonized when excess soluble UNC3866 is added to the lysate. We also blotted for and detected representatives from each of the 4 core families of canonical PRC1 (**Fig. 4.1b** and **4.1c**).



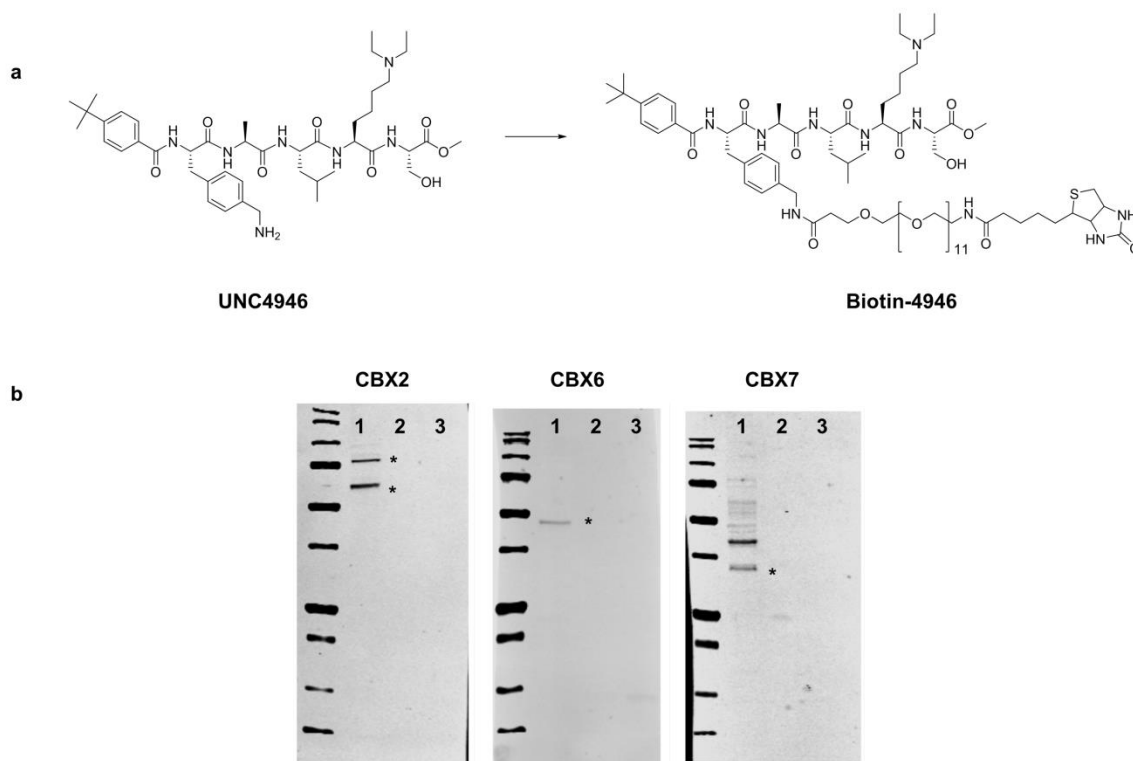
**Figure 4.1.** (a) UNC4195 pull-down studies from PC3 cell lysates. (b) Core components of PRC1 and their respective family members. (c) Pull-down of representative PRC1 components from PC3 cell lysates with UNC4195.

The recent finding that EED, an H3K27me3 reader once thought to only participate in PRC2, can participate in CBX-containing PRC1 complexes intrigued us<sup>57</sup>; therefore, we investigated the ability of UNC4195 to pull-down EED in PC3 cells. Three of the four mammalian isoforms of EED (EED1, -3 and -4)<sup>116</sup> were detected in the PC3 lysate input; however, UNC4195 pulled-down only EED3 and EED4, both of which could be competed away through the addition of UNC3866 to the lysate (**Fig. 4.1c**). Additionally, because EED binds H3K27me3 via its WD40 Kme reading domain, we confirmed that UNC3866 does not bind directly to EED (**Fig. 4.2a** and **4.2b**), indicating that pull-down is mediated through incorporation of EED into PRC1. We next examined the presence of EED in PRC2 using the EZH1 and EZH2 biotinylated pull-down reagent, UNC2399<sup>23</sup>. Interestingly, we detected only EED4 in PRC2 (**Fig. 4.2c**). It has been shown that the incorporation of different isoforms of EED into PRC2 can dictate the *in vitro* histone substrate of PRC2<sup>117</sup>, and hence our discovery that EED incorporates into PRC1 in an isoform-dependent fashion provides the basis for future work to examine the effects of these isoforms on PRC1 function. Furthermore, the absence of EED1 in either PRC1 or PRC2 raises questions as to the binding partners and function of this isoform in PC3 cells.



**Figure 4.2. UNC3866 does not bind to EED as assessed by ITC** (a) Titration of UNC3866 into a solution of EED. (b) Titration of EED into a solution of UNC3866. (c) EED incorporation into PRC2. Only EED4 is detected in PRC2 following pull-down in PC3 cell lysates with UNC2399, a biotinylated derivative of the EZH1/2 inhibitor, UNC1999<sup>23</sup>. Soluble UNC1999 was added to the lysate and successfully prevents chemiprecipitation of EED4.

Our studies in **Table 3.8** indicated that the 4-position of the phenylalanine of UNC3866 is able to accommodate larger functionalities. To this end, we generated a biotinylated derivative of UNC4946 as an alternative pull-down reagent with the aim of accessing CBX2 and -6 in our pull-down studies. Surprisingly, we found that not only was **Biotin-4946** unable to pull-down CBX2 and -6, it also lost the ability to pull-down CBX7, indicating that pull-down of these endogenous proteins from cell lysates is highly sensitive to the positioning of the chemical tag (**Fig. 4.3**). It is possible that biotinylation of the ligand produces a steric clash in the context of specific intact PRC1 complexes that can't be anticipated from CBX domain co-crystal structures.



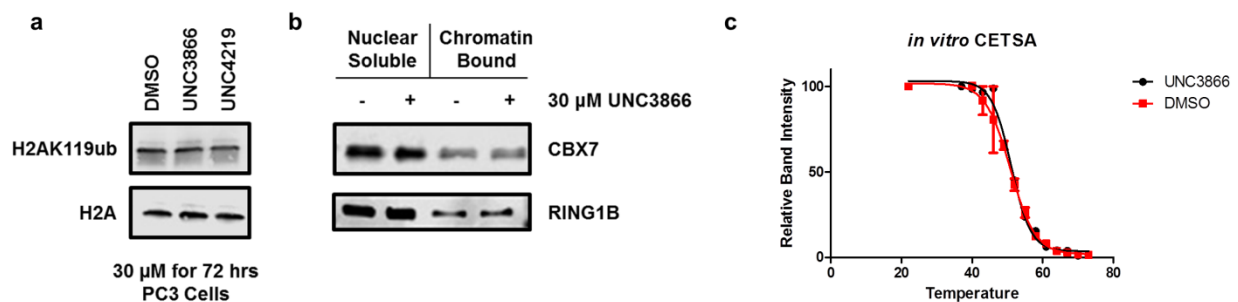
**Figure 4.3.** (a) structures of UNC4946 and Biotin-4946. (b) **Western blots from pulldown Experiments with biotin-47.** Target protein bands are marked with an \*. For CBX2, two isoforms are known to exist, resulting in the presence of two CBX2 bands. For each experiment, lane 1 = input, lane 2 = biotin-47 pull-down and lane 3 = biotin-47 pull-down + 100  $\mu$ M UNC3866 (1).

#### *UNC3866 mechanism of action studies*

Having demonstrated that UNC3866 is capable of engaging full-length, endogenous CBX4/7/8 and CDYL1b/2, we sought to demonstrate dose-dependent modulation of CBX7 in a cellular context. As discussed, the non-enzymatic nature of chromodomains makes demonstrating cellular engagement of these proteins challenging. However, we reasoned that UNC3866 may be able to modulate the E3 ligase activity of PRC1. Therefore, we evaluated the ability of UNC3866 modulate monoubiquitination of H2AK119 (H2AK119ub) by PRC1 and found that it was unable to do so (**Fig. 4.4a**). To understand this result, we evaluated whether UNC3866 is capable of displacing CBX7 and/or RING1B from chromatin and again found no effect from UNC3866 treatment (**Fig. 4.4b**). These results were not surprising for several reasons. First, the presence of non-canonical PRC1 complexes that do not contain CBX



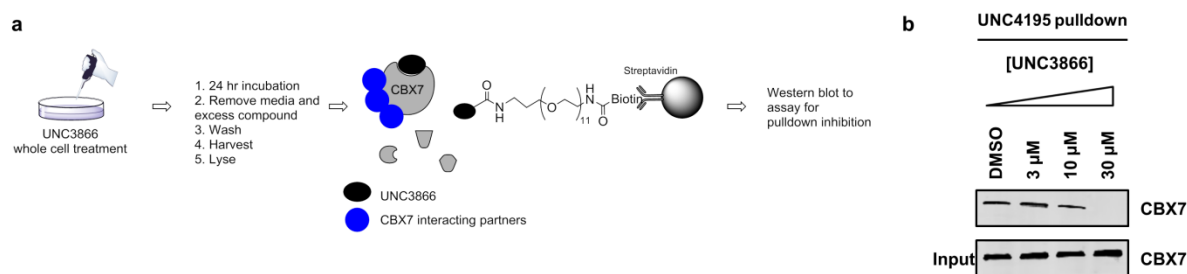
proteins<sup>89</sup> makes monoubiquitination of H2AK119 an unreliable readout. Second, it has been shown previously that the E3 ligase components (RING1b and PCGF proteins) maintain strong affinity for nucleosomes in the absence of CBX proteins<sup>13</sup>. Finally, previous studies have shown that both inactivating mutations and deletions of the chromodomain of CBX proteins have no effect on the association of PRC1 with chromatin<sup>96</sup>. While FRAP has been previously applied to study the effects of antagonists on chromatin readers<sup>8, 75</sup>, previous studies have shown that mutation and deletion of CBX chromodomains has no detectable effect on their FRAP-based cellular mobility<sup>81</sup>, ruling out this technique as a part of mechanism of action studies. Additionally, we failed to detect thermostabilization of CBX7 by UNC3866 in cell lysates (**Fig. 4.4c**), likely because the chromodomain comprises a relatively small portion (~25%) of the full-length protein, which precludes evaluation of cellular target engagement via the Cellular Thermal Shift Assay (CETSA)<sup>115</sup>.



**Figure 4.4.** (a) Western blot analysis of global H2AK119ub levels in PC3 cells following DMSO, UNC3866 or UNC4219 treatment. (b) Western blot analysis following subcellular fractionation of nuclear soluble and chromatin bound fractions of whole cells treated with DMSO or UNC3866 for 24 hours. This result was the same in both isolated nuclei and isolated chromatin treated with UNC3866 (data not shown). (c) UNC3866 does not thermostabilize CBX7 in PC3 cells lysates. CETSA was performed in PC3 cell lysates using DMSO or 30  $\mu$ M of UNC3866. No change in the  $T_M$  of CBX7 was observed (N=3).

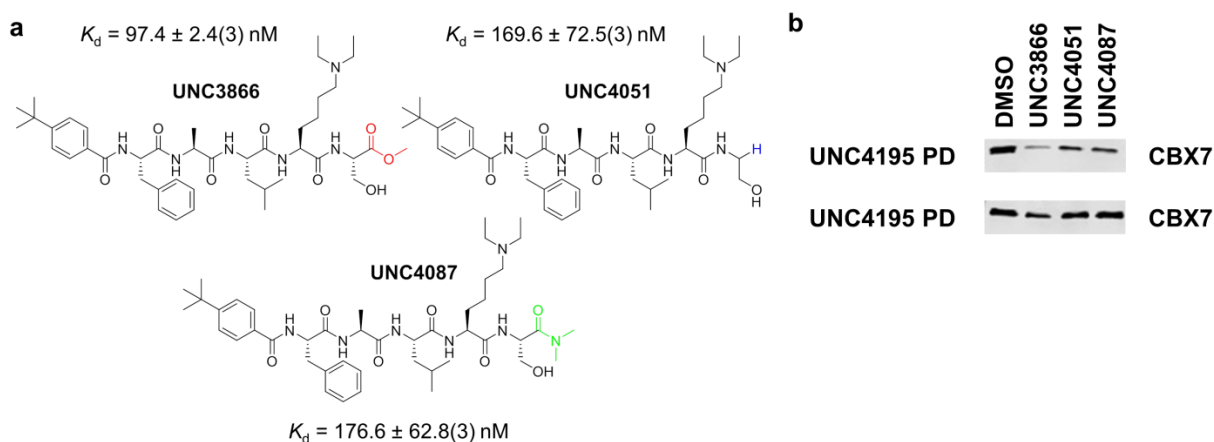
Our struggles to demonstrate cellular target engagement through established techniques prompted us to pursue an alternative approach. We hypothesized that competition-based chemiprecipitation experiments could be utilized to demonstrate cellular target engagement, with the

expectation that pretreatment with UNC3866 would block chemiprecipitation (**Fig. 4.5a**). Whole PC3 cells were incubated with varying concentrations of UNC3866 for 24 hours, after which excess media and compound were removed and the cells were washed, harvested with trypsin and lysed. Encouragingly, CBX7 chemiprecipitation with UNC4195 was completely inhibited in lysates from cells pretreated with 30  $\mu$ M UNC3866, indicating that UNC3866 effectively engages CBX7 in cells at this concentration (**Fig. 4.5b**).



**Figure 4.5.** (a) Cartoon schematic of competition pull-down assay. (b) Incubation of intact PC3 cells with UNC3866 for 24 hours inhibits pull-down of CBX7 with UNC4195.

We sought to determine whether any of our C-terminal derivatives of UNC3866 (**Table 3.8**, **Fig. 4.6a**) were better able to engage the chromodomain of CBX7. While these molecules do display an approximately 2-fold reduction in their potency for CBX7 relative to UNC3866 (**Table 3.8** and **Fig. 4.5a**), we hypothesized that the lack of a C-terminal methyl ester on these molecules, which eliminates their susceptibility to hydrolysis by esterases, could make them more metabolically stable and possibly more membrane permeable, thereby increasing their cellular efficacy. We tested these compounds in our competition pull-down assay (**Fig. 4.5a**) at a concentration of 30  $\mu$ M and compared their ability to inhibit CBX7 pull-down by UNC4195 (**Fig. 4.6b**). Surprisingly, pull-down was reduced most in cells pretreated with UNC3866, indicating that UNC3866 is the most cellularly efficacious of these three compounds.



**Figure 4.6.** (a) Structures of additional CBX7 antagonists tested in our competition pull-down assay. (b) UNC3866 is the most potent cellular CBX7 antagonist of the compounds tested.

