PROBING THE MECHANISM OF BINDING AND RECOGNITION OF METHYLATED LYSINE

Amber Lynn Koenig

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 Department of Chemistry (Organic) in the College of Arts and Sciences

Chapel Hill
2016

Approved by:

Marcey Waters

Jeffrey Johnson

Simon Meek

Eric Brustad

Michel Gagné

ABSTRACT

Amber Lynn Koenig: Probing the Mechanism of Binding and Recognition of Methylated lysine
(Under the direction of Marcey Waters)

Lysine methylation is an important posttranslational modification that is responsible for the proper regulation of gene expression. The misregulation of these methylation marks has been linked to various diseases. Proteins that are involved in the regulation and recognition of these marks are emerging therapeutic targets. Detailed understanding of the mechanism employed by these proteins to recognize their natural substrates would provide valuable information for the development of probes with the necessary affinity and specificity required to provide activity and avoid off target effects.

Cation-π interactions are thought to be one of the major noncovalent interactions contributing to methylated lysine recognition. Here we have demonstrate that two tyrosine residues present in the binding pocket of the reader protein heterochromatin protein 1 (HP1) show differential contributions to trimethyllysine recognition. By incorporating unnatural amino acids containing substitutions on the aromatic rings, we tune the ability of these residues to participate in cation-π interactions, which influences overall binding affinity. We demonstrate a clear linear free energy relationship (LFER) at both tyrosine positions of different magnitudes.

In order to probe cation-π interactions with tryptophan mutations, we report synthetic methods for incorporation of unnatural amino acids that are not amenable to *in vivo* unnatural

amino acid mutagenesis. We demonstrate the synthesis of fmoc-protected fluorinated tryptophan for use in solid phase peptide synthesis, as well as an improved method for synthetically accessing long peptide or short protein sequences. By acetyl capping after coupling steps of solid phase peptide synthesis we have eliminated the possibility of deletion products arising from incomplete coupling reactions. By adding a polyhistidine tag at the N-terminus, we greatly simplified purification by incorporating an affinity tag that allows the isolation of only the fully synthesized protein. Furthermore, the function of the synthetic protein was confirmed by performing a binding assay with its native substrate.

Lastly, we discuss ongoing efforts to expand studies to other reader proteins, as well as other substrates, including dimethyllysine, inhibitors, and a neutral analog. By studying other proteins that contain different binding pockets, such as lower methylation state readers that incorporate salt bridges, we can develop a broader and more complete understanding of the mechanism of recognition of these post-translational modifications. This information not only provides information for therapeutic design, but a fully characterized system for studying cation-π interactions can be a useful system for testing computational methods aimed at modeling these binding interactions for medicinal chemistry purposes.

ACKNOWLEDGEMENTS

First, I would like to thank Marcey Waters for her mentorship and guidance over the past five years. She has truly created a supportive and encouraging lab environment. Every meeting with her left me excited about chemistry and often resulted in new, fun, and exciting ideas she allowed me to pursue. I am truly grateful for the opportunity to work on so many interesting and diverse projects during my time here.

I would also like to thank my committee members, Jeff Johnson, Simon Meek, Eric Brustad, and Mike Gagné, as well as Dave Nicewicz for serving on my orals committee. I appreciate all the support in guidance you all have provided me over the years: in the classroom, by providing me with opportunities for teaching experience, acting as job references, and providing friendly and encouraging smiles on a daily basis. I am grateful for the fantastic faculty I have gotten to work with at Carolina.

Special thanks to my wonderful collaborators, Stef Baril and Eric Brustad, for being awesome to work with. All of your patience and willingness to work with us even when we requested crazy amounts of protein is truly appreciated. Through all the struggles and roadblocks we encountered, I always knew I could count on words of encouragement, someone to complain with, or a baked good when we needed a break.

I owe many thanks to the faculty and staff of various facilities at UNC, especially Nely Dicheva of Proteomics, and Ashutosh Tripathy, the Director of the Macromolecular Interactions Facility. You have provided me valuable direction and advice throughout my work on various projects and instruments.

I am extremely thankful for all of the members of the Waters lab, both past and present, for a truly supportive and welcoming work environment. You have all provided guidance and friendship over my time here. I'm thankful for all the conversations about science, late nights in lab together, volunteer partners, drinking buddies, sports teammates, travel mates, partners in crime, and even short-term roommates. I could not have asked for a better group of people to work with.

Finally, I'd like to thank all of my family and friends, both in Chapel Hill and from back home. I could not imagine going through this process without an amazing network of people to lean on and help me through my stressful days. The support and encouragement I've received from everyone along the way is truly appreciated, and I can't wait to enter this next chapter with all of you.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF SCHEMES

LIST OF ABBREVIATIONS AND SYMBOLS

Ac	acetyl
ACh	acetylcholine
ACN	acetonitrile
Ala, A	alanine
Arg, R	arginine
aRMe2	asymmetric dimethyl arginine
Asn, N	asparagine
Asp, D	aspartic acid
Bn	benzyl
Boc	tert-butyloxycarbonyl
CD	circular dichroism
Cys, C	cysteine
d	doublet
DCM	dichloromethane
dd	doublet of doublet
DFT	density functional theory
DIPEA	N,N'-diisopropylethylamine
DMF	dimethylformamide
DNA	deoxyribonucleic acid
DODT	2,2′-(Ethylenedioxy)diethanethiol
DTT	dithiothreitol
EC50	half maximal effective concentration

EDT	1,2-ethanedithiol
ESI	electrospray ionization
FAM	5(6)-carboxyfluorescein
Fmoc	9-fluorenylmethoxycarbonyl
FPLC	fast protein liquid chromatography
Gln, Q	glutamine
Glu, E	glutamic acid
Gly, G	glycine
HBTU	O-(benzotriazol-1-yl)-N,N,N’N’-tetramethyluronium hexafluorophosphate
His, H	histidine
HKMT	histone lysine methyltransferase
HOBt	1-hydroxybenzotriazole
HP1	heterochromatin protein 1
HPLC	high performance liquid chromatography
HX	histone protein X
Ile, I	isoleucine
ITC	isothermal titration calorimetry
K(Me)	monomethyllysine
K(Me)2	dimethyllysine
K(Me)3	trimethyllysine
K_d	dissociation constant
LC-MS	liquid chromatography – mass spectrometry
Leu, L	leucine

LFER	linear free energy relationship
Log P	octanol-water partition coefficient
Lys, K	lysine
m	multiplet
MALDI	matrix-assisted laser desorption ionization
MeOH	methanol
Met, M	methionine
MHz	megahertz
mol	mole
MR	molar refractivity
MRE	mean residue ellipticity
MS	mass spectrometry
MTDB	7-methyl-1,5,7-Triazabicyclo[4.4.0]dec-1-ene
n-BuLi	n-butyllithium
nAChR	nicotinic acetylcholine receptor
NBS	N-Bromosuccinimide
NCL	native chemical ligation
NMP	N-Methyl-2-pyrrolidone
NMR	nuclear magnetic spectroscopy
NTA	nitrilotriacetic acid
PDB	Protein Data Bank
Phe, F	phenylalanine
Pro, P	proline

PTM	posttranslational modification
PyBOP	(benzotriazol-1-yl-oxytripyrrolidinophosphonium hexafluorophosphate)
RMe	monomethyl arginine
RNA	ribonucleic acid
s	singlet
SAH	S-adenoxyl-L-homocysteine
SAM	S-adenosyl-L-methionine
SAR	structure activity relationship
Ser, S	serine
SPPS	solid phase peptide synthesis
SPR	surface plasmon resonance
sRMe2	symmetric dimethyl arginine
t	triplet
tBu	tert-butyl
TCEP	(*tris*(2-carboxyethyl)phosphine)
TFA	trifluoroacetic acid
THF	tetrahydrofuran
Thr, T	threonine
TIPS	triisopropylsilane
tRNA	transfer ribonucleic acid
Trp, W	tryptophan
Trt	trityl
Ts	*p*-toluenesulfonyl

Tyr, Y	tyrosine
UV	ultra violet
Val, V	valine
WT	wild type

CHAPTER 1: RECOGNITION OF METHYLATED LYSINE AND CATION-π INTERACTIONS

1.1 Purpose of this work

With the emergence of trimethyllysine reader proteins becoming important therapeutic targets, an in depth understanding of their mechanism of recognition of their substrates will be advantageous to drug design. Information on the contribution of each residue to the binding interaction, as well as a broad comparison of reader proteins with different binding pockets can guide the design of inhibitors or probes for broad classes of reader proteins, as well as specific readers. While there have been some identified inhibitors of these proteins, it remains unknown whether they bind by the same cation-π mechanism as the natural substrate, or by an alternative mechanism.

The purpose of this work is to further probe reader protein binding mechanisms to provide a detailed understanding of the nature of these cation-π interactions. A systematic protein structure activity relationship (protein SAR) study, by systematically altering the binding pocket of reader proteins is used to provide a detailed description on the balance of forces in the binding of trimethyllysine.

1.2. Epigenetics

The human body consists of trillions of cells, all of which are genetically identical.[1] Despite containing identical DNA sequences, our cells are highly differentiated and specialized for diverse and specific functions. Cells with the same genotypes can show different phenotypes.[2] Therefore mechanisms must be in place that

control and regulate gene expression in order to produce the diverse sets of phenotypes possible using the same genomic sequence. As a result, the field of epigenetics, the study of these differences in gene expression without differences in the nucleic acid sequence, has become increasingly popular.[3] Moreover, the mechanism for regulating gene expression and silencing has risen in importance from a therapeutic standpoint, as the misregulation of expression is implicated in many diseases.[4]

1.2.1 Chromatin Structure

The extent of DNA packaging is responsible for whether or not a gene is accessible and active or if it is silenced.[5] In order to condense the large amount of DNA that must fit within the nucleus, the DNA is highly condensed into chromatin. One copy of the DNA in the human genome, fully stretched out, would measure over two meters in length, while the nucleus of our cell is only about 10 μM in diameter.[6] About 147 base pairs of the DNA strand is wrapped around eight histone proteins forming a nucleosome (Figure 1.1). These nucleosomes are further compacted together into chromatin.[6] There are two different types of chromatin. Heterochromatin is tightly compacted and genes present in heterochromatin are inaccessible and silenced, while euchromatin is more loosely packaged and active (Figure 1.2).[5]

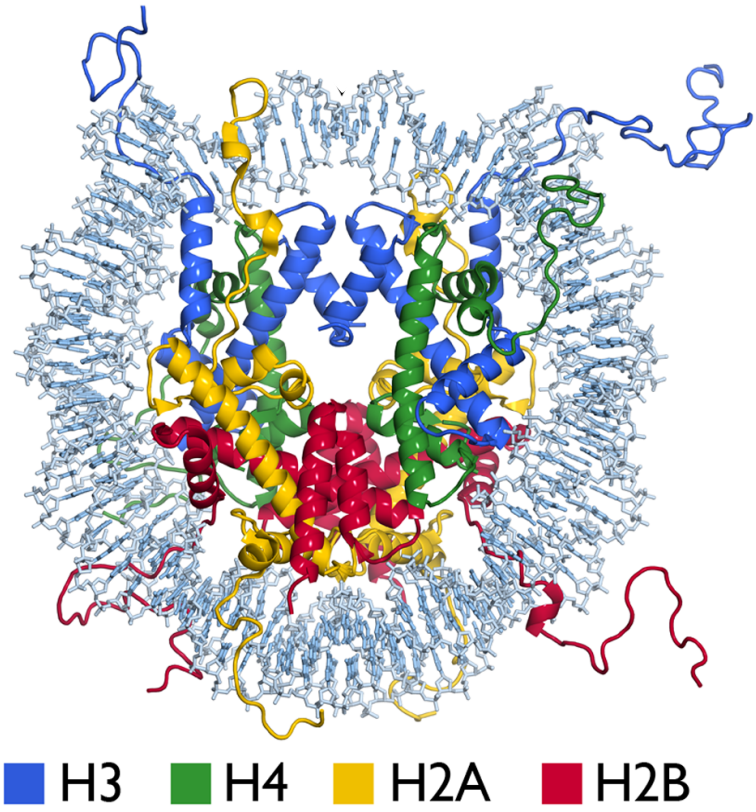


Figure 1.1. Crystal structure of DNA wrapped around histone protein octamer forming a nucleosome with tails extending outward.[6] Reprinted with permission from *Chem. Rev*. **2015,** *115*, 2255- 2273. Copyright 2015 American Chemical Society.

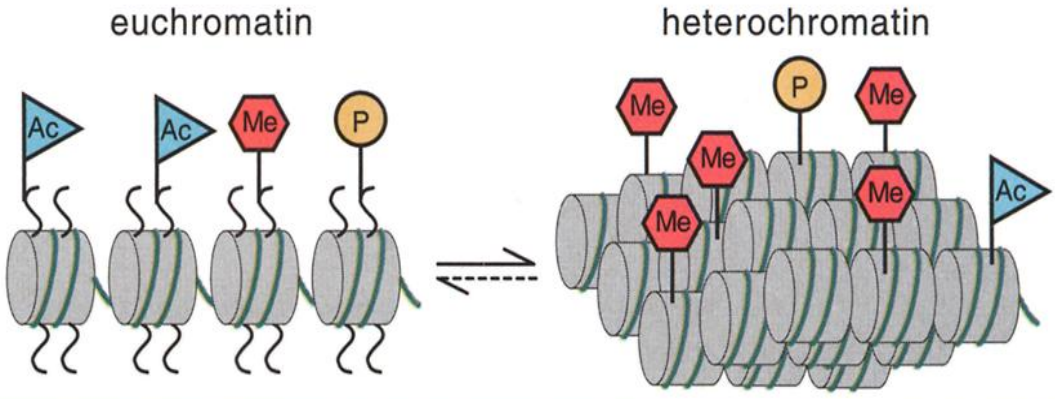


Figure 1.2. Active euchromatin (left) and inactive heterochromatin (right), with representative modifications often seen on each.[5] From *Science*, **2001**, *293*, 1074-1080. Reprinted with permission from AAAS.

1.2.2 Posttranslational Modifications (PTMs)

The formation of heterochromatin and euchromatin is controlled by a complex series of posttranslational modifications (PTMs) on the histone proteins. PTMs are covalent modifications on the side chains of amino acids that are installed after the protein has been translated from RNA (Figure 1.3). Common examples of these modifications include methylation of lysine (mono, di, and tri), methylation of arginine (mono and di – either symmetrically or asymmetrically), acetylation of lysine, and phosphorylation of serine, threonine, and tyrosine.

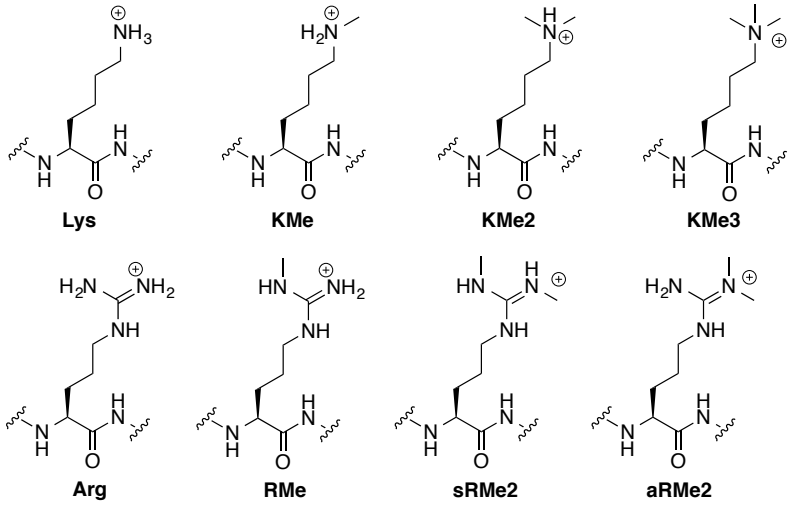


Figure 1.3. Various methylation states of lysine (top) and arginine (bottom) Arginine can be methylated symmetrically (sRMe2) or asymmetrically (aRMe2).

These modifications are found on histone proteins – most abundantly on the unstructured tail regions on the histone N-terminus. The nucleosome consists of an octomer of histone proteins, containing two each of histone proteins H2A, H2B, H3 and H4, while H1 links adjacent nucleosomes.[6] Posttranslational modifications marks are regulated and recognized by proteins known as "writers," "readers," and "erasers." Posttranslational modifications work in tandem with one another to form what has been

termed the "histone code."[5] The modifications regulate chromatin formation both by altering interactions with DNA and between nucleosomes, as well as recruiting other proteins and enzymes involved in chromatin remodeling.[7] They can also work together in the regulation of other modifications, known as histone "cross talk."[8]

1.3 Lysine methylation

Lysine methylation, while one of the most common posttranslational modifications, has not been extensively studied until recent years. Unlike acetylation and phosphorylation, which change the charge of the side chain, lysine methylation maintains the same overall side chain charge. Methylation is site-specific and the different levels (mono, di, and trimethylation) can effect gene expression differently, leading to added complexity with methylation compared to other posttranslational modifications. There have been various assays developed to map out the methylation pattern and determine the proteins that recognize these specific marks. The most characterized histone tail in terms of PTM's has been the H3 tail. There are at least 4 lysine positions that can be methylated on the H3 tail (K4, K9, K27, and K36).[9] The regulation of these marks are important for cell cycle regulation, DNA damage response, and development and differentiation.[10]

The location of these modifications within chromatin are important for proper regulation of gene expression, however determining how specific enzymes are recruited to certain histones is an active area of research. Studies have identified several DNA sequences found to recruit certain modifying enzymes to target areas.[10]The proper levels and activities of these enzymes and reader proteins are important in maintaining proper chromatin state and gene expression.

1.3.1 Writers: Lysine Methyltransferases

The first histone lysine methyltransferase (HKMT) to be discovered was SUV39H1, which methylates H3K9.[11] The vast majority of HKMTs contain a domain known as the SET domain (derived from the proteins it was identified in: Drosophila Su(var)3-9 and 'Enhancer of zeste' proteins).[12] These enzymes catalyze the transfer of a methyl group from S-adenosylmethionine (SAM) to the amino group of the lysine side chain (Figure 1.4). Most of these enzymes are very specific for a single site on a histone tail, and can modify the ammonium to specific degree of methylation (mono-, di-, or tri-). Often times this selectivity for methylation states is caused by a specific amino acid in the active site of the enzyme.[7] For example, one enzyme that can form trimethyllysine (DIM5), contains a phenylalanine in the active site.[13,14] Another enzyme, SET7/9, contains the slightly larger tyrosine residue, and can only accommodate the monomethyl product.[15] This dependence on phenylalanine or tyrosine to control the degree of methylation is thought to be general to these SET domain enzymes.[16,17]

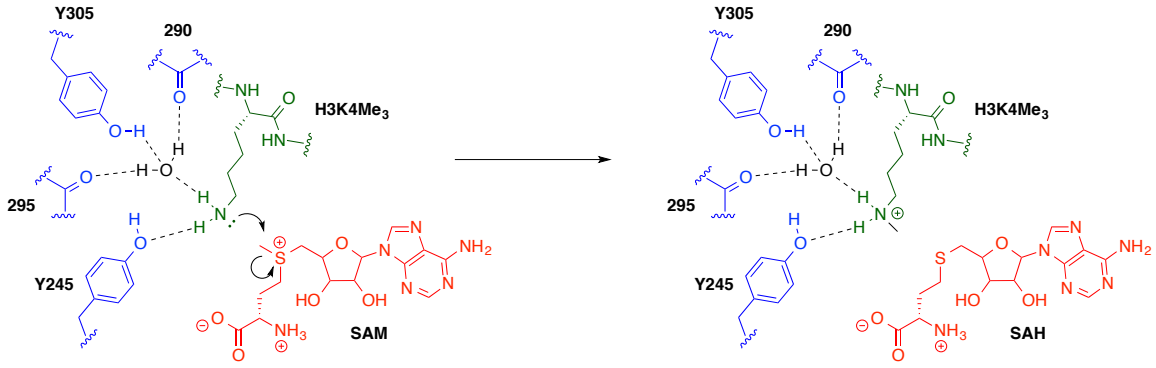


Figure 1.4. Mechanism of SET7/9 (blue) methylation of lysine (green) with cofactor SAM (red), to form monomethyllysine and S-adenosylhomocysteine (SAH).[15]

1.3.2 Erasers: Lysine Demethylases

It was initially believed that lysine methylation was a permanent mark, and there were no mechanisms for removal.[7] In 2004, the first lysine demethylase (LSD1) was discovered, which led to the conclusion that lysine methylation is a dynamic process.[18] This enzyme, however, requires a protonated nitrogen in order to function, and thus, cannot demethylate trimethyllysine. Another class of enzymes that contain a catalytic jumonji domain was discovered in 2006, which was capable of demethylating trimethyllysine.[19] These enzymes utilize a radical mechanism for demethylation of trimethyllysine. Histone methylation, while reversible, has a lower turnover rate than some other modifications. Furthermore, some sites need to be maintained, and different sites have different rates of turnover.[10]

1.3.3 Readers Proteins

Reader proteins are proteins that specifically recognize certain epigenetic marks leading to gene expression, gene silencing, or regulation of other modifications.[10] Some methyl binding domains commonly involved in the recognition of these marks include PHD, WD40, CW, and PWWP domains, as well as the "royal family" which includes tudor, chromo, MBT domains.[20] As the level of methylation increases, the number of hydrogen bond donors on the sidechain decreases (with trimethyllysine unable to form a hydrogen bond), and the size and hydrophobicity increase. This allows for the specific recognition of different methylation states.[20]

There are two primary modes that reader proteins typically employ in order to bind to methylated lysine (Figure 1.5). In a cavity insertion recognition mode, such as that found in 53BP1 tandem tudor domains,[21] the methylated lysine is buried deep into a

protein cavity. This mode is often present with readers of lower methylation states, and may also provide a way of discrimination of different states based on size.[20] Proteins that act via a surface groove recognition mode have a binding pocket that is more accessible, providing slightly less specificity for methylation states. Domains that recognize methylated lysine contain an "aromatic box" consisting of 2-4 aromatic residues that bind to the cation, presumably via cation-π interactions, as described below.[22]

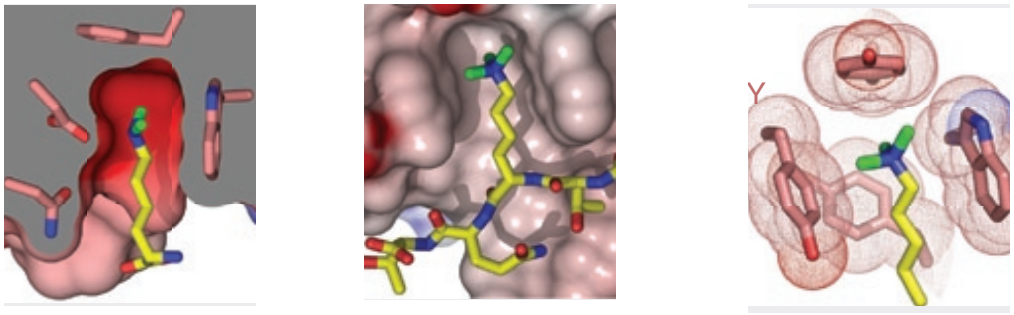


Figure 1.5. Depiction cavity insertion recognition (left), surface groove recognition (center) and aromatic box motif (right). Reprinted by permission from Macmillan Publishers Ltd: *Nat. Struct. Mol. Biol.*, **2007**, 14, 11, 1025-1040., Copyright 2007.

1.4 Cation-π Interactions

Cation-π interactions are prevalent across various biological systems are particularly involved in protein-protein interactions and protein folding.[23] While other intramolecular interactions such as salt bridges, hydrogen bonding, and the hydrophobic effect have been known and well studied in biological systems, advances demonstrating the prevalence of cation-π interactions has been more recent.[23,24]

1.4.1 Description of Cation-π Interactions

Cation-π interactions are an electrostatic interaction between a positively charged cation and the partial negative face of an aromatic ring, arising from the quadrupole moment.[23] The dipole between the hydrogen atoms and the sp^2 hybridized carbon atoms in the ring create the quadrupole moment of aromatic compounds. While these interactions are commonly seen in aromatic rings, the aromaticity is not necessary for the electrostatic interaction. [25]

These interactions were first seen in gas phase experiments where it was observed that the interaction between metal cations with benzene was stronger than their interaction with water. The ΔH of the interaction between potassium ions and benzene, for example, was -19.2 kcal/mol, while the interaction between potassium and water was -17.9 kcal/mol.[26] Trimethylamonium cations also exhibited a difference with a delta H of 9.4 kcal/mol with benzene and 9.0 kcal/mol with water.[27] The cation-π interaction between these metal cations and benzene follow the traditional electrostatic model with $Li^+ > Na^+ > K^+ > Rb^+$.[26,28–30]

While these initial observations were all in the gas phase, the cation-π interaction is also favorable in aqueous systems. Unlike amino acids that form salt-bridges and hydrogen bonds, which are often well solvated and prefer to interact with water than each other, the additional driving force for this interaction, which also includes van der Waals interactions and the hydrophobic effect, make cation-π interactions favorable even in aqueous systems, with countless examples in biological systems.[24,31] Solvation of cations certainly influences the interaction. Cations that are well solvated face a larger

desolvation cost to participate in cation-π interactions. This can be observed in aqueous systems where the interaction of benzene with K^+ is greater than with Na^+.[30]

Within proteins, positively charged lysine and arginine residues interact with aromatic phenylalanine, tyrosine, and tryptophan. These cations are often seen with the alkyl chains packed against the aromatic face, for example the methyl groups of of methylated lysine or the ε-methylene of unmethylated lysine, as this is where the partial positive charge actually lies (Figure 1.6). Analysis of the PDB estimates that there is one cation-π interaction for every 77 amino acid residues, and about 26% of all tryptophan residues are involved in cation-π interactions.[23,32] In intermolecular protein-protein interactions, cation-π interactions were present in about half of all complexes and one third of all homodimers.[32]

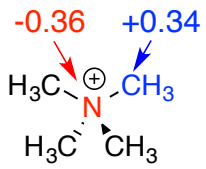


Figure 1.6. Location of partial charges on a quaternary ammonium

1.4.2 Substituent Effects on Cation-π Interactions

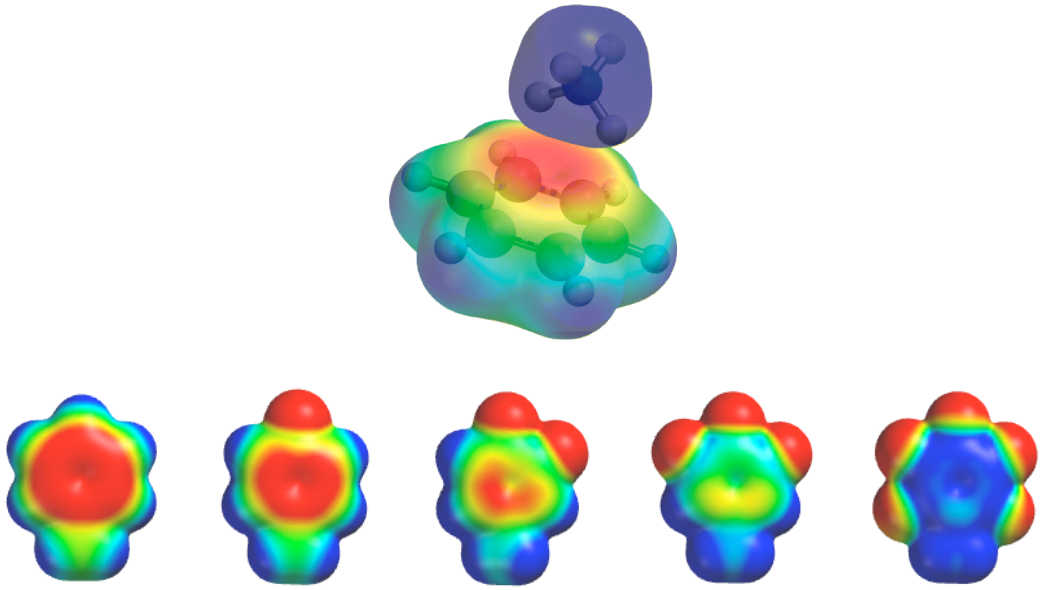


Figure 1.7. Electrostatic potential maps of a cation-π interaction between benzene and ammonium cation (top) and toluene with increasing amounts (1-4) of fluorine substitution.[23] Adapted from Dougherty, D.A. *Acc. Chem. Res*. **2013**, 46, 4, 885-893. Copyright 2013 American Chemical Society

While the net interaction energy is dependent on several forces, including van der Waals and the hydrophobic effect, the variance in the strength of cation-π interactions result from differences in electrostatic interactions.[33] With aromatic rings, this can be estimated quite well using electrostatic potential maps (Figure 1.7).[34] Electron withdrawing substituents reduce the negative electrostatic potential of the ring. This correlates nicely with the reduced binding energy measured for Na^+ in the gas phase. These energies also roughly correlate with σ_{meta} values, as resonance effects do not play a role in cation-π binding energies.[31,33]

Recent computation studies have provided an alternative model for substituent effects on cation-π interactions.[25,35–37] The results suggest that differences in energies are

primarily the result of through space interactions with the dipole between the carbon and the substituent, rather than a change in the electron density of the π-cloud over the center of the ring.[35,36]

1.4.3 Studying Cation-π Interactions in Biological Systems

There have been many model systems developed to study and quantify cation-π interactions, estimating them to be worth about 1-2 kcal/mol/ring.[23] Among the first small molecule model systems that pointed to the potential of cation-π interactions were cyclophane receptors.[38,39] These receptors were known to bind small aromatic compounds, but were also found to bind quaternary ammonium ions in aqueous systems.[40] Since then, a variety of synthetic receptors have been developed to bind quaternary ammonium ions, and more specifically, cationic methylated amino acid residues and peptides.[41–54]

Peptide models have incorporated cation-π interactions to increase structural stability as well as to quantify the strength of the interaction.[55] For example, incorporation of an Arg/Lys and Trp pair at the *i*, *i*+4 positions of an α-helix stabilized the structure by 0.4 kcal/mole, while the interaction of Phe with either Lys or Arg was weaker, but still stabilizing by about 0.1 to 0.2 kcal/mol.[56–58]

β-hairpin peptides containing cross-strand cation-π interactions have provided further insight into magnitude of these interactions.[59–61] The cross-strand interaction between cationic lysine and arginine and aromatic amino acids tryptophan and phenylalanine stabilized the hairpins by about 0.2 to 0.5 kcal/mol.[61]

Studies with these hairpins were also done with methylated lysine and arginine residues. Increasing amounts of methylation on these residues increased the magnitude

of this interaction in aqueous solutions.[62–64] Unmethylated lysine is protonated and therefore positively charged at neutral pH, and so it is still capable of participating in cation-π interactions. However, increased amounts of methylation increases the interaction, presumably due to decreasing the desolvation cost of the cation and increased surface area for van der Waals interactions.

1.4.4 Using Neutral Analogs to Probe Cation-π Interactions

One approach that has been frequently used to study cation-π between aromatic amino acid residues and trimethyllysine has been the replacement of the quaternary ammonium of the trimethyllysine with a neutral isostere, a tert-butyl group. The analog has similar size (and van der Waals interactions) and increased hydrophobicity but lacks the positive charge. By replacing trimethyllysine with these neutral analogs in systems, the role of the cation can be further investigated. This substitution has been used in various systems including host-guest, protein-peptide, protein-inhibitor, and protein-nucleotide to evaluate the contribution of electrostatics to the interaction.[22,40,65,66] Interestingly, the results of these studies frequently measure very similar values, usually around 2.5 kcal/mol for the contribution of the cation to binding (Figure 1.8).

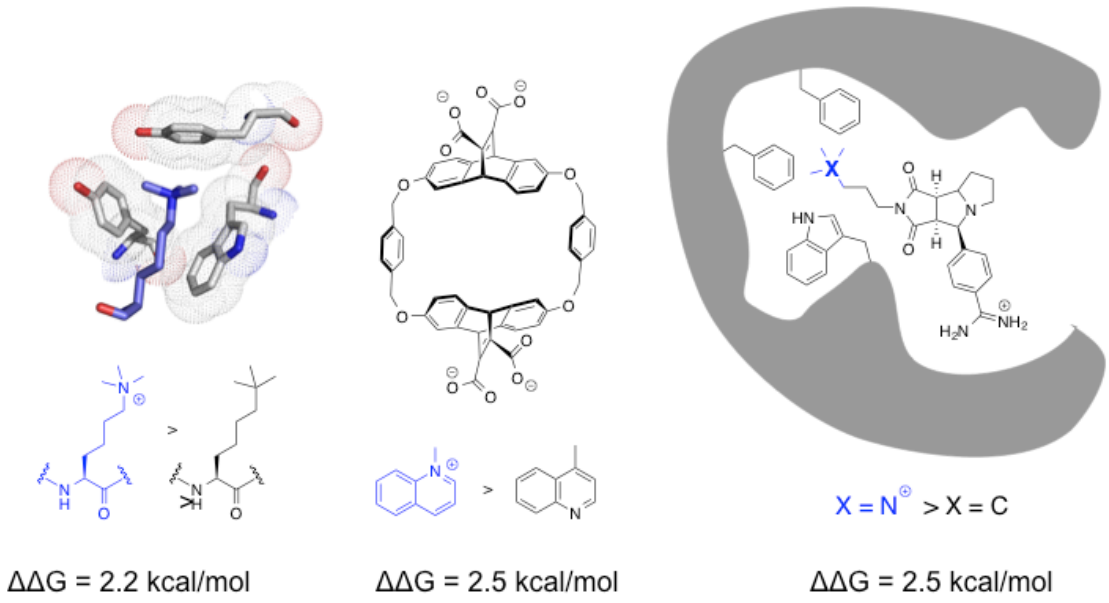


Figure 1.8. Neutral isosteres used to determine electrostatic contributions of cation-p interactions. Interactions between reader proteins and trimethylysine (left), macrocycle and guest (center), and inhibitor of factor XA (right), are examples of systems studied with observed ΔΔG values around 2.5 kcal/mol. Adapted with permission from *J. Am. Chem. Soc.* **1988**, *110* (6), 1983–1985, Copyright 1988 American Chemical Society; *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11184–11188, Copyright 2007 National Academy of Sciences, USA, Copyright 1988 American Chemical Society; *Angew. Chemie - Int. Ed.* **2009**, *48* (4), 811–814, Copyright 2009, John Wiley and Sons.

One of the simplest examples using a neutral analog to demonstrate importance of the cation is a small molecule that can fold and unfold.[67] An indole ring connected with a linker to a quaternary ammonium, can be compared to the molecule with a tert-butyl group in place of the ammonium (Figure 1.9). X-ray analysis demonstrated that the compound containing the ammonium was folded in a way that placed the ammonium near the indole ring. The neutral analog, however, was unfolded. These differences demonstrate the preference for the cation to interact with the indole ring.

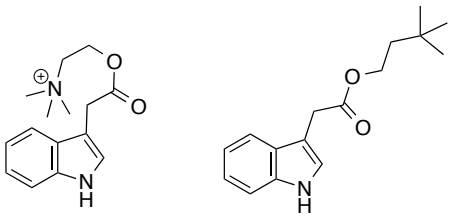


Figure 1.9. Indole linked to cationic ammonium group adopted a folded comformation, while the neutral analog remained unfolded.[76]

Factor XA is a serine protease involved in blood coagulation that has been similarly probed. It has three aromatic residues, Trp, Tyr, and Phe, that are proposed to be involved in a cation-π interaction with an ammonium group of an inhibitor. [68] The inhibitor was synthesized with a neutral tert-butyl group in place of the ammonium group and the binding affinities were compared. Using this method, the electrostatics contributed 2.8 kcal/mol to the binding.[69]

One of the most well studied examples of cation-π interactions between proteins and nucleic acids has been with 7-methyl guanine. This methylation is an important part of RNA processing, and forms the 5' cap. Various proteins with different functions recognize the 5' cap.[70] One such protein, eukaryotic initiation factor 4E, binds to the 5' cap and triggers translation. Crystal structure analysis shows the methylated guanine base sandwiched between two tryptophan residues.[70–72] The positive charge of the 7-methyl G is thought to provide increased stability over π-π stacking interactions. Computation studies were done with the neutral analog 7-methyl D, which is capable of the stacking interaction but not the cation-π interaction (Figure 1.10). The binding

affinity was calculated to be reduced by 2.8 kcal/mol.[73]

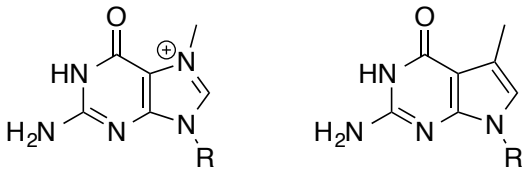


Figure 1.10. Methylated bases for 7-methyl G (m^7G) and neutral analog 7-methyl D (m^7D).

1.5 Reader Proteins and Cation-π Interactions

As many reader proteins are thought to participate in cation-π interactions in the recognition of post-translational modifications, these methods can also be used to begin to probe and understand the mechanism of binding. Computation and experimental studies were performed with reader proteins binding to histone peptides containing either trimethyllysine or the neutral, tert-butyl containing analog.[74] Most of the reader proteins showed significant reduction in binding affinities with the neutral analog, often on the order of 2 kcal/mol, similar in magnitude to that observed in other systems.

1.5.1 Heterochromatin Protein 1

The first reader protein to be crystalized was the chromodomain of heterochromatin protein 1 (HP1) from *Drosophila melanogaster*.[75] This protein binds to di- and trimethyllyine 9 on histone 3 (H3K9Me3). This binding is associated with heterochromatin formation and epigenetic silencing.

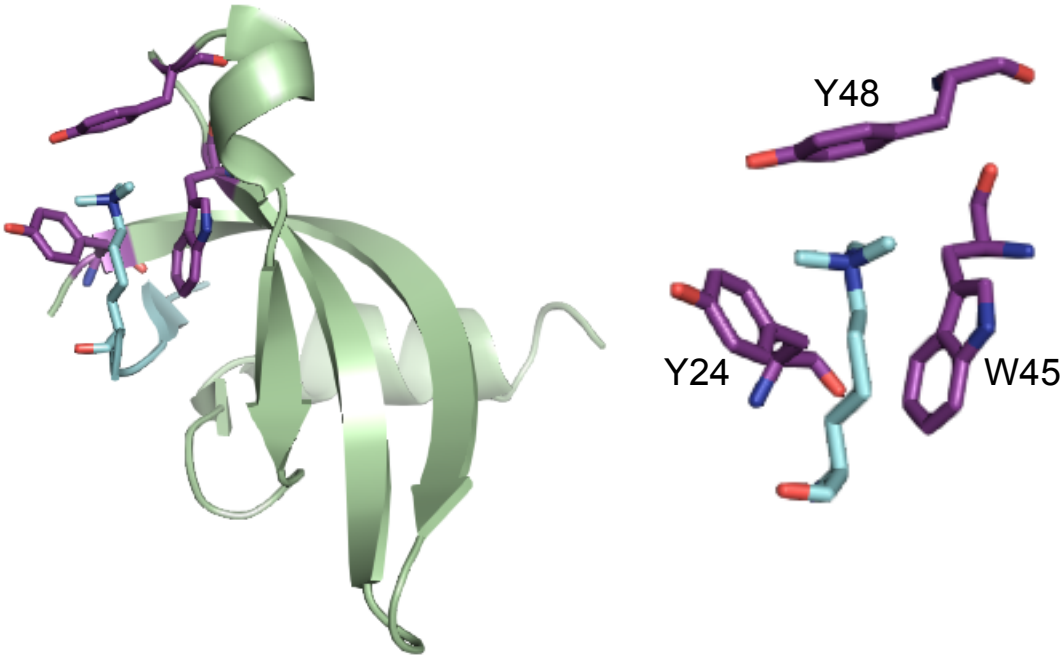


Figure 1.11. Crystal structure of HP1 (green) bound to H3K9Me3 (cyan) with binding pocket highlighted in purple (left), and zoom in of binding pocket with aromatic residues labeled (right). PDB: 1KNE[75]

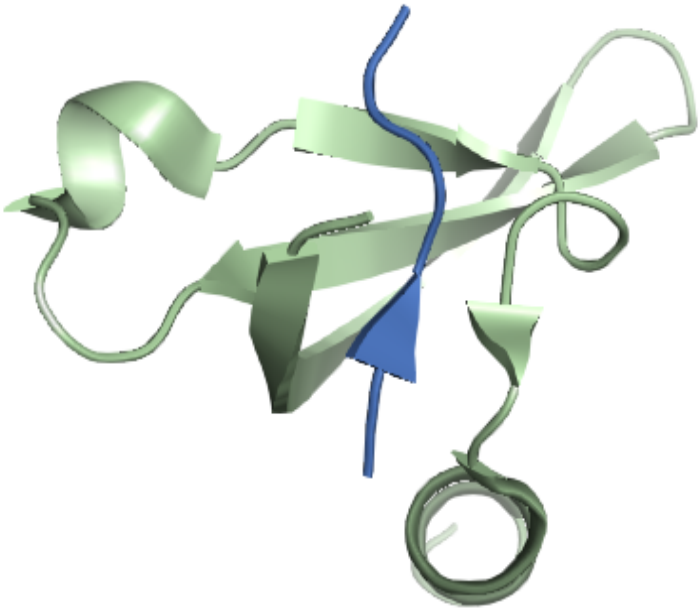


Figure 1.12. Three stranded β-sheet formed between the H3 tail (blue) and HP1 (green). PDB: 1KNE[75]

It has been shown that the chomodomain alone of HP1 is sufficient for the binding interaction with H3K9Me3. The chromodomain was crystalized with H3 peptides (residues 1-15: ARTKQTARK(Me_X)STGGKA) containing both di- (KMe2) and trimethyllysine (KMe3) at position 9. The crystal structure revealed a binding pocket consisting of three aromatic residues. Two tyrosine residues and a tryptophan residue surround the cationic ammonium (Figure 1.11), while a glutamic acid forms a water mediated salt bridge with KMe2. Residues 5-8 (QTAR) of the H3 peptide form a three standed β-sheet with the chromodomain (Figure 1.12), and serine 10 forms a hydrogen bond. Binding affinities were measured by isothermal titration calorimetry (ITC). The dissociation constant (K_d) for HP1 with H3 peptide with KMe3 was 2.5 μM, while the K_d with KMe2 was around 7 μM. Mutation of any of the aromatic residues in the binding pocket to alanine drastically reduced binding.[75]

Further evidence for cation-π interactions comes from previous work done in the Waters' lab (Figure 1.8, left panel).[22] By using the neutral analog approach, the neutral residue was both incorporated into a β-hairpin, as well as an H3 peptide. The neutral analog destabilized the hairpin folding compared to trimethyllysine, and, when incorporrated in the H3 peptide, exhibited reduced binding affinity with HP1 on the order of 2 kcal/mol. This was the initial evidence for the critical role of the cation in this binding interaction.

REFERENCES

(1) Bianconi, E.; Piovesan, A.; Facchin, F.; Beraudi, A.; Casadei, R.; Frabetti, F.; Vitale, L.; Pelleri, M. C.; Tassani, S.; Piva, F.; Perez-Amodio, S.; Strippoli, P.; Canaider, S. *Ann Hum Biol* **2013**, *40* (6), 301–4460.

(2) Esteller, M. *N. Engl. J. Med.* **2008**, *358*, 1148–1159.

(3) Egger, G.; Liang, G.; Aparicio, A.; Jones, P. a. *Nature* **2004**, *429* (6990), 457–463.

(4) Liu, Y.; Liu, K.; Qin, S.; Xu, C.; Min, J. *Pharmacol. Ther.* **2014**, *143* (3), 275–294.

(5) Jenuwein, T.; Allis, C. D. *Science (80-.).* **2001**, *293* (5532), 1074–1080.

(6) McGinty, R. K.; Tan, S. *Chem. Rev.* **2015**, *115* (6), 2255–2273.

(7) Bannister, A. J.; Kouzarides, T. *Cell Res.* **2011**, *21* (3), 381–395.

(8) Fischle, W.; Wang, Y.; Allis, C. D. *Curr. Opin. Cell Biol.* **2003**, *15* (2), 172–183.

(9) Lachner, M.; O'Sullivan, R. J.; Jenuwein, T. *J. Cell Sci.* **2003**, *116* (Pt 11), 2117–2124.

(10) Greer, E. L.; Shi, Y. *Nat. Rev. Genet.* **2012**, *13* (5), 343–357.

(11) Rea, S.; Eisenhaber, F.; O'Carroll, D.; Strahl, B. D.; Sun, Z. W.; Schmid, M.; Opravil, S.; Mechtler, K.; Ponting, C. P.; Allis, C. D.; Jenuwein, T. *Nature* **2000**, *406* (6796), 593–599.

(12) Dillon, S. C.; Zhang, X.; Trievel, R. C.; Cheng, X. *Genome Biol.* **2005**, *6* (8), 227.

(13) Tamaru, H.; Zhang, X.; McMillen, D.; Singh, P. B.; Nakayama, J.; Grewal, S. I.; Allis, C. D.; Cheng, X.; Selker, E. U. *Nat. Genet.* **2003**, *34* (may), 75–79.

(14) Zhang, X.; Yang, Z.; Khan, S. I.; Horton, J. R.; Tamaru, H.; Selker, E. U.; Cheng, X. *Mol. Cell* **2003**, *12* (1), 177–185.

(15) Xiao, B.; Jing, C.; Wilson, J. R.; Walker, P. A.; Vasisht, N.; Kelly, G.; Howell, S.; Taylor, I. A.; Blackburn, G. M.; Gamblin, S. J. *Nature* **2003**, *421*, 652–656.

(16) Cheng, X.; Collins, R. E.; Zhang, X. *Annu. Rev. Biophys. Biomol. Struct.* **2005**, *34*, 267–294.

(17) Collins, R. E.; Tachibana, M.; Tamaru, H.; Smith, K. M.; Jia, D.; Zhang, X.; Selker, E. U.; Shinkai, Y.; Cheng, X. *J. Biol. Chem.* **2005**, *280* (7), 5563–5570.

(18) Shi, Y.; Lan, F.; Matson, C.; Mulligan, P.; Whetstine, J. R.; Cole, P. A.; Casero, R. A.; Shi, Y. *Cell* **2004**, *119* (7), 941–953.

(19) Tsukada, Y.; Fang, J.; Erdjument-Bromage, H.; Warren, M. E.; Borchers, C. H.; Tempst, P.; Zhang, Y. *Nature* **2006**, *439* (February), 811–816.

(20) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14* (11), 1025–1040.

(21) Botuyan, M. V.; Lee, J.; Ward, I. M.; Kim, J. E.; Thompson, J. R.; Chen, J.; Mer, G. *Cell* **2006**, *127* (7), 1361–1373.

(22) Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11184–11188.

(23) Dougherty, D. A. *Acc. Chem. Res.* **2013**, *46* (4), 885–893.

(24) Mahadevi, A. S.; Sastry, G. N. *Chem. Rev.* **2012**, *113* (3), 2100–2138.

(25) Wheeler, S. E.; Bloom, J. W. G. *J. Phys. Chem. A* **2014**, *118* (32), 6133–6147.