#### Identification of the active species

Having demonstrated that UNC3866 is capable of antagonizing the chromodomain of CBX7, we sought to confirm that UNC3866 was in fact the active species in cells. The high concentration of UNC3866 relative to its *in vitro*  $K_d$  suggested that the permeability of UNC3866 is low. This is unsurprising as peptidic inhibitors are often plagued by limited membrane permeability<sup>118</sup>. Therefore, we assessed the permeability of UNC3866 using a Caco-2 assay and found it to be low (**Fig. 4.7**), consistent with our need for relatively high compound concentrations for cellular studies. Additionally, as mentioned previously, we noted the potential for UNC3866 to serve as a pro-drug if the C-terminal methyl ester is susceptible to hydrolysis by intracellular esterases; therefore, we confirmed through a combination of pull-down and ITC experiments that the corresponding C-terminal acid, UNC4007, is still capable of engaging PRC1 chromodomains (**Fig. 4.8a** and **4.8b**). Next, we quantified the intracellular concentration of both UNC3866 and UNC4007 by LC-MS/MS (**Fig. 4.8c**). Treatment for 24 hours at 30  $\mu\text{M}$  UNC3866 resulted in an intracellular concentration of  $1.4 \pm 0.3 \mu\text{M}$ , corresponding to ~5% of the extracellular concentration. The UNC3866 intracellular concentration was approximately 7-fold higher than the intracellular concentration of UNC4007, indicating that hydrolysis of UNC3866 is limited under these conditions. This data correlates well with our pull-down results in **Fig. 4.5b**, wherein pull-down is reduced only slightly at 10  $\mu\text{M}$  and completely at 30  $\mu\text{M}$ . These concentrations would be expected to yield intracellular concentrations of

UNC3866 of 500 nM and 1.4  $\mu$ M, which are approximately 5-and 10-times the  $K_d$  of UNC3866 for the chromodomains of CBX4 and -7, respectively.

#### CELL BATCH QUALITY CONTROL RESULTS

##### Description of Caco-2 Plates Used and Permeability Characterization

Plates	TW12	
Seed Date	9/17/14	
Passage Number	60	
Age at QC (Days)	20	
Age at Experiment (Days)	23	Acceptance Criteria
Atenolol $P_{app}$ , $10^{-6}$ cm/s	0.33	< 0.5
Propranolol $P_{app}$ , $10^{-6}$ cm/s	18.1	10-30
Digoxin (A-to-B) $P_{app}$ , $10^{-6}$ cm/s	0.84	N/A
Digoxin (B-to-A) $P_{app}$ , $10^{-6}$ cm/s	13.8	N/A
Digoxin Efflux Ratio	16.4	> 10.0

#### EXPERIMENTAL RESULTS

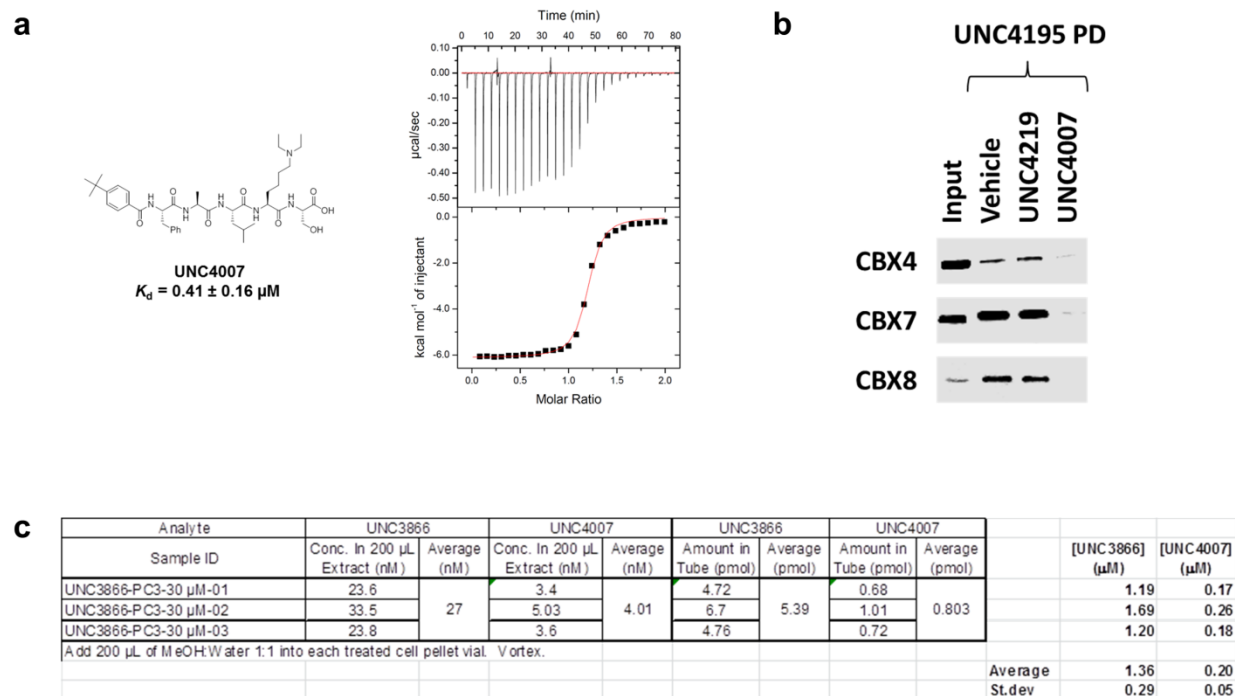
Test Article	Direction	Recovery (%)	$P_{app}$ ( $10^{-6}$ cm/s)			Efflux Ratio	Absorption Potential Classification	Significant Efflux
			R1	R2	AVG			
UNC3866	A-to-B	87	0.08	0.06	0.07	1.9	Low	No
	B-to-A	91	0.14	0.13	0.13			

#### INTERPRETATION AND ADVANCEMENT POTENTIAL

Absorption Potential Classification:  $(P_{app} \text{ A-B}) < 1.0 \times 10^{-6}$  cm/s: **Low**  
 $(P_{app} \text{ A-B}) \geq 1.0 \times 10^{-6}$  cm/s: **High**

Significant Efflux: Efflux ratio  $\geq 2.0$  and  $(P_{app} \text{ B-A}) \geq 1.0 \times 10^{-6}$  cm/s

**Figure 4.7.** UNC3866 exhibits low membrane permeability as measured by the Bi-Directional Caco-2 Permeability Assay. Assay performed by Absorption Systems LP; Exton, PA.



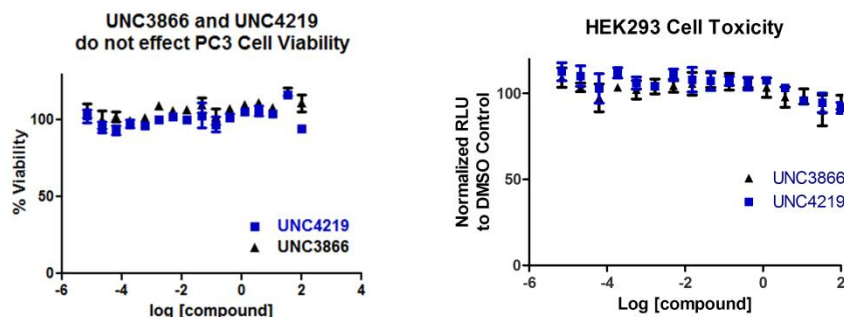
**Figure 4.8.** UNC4007 interacts with PRC1 chromodomains. **(a)** Structure of UNC4007 which binds CBX7 with a  $K_d$  of  $0.41 \pm 0.16 \mu\text{M}$ . The ITC curve shown is representative of the binding interaction of UNC4007 to CBX7. The  $K_d$  is reported as the average of 3 replicates  $\pm$  the standard error of the mean. **(b)** When added to PC3 cell lysates, excess soluble UNC4007 inhibits pull-down of CBX4, -7 and -8 while UNC4219 does not. **(c)** Intracellular concentration of UNC3866 and UNC4007 following treatment of PC3 cells with 30  $\mu\text{M}$  UNC3866 for 24 hrs. Intracellular concentration was calculated using the following equation with 2.9 pL used as an estimate for the volume of PC3 cells based on a previous report<sup>119</sup>.

$$\frac{\text{pmol per tube}}{\text{average cell count of the 3 samples}} = \frac{\text{pmol}}{\text{cell}} \div 2.9 \text{ pL} = [\text{Intracellular}] (M) * 1,000,000 = [\text{Intracellular}] (\mu\text{M})$$

#### Utility of UNC3866 as a chemical probe

Having demonstrated that UNC3866 is sufficiently membrane permeable to engage the chromodomain of CBX7 in a cellular context, we next focused on establishing a biological consequence of treating cells with UNC3866. The high levels of CBX proteins in PC3 cells, coupled with the demonstrated ability of CBX7 overexpression to confer a growth advantage in these cells<sup>16, 43</sup>, encouraged us to investigate the effects of UNC3866 on PC3 cell proliferation. As a prelude to these studies, we confirmed that there were no toxic effects induced upon treatment with UNC3866 up to 100

$\mu\text{M}$  as assessed by CellTiter-Glo® in PC3 and HEK293T cells (**Fig. 4.9**). Similarly, UNC4219 exhibited no observable toxicity within the same concentration range. HEK293T cells were used to evaluate UNC3866 toxicity in a non-cancerous cell line.

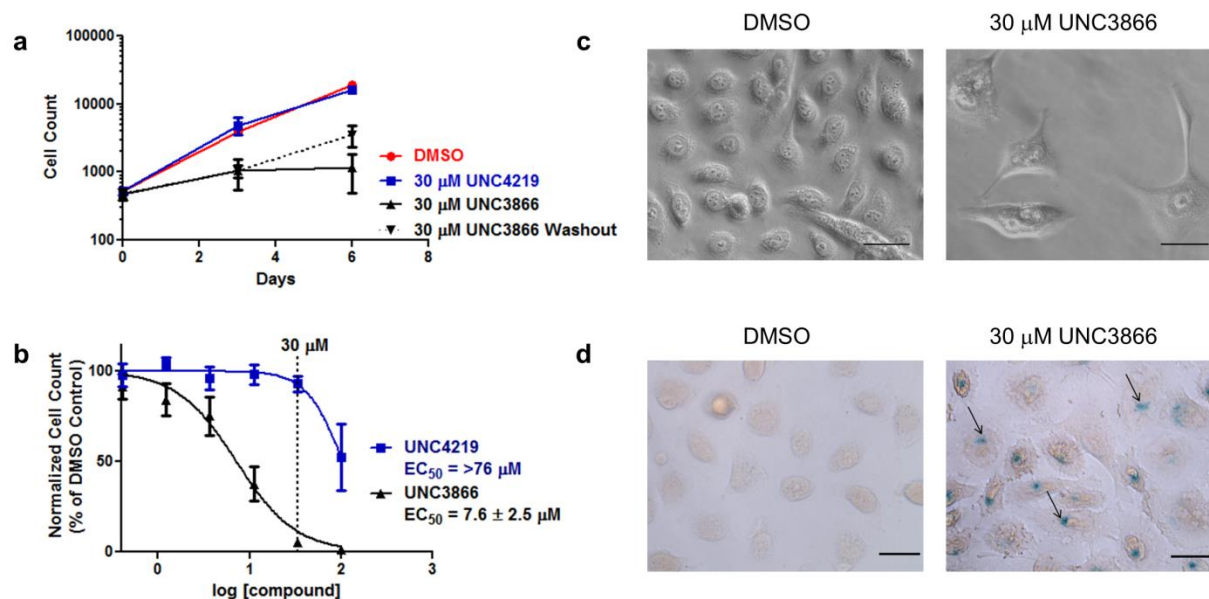


**Figure 4.9.** UNC3866 and UNC4219 display no toxicity up to 100 $\mu\text{M}$  in PC3 and HEK293T cells as assessed by CellTiter-Glo®.

To assess effects on proliferation, PC3 cells were treated with UNC3866 and the negative control, UNC4219, at 30  $\mu\text{M}$ , as this concentration was shown to effectively engage CBX7 in cells (**Fig. 4.5b**). Consistent with the known ability of CBX7 to enhance proliferative capacity in PC3 cells<sup>14, 43</sup>, we observed a dramatic decrease in cell proliferation upon UNC3866 treatment, resulting in almost no change in cell count after 3 and 6 days, while cells treated with UNC4219 or DMSO exhibited exponential cell growth (**Fig. 4.10a**). Further, we observed that cells resumed proliferating when washed and not replenished with UNC3866 after 3 days (**Fig. 4.10a**, dotted line), indicating that the compound effects are at least partially reversible. Dose dependency was next examined for UNC3866 and UNC4219, and a significant reduction in PC3 cell count was only observed for UNC3866 treated cells, resulting in an  $\text{EC}_{50}$  of  $7.6 \pm 2.5 \mu\text{M}$  (**Fig. 4.10b**). We also observed that PC3 cells treated with UNC3866 exhibited an enlarged and flattened morphology (**Fig. 4.8c**), suggesting that these cells had become senescent<sup>120</sup>. We checked for the presence of senescence-associated  $\beta$ -galactosidase (SA- $\beta$ -gal) and observed a dramatic increase in the number of cells positive for SA- $\beta$ -gal following treatment with 30  $\mu\text{M}$  UNC3866 (**Fig. 4.10d**). This was an intriguing observation because CBX7 was originally identified in a genetic screen for

genes capable of bypassing replicative senescence in Human Prostate Epithelial Cells (HPrEC)<sup>15</sup>.

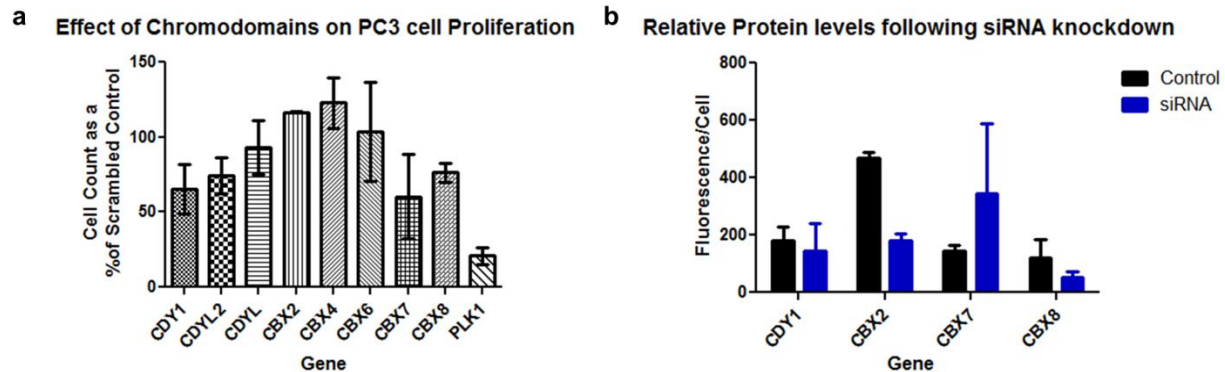
Further, the same study observed that CBX7 knockdown is capable of inducing senescence.



**Figure 4.10.** (a) Treatment of PC3 cells with 30  $\mu$ M UNC3866 inhibits cell proliferation. UNC4219 treatment has no effect on PC3 cell proliferation. For washout experiments (dotted line), UNC3866 was not replenished at day 3. Error bars represent the SD of that data point (n=3, single biological replicate with 3 technical replicates). The data is representative of at least 4 biological replicates with varying plating densities. (b) UNC3866 inhibits PC3 cell proliferation in a dose-dependent manner after exposure for 6 days (n=8, 2 biological replicates with 3 technical replicates and 1 biological replicate with 2 technical replicates). The EC<sub>50</sub> is reported as the 95% confidence interval. UNC4219 was used to control for off-target effects, which were not evident at concentrations below 100  $\mu$ M. The EC<sub>50</sub> for UNC4219 is >76  $\mu$ M, which is the lower bound of the 95% confidence interval (n=6, 3 biological replicates with 2 technical replicates each). Error bars represent the SEM of that data point. (c) UNC3866 induces a senescent-like morphology in PC3 cells and (d) expression of SA- $\beta$ -gal. Arrows indicate examples of positive SA- $\beta$ -gal expression. Cell morphology was assessed using phase contrast microscopy (40X). SA- $\beta$ -gal expression was assessed using bright-field microscopy (40X). The images are representative of 3 biological replicates and the black scale bars are equal to 60  $\mu$ m.