(26) Sunner, J.; Nishizawa, K.; Kebarle, P. *J. Phys. Chem.* **1981**, *85* (13), 1814–1820.

(27) Meot-Ner, M.; Deakyne, C. a. *J. Am. Chem. Soc.* **1985**, *107* (2), 474–479.

(28) Taft, R. W.; Anvia, F.; Gal, J.-F.; Walsh, S.; Capon, M.; Holmes, M. C.; Hosn, K.; Oloumi, G.; Vasanwala, R.; Yazdani, S. *Pure Appl. Chem.* **1990**, *62* (1), 17–23.

(29) Guo, B. C.; Purnell, J. W.; Castleman, A. W. *Chem. Phys. Lett.* **1990**, *168* (2), 155–160.

(30) Kumpf, R. A.; Dougherty, D. A. *Science (80-.).* **1993**, *261* (5129), 1708–1710.

(31) Ma, J. C.; Dougherty, D. A. *Chem. Rev.* **1997**, *97* (5), 1303–1324.

(32) Gallivan, J. P.; Dougherty, D. a. *Proc. Natl. Acad. Sci. U. S. A.* **1999**, *96* (17), 9459–9464.

(33) Mecozzi, S.; West, A. P.; Dougherty, D. A. *J. Am. Chem. Soc.* **1996**, *118* (9), 2307–2308.

(34) Mecozzi, S.; West, A. P.; Dougherty, D. A. *Proc. Natl. Acad. Sci. U. S. A.* **1996**, *93* (20), 10566–10571.

(35) Sherrill, C. D. *Acc. Chem. Res.* **2013**, *46* (4), 1020–1028.

(36) Wheeler, S. E.; Houk, K. N. *J. Am. Chem. Soc.* **2009**, *131* (9), 3126–3127.

(37) Wheeler, S. E. *Acc. Chem. Res.* **2012**, *46* (4), 1029–1038.

(38) Shepodd, T. J.; Petti, M. A.; Dougherty, D. A. *J. Am. Chem. Soc.* **1986**, *108* (19), 6085–6087.

(39) Diederich, F.; Dick, K. *J. Am. Chem. Soc.* **1984**, *106* (26), 8024–8036.

(40) Petti, M. A.; Shepodd, T. J.; Barrans, R. E. J.; Dougherty, D. A. *J. Am. Chem. Soc.* **1988**, *110* (1), 6825–6840.

(41) Murayama, K.; Aoki, K. *Chem. Commun.* **1997**, No. 1, 119–120.

(42) Murayama, K.; Aoki, K. *Chem. Commun.* **1998**, No. 12, 607–608.

(43) Shivanyuk, A.; Rebek, J. *Proc. Natl. Acad. Sci.* **2001**, *98* (14), 7662–7665.

(44) Koh, K. N.; Araki, K.; Ikeda, a; Otsuka, H.; Shinkai, S. *J. Am. Chem. Soc.* **1996**, *118* (4), 755–758.

(45) Lhoták, P.; Shinkai, S. *J. Phys. Org. Chem.* **1997**, *10* (5), 273–285.

(46) Daze, K. D.; Pinter, T.; Beshara, C. S.; Ibraheem, A.; Minaker, S. A.; Ma, M. C. F.; Courtemanche, R. J. M.; Campbell, R. E.; Hof, F. *Chem. Sci.* **2012**, *3* (9), 2695.

(47) Hof, F.; Trembleau, L.; Ullrich, E. C.; Rebek, J. *Angew. Chemie - Int. Ed.* **2003**, *42* (27), 3150–3153.

(48) Gamal-Eldin, M. a; Macartney, D. H. *Org. Biomol. Chem.* **2013**, *11* (3), 488–495.

(49) Corbett, P. T.; Leclaire, J.; Vial, L.; West, K. R.; Wietor, J.-L.; Sanders, J. K. M.; Otto, S. *Chem. Rev.* **2006**, *106* (9), 3652–3711.

(50) Gagné, M. R.; Ghosh, S.; Ingerman, L. a.; Frye, A. G.; Lee, S. J.; Waters, M. L. *Org. Lett.* **2010**, *12* (5), 1860–1863.

(51) Pinkin, N. K.; Waters, M. L. *Org. Biomol. Chem.* **2014**, 7059–7067.

(52) E. Beaver, J.; C. Peacor, B.; V. Bain, J.; James, L. I.; L. Waters, M. *Org. Biomol. Chem.* **2015**, *13* (11), 3220–3226.

(53) James, L. I.; Beaver, J. E.; Rice, N. W.; Waters, M. L. *J. Am. Chem. Soc.* **2013**, *135* (17), 6450–6455.

(54) Beaver, J. E.; Waters, M. L. *ACS Chem. Biol.* **2016**, *11* (3), 643–653.

(55) Waters, M. L. *Biopolymers* **2004**, *76* (5), 435–445.

(56) Shi, Z.; Olson, C. A.; Kallenbach, N. R. *J. Am. Chem. Soc.* **2002**, *124* (13), 3284–3291.

(57) Andrew, C. D.; Bhattacharjee, S.; Kokkoni, N.; Hirst, J. D.; Jones, G. R.; Doig, A. J. *J. Am. Chem. Soc.* **2002**, *124* (43), 12706–12714.

(58) Tsou, L. K.; Tatko, C. D.; Waters, M. L. *J. Am. Chem. Soc.* **2002**, *124* (50), 14917–14921.

(59) Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2006**, *128* (41), 13586–13591.

(60) Andersen, N. H.; Olsen, K. A.; Fesinmeyer, R. M.; Tan, X.; Hudson, F. M.; Eidenschink, L. A.; Farazi, S. R. *J. Am. Chem. Soc.* **2006**, *128* (18), 6101–6110.

(61) Tatko, C. D.; Waters, M. L. *Protein Sci.* **2003**, *12* (11), 2443–2452.

(62) Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2005**, *127* (18), 6518–6519.

(63) Hughes, R. M.; Waters, M. L. *J. Am. Chem. Soc.* **2006**, *128* (39), 12735–12742.

(64) Hughes, R. M.; Benshoff, M. L.; Waters, M. L. *Chem. - A Eur. J.* **2007**, *13* (20), 5753–5764.

(65) Shepodd, T. J.; Petti, M. A.; Dougherty, D. A. *J. Am. Chem. Soc.* **1988**, *110* (6), 1983–1985.

(66) Salonen, L. M.; Bucher, C.; Banner, D. W.; Haap, W.; Mary, J. L.; Benz, J.; Kuster, O.; Seiler, P.; Schweizer, W. B.; Diederich, F. *Angew. Chemie - Int. Ed.* **2009**, *48* (4), 811–814.

(67) Aoki, K.; Murayama, K.; Nishiyama, H. *J. Chem. Soc. Chem. Commun.* **1995**, *764* (21), 2221.

(68) Maignan, S.; Guilloteau, J. P.; Choi-Sledeski, Y. M.; Becker, M. R.; Ewing, W. R.; Pauls, H. W.; Spada, A. P.; Mikol, V. *J. Med. Chem.* **2003**, *46* (5), 685–690.

(69) Schärer, K.; Morgenthaler, M.; Paulini, R.; Obst-Sander, U.; Banner, D. W.; Schlatter, D.; Benz, J.; Stihle, M.; Diederich, F. *Angew. Chemie - Int. Ed.* **2005**, *44* (28), 4400–4404.

(70) Marcotrigiano, J.; Gingras, a C.; Sonenberg, N.; Burley, S. K. *Nucleic Acids Symp. Ser.* **1997**, *89* (36), 8–11.

(71) Brown, C. J.; McNae, I.; Fischer, P. M.; Walkinshaw, M. D. *J. Mol. Biol.* **2007**, *372* (1), 7–15.

(72) Tomoo, K.; Shen, X.; Okabe, K.; Nozoe, Y.; Fukuhara, S.; Morino, S.; Ishida, T.; Taniguchi, T.; Hasegawa, H.; Terashima, A.; Sasaki, M.; Katsuya, Y.; Kitamura, K.; Miyoshi, H.; Ishikawa, M.; Miura, K. *Biochem. J.* **2002**, *362*, 539–544.

(73) Guimarães, C. R. W.; Kopecky, D. J.; Mihalic, J.; Shen, S.; Jeffries, S.; Thibault, S. T.; Chen, X.; Walker, N.; Cardozo, M. *J. Am. Chem. Soc.* **2009**, *131* (50), 18139–18146.

(74) Kamps, J. J. a. G.; Huang, J.; Poater, J.; Xu, C.; Pieters, B. J. G. E.; Dong, A.; Min, J.; Sherman, W.; Beuming, T.; Matthias Bickelhaupt, F.; Li, H.; Mecinović, J. *Nat. Commun.* **2015**, *6*, 8911.

(75) Jacobs, S. A.; Khorasanizadeh, S. *Science (80-.).* **2002**, *295* (5562), 2080–2083.

(76) Aoki, K.; Murayama, K.; Nishiyama, H. *J. Chem. Soc. Chem. Commun.* **1995**, *764* (21), 2221.

CHAPTER 2: PROBING THE CONTRIBUTION OF TYROSINE RESIDUES TO CATION-Π INTERACTIONS IN HP1

2.1 Background

Heterochromatin Protein 1 (HP1) has been extensively studied, including mutation studies and crystal structure analysis, making it a good model system to initiate studies designed to tune the electronics of cation-π interactions.[1–3] There is strong evidence for cation-π interactions already, however the contributions of the individual tyrosine residues is unknown.[4,5] Detailed mechanistic information about the contribution of tyrosine provides information useful about the recognition of their native substrates, providing more guidelines developing drugs that bind to these types of reader proteins with similar affinity and specificity. Tuning the electronics of aromatic residues by incorporating electron withdrawing groups on the aromatic ring is an approach that has been used in a variety of systems to identify important cation-π interactions, and can be applied to the tyrosine residues of HP1.[6]

2.1.1 Tuning Cation-π Interactions of Ligand Gated Ion Channels

Some of the most extensively studied systems are the nicotinic acetylcholine receptors (nAChR). [7–12] These proteins belong to a superfamily of proteins that include receptors for serotonin, γ-aminobutyric acid, and glycine.[9] The nAChR is a ligand gated ion channel that binds to the quaternary ammonium of acetylcholine. This channel also binds to nicotine, which also contains an ammonium ion, and the interaction is the chemical basis for nicotine addiction.[9] While there are no crystal structures available of the membrane bound protein, it is thought to adopt a structure similar to other AChE binding proteins (Fig. 2.1).[8,12]

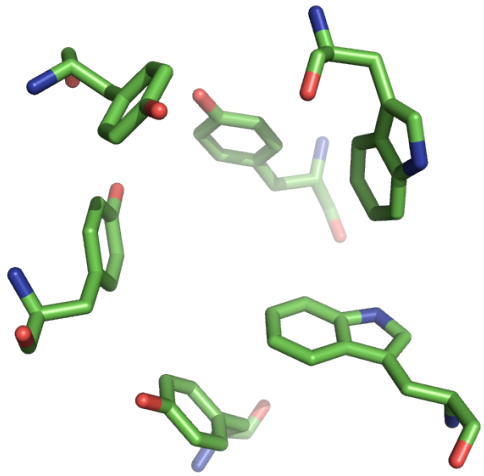


Figure 2.1. Aromatic residues of a binding region of an ACh binding protein. PDB:1I9B.[12]

The binding site was found to contain a large number of aromatic residues, suggesting that cation-π interactions may play an important role in the recognition of the ammonium substrates.[13] In order to test for the presence of important cation-π interactions, mutations were made to the aromatic residues that have potential for participating in this binding interaction. If the residue is participating in a cation-π interaction, it should exhibit similar substituent effects to those calculated in the gas phase. Incorporation of electron withdrawing groups weakens the ability of the residue to participate in cation-π interactions, and should therefore reduce the overall binding affinity if a cation-π interaction is present. A linear free energy relationship will be observed if there is a liner correlation between log of the K_a values (or ΔG) plotted against the calculated gas phase ΔG values. Linear correlation demonstrates a linear free energy relationship and suggests the same mechanism of binding. This is a common approach for probing mechanism of reactions and interactions that can be tuned with electronics. The simplest way to achieve this is by the incorporation of increasing numbers of fluorine atoms on the ring (Figure 2.2).

Fluorine is often the optimum choice as it is the most electronegative atom, and, as a single, small atom, it typically has minimal steric effects.

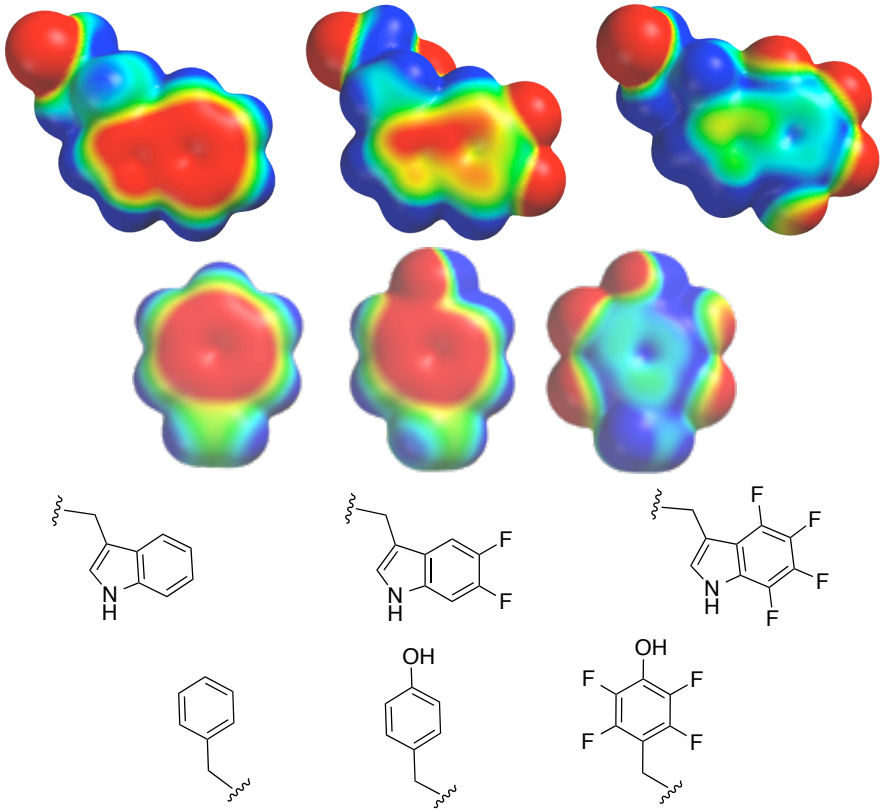


Figure 2.2. Electrostatic potential maps of tryptophan with 0, 2, and 4 fluorine atoms (top), phenylalanine (bottom left), tyrosine (bottom center), and tetrafluorotyrosine (bottom right). Electrostatic potential maps were generated with MacSpartan: HF/6-31G*; isodensity value = 0.02; range = -50 (red, electron rich) to 50 kcal/mol (blue, electron poor).

As the nAChR system contains a membrane bound protein, it is not possible to measure direct binding affinities. Instead, EC_{50} values were measured under the assumption that decreased binding of the ACh leads to decreased activity. Using this method the potential cation-π interactions were probed at all aromatic residues in the binding region. Interestingly, no Tyrosine residues exhibited a correlation in activity upon mutation with electron poor aromatic residues, and only one tryptophan residue demonstrated a linear free energy relationship.[8,9] This

interaction is also seen in the recognition of nicotine by these receptors in the brain.[8] However, in the muscle-type acetylcholine receptors, this cation-π interaction with tryptophan is absent, and thus, the binding energy for nicotine is greatly reduced. Without this weakening of the interaction, nicotine would stimulate the muscle receptors in addition to the brain and have detrimental consequences. This study provided detailed mechanistic insight into the mechanism of nicotine addiction and the basis for selectivity of this interaction.

Several other systems that contain a cation-π interaction have been studied using this approach. Other ligand gated-ion channels have also been shown to participate in important cation-π interactions,[14] as well as GABA,[15] glycine,[16] and 5-HT_3[17] receptors, and have aided in the development of agonists and further biochemical studies of these receptors.[18–20]

2.1.2 Tuning Cation-π Interactions of Enzymes

Terpene synthase enzymes have commonly been thought to rely on cation-π interactions to stabilize their carbocation intermediates. Unnatural amino acid mutagenesis studies replacing aromatic residues with either fluorinated residues[21,22] or phenyl alanine derivatives containing para-substituted electron withdrawing groups[23] led to decreased activity of the enzymes and reduced product formation for several aromatic residues (Figure 2.3). However, like the ligand gated ion channels, it is unclear whether structural changes are impacting binding of the substrates.

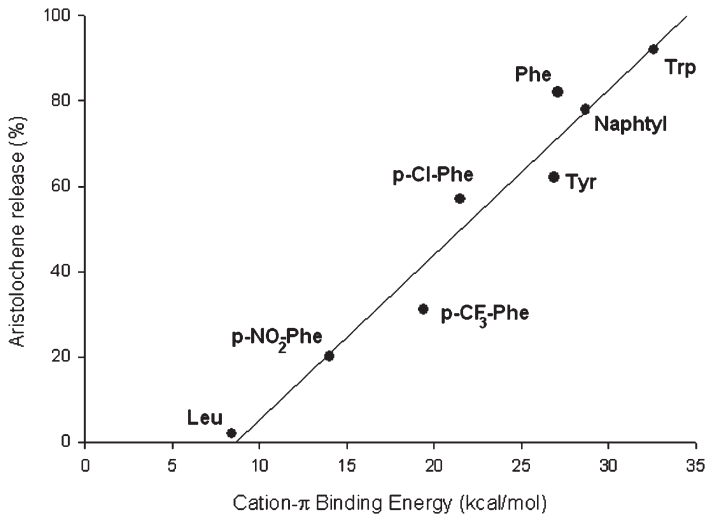


Figure 2.3. Linear free energy relationship between activity of Aristocholene Synthase upon mutation and calculated cation-π binding energies. Reproduced with permission from *J. Am. Chem. Soc.* **2011**, *133* (35), 13906–13909. Copyright 2011, American Chemical Society.

This approach can help develop a detailed understanding for the mechanism of binding selectivity reader proteins have for their substrates. As there are three residues in the aromatic pocket of HP1, it is possible that not all of them contribute to the interaction, or contribute to the interaction by different magnitudes. Additionally, as the nAChR system did not have tyrosine residues that exhibited a linear free energy relationship, it is unknown whether the same will be seen with HP1. Initial mutagenesis studies comparing the two tyrosine residues provide a deeper understanding of the role these residues play in the binding of trimethyllysine.

2.2 System Design

2.2.1 Unnatural Amino Acid Mutagenesis

Because the measurement of binding interactions requires over-expression and isolation of the protein of interest, unlike the measurements done on Dougherty's ACh receptors, similar methods for incorportation of unnatural amino acid mutagenesis cannot be used, as only single cells are amenable to this approach because chemical synthesis is required to link tRNA

molecules to the desired unnatural amino acid. For larger scale mutagenesis, the cells machinery is used to incorporate the unnatural amino acid.[24–26] To accomplish this, orthoganol tRNA tRNA synthetase pair is used to load an unnatural amino acid onto a tRNA in cells. To ensure site-specific incorporation, and Amber (TAG) stop codon is used, as it is the least common stop codon in the bacterial genome.[25] Synthetases capable of loading unnatural amino acids selectively onto a tRNA that decodes this amber codon have been developed for a variety of amino acids. By incorporating this amber codon at the position of mutation, these orthogonal pairs will incorporate the unnatural amino acid at that position instead of stopping, provided the amino acid is given to the cells, allowing for selective and site-specific mutation to an unnatural amino acid.

2.2.2 HP1 Mutants

These methods currently do not allow for the efficient and incorportation of fluorinated phenylalanine, as it is too similar to natural phenylalanine and native synthetases will recognize it, leading to global incorporation at any phenylalanine site. Site specific incorporation studies for fluorinated phenylalanine incorporation are in their initial stages.[27] Instead, the use of other para-substituted phenyalanines with common electron withdrawing groups were used to probe the cation-π interactions at the tyrosine positions (Figure 2.4).

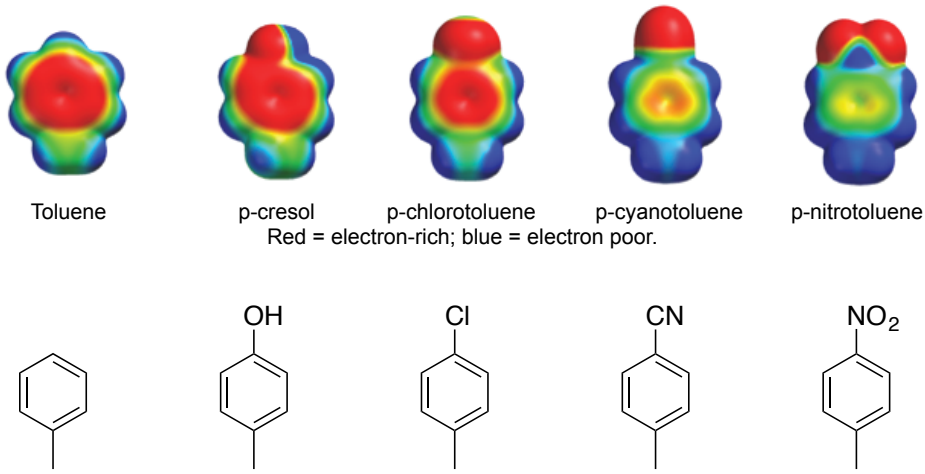


Figure 2.4. Electrostatic potential maps of some of the mutant amino acid residues incorporated into the binding pocket of HP1 by unnatural amino acid mutagenesis. Electrostatic potential maps were generated with MacSpartan: HF/6-31G*; isodensity value = 0.02; range = -50 (red, electron rich) to 50 kcal/mol (blue, electron poor).

Various mutations were incorporated into the binding pocket of HP1 in both of the tyrosine residue positions (Y24 and Y48) of the HP1 binding pocket by Stef Baril in the Brustad lab at UNC. The mutations and the calculated cation-π binding energies of the corresponding substituted benzene are summarized in Table 2.1.

Table 2.1. Mutations of HP1 chromodomain and the calculated cation-π binding energies of the substituent on benzene

Entry	Protein	Cation-π Energy[28] (kcal/mol)	Expected Mass	Observed Mass[a]
1	Wild Type	15.2	8569.30	8569.59
2	Y24pMeF	16.4	8567.46	8567.11
3	Y24F	15.9	8553.31	8553.60
4	Y24pCF_3F	7.2	8621.42	8622.13
5	Y24pCNF	3.4	8578.43	8578.98
6	Y24pNO_2F	1.6	8598.39	8598.80
7	Y48pMeF	16.4	8567.46	8568.10
8	Y48F	15.9	8553.31	8553.85
9	Y48pCF_3F	7.2	8621.42	8622.09
10	Y48pCNF	3.4	8578.43	8579.01
11	Y48pNO_2F	1.6	8598.39	8598.71

[a]Expressed by Stef Baril, Brustad lab and measured by LC-MS

2.3 Characterization of Structure

2.3.1 Circular Dichroism

Circular dichroism is a technique often used to monitor protein structure. Each type of secondary structure has a unique pattern of maxima and minima, and the CD spectra of HP1, as well as several mutants containing natural amino acid mutations have already been characterized.[3] The spectra have local maxima around 233 nm that is generally attributed to exciton coupling arising from aromatic interactions.[29,30] The minima at 209 nm and 222 nm are primarily reflective of alpha helical secondary structure.

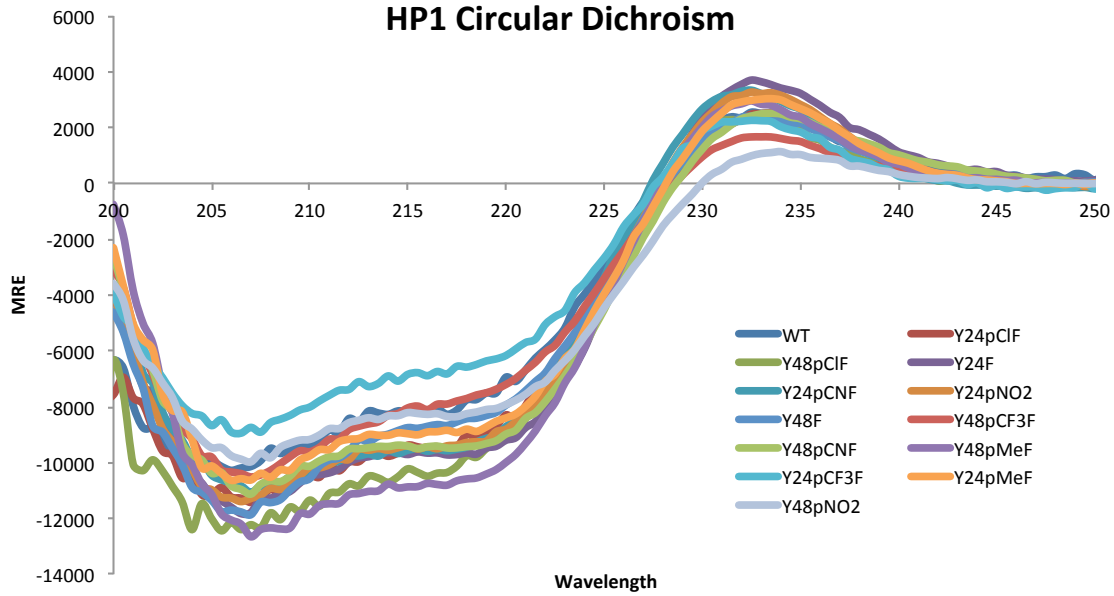


Figure 2.5. Circular dichroism of HP1 and all of the unnatural mutants used for probing cation-π interactions at tyrosine. Experiments done in 10 mM sodium phosphate buffer, pH 7.4, 2 mM DTT at 20°C

These CD structures of all the mutants demonstrate that the mutations did not significantly alter the folding of the chromodomain, and any observed effect we see on binding is not due to misfolded protein (Figure 2.5).

2.3.2 X-Ray Structures

Further evidence for this comes from x-ray analysis performed by Stef Baril in the Brustad lab. Crystal structures were solved for mutants Y24F and Y24NO_2 (Figure 2.6).

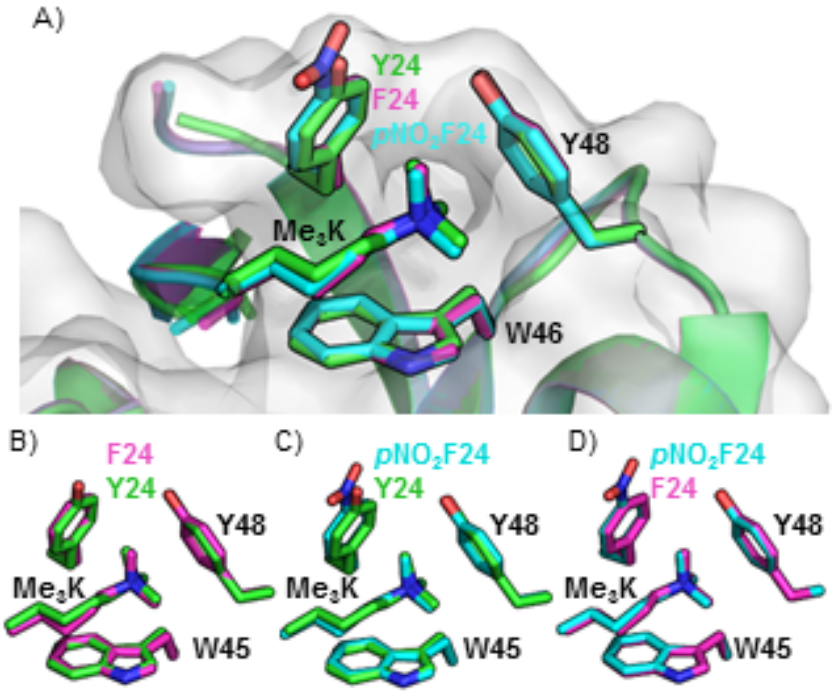


Figure 2.6. A) Crystal structure of Y24pNO_2F (cyan) and Y24F (magenta) overlayed with WT (green) indicating that binding pocket remains in tact with the mutated residue. B) overlay of Y24F and WT C) overlay of Y24pNO_2F and WT D) overlay of Y24pNO_2F and Y24F mutants.

The structures show the aromatic pocket remains in a similar position when either phenylalanine or the unnatural para-nitro amino acid is inserted in the Y24 position. The para-nitro substitution is both the largest mutation sterically, and the mutation that caused the largest disruption in binding, indicating that any of the effects observed in binding is not a result of any change in structure of the binding pocket. The overlay of the mutants with the WT binding pocket indicates very little perturbation of structure.

2.4. Binding Affinities of HP1 Mutants

2.4.1 Fluorescence Anisotropy

Initial binding measurements of HP1 and mutants were done by fluorescence anisotropy and based on previous work in the lab (Table 2.2).[3] This method measures relies on changes of polarization of light observed upon binding.[31] One binding partner is attached to a fluorophore, in this case the H3K9Me3 peptide. When the peptide is tumbling in solution feely, upon excitation with plane polarized light, the tumbling will cause the light emitted back to become scrambled. When HP1 is added to the solution, the binding interaction slows down this tumbling, and the light emitted back will retain some of the polarization. The amount of polarization remaining can be measured for different concentrations of HP1 and a binding curve can be generated and analyzed by curve fitting software. Initial measurements of wild type HP1 gave a Kd of 20 μM, comparable to those previously reported in the same conditions.[3]

Table 2.2. Fluorescence polarization binding data for wild type HP1 and mutants

Entry	Protein	σ_{meta}	K_d (μM)	ΔG (kcal/mol)
1	Wild Type	0.12	20±3	-6.4±0.1
2	Y24F	0	15±4	-6.6±0.1
3	Y24pClF	0.37	24±5	-6.3±0.1
4	Y24pCNF	0.56	43±7	-5.9±0.1

2.4.1.1 Linear Free Energy Relationships

Initial mutation studies were done with mutating position 24 as the yields of the mutants were much higher than position 48 (Stef Baril, unpublished results). Binding measurements of several mutant proteins followed the trend expected if the tyrosine residue was participating in a cation-π interaction. To analyze the data, it can be plotted against values that represent the

relative electronics of the ring. One such constant, σ_{meta} is derived from the relative amounts of deprotonation of benzoic acid with various substituents in the meta position. Electron withdrawing groups lead to a greater extent of deprotonation, and have a greater σ_{meta} value.[32] Cation-π interactions exhibit a linear free energy relationship when various substituted aromatics are plotted against σ_{meta} rather than the para position, indicating resonance effects do not contribute to the variation in strengths of cation-π interactions. When the HP1 mutants are plotted against σ_{meta} a clear linear free energy relationship was observed (Figure 2.7).

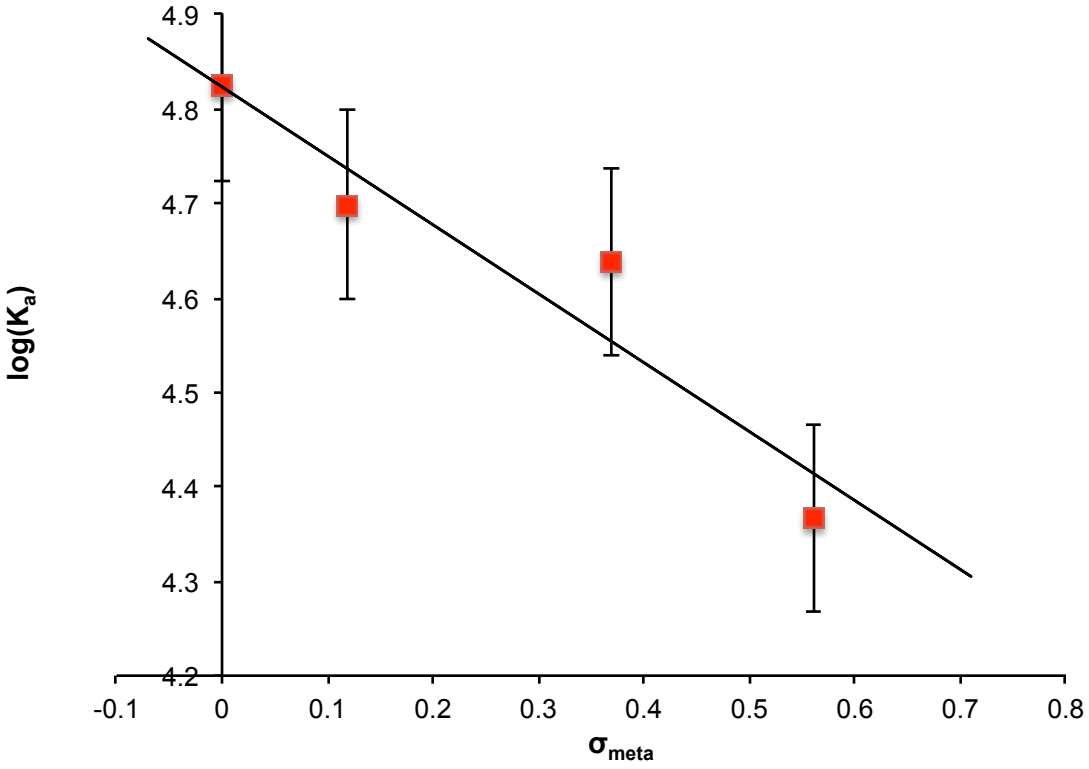


Figure 2.7. Linear free energy relationship between binding data of position 24 mutants of HP1. Slope = -0.73; $R^2 = 0.90$.

As subsequent binding experiments were performed using this method, batch to batch inconsistencies were observed. While the linear relationship remained similar, the actual binding

affinities fluctuated greatly and experiments performed on the different days could not be compared to each other.

Additionally, the same batch of protein gave relatively inconsistent data when several runs were done on the same day (Figure 2.8).

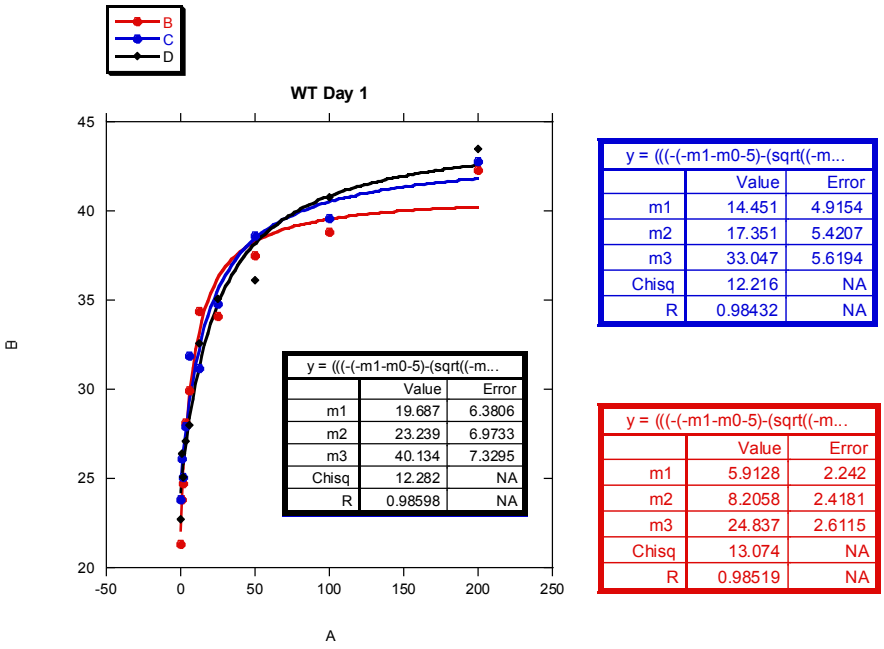


Figure 2.8. Three runs performed on the same batch of protein on the same day. The range of measured K_d values (shown in the table as m2) was 8 μM to 23 μM.

The inconsistencies with this method were attributed to inaccurate protein concentrations of the stock solutions and error in dilutions. This led to the exploration of other methods that do not rely on accurate protein concentration for calculation of binding constants.

2.4.2 Isothermal Titration Calorimetry

Isothermal titration calorimetry (ITC) has previously been used to measure binding of HP1 to H3 peptides as well.[2] ITC is performed by titration of the peptide into a cell containing a solution of protein. The instrument measures the heat released by with each injection. Consecutive injections are performed until the protein saturates. The heat released can be plotted

relative to molar equivalents of titrant to give a binding isotherm, directly measuring K_d and ΔH, which can be used to calculate ΔS, independent of protein concentration. Using this method, binding measurements to $H3K9Me_3$ were performed.

This time, batch to batch consistency was observed, and runs from different batches, and different days could be compared (Figure 2.9). Binding data from this method is summarized in Table 2.3.

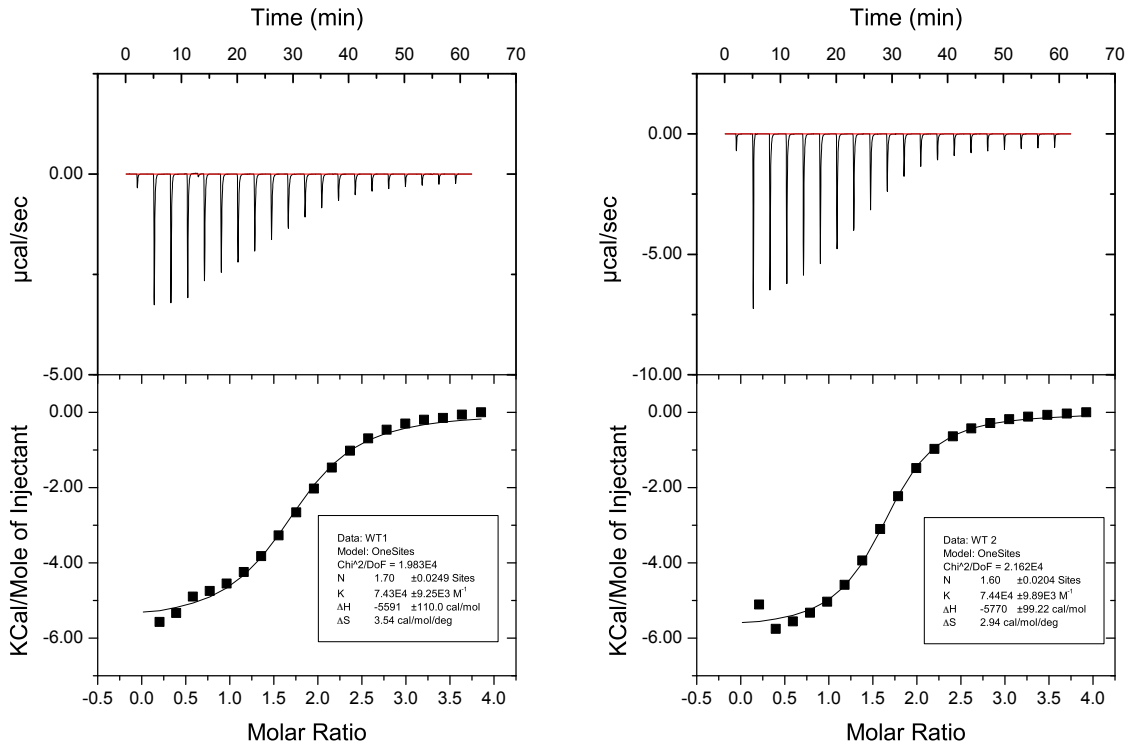


Figure 2.9. Two runs of WT HP1 with $H3K9Me_3$ on two different days with two different batches of protein.

Table 2.3. Binding constants for HP1 mutants as measured by ITC

Entry	Protein	Cation-π Energy[28] (kcal/mol)	K_d (μM)[a]	ΔG[a] (kcal/mol)	C Value[b]	[M]:[L][d]
1	Wild Type	15.2	14±1	-6.58±0.07	17.9	1:20.0
2	Y24pMeF	16.4	18±0.2	-6.44±0.1	13.5	1:12.3
3	Y24F	15.9	19±0.2	-6.41±0.06	8.4	1:18.75
4	Y24pClF	N/A	22±0.4	-6.32±0.06	11.4	1:19.8
5	Y24pCF_3F	7.2	54±0.4	-5.79±0.07	4.6	1:20.0
6	Y24pCNF	3.4	103±14	-5.42±0.1[c]	2.4	1:20.0
7	Y24pNO_2F	1.6	141±11	-5.23±0.16	1.4	1:25.6
8	Y48pMeF	16.4	17±1	-6.45±0.04[c]	13.3	1:20.7
9	Y48F	15.9	17±2	-6.47±0.09[c]	14.7	1:20.0
10	Y48pClF	N/A	19±2	-6.41±0.06[c]	15.2	17.1
11	Y48pCF_3F	7.2	25±2	-6.24±0.1	10.0	1:20.0
12	Y48pCNF	3.4	43±4	-5.93±0.07[c]	5.5	1:21.0
13	Y48pNO_2F	1.6	62±6	-5.71±0.1	4.1	1:19.8

[a]Values are an average of 3 runs unless otherwise noted. Errors calculated from error values given by fitting software. One-site binding used to fit data. [b]Calculated using the formula $c = \frac{[M]}{k_d}$ where [*M*] is the concentration of protein. N=1 is assumed. [c]Average of 2 runs. [d]Ratio of concentration of protein to ligand. Higher amounts of H3 ligand needed for accurate curve fitting of runs with lower C values (weaker binding proteins).

2.4.2.1 Linear Free Energy Relationships

Mutation studies at both positions revealed a linear relationship between σ_{meta} and $\log(K_a)$, indicating that both of these positions are participating in a tunable cation-π interaction. Interestingly, postion 24 shows a greater effect than position 48, implying that the cation-π interaction at that residue is stronger (Figure 2.10).

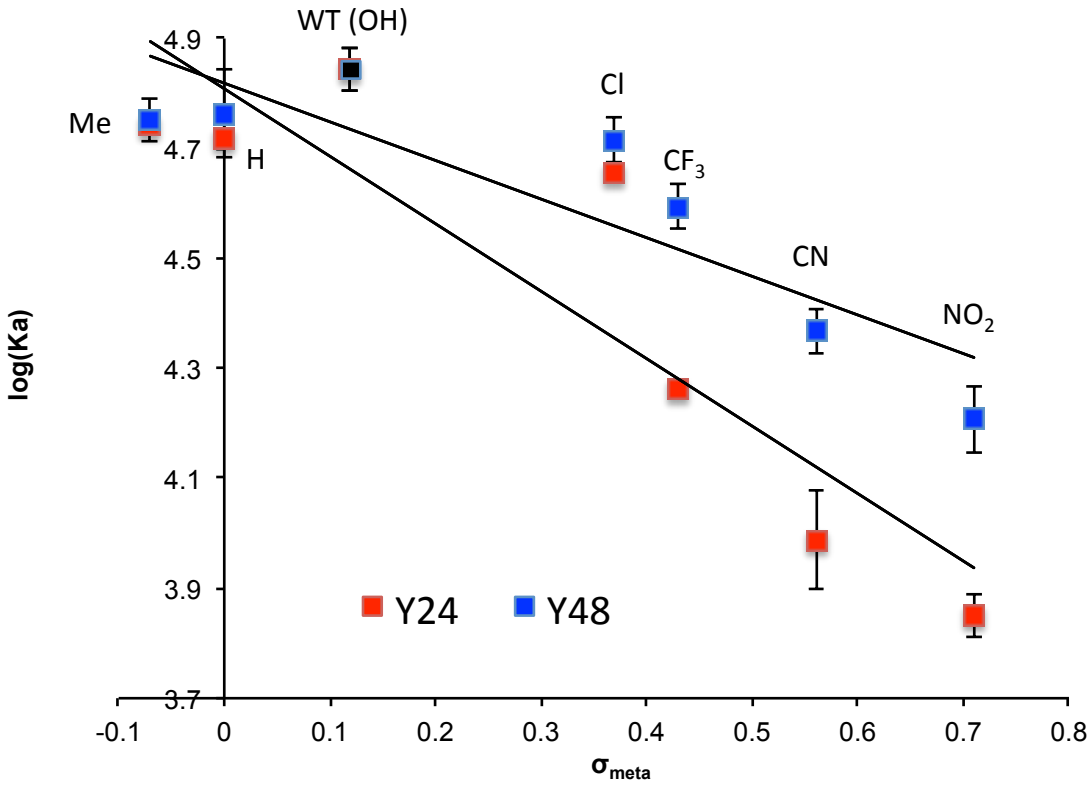


Figure 2.10. Plot of $\log(K_a)$ vs σ_{meta} values.[32] Y48: slope = -0.703, $R^2 = 0.772$; Y24: slope = -1.23, $R^2 = 0.812$.

While σ_{meta} has been shown to provide a good rough correlation for cation- π interactions, typically calculated gas phase cation- π energies have been used to demonstrate cation- π energies in linear free energy relationships.[6,21] Using values calculated from electrostatic potential maps (Figure 2.11), as well as DFT calculations (Figure 2.12) of gas phase interactions between Na+ and substituted benzene,[28] an even stronger correlation was observed. Again, position 24 displays a greater reduction in binding affinity upon incorporation of the electron poor aromatic residues, with a slope about two fold less than the slope observed for position 48.

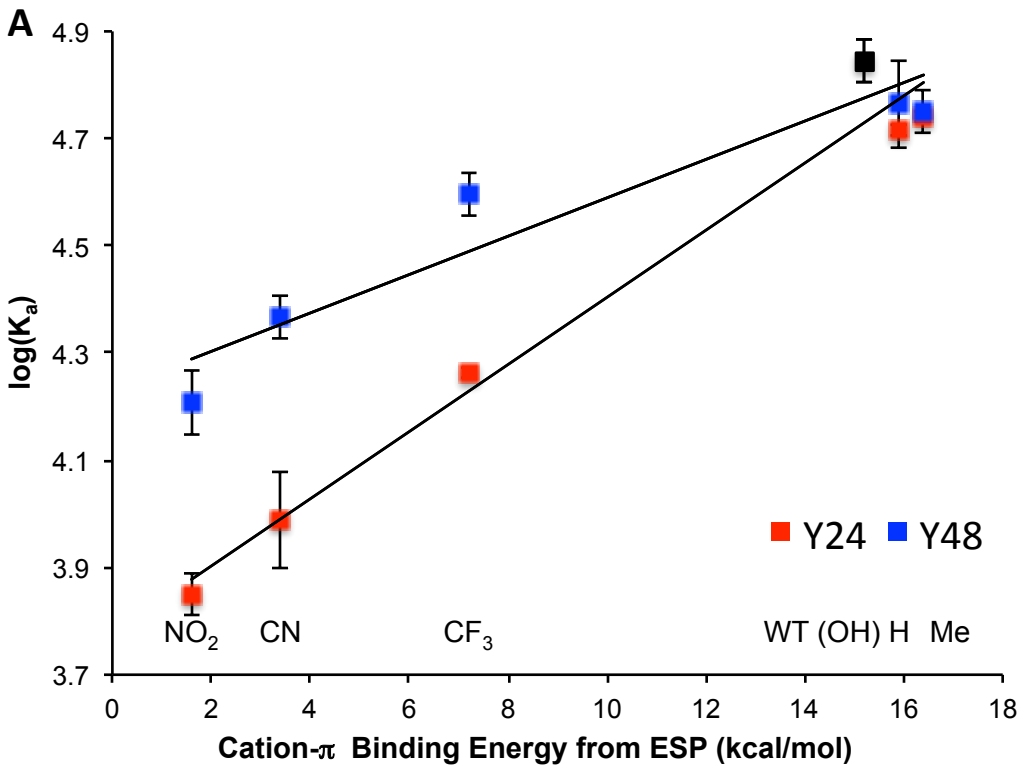


Figure 2.11. Relationship between $\log(K_a)$ of H3K9Me$_3$/HP1 interactions and cation-π binding energies calculated using electrostatic potential maps.[28] Y24: slope = 0.063, R^2 = 0.975; Y48: slope = 0.036, R^2 = 0.910.