We attempted an siRNA screen with each of the UNC3866 target chromodomains to determine if knockdown of any of these chromodomains could recapitulate the anti-proliferative effect of UNC3866 (**Fig. 4.11**). PC3 cells treated with siRNA (up to 100 nM) targeting each protein hinted that knockdown of CDY1, CDYL2 and to some extent CBX7 had a slight effect on PC3 cell proliferation. However, quantification of the protein levels by In-Cell Western blotting (**Fig. 4.11b**) of all targets for which we had compatible antibodies revealed that we were unable to obtain significant knockdown of any chromodomain except CBX2, rendering the results in **Fig. 4.11a** meaningless. Nonetheless, while we could not confirm the on-target effects of UNC3866 through this siRNA screen, the intracellular concentration of UNC3866 at its EC<sub>50</sub> in PC3 cells is estimated to be 340 nM (**Fig. 4.8** and **Fig. 4.10**), which is adequate to fully engage CBX4 and -7 based on the measured ~100 nM *K<sub>d</sub>*'s versus these targets (**Fig. 3.2a**). Although some contribution of other chromodomains bound by UNC3866 to this antiproliferative phenotype cannot be ruled out, the concordance of this result with previously published CBX7 genetic manipulation is striking<sup>15, 17</sup>.

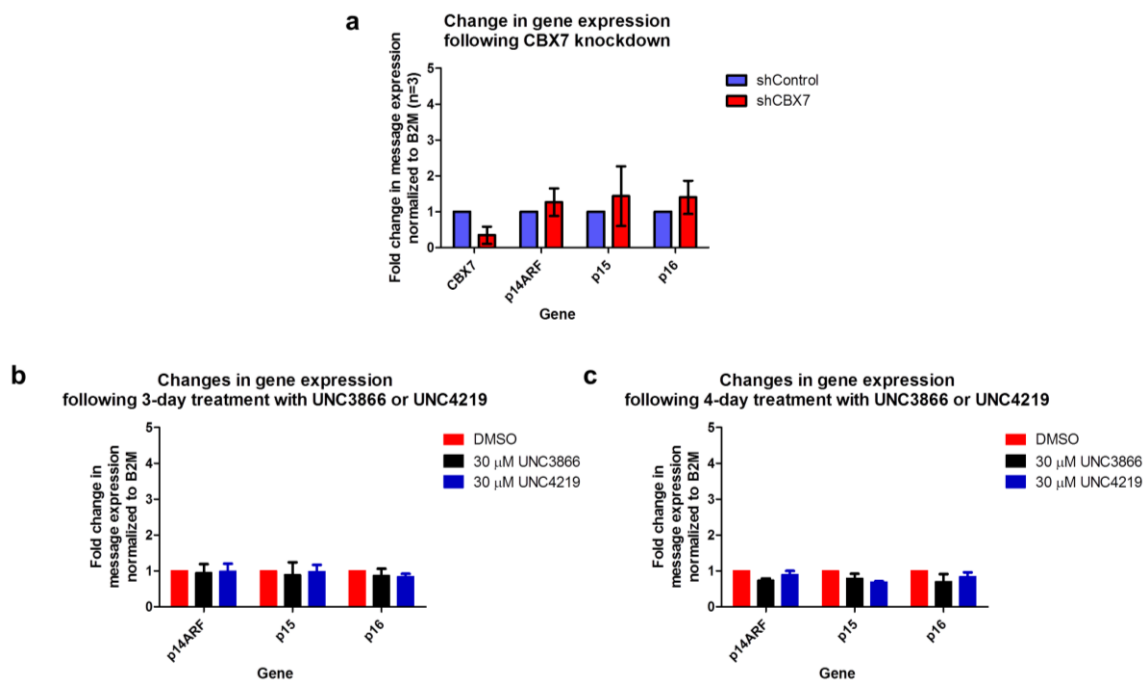


**Figure 4.11.** (a) Relative cell counts of PC3 cells 96-hours post siRNA exposure. (b) Relative protein levels of siRNA targets.

While it has been suggested that CBX7 controls cellular proliferation, especially in prostate tissue, through regulation of the *Ink4a/ARF* locus<sup>15</sup>, its actual role in regulating this locus is controversial<sup>15, 19, 87</sup>. For example, CBX7 has been shown to enhance proliferative capacity without negatively regulating this locus<sup>19</sup>, and negative regulation of this locus does not necessarily enhance proliferative capacity<sup>80</sup>.



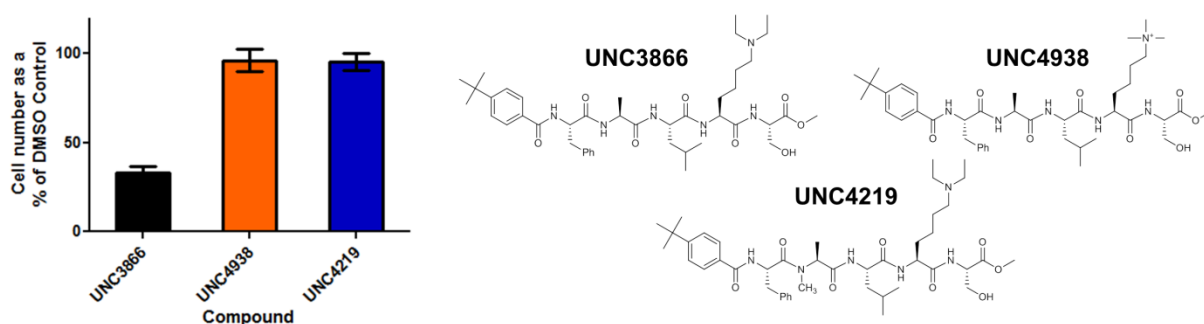
Further, there are conflicting reports on the regulation of this locus by CBX7 in PC3 cells<sup>15, 16, 18, 43</sup>. Nonetheless, we analyzed the expression of this locus in response to knockdown of CBX7 by shRNA (>60% mRNA reduction) in PC3 cells and did not observe a statistically significant change in expression of the *Ink4a/ARF* locus under these conditions. Further, evaluation of the effects of UNC3866 on expression of this locus showed no change at 30  $\mu$ M, ruling out *Ink4a/ARF* regulation in the antiproliferative effects of UNC3866 (**Fig. 4.12**). This was not entirely surprising as there is considerable evidence indicating that this locus is deeply repressed in PC3 cells through DNA hypermethylation<sup>121-124</sup>. Further, methylation of this locus is unaffected by both DNMT knockdown and inhibition with 5-AzaC, suggesting that this gene may be irreversibly repressed in PC3 cells<sup>121-123</sup>.



**Figure 4.12.** Analysis of mRNA transcript levels from the *Ink4a/ARF* locus by RT-qPCR following CBX7 knockdown in PC3 cells (**a**) and treatment of PC3 cells with UNC3866 or UNC4219 (**b**) for 3 (left) or 4 (right) days. For the CBX7 knockdown experiments, the data are plotted as the mean  $\pm$  the standard deviation of 3 biological replicates, each done in duplicate. Values from compound treated cells are the average of 2 technical replicates from a single biological replicate  $\pm$  the standard error of the mean.

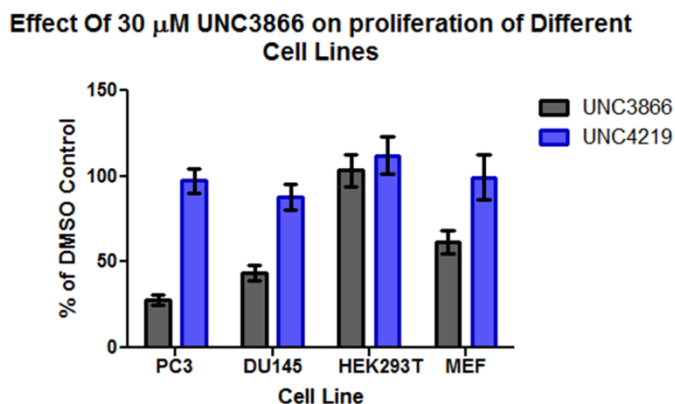
Next, we sought to compare the cellular activity of UNC4938 to UNC3866 since we initially hypothesized that replacement of the Kme3 was required for passive membrane permeability. We first measured the intracellular concentration of UNC4938. While treatment at a 30  $\mu\text{M}$  extracellular dose of UNC3866 for 24 hours results in an intracellular concentration of  $1.4 \pm 0.3 \mu\text{M}$  in PC3 cells, under the same conditions, we observed an intracellular concentration of  $0.61 \pm 0.2 \mu\text{M}$  of UNC4938. It was surprising that incorporation of trimethyl-lysine resulted in only an approximately 2-fold decrease in the permeability of the compound, and we were intrigued to investigate whether UNC4938 was functionally available for chromodomain binding in cells since it is known that highly charged compounds may accumulate in endosomes and lysosomes<sup>99, 100</sup>.

Having shown that inhibition of PC3 cell proliferation by UNC3866 is predominantly accomplished through antagonism of polycomb CBX chromodomains, we wanted to assess the effect of Kme3 substitution on the functional activity of our peptidic antagonists. Therefore, we compared the anti-proliferative effects of UNC3866 and UNC4938 to our negative control analog of, UNC4219 (**Fig. 4.13**). Despite having nearly identical *in vitro* potencies for CBX chromodomains as UNC3866 (**Table 3.7**) and only a 2-fold reduction in permeability, UNC4938 had no effect on PC3 cell proliferation, analogous to the negative control (**Fig. 4.13**), supporting our hypothesis that Kme3 replacement is required for cellular activity. We suspect that the lack of biological activity observed upon treatment with UNC4938 may be the result of endosomal and lysosomal trapping of the molecule<sup>99, 100</sup>.



**Figure 4.13.** Effect of compounds UNC3866, UNC4938, and UNC4219 on PC3 cell proliferation. Cells were treated for 72 hours with the indicated compounds (30  $\mu\text{M}$ ) or DMSO, fixed and DAPI stained. Cell counts were numerated using high content microscopy and normalized to DMSO control ( $n \geq 4$ ).

Finally, we tested the anti-proliferative effects of UNC3866 across a panel of multiple cell lines (**Fig. 4.14**). We tested another human prostate cancer cell line, DU-145 cells, HEK293T cells and MEF cells and compared them to the anti-proliferative effects of UNC3866 in PC3 cells. UNC3866 is most potent in PC3 cells, but still has a substantial anti-proliferative effect in DU-145 cells. We observed no effect on proliferation in HEK293T cells, which was not surprising as these cells have previously shown to proliferate in a CBX7-independent fashion. Finally, we tested MEF cells to determine if UNC3866 is able to affect growth in non-human cell line known to have its proliferation regulated by CBX7. Consistent with the role of CBX7 in MEFs<sup>15</sup>, we found that UNC3866 reduces proliferation in these cells. Further, this strongly suggests that UNC3866 is able to engage mouse CBX7, which was to be expected as mouse and human CBX7 are identical in their chromodomain regions.



**Figure 4.14.** Varying effects of UNC3866 on cellular proliferation.

#### *Pharmacokinetic studies with UNC3866*

We examined the pharmacokinetic properties of UNC3866 in male Swiss Albino mice in order to determine the potential utility of UNC3866 *in vivo*. We investigated the effects of dosing both intravenously (IV) and intraperitoneally (IP). Mice were dosed IV at 3 mg/kg and IP at 10 mg/kg. The concentration of both UNC3866 and UNC4007 were quantified at multiple time points. When given intravenously, UNC3866 is rapidly eliminated with an observed half-life of 0.09 hours (5.4 minutes) and does not significantly accumulate in body tissues ( $V_{ss} = 0.19$  L/kg). Encouragingly, UNC3866 does have moderate clearance (26 mL/min/kg). When administered intraperitoneally, UNC3866 showed 25% bioavailability and is detectable in the blood of mice for up to 4 hours with maximum levels occurring at 2

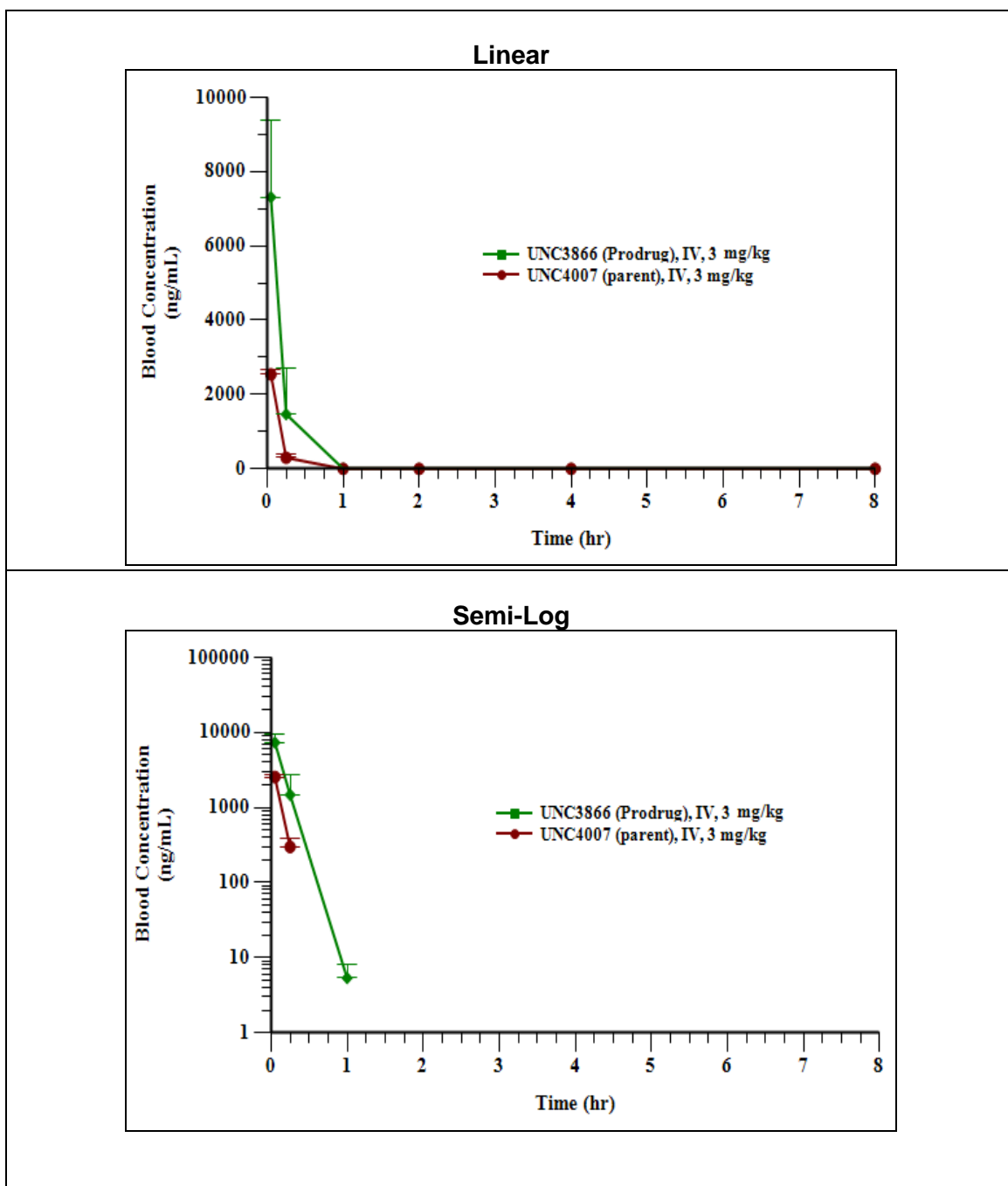
hours. Additionally, UNC3866 was the dominate species present over UNC4007 at all time points tested, further highlighting the stability of the UNC3866 methyl ester. While these PK results are promising for a peptidic compound, the use of UNC3866 *in vivo* may be limited because of the high circulating levels required for intracellular target engagement due to its poor cell permeability. The potential utility of UNC3866 at higher doses for *in vivo* experiments is currently under investigation.

**Table 4.1.** Mean pharmacokinetic parameters of UNC3866 (Prodrug) in blood following a single intravenous (Dose: 3 mg/kg) and intraperitoneal (Dose: 10 mg/kg) administration to male Swiss Albino mice

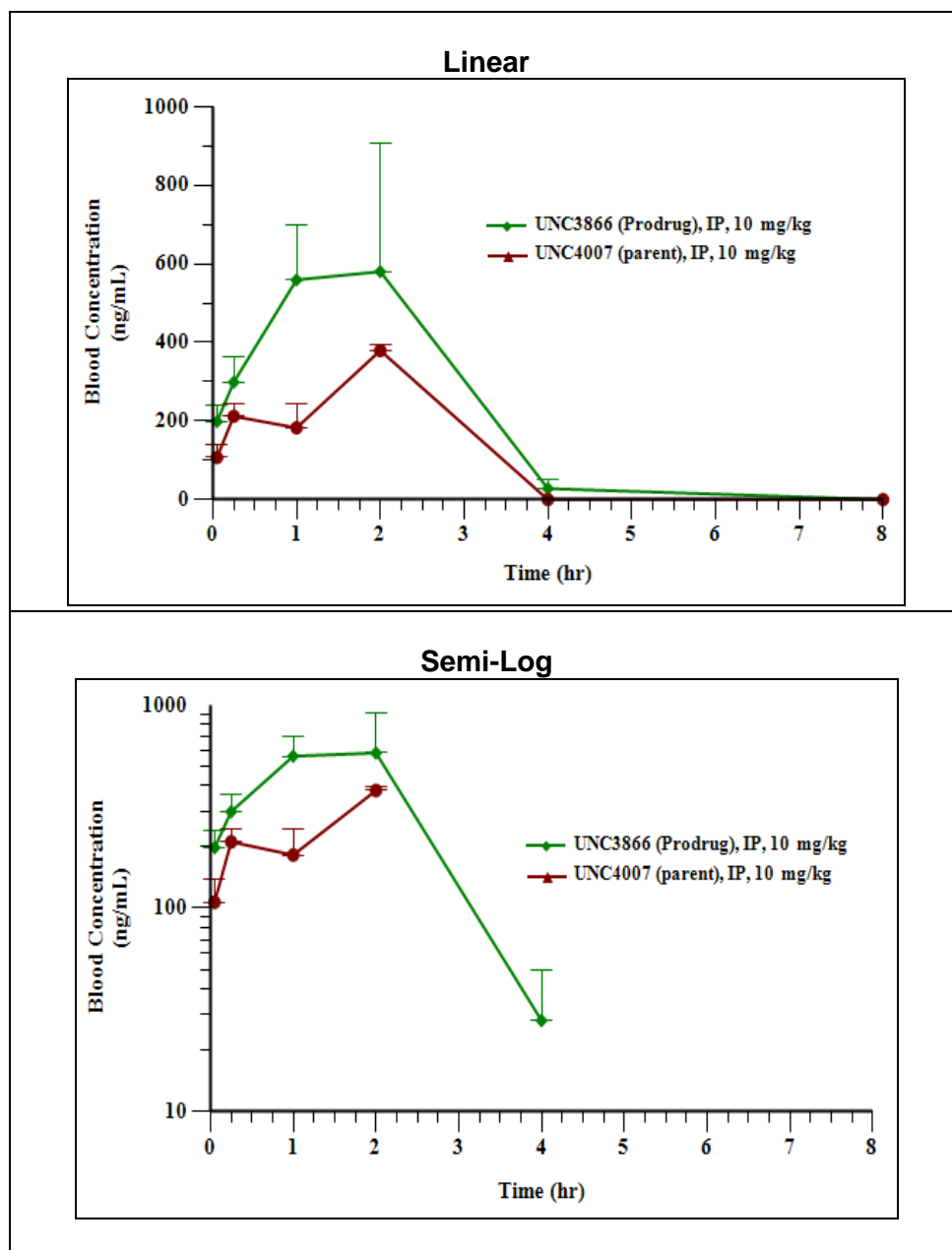
Route	Compound (Dose)	T <sub>max</sub> (hr)	<sup>a</sup> C <sub>0</sub> /C <sub>max</sub> (ng/mL)	AUC <sub>last</sub> (hr*ng/mL)	AUC <sub>inf</sub> (hr*ng/mL)	T <sub>1/2</sub> (hr)	CL (mL/min/kg)	V <sub>ss</sub> (L/kg)
i.v.	UNC3866 (Prodrug)	-	10911.87	1887.25	1887.95	0.09	26.48	0.19
	UNC4007 (Parent)	-	NC	NC	NC	NC	NC	NC
i. p.	UNC3866 (Prodrug)	2.00	580.01	1553.89	NC	-	-	-
	UNC4007 (Parent)	2.00	379.34	463.57	NC	-	-	-

a - back extrapolated conc. for i.v. group

NC: Not calculated due to insufficient data



**Figure 4.15.** Mean blood concentration-time profiles of UNC3866 (Prodrug) and UNC4007 (parent) following a single intravenous administration to male Swiss Albino mice (Dose: 3 mg/kg)

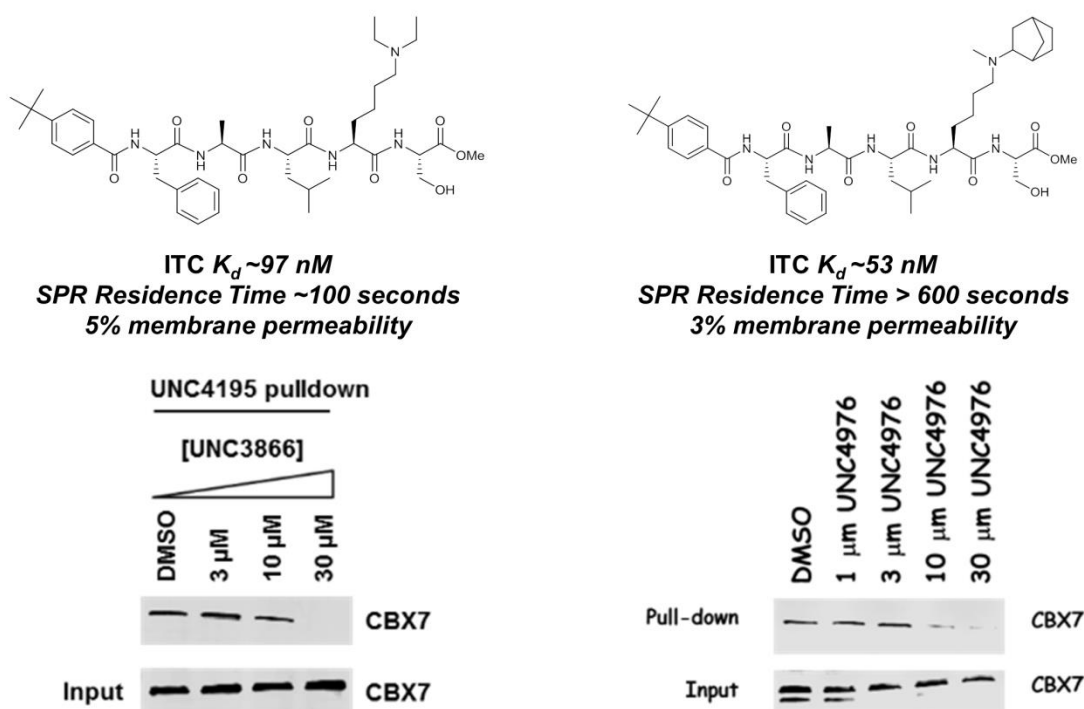


**Figure 4.16.** Mean blood concentration-time profiles of UNC3866 (Prodrug) and UNC4007 (parent) following a single intraperitoneal administration to male Swiss Albino mice (Dose: 10 mg/kg)

#### *The importance of compound residence time on CBX7 antagonism*

We were intrigued to determine whether the increased residence time of UNC4976 with CBX7 relative to UNC3866 significantly effects its interaction with CBX7 in a cellular system. First, we quantified the intracellular concentration of UNC4976 in PC3 cells following a 30  $\mu$ M dose for 24 hours. We found

that the intracellular concentration of UNC4976 was slightly nearly half of that of UNC3866 ( $0.78 \pm 0.3 \mu\text{M}$  vs  $1.4 \pm 0.3 \mu\text{M}$ , respectively). Interestingly however, pull-down with UNC4195 was more strongly antagonized in cells pretreated with UNC4976 relative to UNC3866 (**Fig. 4.17**), despite the reduced permeability and only slightly improved binding affinity (UNC4976-CBX7  $K_d = 53 \text{ nM}$  vs UNC3866-CBX7  $K_d = 97 \text{ nM}$ ) Nearly complete pull-down inhibition was observed at the  $10 \mu\text{M}$  dose of UNC4976, relative to UNC3866, which requires a  $30 \mu\text{M}$  dose in order to completely block pull-down. Further studies are underway to continue to optimize UNC4976 and comparing its cellular efficacy relative to UNC3866, but this result strongly suggests that the development of more cellularly efficacious CBX7 antagonists depends more strongly on off-rate optimization than optimization of the binding constant per se.



**Figure 4.17.** Comparison of the ability of UNC4976 to antagonize CBX7 in cells compared to UNC3866.

## CONCLUSIONS

UNC3866 is a highly selective, cellularly active antagonist of PRC1 chromodomains. Our studies have demonstrated that UNC3866 is capable of engaging CBX4/7/8-containing canonical PRC1 complexes. It is currently unclear whether CBX2/6-containing PRC1 complexes are simply inaccessible to pull-down, or are inaccessible to UNC3866 in general. Further, the precise mechanism by which

UNC3866 antagonism of PRC1 chromodomains modulates their signaling without displacing them from chromatin is currently unclear. Further studies are needed to elucidate whether UNC3866 works by inhibiting the interaction of PRC1 chromodomains with non-histone proteins or whether it exerts its effects through changes in the genomic distribution of PRC1.

The ability of UNC3866 to inhibit proliferation with little apparent toxicity, especially in PC3 cells, is encouraging for the development of CBX antagonists as oncological therapeutic agents. While UNC3866 has modest PK properties, its utility is likely to be limited to cellular systems. Therefore, evaluation of the efficacy and safety of CBX antagonists as therapeutic agents *in vivo* will require the continued development of CBX antagonists with improved permeability and PK properties. Nonetheless, UNC3866 represents an important first step toward these molecules and is a reasonable starting point for further optimization.

While our studies have provided strong support that the phenotypic effects of UNC3866 are exerted through its ability to modulate CBX7, future studies will focus on providing stronger evidence for direct modulation of CBX7 by UNC3866 and its derivatives. Preliminary studies which will be disclosed in future publications have provided much more powerful evidence demonstrating the ability of UNC3866 to antagonize CBX7-mediated repression in cell-based assays. Further, the results of these studies corroborate our findings in **Fig. 4.5**, wherein the methyl ester of UNC3866 is the preferred functionality at the C-terminus of this peptidic antagonist. Finally, these same studies suggest that UNC4976, while only approximately 2-fold more potent for CBX7 than UNC3866 *in vitro*, is > 15-fold more potent at antagonizing CBX7-mediated repression in cells. This result, coupled with our pull-down studies in **Fig. 4.17**, strongly indicates that the efficacy of CBX7 antagonists in biological systems is intimately tied to their dissociative kinetics and that our antagonists will benefit from continued improvements in their target residence time.

## METHODS

### *Cell culture and lysis*

PC3 cells were obtained from ATCC® (CRL-1435™) through the UNC Lineberger Tissue Culture Facility. Cells were cultured using GIBCO® DMEM/F12 (Ham), [+] L-Glutamine, and [+] 15 mM HEPES media.



Cells were trypsinized using 0.25% trypsin. Lysis was performed using Cytobuster™ protein extraction reagent supplemented with protease inhibitors and Benzonase® (used at 25 U/mL). For analysis of H2AK119ub, 50 µM of the deubiquitinase inhibitor, PR-619, was added to the lysis buffer. Samples were incubated at 37°C for 10 minutes, followed by incubation at RT on a rotator for 20 minutes. The samples were spun down and the supernatant collected and transferred to a clean Eppendorf tube. Protein concentrations were quantified using the Bradford protein assay.

#### *Pull-down studies*

Cells were cultured on a T175 tissue culture flask until reaching ~80-90% confluency. Following trypsinization, the pellet was washed twice with PBS and lysed with 500 µL of lysis buffer. When applicable, vehicle control or 100 µM UNC3866 was added to an aliquot of 1000 µg of total protein and diluted to 500 µL with 20 mM Tris (pH 8)/150 mM NaCl/0.1% Tween-20 (TBST). Pull-down reagents were bound to Streptavidin M-270 Dynabeads® by rotating 30 µL of beads with a 20-fold excess of pull-down reagent for 30 minutes at room temperature in. Unbound pull-down reagent was then removed and the beads were washed 3X with 200 µL of TBST. The lysate was then transferred to an Eppendorf tube containing 30 µL of Streptavidin M-270 Dynabeads® that had been pre-bound to UNC4195 or 2399. The mixture was rotated overnight at 4°C. The following morning, the depleted lysate was removed and the beads were washed 3X with 300-500 µL of TBST. The beads were re-suspended with 30 µL of 1X Laemmli sample buffer and heated at 95°C for 3 minutes. Fifteen microliters of the sample was then loaded into a gel and analyzed via western blotting using the appropriate primary antibody (**Table 4.2**) and detected using a LI-COR Odyssey® instrument and the appropriate fluorescent secondary antibodies (diluted 1:10,000 in PBST). One percent of input was used for western blotting of all proteins except CBX8 and HPH2. For both of these proteins, 3% of sample input was used.

**Table 4.2.** Primary antibodies used for western blot detection.

Antibody	Dilution	Vendor (catalog number)
Rabbit $\alpha$ CBX2	1:5000 in PBST	Abcam (ab184968)
Rabbit $\alpha$ CBX4	1:5000 in PBST	Abcam (ab174300)
Rabbit $\alpha$ CBX6	1:5000 in PBST	Abcam (ab195235)
Rabbit $\alpha$ CBX7	1:5000 in PBST	Abcam (ab21873)
Rabbit $\alpha$ CBX8	1:5000 in PBST	Abcam (ab182627)
Rabbit $\alpha$ RING1B	1:5000 in PBST	Abcam (ab101273)
Rabbit $\alpha$ HPH2	1:2000 in PBST	Abcam (ab168441)
Rabbit $\alpha$ H2AK119ub	1:5000 in PBST	Cell Signaling (8240)
Mouse $\alpha$ BMI-1	1:5000 in PBST	Active Motif (39993)
Mouse $\alpha$ EED	1:5000 in PBST	EMD Millipore (17-663)

#### *Cell Proliferation Assay*

PC3 cells were seeded at 200 cells/well into 24-well plates (Costar #3524). Cells were allowed to adhere overnight. The media (DMEM supplemented with 10 % FBS) was then exchanged with fresh media containing DMSO, UNC3866 or UNC4219. On day three, the media were exchanged with fresh media containing DMSO, UNC3866 or UNC4219. For dose-response studies, the EC<sub>50</sub> was derived from a six-point titration ranging from 100  $\mu$ M to 0.4  $\mu$ M of UNC3866 or UNC4219. At day 0, 3 or 6, cells were fixed with ice-cold methanol for 30 sec. and rehydrated with PBS. Nuclei of the cells were stained with DAPI (0.05  $\mu$ g/ml) and numerated using High Content Microscopy (Array Scan™ High Content Analysis, Thermo Fisher #NX10002L). For dose-response studies, the cell count of UNC3866- or UNC4219-treated cells was normalized to the average cell count of DMSO-treated cells. The EC<sub>50</sub> was calculated using the “log[inhibitor] vs. the normalized response -- Variable slope” equation in GraphPad Prism 5.