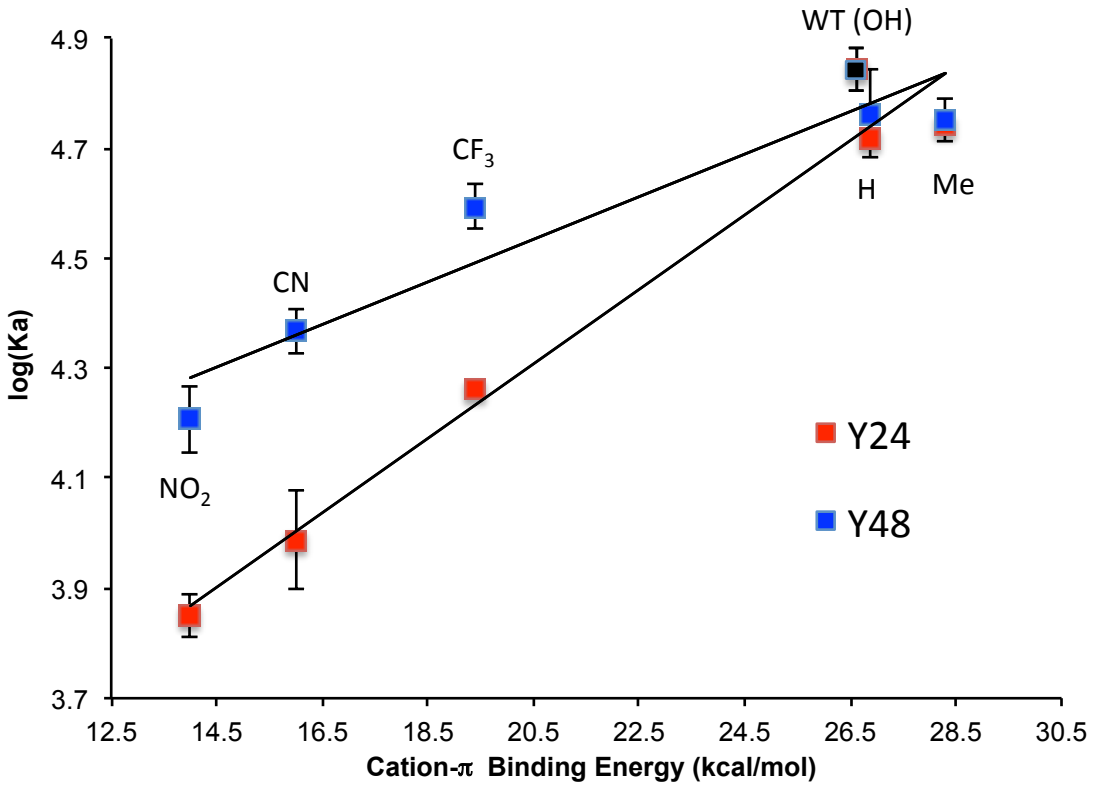


Figure 2.12. Plot of log(K_a) vs. cation- π gas phase binding energies between Na^+ and C6H5X calculated by Wheeler and Houk (Eint(C5H6X)).[28] Y48: slope = 0.0387, R^2 = 0.907; Y24: slope = 0.0677, R^2 = 0.971.

2.4.2.2 Polarizability

The data was also plotted against molar refractivity (MR) which is a measure of polarizability of the substituents (Figure 2.13).[33] The molar refractivity depends on the overall polarization of a sample. By using high frequencies of light, the permanent dipoles of the sample cannot rearrange quickly, and the polarization is due to induced polarization from the electric field of the light. Therefore, MR is directly related to polarizability and frequently used as a constant to represent differences in polarizability. Increased polarizability is associated with increased ability for the substituent to participate in van der Waals interactions. As cation-π interactions include a VDW component, it is possible that observed effects of mutations could be due to changes in polarizability. While there does appear to be a small correlation between the

polarizability of the substituents and their binding affinity, it is in the opposite direction than would be expected. If these interactions were contributing significantly, binding affinities are expected to increase as this parameter increases, while the observed relationship is a decrease in affinity as polarizability increases.

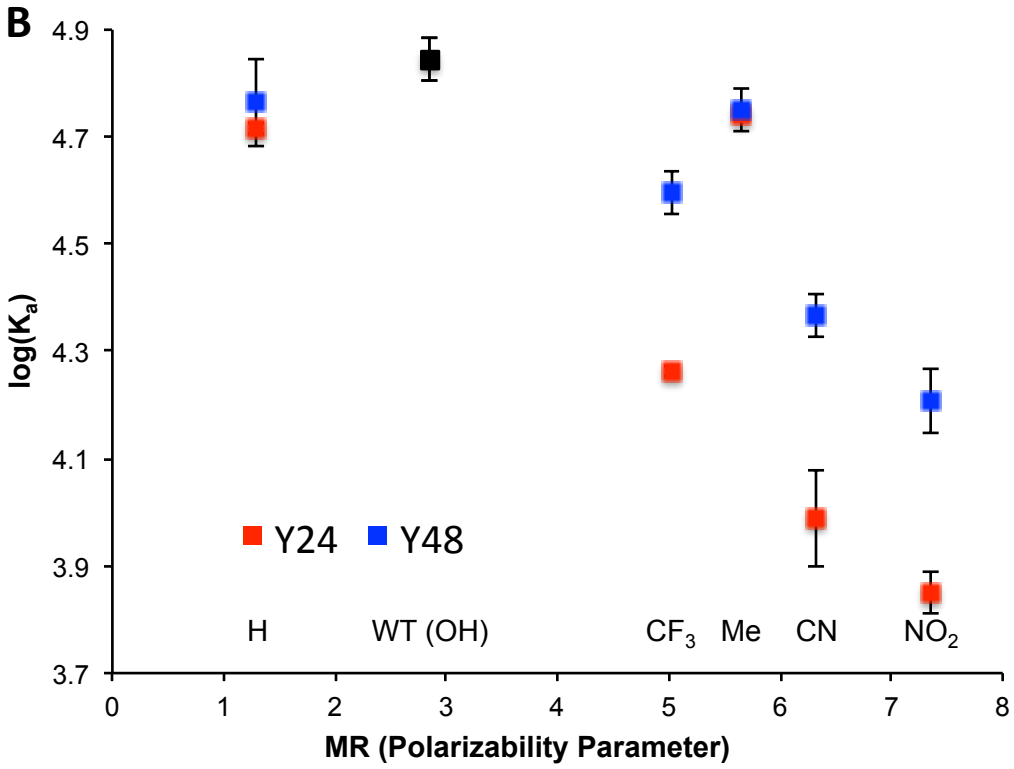


Figure 2.13. Relationship between log(K_a) and polarizability parameter (MR) for Y24X (red) and Y48X (blue) mutants.

2.4.2.3 Hydrophobicity

Another noncovelent interaction that could be contributing to this interaction is the hydrophobic effect. The π parameter is often used as a measure of hydrophobicity. The π parameter is based on Log P, another parameter frequently used as a measure of hydrophobicity, scaled to benzene as 0.[34] The P parameter is the partition coefficient of a compound between octanol and water. The more hydrophobic the compound, the greater it partitions into the

octanol, and has a higher Log P (and π) value.[35] There is no significant correlation between the data and π values, indicating changes in the hydrophobic effect are not responsible for the variation in binding affinities with variation of the substituent (Figure 2.14).

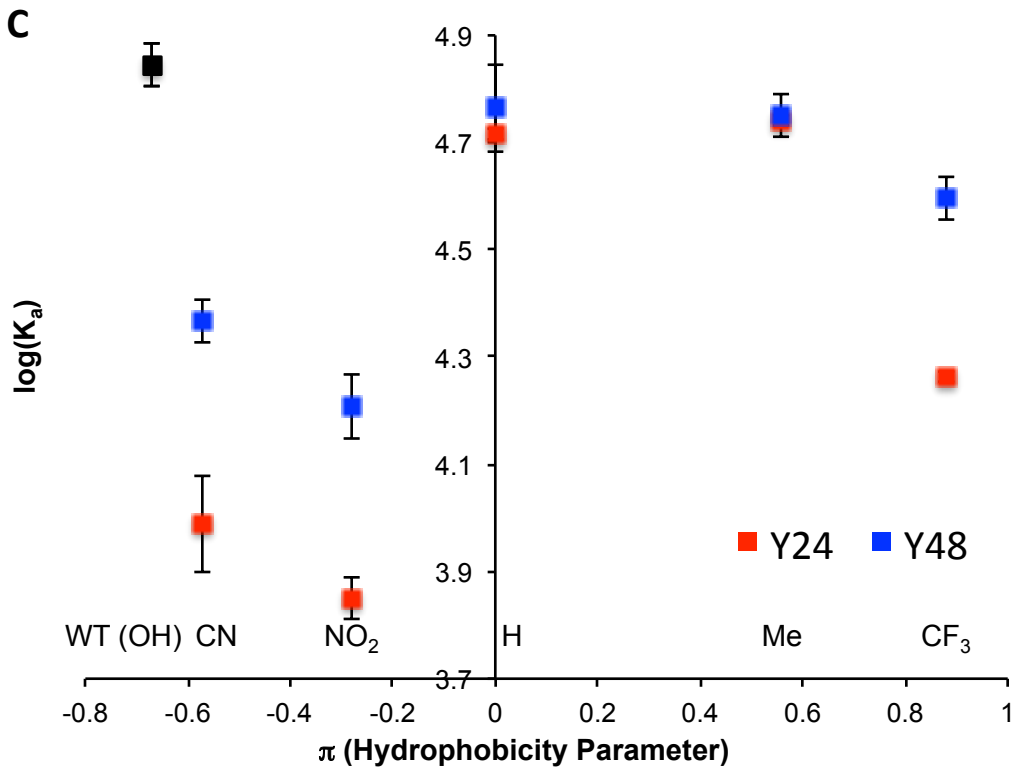


Figure 2.14. Relationship between log(K_a) and hydrophobicity parameter (π).

This data, taken together, strongly indicates the presence of a tunable cation-π interaction between both tyrosine positions of HP1 as the main driving force for this interaction. Moreover, this demonstrates the first example in which differences in magnitude of two cation-π interactions within the same protein binding pocket have been measured using direct binding experiments.

2.4.3 Surface Plasmon Resonance

This trend was confirmed by a second method, using surface plasmon resonance. In this technique, protein is loaded onto the surface of a chip, while a solution of peptide binding partners of different concentrations are flown over the immobilized protein for a set injection time. The amount of material loaded onto the chip effects the reflection angle of light shined on the chip. The amount this angle changes upon binding to the peptide at different concentrations flowing over the surfaced can be used to calculate a binding constant. This method uses significantly less material that ITC as has the potential for a more high throughput study of different peptides. To confirm that the trend was the same, the mutants at position 48 were used and their binding to H3K9Me$_3$ measured (Table 2.4). The data shown was analyzed using either kinetic analysis or equilibrium analysis by the instrument software. Kinetic analysis calculates the on (K_{on}) rates from the beginning of the injection and the dissociation rate (K_{off}) after the injection ends to then calculate binding affinities. Equilibrium analysis plots the raw data for the different concentrations once the binding is constant to create a binding curve and calculate a dissociation constant. Data shown is from the kinetic analysis. While the overall magnitude of the binding constants were different, this can likely be attributed to the different buffer conditions needed to perform the experiments.

Table 2.4. Binding constants of HP1 mutants measured by SPR

Entry	Protein	Cation-π Energy[28] (kcal/mol)	K_d (μM)[a]
1	Wild Type	15.2	0.89±0.02
8	Y48F	15.9	1.5±0.02
9	Y48pCF_3F	7.2	2.0±0.02
10	Y48pCNF	3.4	10[b]
11	Y48pNO_2F	1.6	5.7±0.04

[a]Average of 3 runs. Error calculated using standard deviation. [b]average of 2 runs; no error calculated.

2.4.3.1 Linear Free Energy Relationships

Despite the different in binding constants, the overall trend and linear relationship remained the same (Figure 2.15).

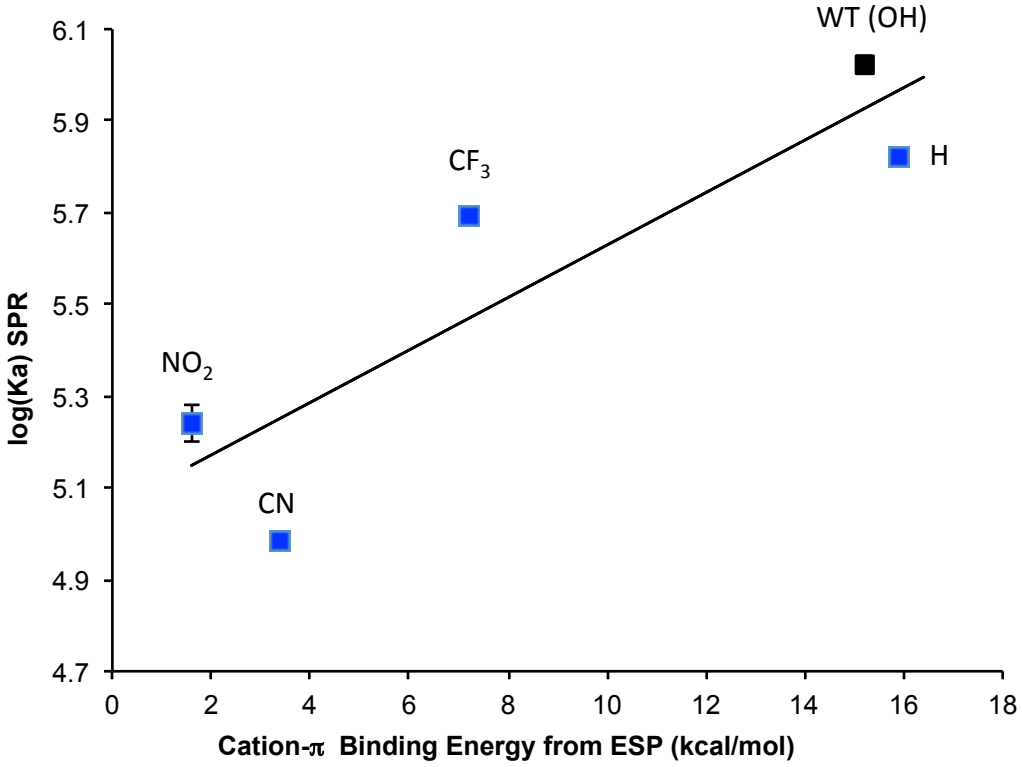


Figure 2.15. SPR data of Y48 HP1mutants plotted against binding energies calculated from electrostatic potential maps.[28] Errors calculated from standard deviation. CN data point in duplicate, the rest in triplicate. Difference in binding constants attributed to different buffer

conditions required for method. Slope = 0.0572, $R^2 = 0.781$.

2.5 Discussion

Tryptophan is the aromatic residue most often seen in cation-π interactions[6] and also the residue that Dougherty has found to be most important in the cation-π recognition of acetylcholine,[8,9] so it was unclear what kind of effect mutation of tyrosine would have on this interaction. Interestingly, not only do both tyrosine residues participate in a tunable cation-π interaction, the data suggests that position 24 is interacting more strongly with the cation than position 48. X-ray structural analysis clearly indicates that differences do not arise from a change in structure.

By examining the crystal structure of the protein, there are some observations supporting these conclusions (Figure 2.16). Position 24 appears to be centered over two of the partially positive methyl groups as well as the methylene, while position 24 appears to be centered over just one methyl group. In a cation-π interaction with trimethyllysine, the alkyl groups are typically packed against the aromatic face, as the partial positive charge lies there, and not on the nitrogen atom.[6]

As this is where the actual partial positive charges of the cation lie, the closer contacts with more of the alkyl groups would be expected to provide a stronger interaction. The actual distances observed in the crystal structure confirm this observation, with position 24 having more close contacts than position 48 (Figure 2.17).

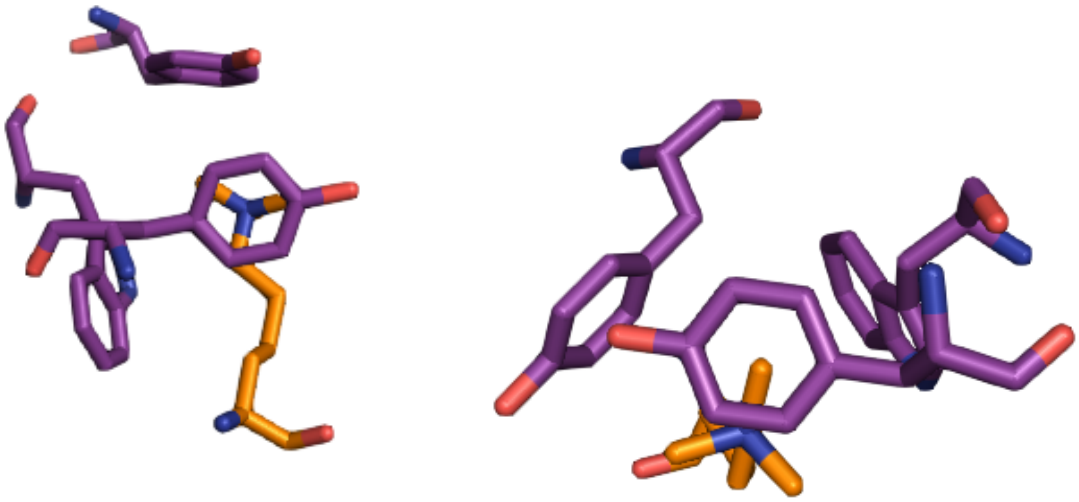


Figure 2.16. Top view of Y24 (left) centered over 2 methyl groups and one methylene compared to top view of Y48 (right) centered over one methyl group.

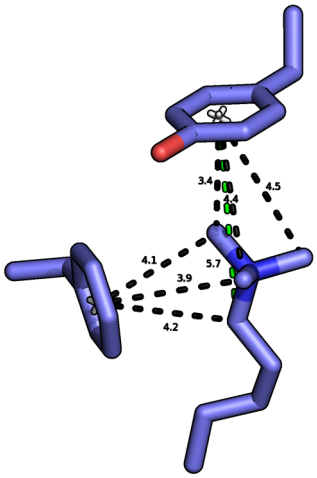


Figure 2.17. Distances between tyrosine residues and methyl groups and methylene of trimethyllysine.

While there have been several systems studied by this same approach, the vast majority of them are functional measurements (either enzyme activities or EC_{50} values) rather than actual binding energy measurements. Thus, there are not many systems with which to compare the relative decrease in binding affinities. One example measures substituent effects in organic solvent, which interestingly resulted in a similar $\Delta\Delta G$ to that observed for the nitro substituent in the Y24

mutations (1.4 kcal/mol).[36] Many of the studies involving ΔG measurements involve removing the cation, often resulting in a loss of about 2.5 kcal/mol in binding affinity, which is also similar to what is observed for HP1.[5] Other studies have estimated the strength of a cation-π interaction at around 0.5-1 kcal/mol/ring.[37] The measured ΔG values measured are in a similar range as other methods used to quantify cation-π interactions, however, this is the first time different magnitudes of cation-π interactions within the same binding pocket have been directly measured.

This is the first example of tuning of cation-π interactions in histone reader proteins. As many share a similar aromatic box motif in their binding pocket, further studies can reveal whether this is a trend common among other reader proteins, or unique to HP1. As some reader proteins have only a partial formed aromatic box that completely forms upon binding to KMe_3, they may show different dependences or non-linear relationships, allowing for the potential to provide further insights into the binding mechanism of this class of proteins. Detailed understanding of the binding mechanism for this class of proteins provides valuable information for probe design of these emerging therapeutic targets. By understanding how a protein recognizes its natural substrate with the necessary affinity and selectivity for biological processes, we provide a new framework for the design of probes with the necessary affinity and selectivity for therapeutic use. Additionally, computational modeling has become a tool for more efficient drug design, this highlights the need for accurately modeling cation-π interactions. This system can provide a useful tool for validating such methods.

2.6 Experimental

2.6.1 Peptide Synthesis.

2.6.1.1 H3 1-15 for ITC

H3K9Me3 (ARTKQTARK(Me)$_3$STGGKAY) was synthesized by hand using Fmoc protected amino acids and Rink Amide resin on a 0.06 mmol. The amino acid residues were activated with HBTU (O-benzotriazole-N, N, N', N',-tetramethyluronium hexafluorophosphate) and HOBT (N-hydroxybenzotriazole) in the presence DIPEA (diisopropylethylamine) in DMF (N,N-dimethylformamide). Double couplings of 30 minutes each were used. Deprotections were carried out in 20% piperidine in DMF, twice for 15 minutes each.

Dimethyllysine was coupled using 2 equivalents of Fmoc-Lys(Me)2-OH□HCl for 5 hours. Immediately after coupling, the resin was washed with DMF and the residue was methylated to form trimethyllysine with 7-methyl- 1,5,7-triaza-bicyclo[4.4.0]dec-5-ene (MTDB, 10.8 uL, 0.075 mmol) and methyl iodide (37.4 uL, 0.6 mmol) in DMF (5mL) for 5 hours. The resin was washed with DMF and peptide synthesis was continued as normal.

Cleavage from the resin was performed in 95:2.5:2.5 trifluoroacetic acid (TFA):water:triisopropylsilane (TIPS) for 4 hours. The TFA was evaporated and products were precipitated with cold diethyl ether. The peptides were extracted with water and lyophilized. Peptides were purified by reversed phase HPLC using a C-18 semipreparative column and a gradient of 0 to 100% B in 60 minutes, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. The purified peptides were lyophilized. The peptide was desalted for ITC using a Sephadex G-24 column from GE in water and lyophilized to a powder and identity was confirmed by MALDI mass spectrometry. Calculated: 1765.02 Observed: 1765.95.

2.6.1.2 H3 3-15 for Anisotropy

H3K9Me3 (FAM-TKQTARK(Me)$_3$STGGKA) peptide containing N-terminal 5,6-carboxyflurosceine (FAM) was synthesized for fluorescence anisotropy using the same procedure. Prior to cleavage from the resin, FAM was coupled to the N-terminus using 2 eq FAM, 2 eq PyBOP, 2 eq HOBT, and 4 eq DIPEA. The coupling was bubbled with N_2 overnight in a foil wrapped reaction flask and kept in the dark. Cleavage from the resin and purification were performed as previously described. Peptide was used for experiments without desalting, and concentration was determined using UV-Vis absorbance of FAM at 492 nm (ε = 78000 $M^{-1}cm^{-1}$).

2.6.2 Circular Dichroism of HP1 Mutants

CD experiments were performed using an Applied Photophysics Chiroscan Circular Dicroism Spectrophotometer. Spectra were obtained with 30 μM chromodomain in 10 mM sodium phosphate buffer, pH 7.4 with dithiothreitol (DTT) 2 mM at 20°C. All scans were corrected by subtracting the buffer. The mean residue ellipticity was calculated using the equation $\theta = \frac{signal}{10lc}\frac{1}{r}$ where θ is MRE, signal is CD signal, l is path length, c is protein concentration, and r is the number of amino acid residues.