#### *Subcellular fractionation experiments*

PC3 Cells were cultured on a 10-cm dish. For whole cell treatments, compounds were added when cells reached ~70% confluency. For treatment of isolated cellular compartments, cells were harvested after reaching approximately 80-90~ confluency. Experiments were performed with Subcellular Protein Fractionation Kit for Cultured Cells from ThermoFisher (catalog number 78840). The manufacturer's recommended protocol was followed except when treating isolated nuclei and chromatin with compound.

In these cases, a 30-minute compound incubation step was added in which the appropriate cellular compartment was rotated at room temperature for 30-minutes with DMSO, UNC3866 or UNC4219.

#### *Compound permeability studies*

PC3 cells were cultured until reaching ~80% confluency in a T75 cell culture flask. The media was exchanged with 10 mL fresh media (0.3% DMSO) containing 30  $\mu$ M compound. Cells were incubated with compound for 24 hours. After 24 hours, cells were washed with PBS, trypsinized, and wash again with PBS. The resulting cell pellets (3) were flash frozen with liquid nitrogen and submitted to Drumetix Laboratories (Greensboro, NC) for analysis by LC-MS/MS.

#### *CellTiter-Glo Luminescent Cell Viability Assay*

The effect of UNC3866 and UNC4219 on cell viability was determined using a CellTiter-Glo™ ATP detection system (Promega #7573). Ten point, 1:3 dilution curves of compounds starting at 100  $\mu$ M final concentration were diluted to 5X final concentration in PBS (vehicle control) and then 5  $\mu$ L were added to 384-well white, clear bottom tissue culture plates (Corning #3707) with a Multimek automated liquid handling device (Nanoscreen, Charleston, SC). Twenty microliters of low passage, subconfluent HEK293T/PC3 cells grown in Dulbecco's Modified Eagle's Medium without phenol red (GIBCO® #31053) and supplemented with 10% Fetal Bovine Serum (GIBCO® #26140) were immediately added at a density of 5,000 cells per well using a Multidrop 384 (Titertek). Cell plates were incubated for 48 hours at 37°C and 5% CO<sub>2</sub>, and then lysed with 25 microliters of CellTiter-Glo™ reagent. Luminescence was read on an Envision platereader (Perkin Elmer) after 15 minutes at room temperature in dim light.

#### *Cellular Thermal Shift Assay (CETSA)*

A PC3 cell pellet consisting of 10<sup>7</sup> cells was re-suspended in 1200  $\mu$ L of PBS complete with protease inhibitors. The suspension was subjected to 3 consecutive freeze-thaw cycles performed with liquid N<sub>2</sub> in order to lyse the cells. Following the final freeze-thaw cycle, the suspension was centrifuged at 20,000 g for 20 minutes at 4°C. The supernatant was collected and the resulting pellet was discarded. The lysate was then divided into two tubes (600  $\mu$ L) each. The volume of each tube was brought up to

650  $\mu$ L with PBS + protease inhibitors. DMSO or 30  $\mu$ M UNC3866 was added to each tube. The tubes were incubated at RT for 30 minutes. The lysates were aliquoted into 12 thin-walled PCR tubes (50  $\mu$ L/tube). The tubes were incubated for 3 minutes at temperatures ranging from 37°C to 73°C in 3° increments in a Mastercycle® nexus gradient qPCR machine. Tubes were then incubated at room temperature for 3 minutes followed by centrifugation at 20,000 g for 20 minutes at 4°C. The supernatant was collected. Ten microliters of 6X Laemmli buffer was added to each tube. The tubes were then heated to 95°C for 3 minutes. Twenty microliters was then loaded into an SDS page gel and analyzed by western blotting.

#### *CBX7 knockdown and quantitative PCR*

PC3 cells were seeded at a density of  $10^5$  cells/well in a 6-well plate and allowed to adhere overnight. The media was then exchanged for 2 mL of media containing polybrene (8  $\mu$ g/mL) and pre-packaged shRNA (200,000 Transducing Units/well) targeting CBX7 (Mission CBX7, SHCLNV (Sigma) TRCN0000019144RNA) or Non-targeting control (Mission pLKO.1-puro Non-Target shRNA, SHC016V (Sigma)). After 20 hours, the transduction mixture was exchanged with fresh media (GIBCO® DMEM/F12 (Ham), [+] L-Glutamine, and [+] 15 mM HEPES) and allowed to incubate for an additional 24 hours. The media was then replaced with media containing 2  $\mu$ g/mL of puromycin and allowed to incubate until selection was complete as evidence by the death of all control cells that were not transduced. RNA was extracted using the RNeasy Plus Mini Kit (Qiagen) and RNA was quantified on the NanoDrop 2000 spectrophotometer (Thermo Fisher Scientific). The reverse transcription reaction was performed with 500 ng RNA and the Superscript III First Strand Synthesis Supermix (Invitrogen #1172-050) according to manufacturer instructions. The resulting cDNA was diluted 1:10, and 2  $\mu$ L was used for each PCR reaction. Quantitative PCR was performed using Roche FastStart Universal SYBR Green Master Mix (Rox) in a 384-well plate format on the ABI 2900HT instrument. Oligo sequences are available upon request.

#### *Phase contrast microscopy images*

PC3 cells were seeded at 200 cells/well into 24-well plates (Costar #3524). Cells were allowed to adhere overnight. The media (DMEM supplemented with 10 % FBS) was then exchanged with fresh media containing DMSO or 30  $\mu$ M UNC3866. On day three, the media were exchanged with fresh media containing DMSO or 30  $\mu$ M UNC3866. At day 6, cells were fixed with ice-cold methanol for 30 sec. and rehydrated with PBS. Samples were imaged using Volocity software and an Olympus IX81 microscope equipped with an Orca ER digital camera.

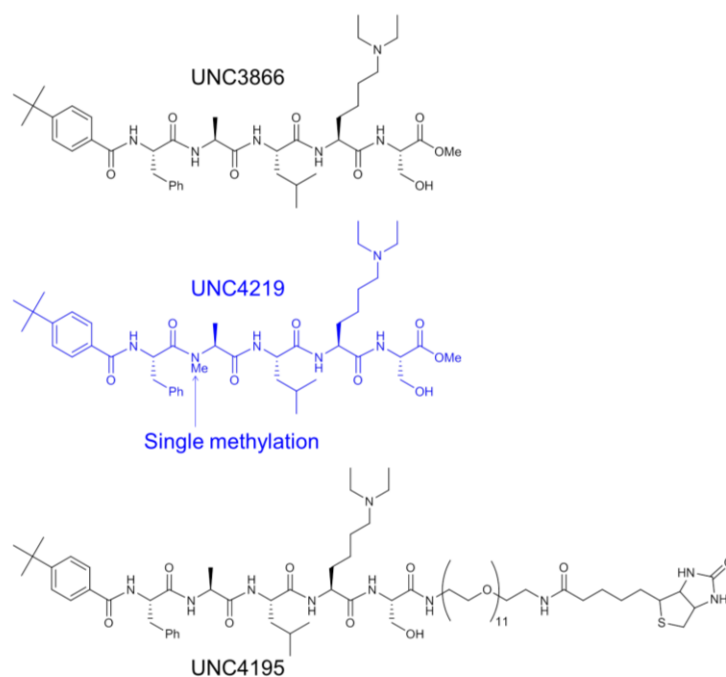
#### *Senescence-associated $\beta$ -Gal staining*

PC3 cells were seeded in a 96-well plate at a density of 100-200 cells/well and allowed to adhere overnight. The media was then replaced with fresh media containing 30  $\mu$ M UNC3866 or DMSO. After 3 days, the media was removed and the cells were fixed and stained using the Senescence  $\beta$ -Galactosidase Kit (#9860) from Cell Signaling Technology® according to the manufacturer's recommendations. Wells were imaged using bright-field microscopy with an Olympus IX81 microscope equipped with an Orca ER digital camera.

## CHAPTER V: FUTURE DIRECTIONS

### SUMMARY OF KEY FINDINGS

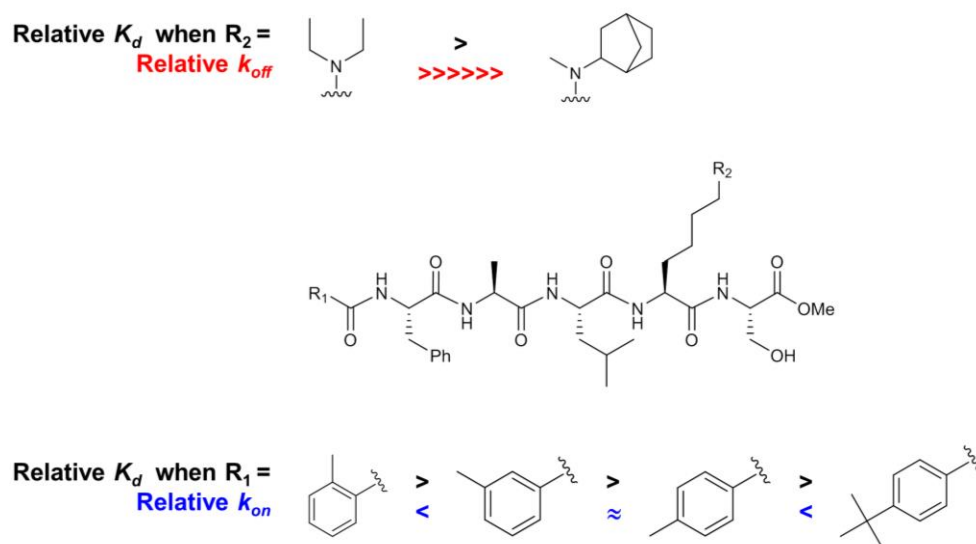
Our efforts to develop antagonists for the chromodomain of CBX7 proved to be both challenging and informative. The primary accomplishment of these studies was the development of a chemical toolkit consisting of UNC3866, a potent antagonist ( $K_d$  of ~100 nM) of the chromodomains of CBX4/7 with modest affinity for the chromodomains of CBX2/6/8, CDY1, CDYL1b and CDYL2; a negative control analog, UNC4219; and UNC4195, a biotinylated analog of UNC3866. We have made these compounds freely available to the research community to be used as a toolset that can be used to study CBX and CDY chromodomains both biochemically and in cell-based systems (**Fig. 5.1**).



**Figure 5.1.** Structures of UNC3866, UNC4219 and UNC4195.

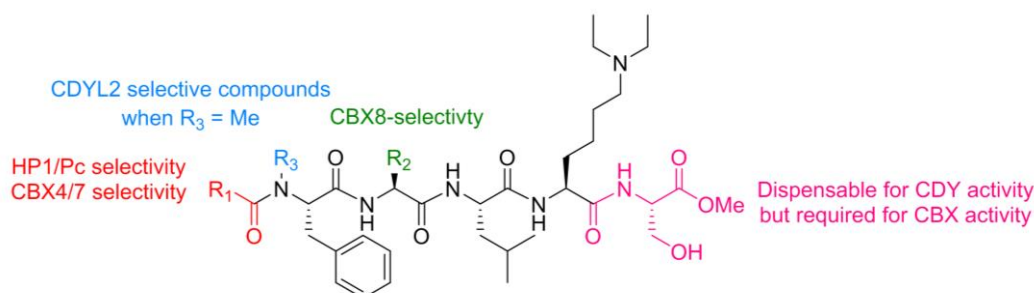
UNC3866 was developed through systematic SAR studies focused on testing our induced fit binding pathway of recognition of H3 by CBX7 based on previous structural studies and our own

molecular dynamics simulations. Our proposed binding pathway hypothesizes that the N-terminus of our peptidic antagonists forms the first specific interaction with CBX7. Further, these interactions are critical for facilitating rearrangement of the chromodomain around the inhibitor and entrance of the chromodomain-antagonist complex into the stably bound state in which the aromatic cage forms last by collapsing around the modified lysine. In agreement with this hypothesis, we found optimization of the N-terminus of our antagonists led to dramatic improvements in their binding affinity primarily through increases in their association rates (as determined by SPR) with CBX7. In further agreement with this hypothesis, we found that appropriate modification of the lysine residue can dramatically slow the dissociation rate of the inhibitor-chromodomain complex (**Fig. 5.2**).



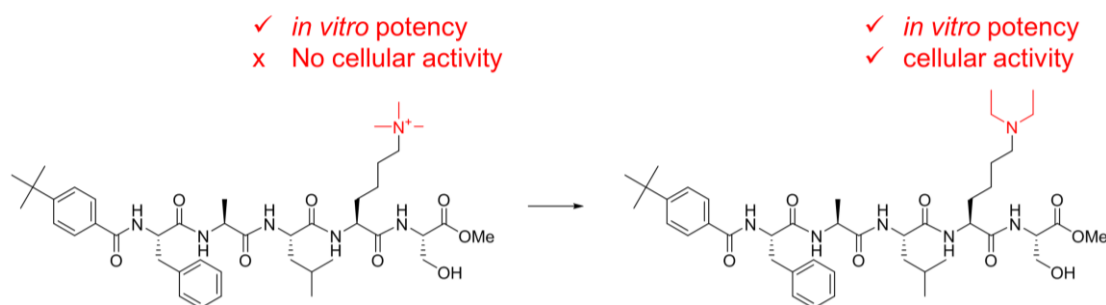
**Figure 5.2.** Relationship between various modifications to our antagonists and their  $K_d$ ,  $k_{on}$ , and  $k_{off}$  for CBX7.

We were able to leverage the high affinity of UNC4195 to each of the UNC3866 target chromodomains and utilize this molecule to develop high throughput assays for CBX4/7/8 and CDYL2. The development of these assays enabled us to begin to understand the selectivity determinants between these different chromodomains. Further, we developed an AlphaScreen® assay for the closely related chromodomain of CBX5 which allowed us to gain valuable insight into the selectivity of UNC3866 for Pc chromodomains over HP1 chromodomains. The results of these studies are summarized in **Fig. 5.3**.



**Figure 5.3.** Summary of selectivity parameters determined from UNC3866 SAR studies

Finally, we demonstrated that UNC3866 is capable antagonizing CBX7 in cells and inhibits PC3 cell proliferation with an  $EC_{50}$  of approximately 7.5  $\mu$ M. Proliferation is inhibited through induction of cellular senescence, a mechanism that is consistent with antagonism of CBX7. Further, cellular activity is dependent upon replacement of the endogenous Kme3 with a non-natural lysine derivative (**Fig. 5.4**). The utility of UNC3866 *in vivo* is likely to be limited, but it represents an important starting point for the development of *in vivo* chemical probes targeting PRC1 chromodomains.



**Figure 5.4.** Kme3 replacement is required for inhibition of proliferation by peptidic CBX7 antagonists.

## FUTURE DIRECTIONS

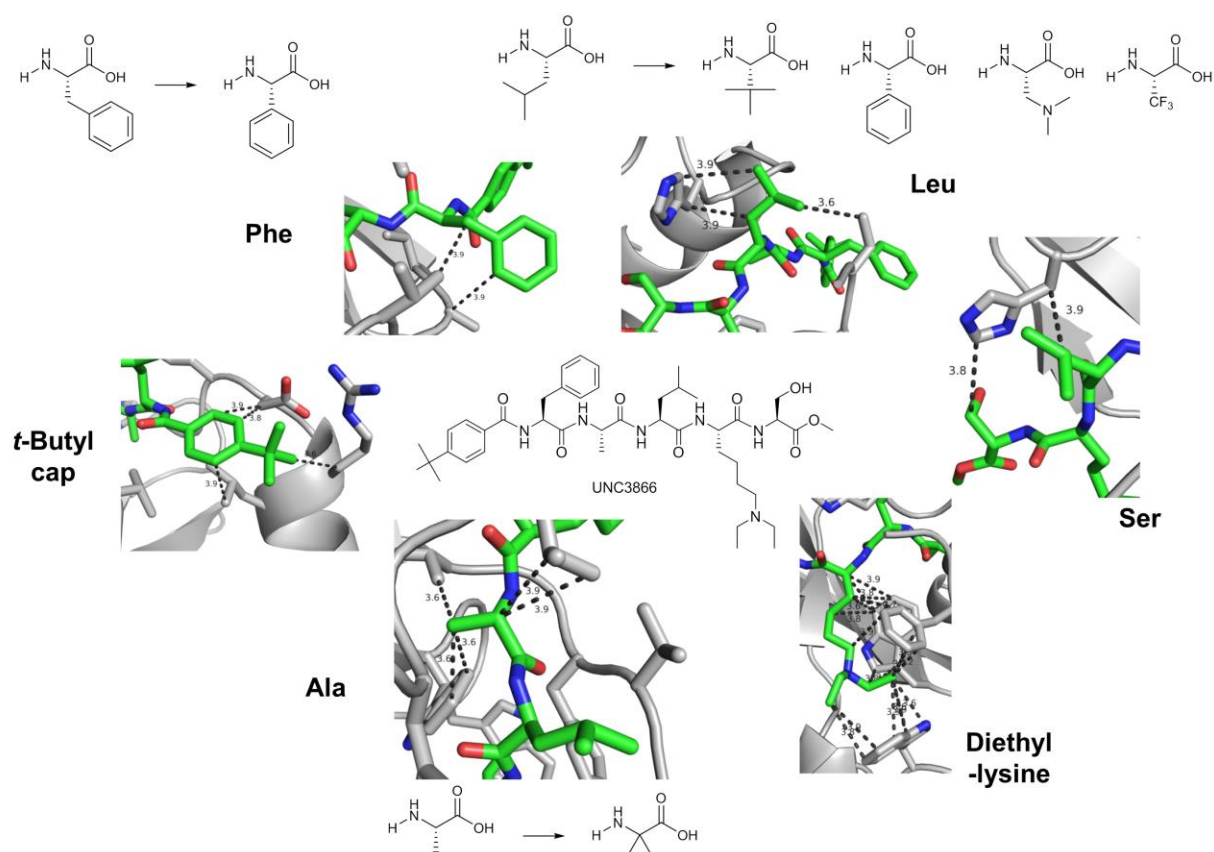
### *Potency and PK optimization of peptidic antagonists*

Efforts to discover antagonists of PRC1 chromodomains, and Kme readers in general, are still in their infancy. Our studies have provided evidence that high-affinity molecules targeting PRC1 chromodomains are not only possible, but may one day serve as therapeutically useful molecules. While methods, such as liposomal packaging<sup>125</sup>, that are known to improve the pharmacokinetic properties of

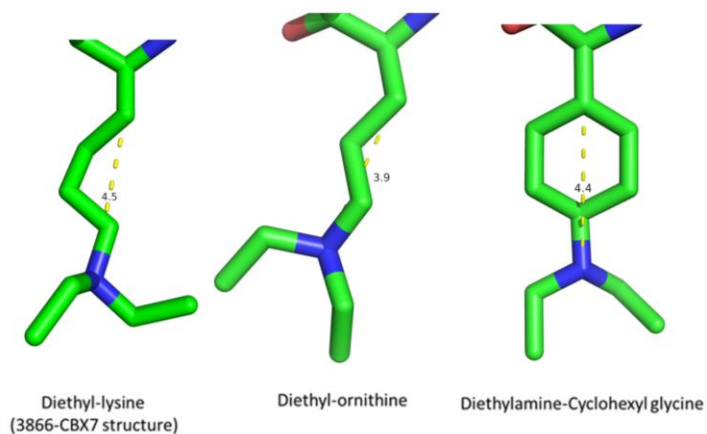


peptidic molecules should be pursued with respect to UNC3866 and especially UNC4976 and its derivatives, continued medicinal efforts should also be pursued with the aim of discovering molecules with higher potency and more drug like properties. For example, continued peptide-based SAR studies could be done with the goal of further improving potency and decreasing the number of natural amino acids present in the antagonist. **Figure 5.5** depicts some modifications that could be made to potentially improve potency. The poor permeability of our antagonists could be overcome by continued improvements in potency provided the permeability is unaffected.

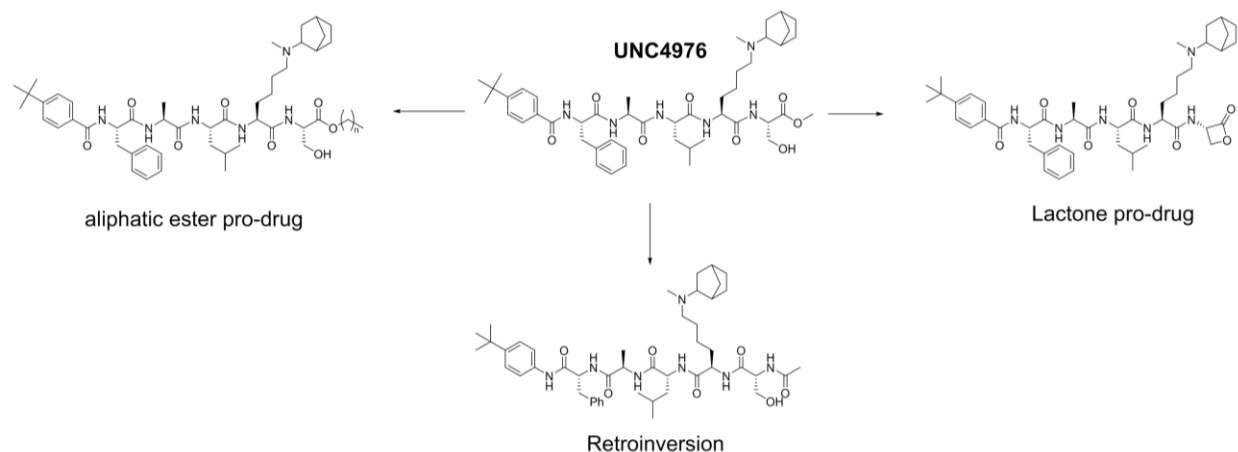
The gauche conformation of the diethyl-lysine side chain when bound to CBX7 is another area of potential optimization. While shortening the side chain to diethyl-ornithine appears to be too short based on the co-crystal structure of UNC3866 bound to CBX7, cyclization may prove effective as this would produce the appropriate spacing (**Fig. 5.6**). Additionally, currently ongoing crystallization studies with UNC4976 and CBX7 may provide valuable insight into the interaction between the methyl, norbornyl-lysine and CBX7, which is currently the preferred Kme3 replacement for future studies. Finally, pro-drug approaches, such as conversion of the C-terminal serine methyl ester to a lactone derivative, or increases in the hydrophobicity of the ester could be employed to improve the permeability of these compounds. Retroinversion peptides<sup>126</sup> also present a viable alternative for improvement of the pharmacokinetic properties of our compounds (**Fig. 5.7**) as well as exploration of various amide isomers<sup>127</sup>.



**Figure 5.5.** van der Waals' interactions between UNC3866 and CBX7. Interactions are depicted as black dashed lines. Possible amino acid replacements to improve these interactions are shown.



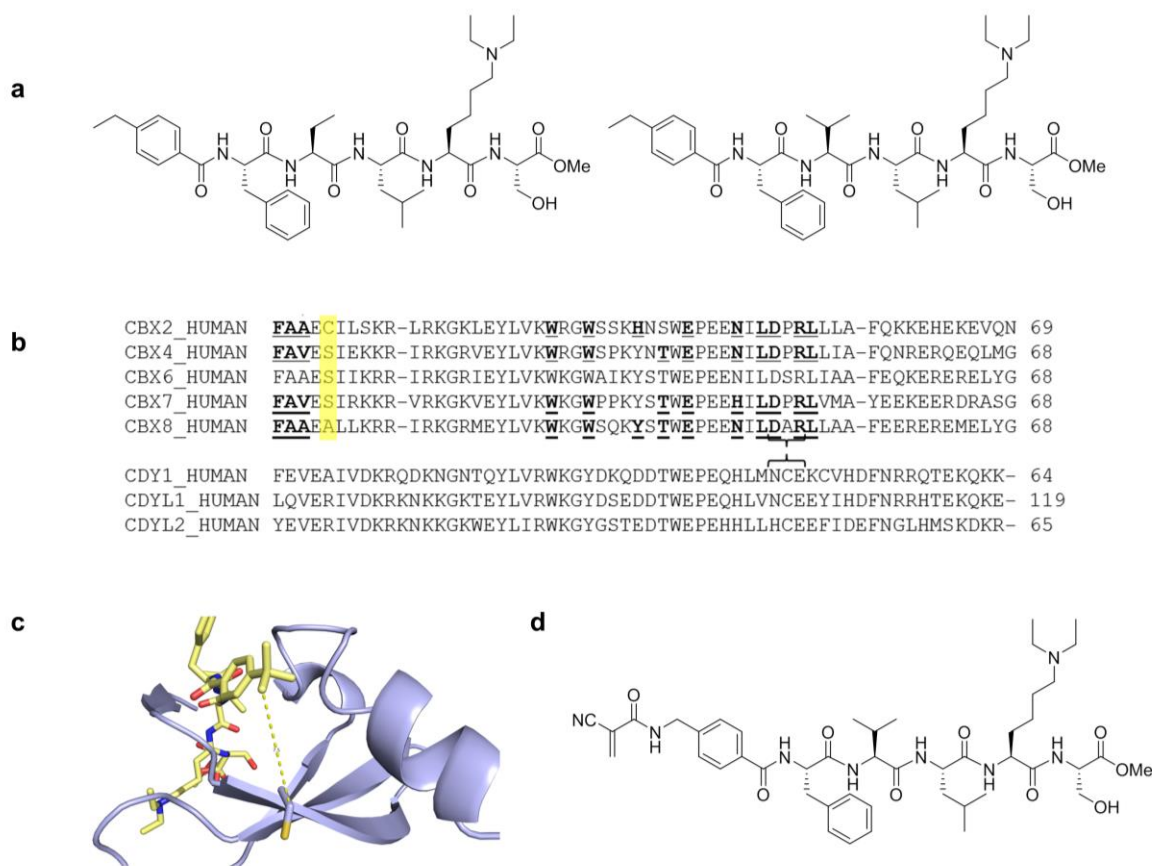
**Figure 5.6.** Strategies for side chain optimization of UNC3866 based on the UNC3866-CBX7 co-crystal structure. Distances from the  $\beta$ -carbon to the nitrogen are shown in Å.



**Figure 5.7.** Strategies for improving the *in vivo* efficacy of UNC4976

#### *Pc CBX selectivity*

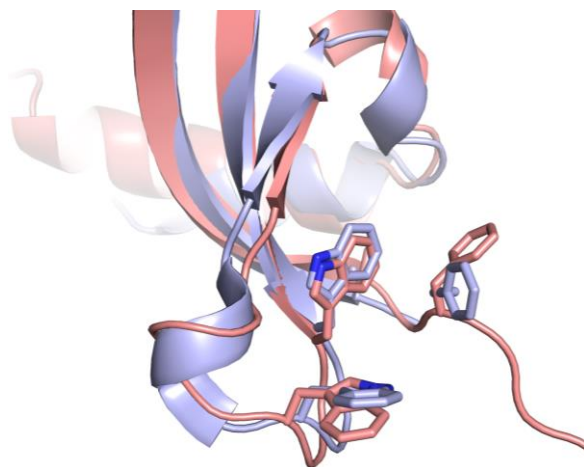
Our structural and SAR studies in Chapter III indicated that larger alkyl substituents tune the selectivity of our antagonists for CBX8 (and likely CBX2 and -6) while reducing the affinity for CBX4/7. Additionally, the degree of alkylation on the aromatic cap of UNC3866 affects binding to a much lesser extent for CBX8 than for CBX4/7. Thus, molecules such as those shown in **Fig. 5.8a** may represent a simple and straightforward approach for CBX antagonists with a preference for CBX2/6/8 over CBX4/7. Interestingly, CBX2 is unique in that it contains a cysteine residue in close proximity to the *t*-butyl group of UNC3866 (~7.5 Å, **Fig. 5.8b and c**). Thus, CBX2 selective compounds may be attainable through the development of covalent reversible antagonists<sup>128</sup> that target this unique residue.



**Figure 5.8.** (a) Structures of potential CBX2/6/8 selective antagonists. (b) Sequence alignment of chromodomain targets of UNC3866. (c) Representation of the CBX2 cysteine residue in close proximity to the *t*-butyl group of UNC3866. (d) Hypothetical covalent reversible CBX2 antagonist.