2.6.3 Fluorescence Polarization Binding Measurements

Fluorescence polarization measurements were performed using a PolarStar omega plate reader by BMG Labtech. Chromodomain solutions of different concentrations were made by performing serial half dilutions of a concentrated stock. Chromodomain (25 μL) was added into 1 μM FAM labeled peptide (25 μM) in a 96 well plate in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, and 2 mM DTT). Samples were allowed to equilibrate at 25°C for 30 minutes. The

samples were analyzed using an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The data were analyzed in kaleidegraph using the equation:

$$r = \left[\left(\frac{p + [c] + K_D \pm \sqrt{(-p-[c]-K_D)^2 - 4(p \cdot [c])}}{2p}\right) \cdot (r_\infty - r_0)\right] + r_0$$

where r is fluorescence anisotropy, r_0 is the anisotropy of the free histone tail, r_∞ is the anisotropy of the fully bound histone tail, ρ is the total concentration of histone peptide, |c| is the total concentration of chromodomain, and K_d is the dissociation constant. The variables r_0, r_∞, and K_d were all treated as floating variables.

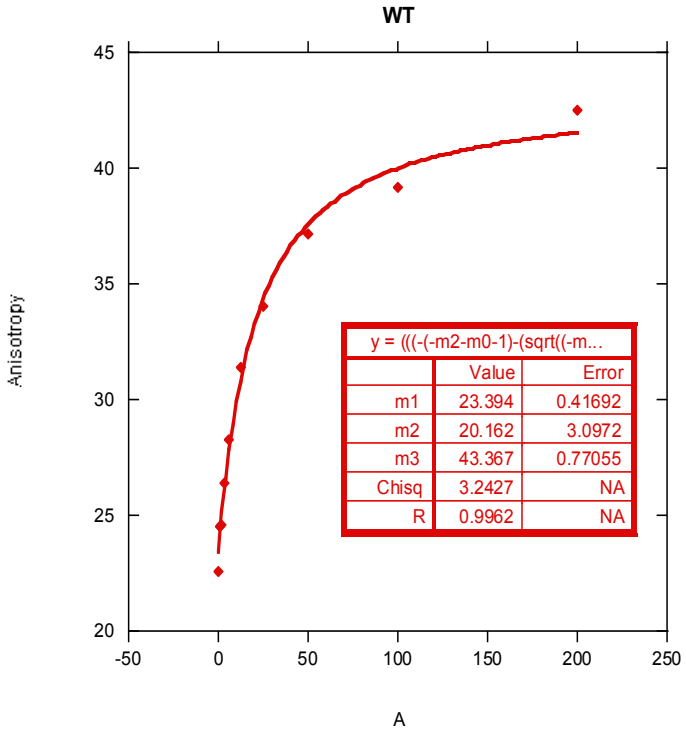


Figure 2.18. Binding curve of Wild Type HP1 titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8, 25 nM NaCl, 2 mM DTT.

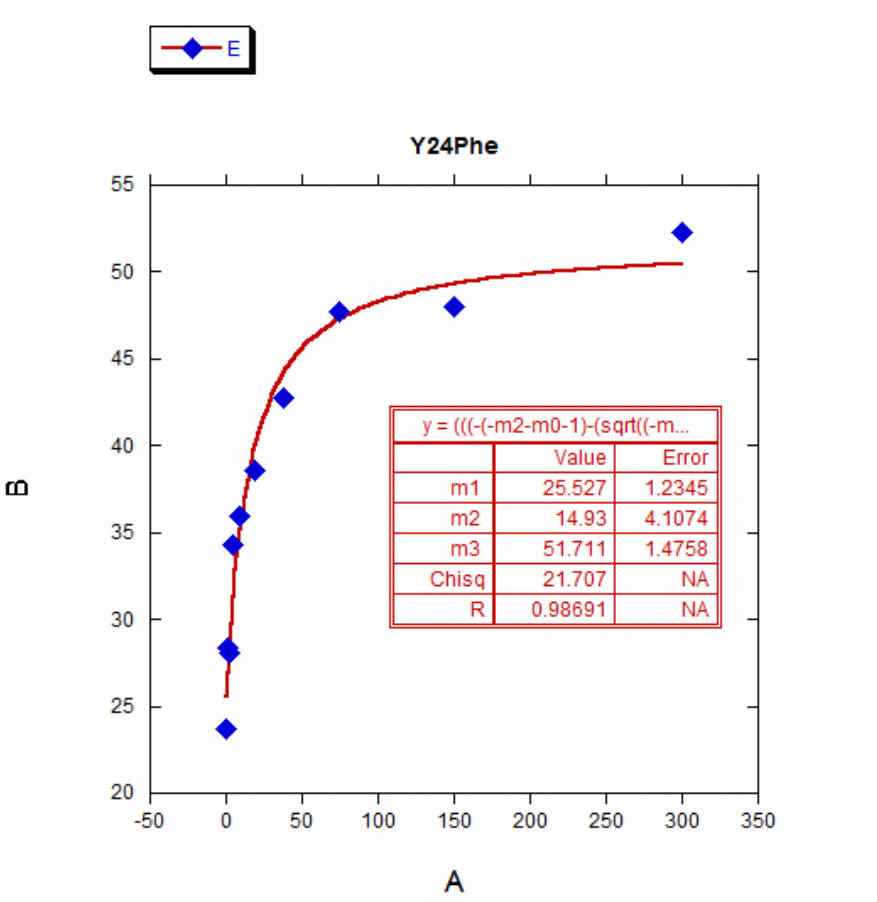


Figure 2.19. Binding curve of HP1 Y24F titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8, 25 nM NaCl, 2 mM DTT.

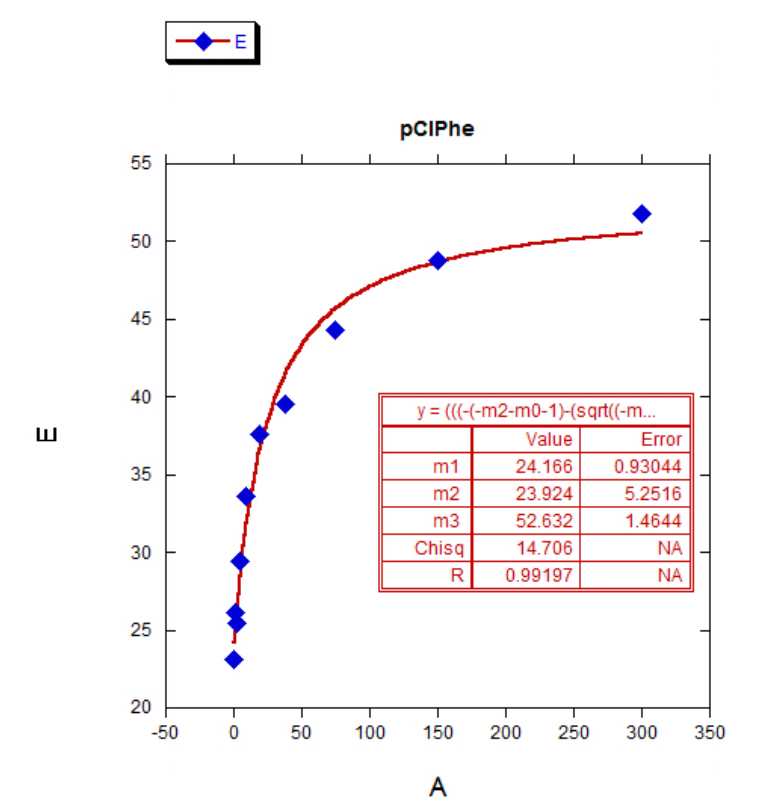


Figure 2.20. Binding curve of HP1 Y24pClF titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8, 25 nM NaCl, 2 mM DTT.

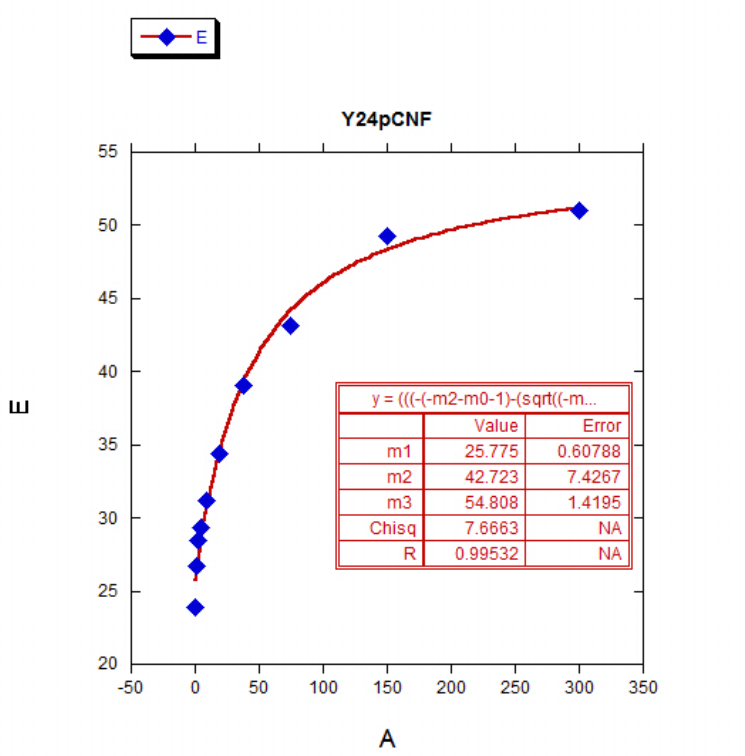


Figure 2.21. Binding curve of HP1 Y24pCNF titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8, 25 nM NaCl, 2 mM DTT.

2.6.4 Isothermal Titration Calorimetry Binding Measurements.

ITC experiments were performed by titrating peptide (2.5-5 mM) into protein (160-250 μM) in 50 mM sodium phosphate buffer, pH 7, 25 mM NaCl, 2 mM TCEP at 25°C using a Microcal AutoITC200. Peptide concentrations were determined by measuring the UV-Vis absorbance at 280 nm, using a NanoDrop2000 with xenon flash lamp, 2048 element linear silicon CCD array detector and 1 mm path length. Heat of dilution subtractions were made using the endpoint. Data was analyzed using a one-site binding model in supplied Origin software. Data shown is average of 3 runs unless otherwise noted. Error is calculated from standard deviation unless otherwise noted.

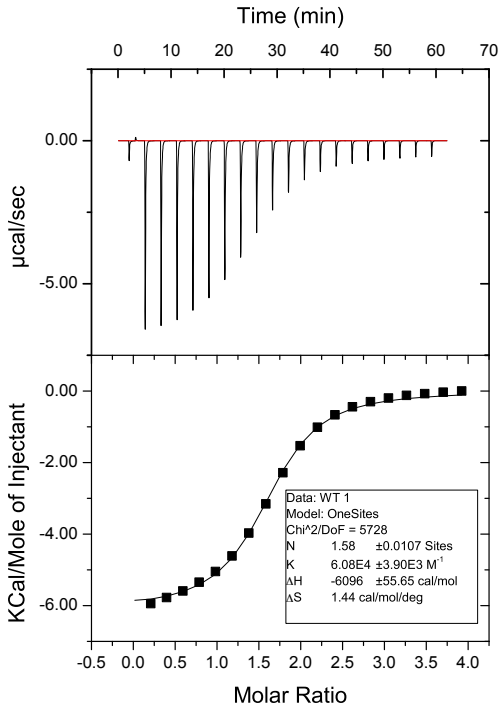


Figure 2.22. Wild type: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 5 mM) titrated into protein (250 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

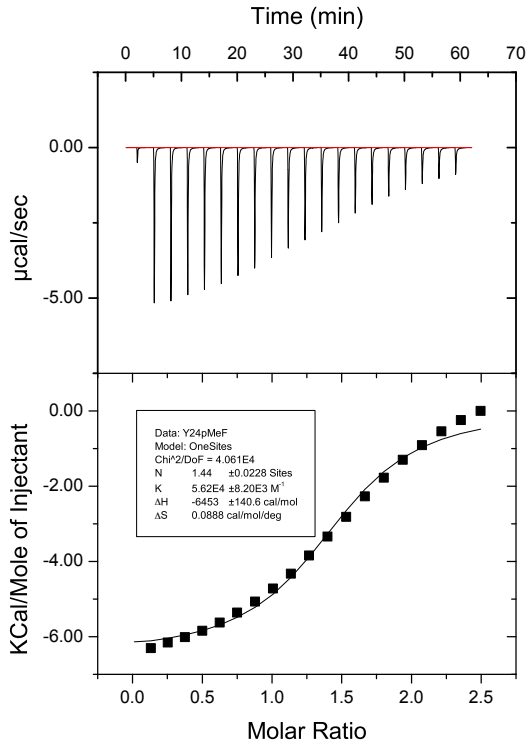


Figure 2.23. Y24pMeF: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 3 mM) titrated into protein (243 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

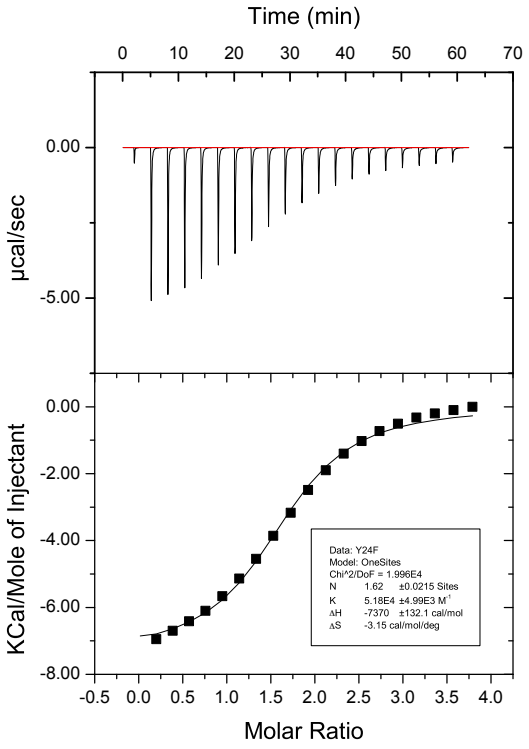


Figure 2.24. Y24F: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 3 mM) titrated into protein (160 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

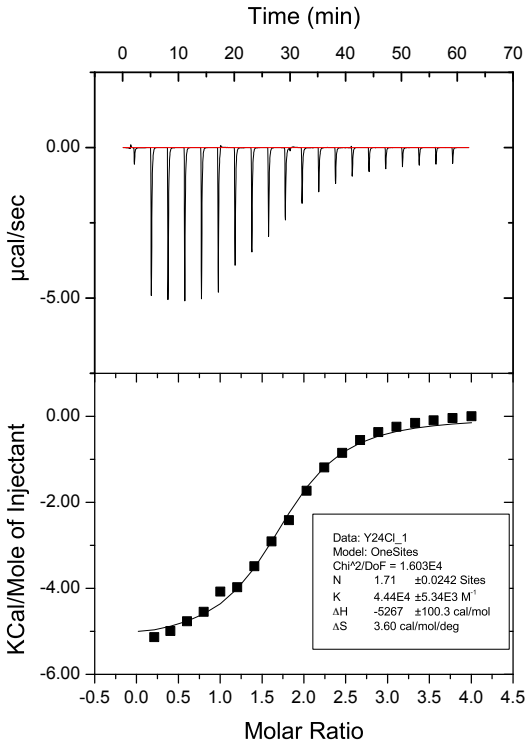


Figure 2.25. Y24pClF: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 4.95 mM) titrated into protein (250 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

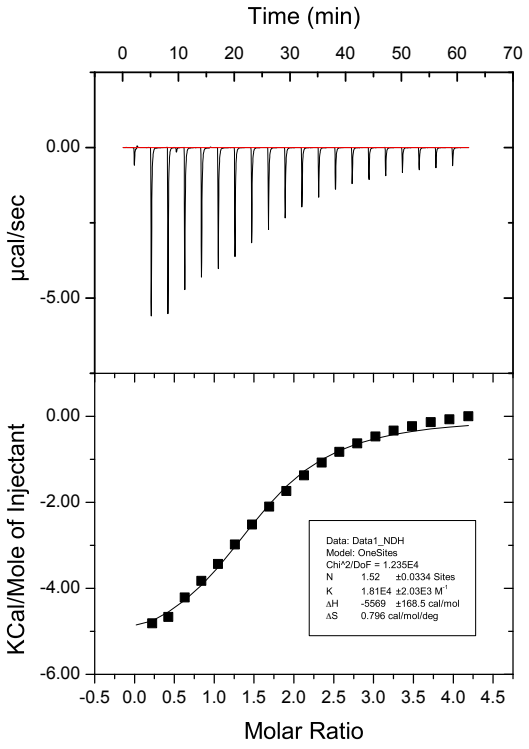


Figure 2.26. Y24pCF$_3$F: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 5 mM) titrated into protein (250 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

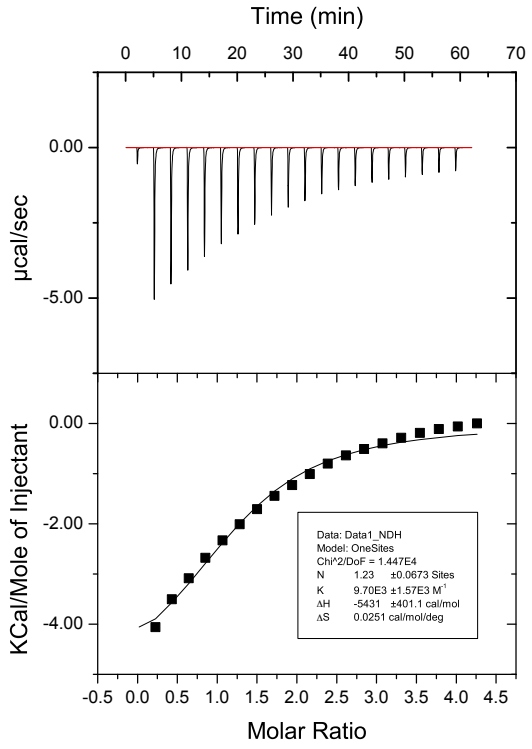


Figure 2.27. Y24pCNF: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 5 mM) titrated into protein (250 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

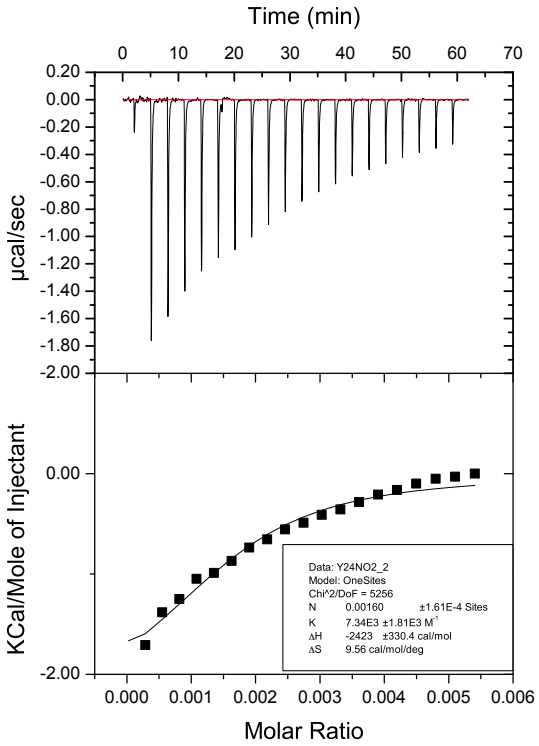


Figure 2.28. Y24pNO$_2$F: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 5 mM) titrated into protein (195 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

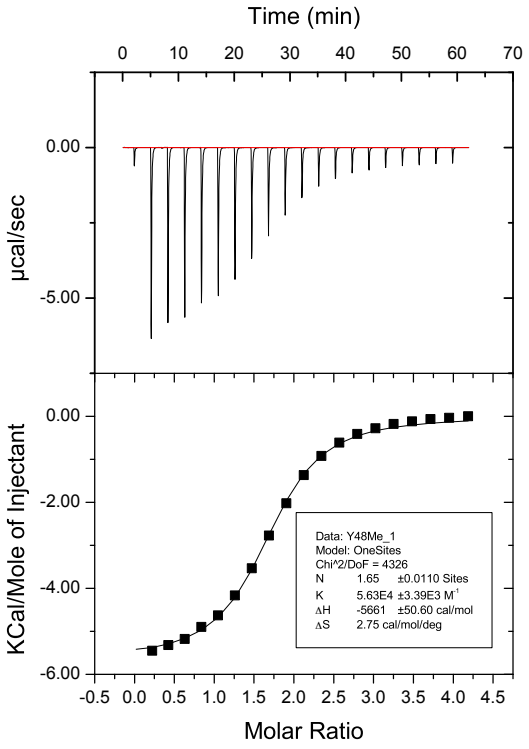


Figure 2.29. Y48pMeF: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 4.95 mM) titrated into protein (239 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

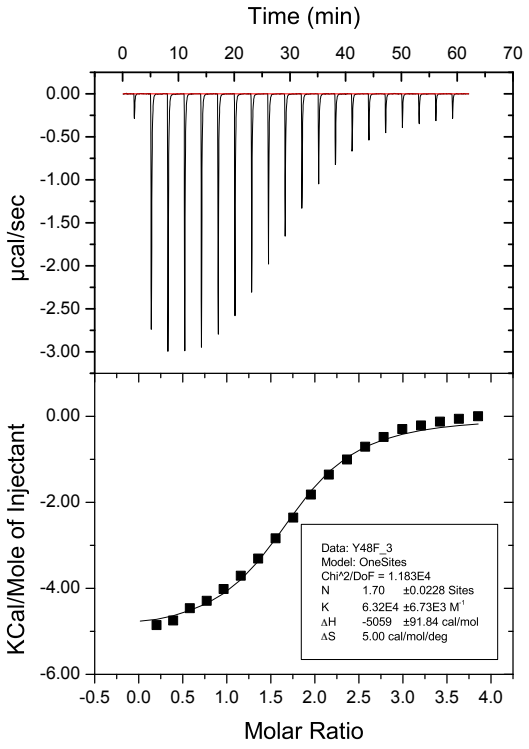


Figure 2.30. Y48F: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 5 mM) titrated into protein (250 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

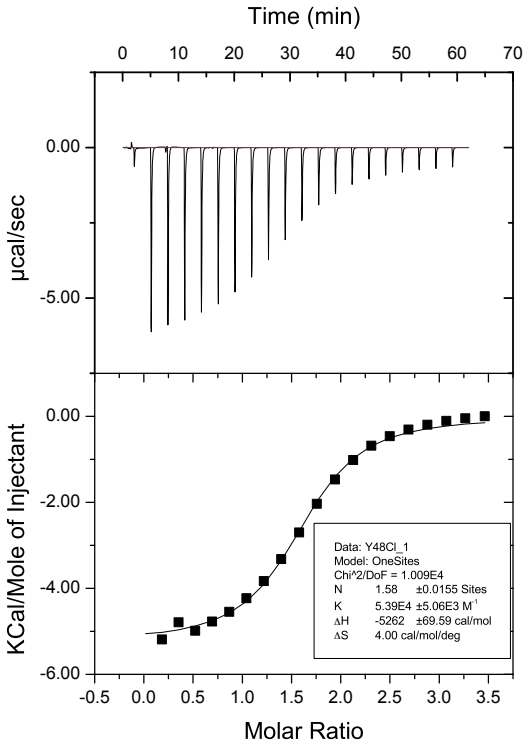


Figure 2.31. Y48pClF: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 4.95 mM) titrated into protein (289 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