#### *Toward small molecule CBX antagonists*

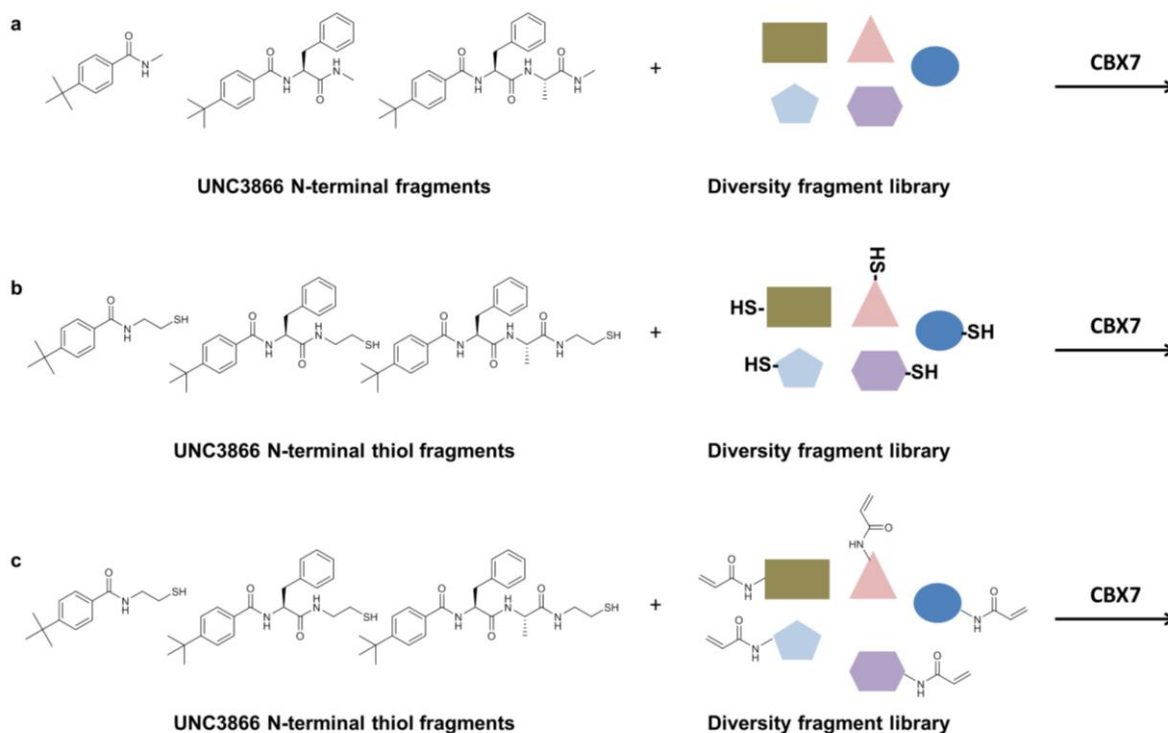
While continued exploration of peptides for use *in vivo* chemical probes for CBX chromodomains presents a promising option, traditional small molecules scaffolds are still the most attractive chemotypes for *in vivo* chemical probes. As mentioned in Chapter III, the chromodomain of CBX2 (likely CBX6 and -8 as well), does not require an induced fit binding pathway like that of CBX7 (**Fig. 5.9**). Hence, a small molecule screening campaign for these chromodomains would be expected to be much more fruitful than a campaign targeting CBX4/7. Additionally, the development of UNC4195 and its application as an AlphaScreen® bait ligand makes a high throughput screen for these proteins a cost-feasible option.



**Figure 5.9.** Overlay of the CBX2 chromodomain in the apo (salmon) and UNC3866-bound conformation (light blue, UNC3866 not shown).

Our studies have highlighted the challenges of developing CBX7 small molecule antagonists. While these studies indicate that the probability of identifying a reasonable small molecule scaffold for SAR studies is low, our cellular studies indicate that such a molecule could be a powerful chemical tool with important therapeutic potential. In order to facilitate the identification of a small molecule scaffold targeting the CBX7 chromodomain, a fragment-based approach could be applied utilizing the N-terminus of UNC3866 in conjunction with a library of small molecule fragments (**Fig. 5.10**). Our binding hypothesis, for which we have provided extensive experimental evidence, indicates that the N-terminus of UNC3866 is responsible for initiating the induced fit binding pathway. Therefore, screening libraries of various fragments in the presence of this fragment of UNC3866 could lead to the identification of fragments that display synergistic antagonism of CBX7 with the N-terminus of UNC3866. The N-terminal fragment of UNC3866 should help to initiate the induced fit binding mechanism and create a binding pocket for the fragments to interact with (**Fig. 5.10a**). Screening could be done through traditional NMR studies, AlphaScreen® or SPR. Libraries could also be designed with reactive functionalities in order to enable a dynamic combinatorial chemistry (DCC) approach<sup>129</sup> (**Fig. 5.10b**). Alternatively, similar to a DCC approach, a library of acrylamide-containing compounds could also be incubated with a thiol-derivative of the N-terminus of UNC3866 in which the solution was analyzed for covalent products formed in the presence of CBX7<sup>130</sup> (**Fig. 5.10c**). While the risk of failure is significant for all three of these approaches,

they present perhaps the most rational and likely approach to the discovery of traditional small molecule antagonists of CBX7. As is always the case, there are numerous other approaches which may lead to an *in vivo* chemical probe of PRC1 chromodomains and only time will tell which, if any, will be effective.



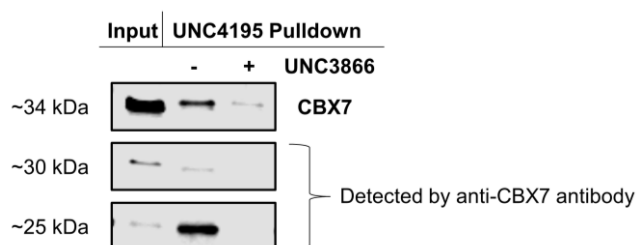
**Figure 5.10** Fragment-based approaches for the development a CBX7 small molecule antagonist.

### Biological studies

The development of UNC3866 provides a novel mechanism by which to interrogate the canonical Polycomb signaling pathway. If the canonical Polycomb signaling pathway holds true, the UNC3866 and EZH2 inhibitors would be expected to act synergistically, especially in inhibiting proliferation of cancerous cell lines that have been shown to depend on either PRC1 or PRC2 for their growth and proliferation. Our own preliminary studies have already begun to argue against this model. In the context of PC3, our own preliminary studies found no antiproliferative effect upon EZH2 inhibition, in agreement with a previous study<sup>24</sup>. The dependence of PC3 cells on CBX7 for their growth, their susceptibility to UNC3866 and resistance to EZH2 inhibition strongly suggests that CBX7 functions independently of PRC2 and

H3K27me3. Further, our own preliminary studies indicate that CBX7 mediated repression is not inhibited by EZH2 antagonism and no synergy is observed between our CBX antagonists and inhibitors of EZH2. However, as discussed in the introduction, the relevance of Polycomb signaling appears to be cell-type dependent. Thus, our molecules may help to identify in which cell types and on what signaling pathways, if any, canonical Polycomb signaling plays a critical role. Intriguingly, the role of EED in PRC1 suggests that antagonists of PRC2 that target EED could be synergistic with our PRC1 antagonists while EZH2 antagonists preliminarily do not appear to be.

We noted in our pull-down studies that UNC3866 in Chapter IV is capable of engaging what appears to be three different isoforms of CBX7 present in PC3 cells (**Fig. 5.11**). Consistent with this, two variant, protein-coding transcripts have been annotated in the Ensembl database. Future studies confirming the identity of these bands may ultimately help shed light on which isoform(s) incorporate into PRC1 and which one(s) are responsible for enhancing PC3 cell proliferation. It is conceivable that the effect of UNC3866 is a result of modulation of a CBX7 isoform not incorporated into PRC1, which could explain our lack of evidence for the canonical polycomb signaling model.



**Figure 5.11.** Putative CBX7 isoforms engaged by UNC3866.

## CONCLUSIONS

The studies described in this dissertation have answered many questions concerning strategies to target CBX chromodomains. Additionally, these studies have begun to answer important biological questions concerning CBX antagonism. However, seemingly endless questions remain, some of which have been highlighted in this chapter. Undoubtedly, many of these “known unknowns” will be answered in time. However, as is nearly always the case in scientific endeavors, continued research into CBX antagonism is likely to reveal numerous “unknown unknowns” and produce more questions than it answers.

## REFERENCES

1. Bianconi, E. *et al.* An estimation of the number of cells in the human body. *Ann Hum Biol* **40**, 463-471 (2013).
2. Turner, B.M. Epigenetic responses to environmental change and their evolutionary implications. *Philos Trans R Soc Lond B Biol Sci* **364**, 3403-3418 (2009).
3. Dolinoy, D.C., Weidman, J.R. & Jirtle, R.L. Epigenetic gene regulation: linking early developmental environment to adult disease. *Reprod Toxicol* **23**, 297-307 (2007).
4. Luger, K. in eLS. (ed. J.W.S. Ltd) (Nature Publishing Group, <http://www.els.net> [doi: 10.1038/npg.els.0001155]; 2001).
5. Strahl, B.D. & Allis, C.D. The language of covalent histone modifications. *Nature* **403**, 41-45 (2000).
6. Jenuwein, T. & Allis, C.D. Translating the histone code. *Science* **293**, 1074-1080 (2001).
7. Arrowsmith, C.H., Bountra, C., Fish, P.V., Lee, K. & Schapira, M. Epigenetic protein families: a new frontier for drug discovery. *Nat. Rev. Drug Discovery* **11**, 384-400 (2012).
8. Filippakopoulos, P. *et al.* Selective inhibition of BET bromodomains. *Nature* **468**, 1067-1073 (2010).
9. Tan, M. *et al.* Identification of 67 histone marks and histone lysine crotonylation as a new type of histone modification. *Cell* **146**, 1016-1028 (2011).
10. Dai, L. *et al.* Lysine 2-hydroxyisobutyrylation is a widely distributed active histone mark. *Nature chemical biology* **10**, 365-370 (2014).
11. Nathan, D. *et al.* Histone sumoylation is a negative regulator in *Saccharomyces cerevisiae* and shows dynamic interplay with positive-acting histone modifications. *Genes & development* **20**, 966-976 (2006).
12. Eskeland, R. *et al.* Ring1B compacts chromatin structure and represses gene expression independent of histone ubiquitination. *Mol Cell* **38**, 452-464 (2010).
13. Bentley, M.L. *et al.* Recognition of UbcH5c and the nucleosome by the Bmi1/Ring1b ubiquitin ligase complex. *The EMBO journal* **30**, 3285-3297 (2011).
14. Tavares, L. *et al.* RYBP-PRC1 complexes mediate H2A ubiquitylation at polycomb target sites independently of PRC2 and H3K27me3. *Cell* **148**, 664-678 (2012).
15. Gil, J., Bernard, D., Martinez, D. & Beach, D. Polycomb CBX7 has a unifying role in cellular lifespan. *Nature cell biology* **6**, 67-72 (2004).
16. Bernard, D. *et al.* CBX7 controls the growth of normal and tumor-derived prostate cells by repressing the Ink4a/Arf locus. *Oncogene* **24**, 5543-5551 (2005).
17. Zhang, X.W. *et al.* Oncogenic role of the chromobox protein CBX7 in gastric cancer. *J Exp Clin Cancer Res* **29**, 114 (2010).
18. Yap, K.L. *et al.* Molecular interplay of the noncoding RNA ANRIL and methylated histone H3 lysine 27 by polycomb CBX7 in transcriptional silencing of INK4a. *Mol Cell* **38**, 662-674 (2010).



19. Klauke, K. *et al.* Polycomb Cbx family members mediate the balance between haematopoietic stem cell self-renewal and differentiation. *Nature cell biology* **15**, 353-362 (2013).
20. Li, G. *et al.* Altered expression of polycomb group genes in glioblastoma multiforme. *PloS one* **8**, e80970 (2013).
21. McCabe, M.T. *et al.* EZH2 inhibition as a therapeutic strategy for lymphoma with EZH2-activating mutations. *Nature* **492**, 108-112 (2012).
22. Kim, W. *et al.* Targeted disruption of the EZH2-EED complex inhibits EZH2-dependent cancer. *Nature chemical biology* **9**, 643-650 (2013).
23. Konze, K.D. *et al.* An orally bioavailable chemical probe of the Lysine Methyltransferases EZH2 and EZH1. *ACS chemical biology* **8**, 1324-1334 (2013).
24. Garapaty-Rao, S. *et al.* Identification of EZH2 and EZH1 small molecule inhibitors with selective impact on diffuse large B cell lymphoma cell growth. *Chemistry & biology* **20**, 1329-1339 (2013).
25. Xu, B. *et al.* Selective inhibition of EZH2 and EZH1 enzymatic activity by a small molecule suppresses MLL-rearranged leukemia. *Blood* (2014).
26. Wang, G.G. *et al.* Haematopoietic malignancies caused by dysregulation of a chromatin-binding PHD finger. *Nature* **459**, 847-851 (2009).
27. Cheng, A.C. *et al.* Structure-based maximal affinity model predicts small-molecule druggability. *Nat Biotechnol* **25**, 71-75 (2007).
28. Santiago, C., Nguyen, K. & Schapira, M. Druggability of methyl-lysine binding sites. *Journal of computer-aided molecular design* **25**, 1171-1178 (2011).
29. Filippakopoulos, P. & Knapp, S. Targeting bromodomains: epigenetic readers of lysine acetylation. *Nat Rev Drug Discov* **13**, 337-356 (2014).
30. Milosevich, N. & Hof, F. Chemical Inhibitors of Epigenetic Methyllysine Reader Proteins. *Biochemistry* **55**, 1570-1583 (2016).
31. Stuckey, J.I. *et al.* A cellular chemical probe targeting the chromodomains of Polycomb repressive complex 1. *Nat. Chem. Biol.* **12**, 180-187 (2016).
32. Knutson, S.K. *et al.* A selective inhibitor of EZH2 blocks H3K27 methylation and kills mutant lymphoma cells. *Nature chemical biology* **8**, 890-896 (2012).
33. Falkenberg, K.J. & Johnstone, R.W. Histone deacetylases and their inhibitors in cancer, neurological diseases and immune disorders. *Nat Rev Drug Discov* **13**, 673-691 (2014).
34. Hashizume, R. *et al.* Pharmacologic inhibition of histone demethylation as a therapy for pediatric brainstem glioma. *Nat Med* **20**, 1394-1396 (2014).
35. Zwergel, C., Stazi, G., Valente, S. & Mai, A. Histone Deacetylase Inhibitors: Updated Studies in Various Epigenetic-Related Diseases. *Journal of Clinical Epigenetics* **2**, 1-15 (2016).
36. [www.clinicaltrials.gov](http://www.clinicaltrials.gov) (July 5, 2016).
37. Kruidenier, L. *et al.* A selective jumonji H3K27 demethylase inhibitor modulates the proinflammatory macrophage response. *Nature* **488**, 404-408 (2012).

38. Simhadri, C. *et al.* Chromodomain antagonists that target the polycomb-group methyllysine reader protein chromobox homolog 7 (CBX7). *J. Med. Chem.* **57**, 2874-2883 (2014).
39. Milosevich, N. *et al.* Selective inhibition of CBX6: a methyllysine reader protein in the polycomb family. *ACS Med. Chem. Lett.* **7**, 139-144 (2016).
40. Ren, C. *et al.* Small-molecule modulators of methyl-lysine binding for the CBX7 chromodomain. *Chem. Biol.* **22**, 161-168 (2015).
41. Dietrich, N. *et al.* Bypass of senescence by the polycomb group protein CBX8 through direct binding to the INK4A-ARF locus. *The EMBO journal* **26**, 1637-1648 (2007).
42. Mohammad, H.P. *et al.* Polycomb CBX7 promotes initiation of heritable repression of genes frequently silenced with cancer-specific DNA hypermethylation. *Cancer Res* **69**, 6322-6330 (2009).
43. Li, Q. *et al.* Polycomb CBX7 directly controls trimethylation of histone H3 at lysine 9 at the p16 locus. *PloS one* **5**, e13732 (2010).
44. Tan, J. *et al.* CBX8, a polycomb group protein, is essential for MLL-AF9-induced leukemogenesis. *Cancer Cell* **20**, 563-575 (2011).
45. Luis, N.M. *et al.* Regulation of human epidermal stem cell proliferation and senescence requires polycomb- dependent and -independent functions of Cbx4. *Cell stem cell* **9**, 233-246 (2011).
46. O'Loghlen, A. *et al.* MicroRNA regulation of Cbx7 mediates a switch of Polycomb orthologs during ESC differentiation. *Cell stem cell* **10**, 33-46 (2012).
47. Wang, B. *et al.* Chromobox homolog 4 is correlated with prognosis and tumor cell growth in hepatocellular carcinoma. *Annals of surgical oncology* **20 Suppl 3**, S684-692 (2013).
48. Shinjo, K. *et al.* 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* **135**, 308-318 (2014).
49. Tang, J. *et al.* Paradoxical role of CBX8 in proliferation and metastasis of colorectal cancer. *Oncotarget* **5** (2014).
50. Bouwman, P. *et al.* 53BP1 loss rescues BRCA1 deficiency and is associated with triple-negative and BRCA-mutated breast cancers. *Nat Struct Mol Biol* **17**, 688-695 (2010).
51. Bartkova, J. *et al.* DNA damage response mediators MDC1 and 53BP1: constitutive activation and aberrant loss in breast and lung cancer, but not in testicular germ cell tumours. *Oncogene* **26**, 7414-7422 (2007).
52. Naruke, Y. *et al.* Alteration of p53-binding protein 1 expression during skin carcinogenesis: association with genomic instability. *Cancer Sci* **99**, 946-951 (2008).
53. Iwabuchi, K. *et al.* Characterization of a cancer cell line that expresses a splicing variant form of 53BP1: separation of checkpoint and repair functions in 53BP1. *Biochem Biophys Res Commun* **376**, 509-513 (2008).
54. Tang, M. *et al.* The malignant brain tumor (MBT) domain protein SFMBT1 is an integral histone reader subunit of the LSD1 demethylase complex for chromatin association and epithelial-to-mesenchymal transition. *J Biol Chem* **288**, 27680-27691 (2013).