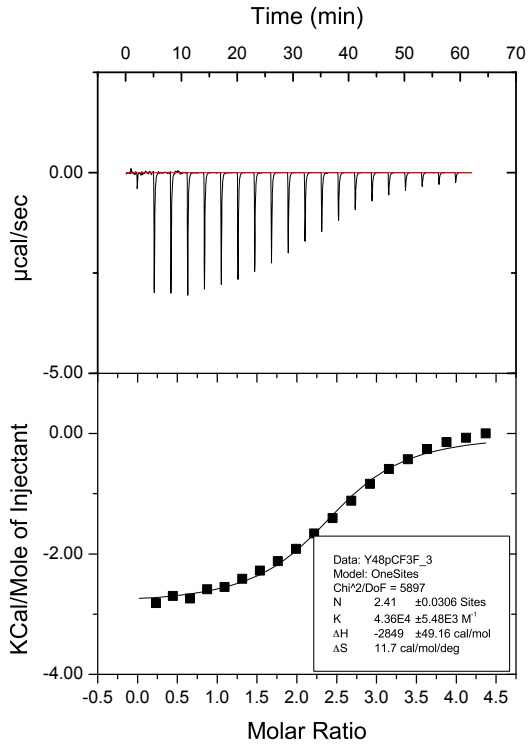


Figure 2.32. Y48pCF$_3$F: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 5 mM) titrated into protein (250 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

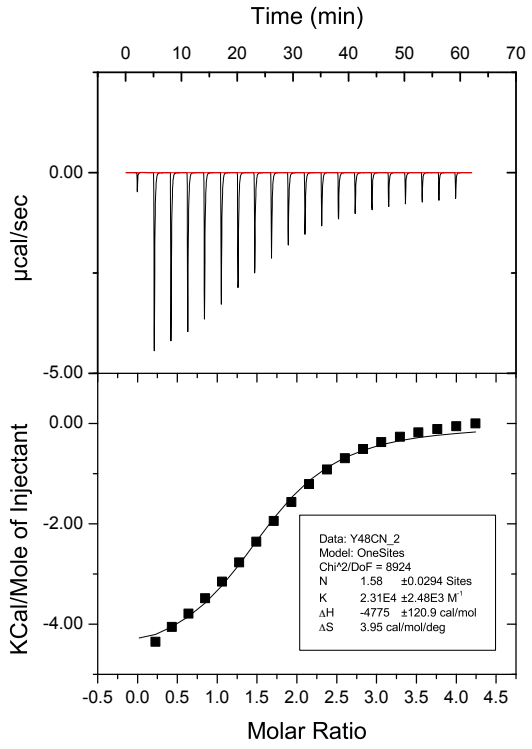


Figure 2.33. Y48pCNF: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 4.95 mM) titrated into protein (236 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

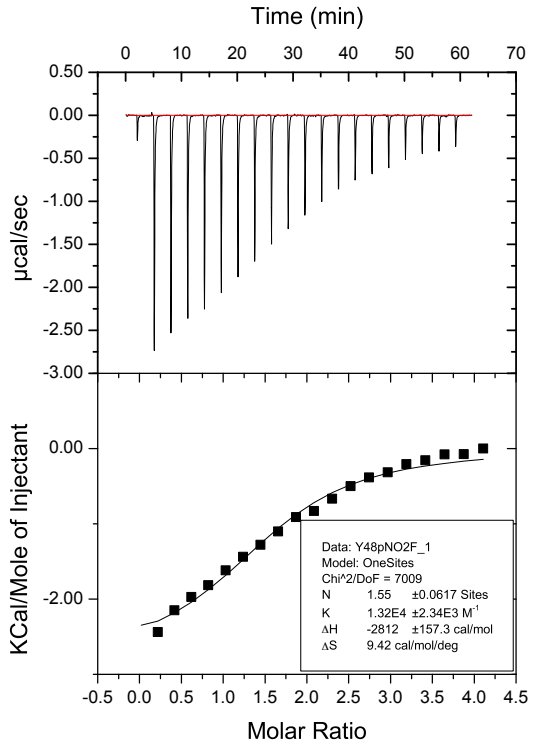


Figure 2.34. Y48pNO$_2$F: H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY; 5 mM) titrated into protein (252 µM) at 25°C in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, 2 mM TCEP fit to one-site binding.

2.6.5 Surface Plasmon Resonance

Binding measurements were performed on Proteon XPR36 surface plasmon resonance (SPR) based biosensor. His-tagged Protein was immobilized on a Bio-Rad ProteOn HTG chip (containing Nickel-NTA). Protein was loaded to approximately 800 RU. Experiments were were done in 50 mM sodium phosphate buffer, pH 7.4, 150 mM NaCl, 2 mM TCEP, and 0.005% Tween 20 at 25°C. Binding measurements were done by flowing H3K9Me$_3$ peptide (ARTKQTARK(Me)$_3$STGGKAY) at 5 concentrations (10 µM, 5 µM, 2.5 µM, 1 µM, 0.5 µM) over chip with immobilized proteins at a flow rate of 100 µL/min for 120 s. Kinetic analysis of

binding data was done using supplied ProteOn software to calculate binding constants.

Measurements were done in triplicate unless otherwise noted.

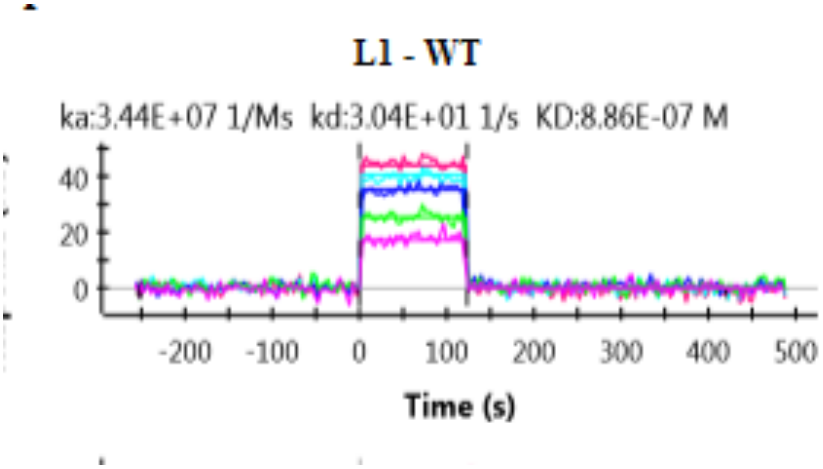


Figure 2.35. Wild Type kinetic analysis of H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY) flown over immobilized protein at 25°C in 50 mM phosphate buffer, pH 7.4, 150 mM NaCl, 2 mM TCEP, 0.005% Tween 20.

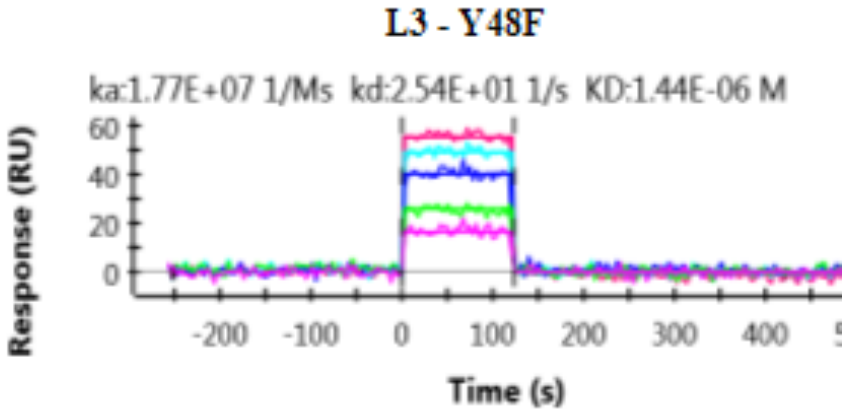


Figure 2.36. Y48F kinetic analysis of H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY) flown over immobilized protein at 25°C in 50 mM phosphate buffer, pH 7.4, 150 mM NaCl, 2 mM TCEP 0.005% Tween 20.

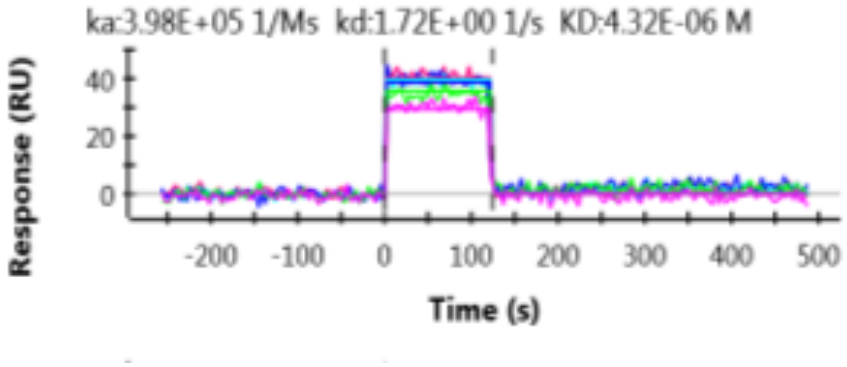


Figure 2.37. Y48pClF kinetic analysis of H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY) flown over immobilized protein at 25°C in 50 mM phosphate buffer, pH 7.4, 150 mM NaCl, 2 mM TCEP 0.005% Tween 20.

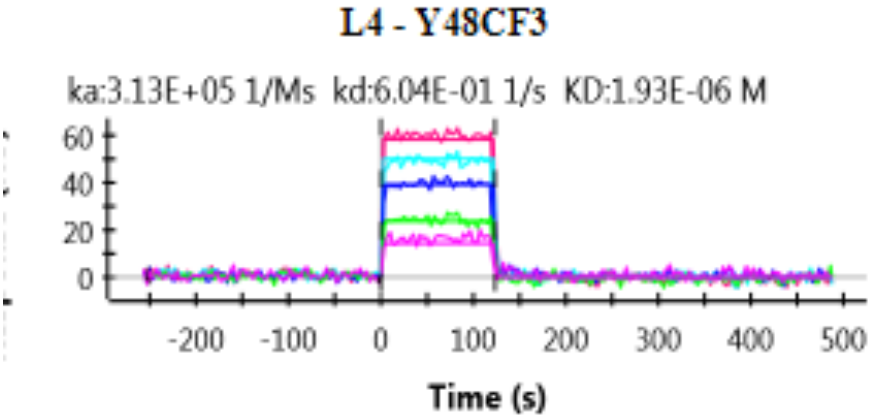


Figure 2.38. Y48pCF$_3$F kinetic analysis of H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY) flown over immobilized protein at 25°C in 50 mM phosphate buffer, pH 7.4, 150 mM NaCl, 2 mM TCEP 0.005% Tween 20.

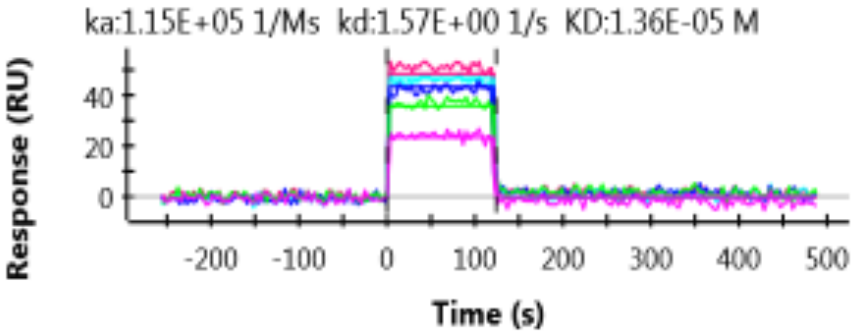


Figure 2.39. Y48pCNF kinetic analysis of H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY) flown over immobilized protein at 25°C in 50 mM phosphate buffer, pH 7.4, 150 mM NaCl, 2 mM TCEP 0.005% Tween 20.

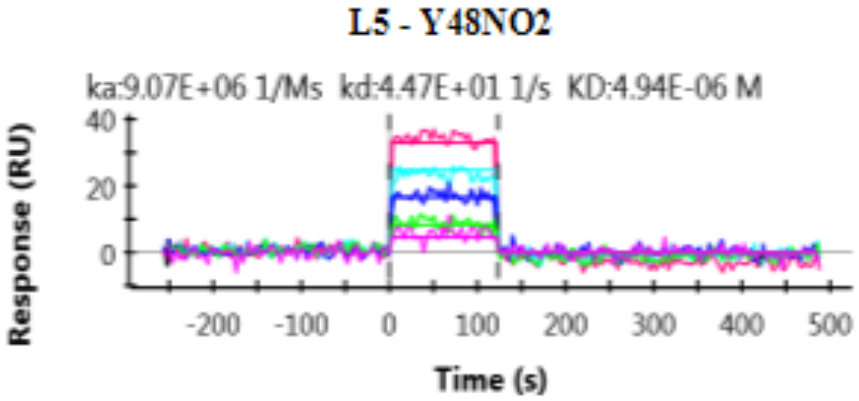


Figure 2.40. Y48pNO$_2$F kinetic analysis of H3K9Me$_3$ (ARTKQTARK(Me)$_3$STGGKAY) flown over immobilized protein at 25°C in 50 mM phosphate buffer, pH 7.4, 150 mM NaCl, 2 mM TCEP 0.005% Tween 20.

REFERENCES

(1) Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11184–11188.

(2) Jacobs, S. A.; Khorasanizadeh, S. *Science (80-.).* **2002**, *295* (5562), 2080–2083.

(3) Eisert, R. J.; Waters, M. L. **2011**, *27599*, 2786–2790.

(4) Jacobs, S. A.; Khorasanizadeh, S. *Science (80-.).* **2002**, *295* (5562), 2080–2083.

(5) Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11184–11188.

(6) Dougherty, D. A. *Acc. Chem. Res.* **2013**, *46* (4), 885–893.

(7) Kearney, P. C.; Nowak, M. W.; Zhong, W.; Silverman, S. K.; Lester, H. A.; Dougherty, D. A. *Mol.Pharmacol.* **1996**, *50* (5), 1401–1412.

(8) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. *Nature* **2009**, *458* (7237), 534–537.

(9) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. a; Dougherty, D. a. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (21), 12088–12093.

(10) Kumpf, R. A.; Dougherty, D. A. *Science (80-.).* **1993**, *261* (5129), 1708–1710.

(11) Nowak, M. W.; Kearney, P. C.; Sampson, J. R.; Saks, M. E.; Labarca, C. G.; Silverman, S. K.; Zhong, W.; Thorson, J.; Abelson, J. N.; Davidson, N. *Science* **1995**, *268*, 439–442.

(12) Brejc, K.; van Dijk, W. J.; Klaassen, R. V; Schuurmans, M.; van Der Oost, J.; Smit, a B.; Sixma, T. K. *Nature* **2001**, *411* (6835), 269–276.

(13) Dougherty, D. A.; Stauffer, D. A. *Science (80-.).* **1990**, *250* (4987), 1558–1560.

(14) Padgett, C. L.; Hanek, A. P.; Lester, H. A.; Dougherty, D. A.; Lummis, S. C. R. *J. Neurosci.* **2007**, *27* (4), 886–892.

(15) Lummis, S. C. R.; Harrison, N. J.; Wang, J.; Ashby, J. a; Millen, K. S.; Beene, D. L.; Dougherty, D. a. *ACS Chem. Neurosci.* **2012**, *3*, 186–192.

(16) Pless, S. A.; Millen, K. S.; Hanek, A. P.; Lynch, J. W.; Lester, H. A.; Lummis, S. C. R.; Dougherty, D. A. *J. Neurosci.* **2008**, *28* (43), 10937–10942.

(17) Duffy, N. H.; Lester, H. A.; Dougherty, D. A. *ACS Chem. Biol.* **2012**, *7* (10), 1738–1745.

(18) Mortensen, M.; Iqbal, F.; Pandurangan, A. P.; Hannan, S.; Huckvale, R.; Topf, M.; Baker, J. R.; Smart, T. G. *Nat. Commun.* **2014**, *5*, 4454.

(19) Van Arnam, E. B.; Dougherty, D. A. *J. Med. Chem.* **2014**, *57* (15), 6289–6300.

(20) Maksay, G.; Bikádi, Z.; Simonyi, M. *J. Recept. Signal Transduct. Res.* **2003**, *23* (2-3), 255–270.

(21) Morikubo, N.; Fukuda, Y.; Ohtake, K.; Shinya, N.; Kiga, D.; Sakamoto, K.; Asanuma, M.; Hirota, H.; Yokoyama, S.; Hoshino, T. *J. Am. Chem. Soc.* **2006**, *128* (40), 13184–13194.

(22) Ito, R.; Hashimoto, I.; Masukawa, Y.; Hoshino, T. *Chem. - A Eur. J.* **2013**, *19* (50), 17150–17158.

(23) Faraldos, J. A.; Antonczak, A. K.; González, V.; Fullerton, R.; Tippmann, E. M.; Allemann, R. K. *J. Am. Chem. Soc.* **2011**, *133* (35), 13906–13909.

(24) Wang, L.; Schultz, P. G. *Angew. Chem. Int. Ed. Engl.* **2004**, *44* (1), 34–66.

(25) Xie, J.; Schultz, P. G. *Nat. Rev. Mol. Cell Biol.* **2006**, *7* (10), 775–782.

(26) Liu, D. R.; Magliery, T. J.; Pastrnak, M.; Schultz, P. G. *Proc. Natl. Acad. Sci. U. S. A.* **1997**, *94* (19), 10092–10097.

(27) He, T.; Gershenson, A.; Eyles, S. J.; Lee, Y.-J.; Liu, W. R.; Wang, J.; Gao, J.; Roberts, M. F. *J. Biol. Chem.* **2015**, *290* (31), jbc.M115.668343.

(28) Wheeler, S. E.; Houk, K. N. *J. Am. Chem. Soc.* **2009**, *131* (9), 3126–3127.

(29) Andersen, N. H.; Olsen, K. A.; Fesinmeyer, R. M.; Tan, X.; Hudson, F. M.; Eidenschink, L. A.; Farazi, S. R. *J. Am. Chem. Soc.* **2006**, *128* (18), 6101–6110.

(30) Cochran, a G.; Skelton, N. J.; Starovasnik, M. a. *Proc. Natl. Acad. Sci. U. S. A.* **2001**, *98* (10), 5578–5583.

(31) Rossi, A. M.; Taylor, C. W. *Nat. Protoc.* **2011**, *6* (3), 365–387.

(32) Hansch, C.; Leo, a; Taft, R. W. *Chem. Rev.* **1991**, *91* (2), 165–195.

(33) Hansch, C.; Leo, A. *Substituent constants for correlation analysis in chemistry and biology*; Wiley-Interscience publication; Wiley: New York, NY, 1979.

(34) Chiou, C. T.; Schmedding, D. W.; Block, J. H.; Manes, M. *J. Pharm. Sci.* **1982**, *71* (11), 1307–1309.

(35) Sangster, J. *Journal of Physical and Chemical Reference Data*. 1989, pp 1111–1229.

(36) Hunter, C. A.; Low, C. M. R.; Rotger, C.; Vinter, J. G.; Zonta, C. *Proc. Natl. Acad. Sci.* **2002**, *99* (8), 4873–4876.

(37) Tatko, C. D.; Waters, M. L. *Protein Sci.* **2003**, *12* (11), 2443–2452.

CHAPTER 3: SYNTHETIC METHODS FOR PROBING CATION-II INTERACTIONS AT TRYPTOPHAN OF HP1

3.1 Cation-π Interactions at Tryptophan

In protein folding and protein-protein interactions, it has been observed that tryptophan participates in cation-π interactions with a higher frequency (about 25% of all tryptophans) than other aromatic residues.[1] They have been shown to be important in protein folding, protein-protein interactions, and interactions with membranes. The importance of one cation-π interaction at tryptophan has been observed in ligand gated ion channels. In binding regions containing multiple aromatic residues, it has often been found that a single tryptophan participates in a tunable cation-π interaction.[2,3]

3.1.1 Tuning Tryptophan Cation-π interactions

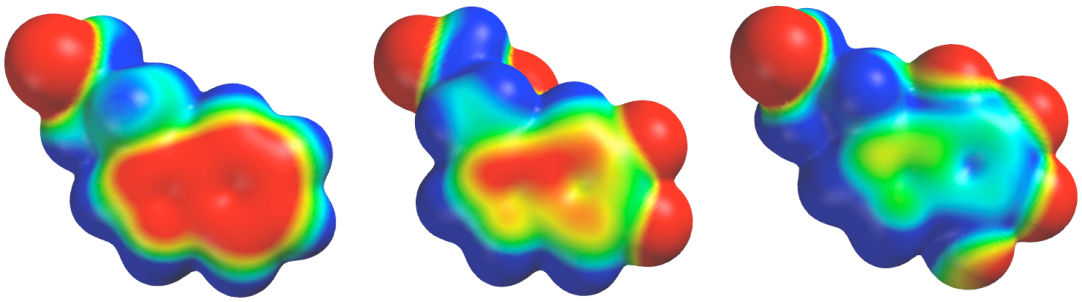


Figure 3.1. Electrostatic potential maps of tryptophan (left) difluorotryptophan (middle) and tetrafluorotryptophan. Electrostatic potential maps were generated with MacSpartan: HF/6-31G*; isodensity value = 0.02; range = -50 (red, electron rich) to 50 kcal/mol (blue, electron poor).

Incorporating fluorine atoms around the indole ring of the tryptophan has been the method most often used for studying these interactions (Figure 3.1). Systematic fluorination provides a method to incrementally tune the electronics of the ring. Tetrafluoroindole (and trifluorophenylalanine) has been used to completely eliminate the electrostatic component of the cation-π interaction, and can be used to potentially quantify the interaction at specific tryptophan residues.[1,4,5] Using this method, replacement of tryptophan residues with tetrafluorotryptophan has led to a 50 to 500 fold decrease in activity, which can be estimated as up to a $-\Delta G$ of 3.7 kcal/mol if binding affinities change by the same ratio as activies.[1] The largest magnitude observed in receptors of this type is a $-\Delta G$ of 5.5 kcal/mol in a glycine binding receptor.[1]

Due to the membrane bound nature of systems studied, functional assays rather than binding measurements have been performed in studies of ligand-gated ion channels. This can be done on single cells, and much less material is used than is required for binding measurements. To incorporate these amino acids, stoichiometric tRNA already loaded with the desired amino acid need to be used. Traditional unnatural amino acid mutagenesis methods have not yet been developed for fluorinated tryptophan, as they are too similar to natural tryptophan and will be recognized by the native synthetase and get incorporated into any proteins naturally containing a tryptophan, as well as all AChR tryptophan residues. In the reader protein heterochromatin protein 1 (HP1), we have shown that two of the tyrosine residues participate in cation-π interactions to different extents; however, to study tryptophan alternative methods need to be developed.

Synthetic methods for the incorporation of fluorinated tryptophan (as well as methods for synthesizing the amino acid itself) are needed for the study of binding interactions at this position. Knowing the contribution of tryptophan to cation-π interaction can help provide a more

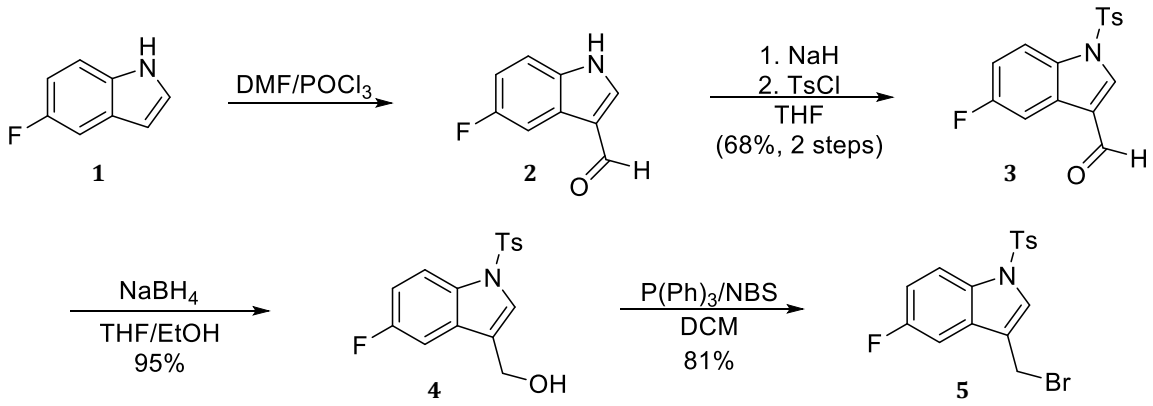
detailed description of the mechanism reader proteins use to recognize their substrates. This can aid in the design of developing therapeutics with the necessary affinity and selectivity.