55. Liu, Y.L. *et al.* Expression and clinicopathological significance of EED, SUZ12 and EZH2 mRNA in colorectal cancer. *J Cancer Res Clin Oncol* **141**, 661-669 (2015).
56. Yu, H. *et al.* PRC2/EED-EZH2 complex is up-regulated in breast cancer lymph node metastasis compared to primary tumor and correlates with tumor proliferation in situ. *PLoS one* **7**, e51239 (2012).
57. Cao, Q. *et al.* The central role of EED in the orchestration of polycomb group complexes. *Nature communications* **5**, 3127 (2014).
58. Frye, S.V. The art of the chemical probe. *Nature chemical biology* **6**, 159-161 (2010).
59. Workman, P. & Collins, I. Probing the probes: fitness factors for small molecule tools. *Chemistry & biology* **17**, 561-577 (2010).
60. Bunnage, M.E., Chekler, E.L. & Jones, L.H. Target validation using chemical probes. *Nature chemical biology* **9**, 195-199 (2013).
61. Arrowsmith, C.H. *et al.* The promise and peril of chemical probes. *Nature chemical biology* **11**, 536-541 (2015).
62. Scott, C.L. *et al.* Role of the chromobox protein CBX7 in lymphomagenesis. *Proceedings of the National Academy of Sciences of the United States of America* **104**, 5389-5394 (2007).
63. Herold, J.M., Ingberman, L.A., Gao, C. & Frye, S.V. Drug Discovery Toward Antagonists of Methyl-Lysine Binding Proteins. *Current Chemical Genomics* **5**, 51-61 (2011).
64. Wagner, T., Robaa, D., Sippl, W. & Jung, M. Mind the methyl: methyllysine binding proteins in epigenetic regulation. *ChemMedChem* **9**, 466-483 (2014).
65. Iwase, S. *et al.* ATRX ADD domain links an atypical histone methylation recognition mechanism to human mental-retardation syndrome. *Nat Struct Mol Biol* **18**, 769-776 (2011).
66. Herold, J.M. *et al.* Structure–activity relationships of methyl-lysine reader antagonists. *Med. Chem. Commun.* **3**, 45-51 (2012).
67. 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.* **283**, 19626-19635 (2008).
68. Kaustov, L. *et al.* Recognition and specificity determinants of the human cbx chromodomains. *J. Biol. Chem.* **286**, 521-529 (2011).
69. Wigle, T.J. *et al.* Screening for inhibitors of low-affinity epigenetic peptide-protein interactions: an AlphaScreen-based assay for antagonists of methyl-lysine binding proteins. *J. Biomol. Screen* **15**, 62-71 (2010).
70. Vega, S., Abian, O. & Velazquez-Campoy, A. A unified framework based on the binding polynomial for characterizing biological systems by isothermal titration calorimetry. *Methods* **76**, 99-115 (2015).
71. Duff, M.R., Jr., Grubbs, J. & Howell, E.E. Isothermal titration calorimetry for measuring macromolecule-ligand affinity. *J Vis Exp* (2011).
72. Myszka, D.G. & Rich, R.L. Implementing surface plasmon resonance biosensors in drug discovery. *Pharmaceutical Science and Technology Today* **3**, 310-317 (2000).

73. Gao, C. *et al.* Biophysical probes reveal a "compromise" nature of the methyl-lysine binding pocket in L3MBTL1. *J Am Chem Soc* **133**, 5357-5362 (2011).
74. Herold, J.M. *et al.* Small-molecule ligands of methyl-lysine binding proteins. *J Med Chem* **54**, 2504-2511 (2011).
75. James, L.I. *et al.* Discovery of a chemical probe for the L3MBTL3 methyllysine reader domain. *Nature chemical biology* **9**, 184-191 (2013).
76. Perfetti, M.T. *et al.* Identification of a fragment-like small molecule ligand for the methyl-lysine binding protein, 53BP1. *ACS Chem. Biol.* **10**, 1072-1081 (2015).
77. Camerino, M.A. *et al.* The structure-activity relationships of L3MBTL3 inhibitors: flexibility of the dimer interface. *Medchemcomm* **4**, 1501-1507 (2013).
78. James, L.I. *et al.* Small-molecule ligands of methyl-lysine binding proteins: optimization of selectivity for L3MBTL3. *J. Med. Chem.* **56**, 7358-7371 (2013).
79. Eissenberg, J.C. Structural biology of the chromodomain: form and function. *Gene* **496**, 69-78 (2012).
80. Pallante, P. *et al.* Loss of the CBX7 gene expression correlates with a highly malignant phenotype in thyroid cancer. *Cancer Res* **68**, 6770-6778 (2008).
81. Ren, X., Vincenz, C. & Kerppola, T.K. Changes in the distributions and dynamics of polycomb repressive complexes during embryonic stem cell differentiation. *Mol. Cell Biol.* **28**, 2884-2895 (2008).
82. Federico, A. *et al.* Chromobox protein homologue 7 protein, with decreased expression in human carcinomas, positively regulates E-cadherin expression by interacting with the histone deacetylase 2 protein. *Cancer Res* **69**, 7079-7087 (2009).
83. Karamitopoulou, E. *et al.* Loss of the CBX7 protein expression correlates with a more aggressive phenotype in pancreatic cancer. *Eur J Cancer* **46**, 1438-1444 (2010).
84. Mansueto, G. *et al.* Identification of a New Pathway for Tumor Progression: MicroRNA-181b Up-Regulation and CBX7 Down-Regulation by HMGA1 Protein. *Genes Cancer* **1**, 210-224 (2010).
85. Pallante, P. *et al.* The loss of the CBX7 gene expression represents an adverse prognostic marker for survival of colon carcinoma patients. *Eur J Cancer* **46**, 2304-2313 (2010).
86. 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. *Molecular & cellular proteomics : MCP* **10**, M110 002642-002641-002623 (2011).
87. Forzati, F. *et al.* CBX7 is a tumor suppressor in mice and humans. *J Clin Invest* **122**, 612-623 (2012).
88. Morey, L. *et al.* Nonoverlapping functions of the Polycomb group Cbx family of proteins in embryonic stem cells. *Cell stem cell* **10**, 47-62 (2012).
89. Morey, L., Aloia, L., Cozzuto, L., Benitah, S.A. & Di Croce, L. RYBP and Cbx7 define specific biological functions of polycomb complexes in mouse embryonic stem cells. *Cell reports* **3**, 60-69 (2013).

90. Clermont, P.L. *et al.* Genotranscriptomic meta-analysis of the Polycomb gene CBX2 in human cancers: initial evidence of an oncogenic role. *Br J Cancer* **111**, 1663-1672 (2014).
91. Ma, R., Zhang, Y., Sun, T. & Cheng, B. Epigenetic Regulation by polycomb group complexes: focus on roles of CBX proteins. *Journal of Zhejiang University-SCIENCE B (Biomedicine & Biotechnology)* **15**, 412-428 (2014).
92. Gil, J. & O'Loughlen, A. PRC1 complex diversity: where is it taking us? *Trends Cell Biol* **24**, 632-641 (2014).
93. Clermont, P.L. *et al.* Polycomb-mediated silencing in neuroendocrine prostate cancer. *Clin Epigenetics* **7**, 40 (2015).
94. Bernstein, E. *et al.* Mouse polycomb proteins bind differentially to methylated histone H3 and RNA and are enriched in facultative heterochromatin. *Mol Cell Biol* **26**, 2560-2569 (2006).
95. Wang, H. *et al.* Role of histone H2A ubiquitination in Polycomb silencing. *Nature* **431**, 873-878 (2004).
96. Vincenz, C. & Kerppola, T.K. Different polycomb group CBX family proteins associate with distinct regions of chromatin using nonhomologous protein sequences. *Proceedings of the National Academy of Sciences of the United States of America* **105**, 16572-16577 (2008).
97. Ren, C. *et al.* Structure-guided discovery of selective antagonists for the chromodomain of polycomb repressive protein CBX7. *ACS Med. Chem. Lett.* **7**, 601-605 (2016).
98. Lipinski, C.A., Lombardo, F., Dominy, B.W. & Feeney, P.J. Experimental and computational approaches to estimate solubility and permeability in drug discovery and development settings. *Advanced Drug Delivery Reviews* **46**, 3-26 (2001).
99. Van Dyke, R.W., Faber, E.D. & Meijer, D.K. Sequestration of organic cations by acidified hepatic endocytic vesicles and implications for biliary excretion. *J. Pharmacol. Exp. Ther.* **261**, 1-11 (1992).
100. Goldman, S.D., Funk, R.S., Rajewski, R.A. & Krise, J.P. Mechanisms of amine accumulation in, and egress from, lysosomes. *Bioanalysis* **1**, 1445-1459 (2009).
101. Hopkinson, R.J. *et al.* Is JmjC oxygenase catalysis limited to demethylation? *Angewandte Chemie* **52**, 7709-7713 (2013).
102. Dickson, B.M. Approaching a parameter-free metadynamics. *Physical Review E* **84** (2011).
103. Hess, B., Kutzner, C., van der Spoel, D. & Lindahl, E. GROMACS 4: Algorithms for Highly Efficient, Load-Balanced, and Scalable Molecular Simulation. *J. Chem Theory Comput.* **4**, 435-447 (2008).
104. Tiwary, P. & Parrinello, M. From Metadynamics to Dynamics. *Physical Review Letters* **111**, 230602 (2013).
105. Voter, A.F. Hyperdynamics: Accelerated Molecular Dynamics of Infrequent Events. *Physical Review Letters* **78**, 3908-3911 (1997).
106. Bussi, G., Donadio, D. & Parrinello, M. Canonical sampling through velocity rescaling. *J Chem Phys* **126**, 014101 (2007).

107. Copeland, R.A. The drug-target residence time model: a 10-year retrospective. *Nature reviews. Drug discovery* **15**, 87-95 (2016).
108. Besnard, J. *et al.* Automated design of ligands to polypharmacological profiles. *Nature* **492**, 215-220 (2012).
109. Kroeze, W.K. *et al.* PRESTO-Tango as an open-source resource for interrogation of the druggable human GPCRome. *Nat Struct Mol Biol* (2015).
110. Weikl, T.R. & von Deuster, C. Selected-fit versus induced-fit protein binding: kinetic differences and mutational analysis. *Proteins* **75**, 104-110 (2009).
111. Barsyte-Lovejoy, D. *et al.* (R)-PFI-2 is a potent and selective inhibitor of SETD7 methyltransferase activity in cells. *Proceedings of the National Academy of Sciences of the United States of America* (2014).
112. Kabsch, W. Xds. *Acta Crystallogr D Biol Crystallogr* **66**, 125-132 (2010).
113. Evans, P.R. & Murshudov, G.N. How good are my data and what is the resolution? *Acta Crystallogr D Biol Crystallogr* **69**, 1204-1214 (2013).
114. McCoy, A.J. *et al.* Phaser crystallographic software. *J Appl Crystallogr* **40**, 658-674 (2007).
115. Martinez Molina, D. *et al.* Monitoring drug target engagement in cells and tissues using the cellular thermal shift assay. *Science* **341**, 84-87 (2013).
116. Montgomery, N.D., Yee, D., Montgomery, S.A. & Magnuson, T. Molecular and functional mapping of EED motifs required for PRC2-dependent histone methylation. *Journal of molecular biology* **374**, 1145-1157 (2007).
117. Kuzmichev, A., Jenuwein, T., Tempst, P. & Reinberg, D. Different EZH2-Containing Complexes Target Methylation of Histone H1 or Nucleosomal Histone H3. *Molecular Cell* **14**, 183-193 (2004).
118. Burton, P.S., Conradi, R.A., Ho, N.F.H., Hilgers, A.R. & Borchardt, R.T. How structural features influence the biomembrane permeability of peptides. *Journal of Pharmaceutical Sciences* **85**, 1336-1340 (1996).
119. Wagner, B.A., Venkataraman, S. & Buettner, G.R. The rate of oxygen utilization by cells. *Free Radic Biol Med* **51**, 700-712 (2011).
120. Kong, Y., Cui, H., Ramkumar, C. & Zhang, H. Regulation of senescence in cancer and aging. *J Aging Res* **2011**, 963172 (2011).
121. Jarrard, D.F. *et al.* Deletional, Mutational, and Methylation Analyses of CDKN2 (p16/MTS1) in Primary and Metastatic Prostate Cancer. *GENES, CHROMOSOMES & CANCER* **19**, 90-96 (1997).
122. Yaqinuddin, A., Qureshi, S.A., Qazi, R. & Abbas, F. Down-regulation of DNMT3b in PC3 cells effects locus-specific DNA methylation, and represses cellular growth and migration. *Cancer Cell Int* **8**, 13 (2008).
123. Yaqinuddin, A., Qureshi, S.A., Qazi, R., Farooq, S. & Abbas, F. DNMT1 silencing affects locus specific DNA methylation and increases prostate cancer derived PC3 cell invasiveness. *J Urol* **182**, 756-761 (2009).
124. Mirochnik, Y. *et al.* Androgen receptor drives cellular senescence. *PloS one* **7**, e31052 (2012).

125. Bruno, B.J., Miller, G.D. & Lim, C.S. Basics and recent advances in peptide and protein drug delivery. *Ther Deliv* **4**, 1443-1467 (2013).
126. Taylor, E.M., Otero, D.A., Banks, W.A. & O'Brien, J.S. Retro-inverso prosaptide peptides retain bioactivity, are stable *in vivo*, and are blood-brain barrier permeable. *J Pharmacol Exp Ther* **295**, 190-194 (2000).
127. Patani, G.A. & LaVoie, E.J. Bioisosterism: a rational approach in drug design. *Chem. Rev.* **96**, 3147-3176 (1996).
128. Bradshaw, J.M. *et al.* Prolonged and tunable residence time using reversible covalent kinase inhibitors. *Nature chemical biology* **11**, 525-531 (2015).
129. Ramstrom, O. & Lehn, J.M. Drug discovery by dynamic combinatorial libraries. *Nat Rev Drug Discov* **1**, 26-36 (2002).
130. Kwarczynski, F.E., Steffey, M.E., Fox, C.C. & Soellner, M.B. Discovery of Bivalent Kinase Inhibitors via Enzyme-Templated Fragment Elaboration. *ACS Med Chem Lett* **6**, 898-901 (2015).