3.2 Synthesis of Fluorinated Tryptophan

While fluorinated tryptophans are commercially available, many are only available as the racemic mixture, or they are prohibitively expensive. Thus, synthetic methods for accessing the desired tryptophan residues were pursued with two undergraduate researchers, Nathan John and Kaitlyn Tsai.

3.2.1 Synthesis of Alkyl Bromide Side Chain

While there are several methods for accessing the chiral amino acids, many of them require the alkyl bromide derivative of the side chain. Previously reported methods for the synthesis of fluorotryptophan beginning with the corresponding indole worked nicely for mono, di, and tetrafluoroindole (scheme 3.1).[6]



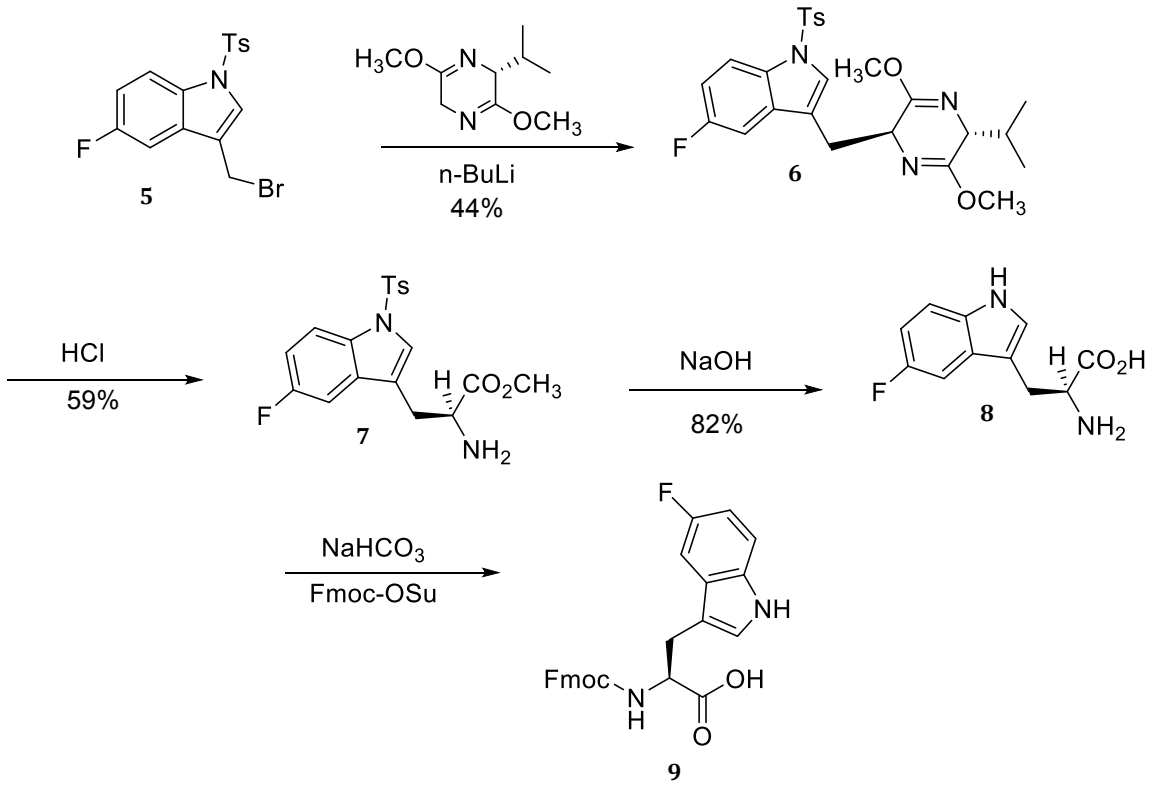
Scheme 3.1. Synthesis of alkyl bromide for tryptophan synthesis starting with the corresponding indole.

Vilsmeier-Haack formylation of the indole was used to install the aldehyde, while the corresponding aldehyde of tetrafluoroindole was commercially available. Subsequent protection of the indole with a tosyl group, and reduction of the aldehyde to the alcohol proceeded cleanly and in high yields. Conversion of the alcohol to the corresponding bromide gave the desired

alkylating agent for amino acid synthesis. Large quantities were stored as the alcohol, as the bromide was less stable and more difficult to work with as it was a moderate lachrymator when completely free of solvent.

3.2.2 Asymmetric Alkylation

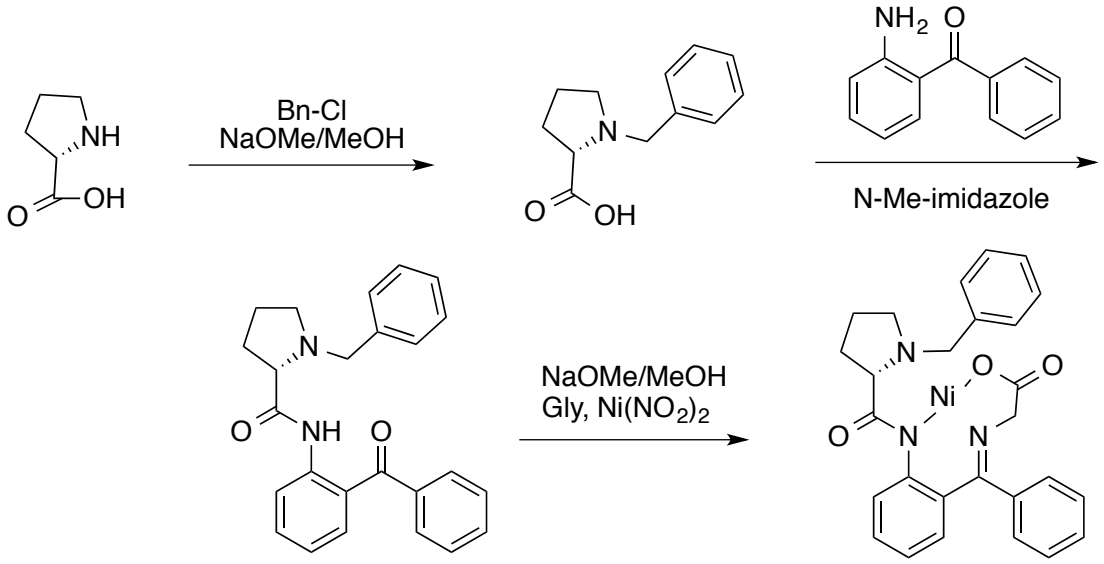
Several strategies were employed for the asymmetric alkylation of glycine with the desired side chain. Previous work in our lab had used the pseudoephedrine auxiliary for the asymmetric synthesis of amino acids.[7] However, when trying to use this approach, low yields of coupling the auxiliary to glycine led to the exploration of other methods. Use of the Schöllkopf auxiliary proved to be more amenable to the tryptophan synthesis (Scheme 3.2).[6] The auxiliary is essentially the dimer of glycine and valine, with the isopropyl group of the valine providing the chiral directing group for the reaction. The lithiation and alkylation proceeded in moderate yields. Subsequent acid and base hydrolysis revealed the tryptophan residue.



Scheme 3.2. Synthesis of tryptophan derivatives using the Schöllkopf auxiliary.[6]

Additional methods showed promise, though were subsequently abandoned in favor of the Schöllhopf method. They may, however, be useful in the synthesis of other amino acids in future studies.

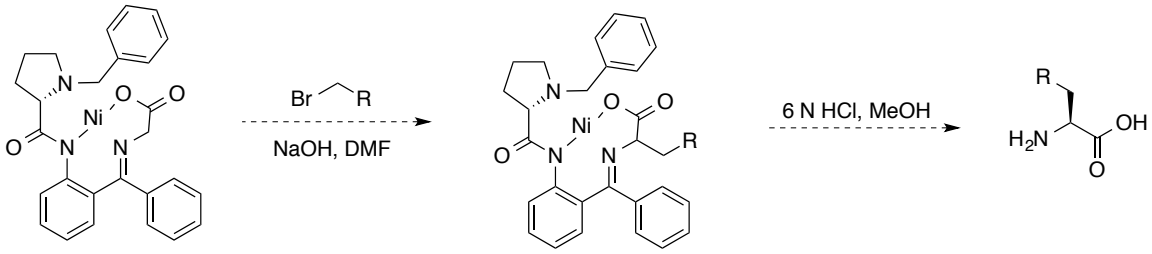
Most notably, the use of proline derived nickel complexes was explored.[8,9] The method was appealing due to the potential of reusing the nickel complex after acid hydrolysis to release the desired amino acid (Scheme 3.3). The complex exhibits a characteristic bright red color, making purification easier as well.



Scheme 3.4. Synthesis of proline derived nickel complex.

Benzylation of proline was done with benzyl chloride in the presence of sodium methoxide. Aminobenzophenone was coupled to the proline to provide the chiral framework. Glycine could then be condensed on and the structure could be complexed with Ni^{2+} to from the bright red crystalline solid for further amino acid synthetic studies.

Asymmetric alkylation was not further pursued, as previous methods were sufficient for amino acid synthesis. However, the alkylation could be done as described in the literature (scheme 3.5), and may be useful for future synthesis of amino acid derivatives.



Scheme 3.5. Asymmetric alkylation of proline derived nickel complex followed by release of the amino acid.

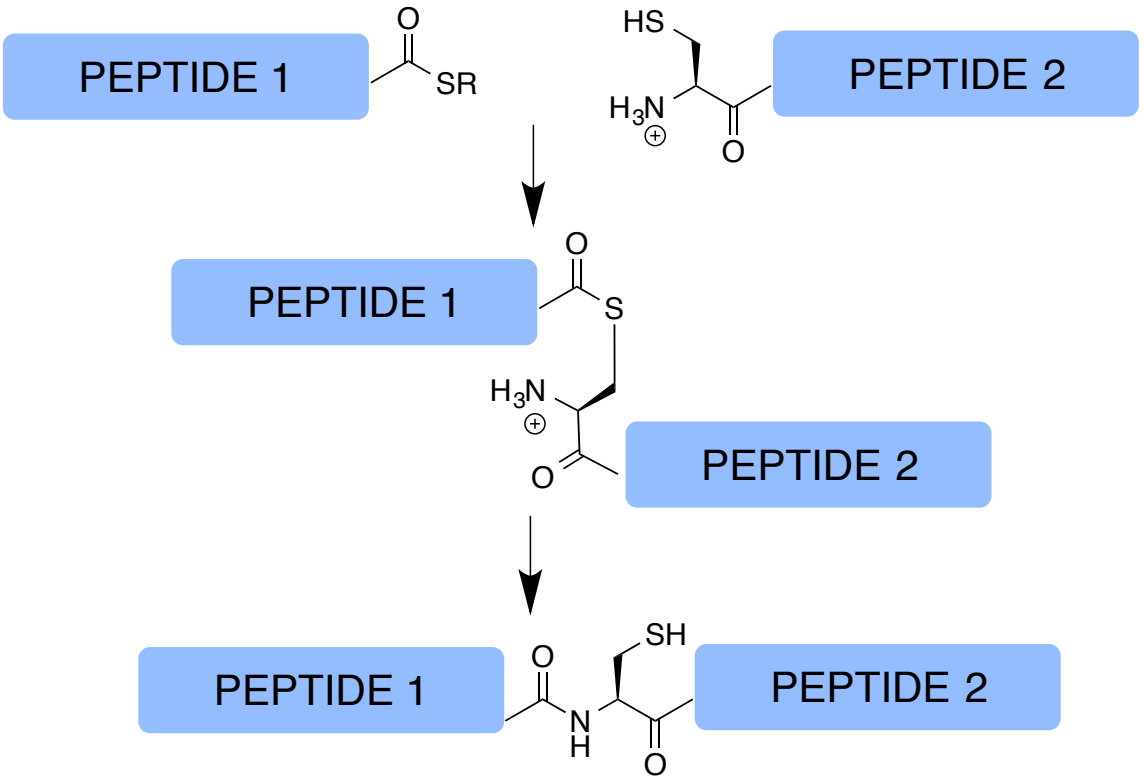
3.3 Native Chemical Ligation

3.3.1 Background

While proteins for studies like this are often made by cellular expression of the protein, the HP1 chromodomain consists of around 70 residues, making it amenable to synthetic methods.[10] To avoid the potential low yields and difficult purification anticipated with linear synthesis, initial studies looked at ligation techniques that can be employed, such as native chemical ligation.[11]

This technique allows for the ligation of two shorter peptides into one long peptide containing the native peptide bond at the ligation site (Scheme 3.6). The ligation requires one peptide to have a thioester at the C-terminus, while the other peptide has a free amine at the N-terminus with a thiol containing amino acid (typically cysteine). The thiol of the cysteine reacts with the thioester. This brings the free amine in close proximity, allowing for S to N acyl transfer to occur, resulting in a native peptide bond. This method has been successful in synthesizing

many large peptides that would have been difficult with linear synthesis.

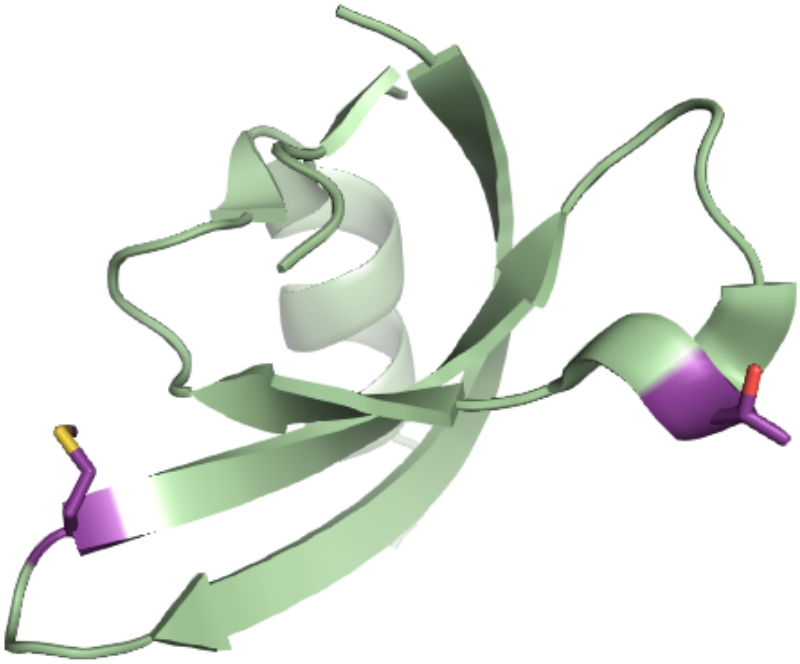


Scheme 3.6. Native chemical ligation between peptide with C-terminal thioester (peptide 1) and N-terminal cysteine (peptide 2).

One of the main disadvantages of this approach is the necessity of a naturally occurring cysteine positioned at a site well suited for breaking apart the peptide. Cysteine has a natural abundance of about 1.2%, limiting the number of proteins for which this approach is useful.[12] There are several ways to expand the scope of this ligation. The simplest is to mutate a residue in the peptide to a cysteine. Alternatively, if there is no residue that can be mutated without changes in structure or function, using a thiol derivative on an amino acid followed by a desulfurization reaction can greatly increase the possible sites. This has been used successfully for ligations at alanine,[13] valine,[12] leucine,[14] threonine,[15] lysine,[16] phenylalanine,[17] and proline.[18]

3.3.2 NCL of HP1

Residues 22-76 of the HP1 chromodomain are structured based on the x-ray structure data (Figure 3.2).



A) EEYAVEKIIDRRVRKGMVEYYLKWKGYPETENTWEPENNLDCQDLIQQYEASRKD
B) EEYAVEKIIDRRVRKGMVEYYLKWKGYPECENTWEPENNLDCQDLIQQYEASRKD
C) EEYAVEKIIDRRVRKGCVEYYLKWKGYPETENTWEPENNLDCQDLIQQYEASRKD

Figure 3.2. Crystal structure of HP1 (top) with proposed sites of mutation to cysteine for ligation in purple. Sequence of HP1 (bottom – residues 22-76) with proposed ligation mutations in red and binding pocket residues in blue.

As the domain does not naturally contain any cysteine residues, sites for potential mutations were identified. Both threonine 51 and methionine 38 were identified as potential sites as they split the peptide into manageable fragments, and were not expected to cause drastic structural changes after mutation to cysteine as they are both solvent exposed, and the side chains

do not appear to be engaging in any interactions important for protein folding. Synthesis of the fragments for the T51C mutation was synthetically problematic, as premature truncation products were always observed and could not be avoided after screening many different conditions including added salts, and different coupling reagents. The M38C mutation was pursued further as it was synthetically less challenging.

The full length HP1 was expressed with the M38C mutation and fluorescence anisotropy was used to confirm the chosen mutation did not affect binding of the natural substrate in any way (Figure 3.3). The K_d of the M38C mutant was measured to be 22 μM, which is within error of the 20 μM previously measured for wild type.

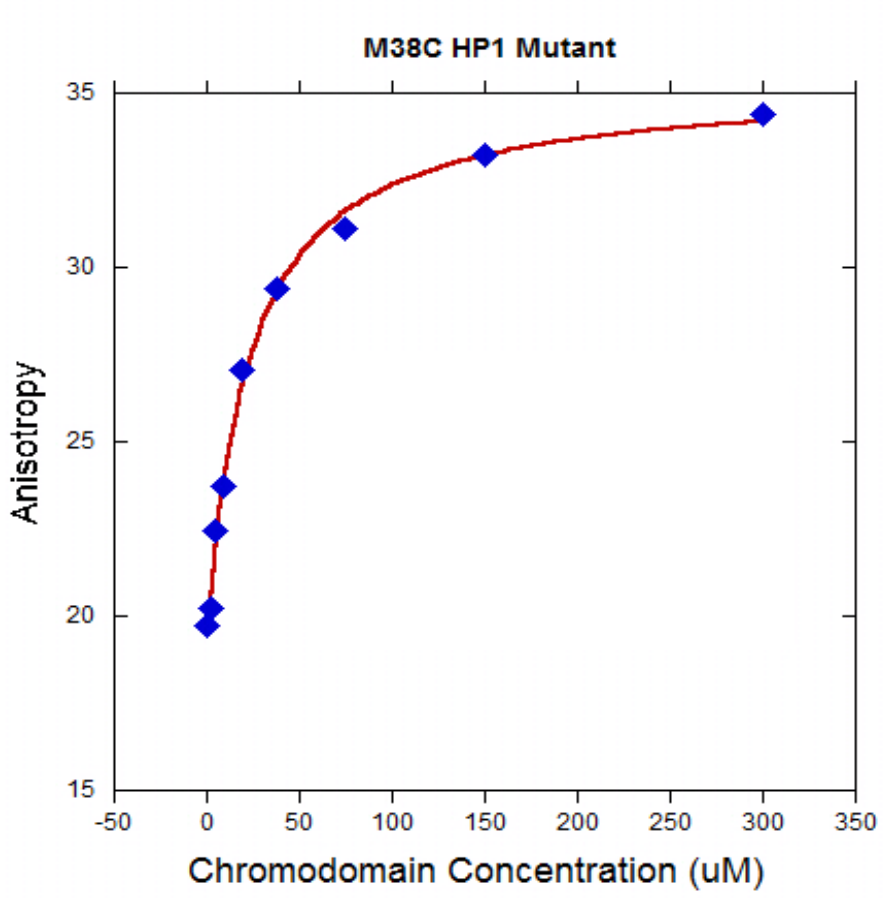
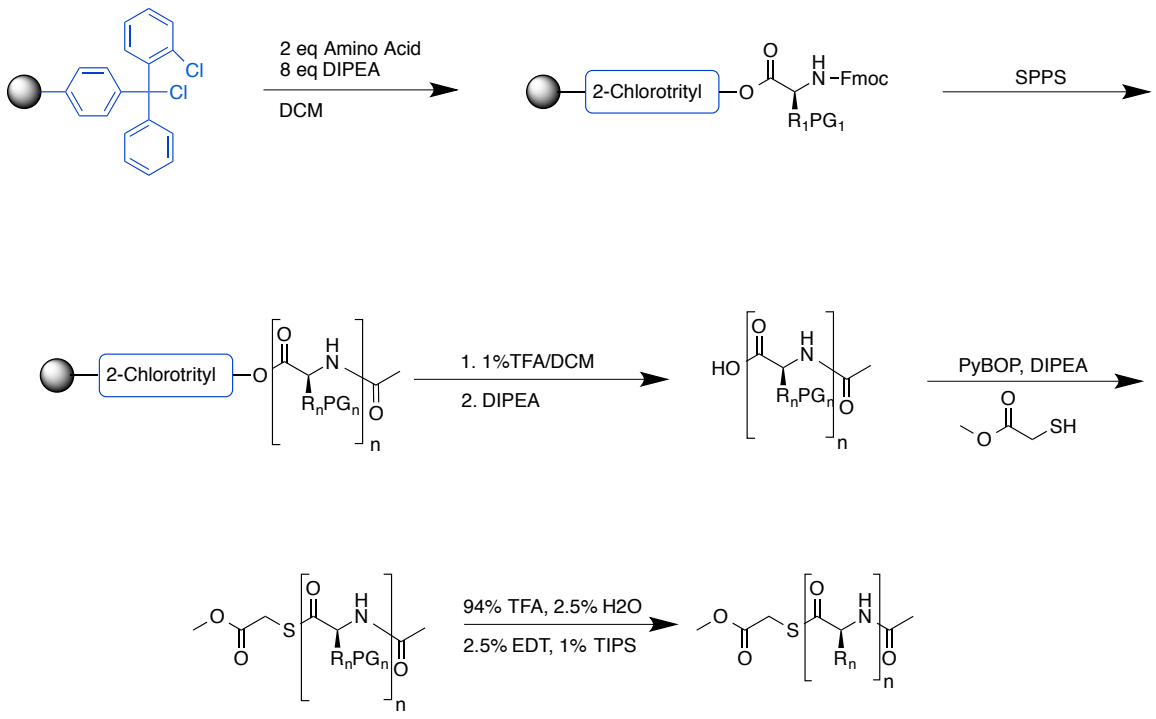


Figure 3.3. Binding curve of M38C HP1 titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8.0, 25 nM NaCl, 2 mM DTT.

3.3.2.1 Thioester Fragment

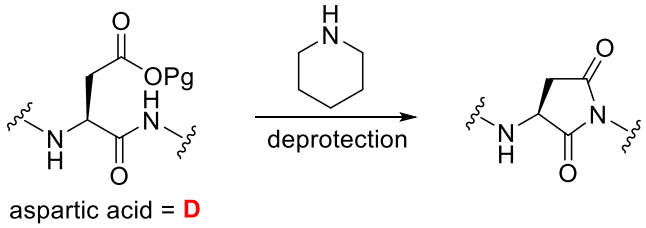
Synthesis of the thioester fragment of HP1 was based on a method previously used.[19] In order to synthesize the thioester at the C-terminus, the peptide needed to be cleaved from the resin with the side chain protecting groups still in place. To achieve this, 2-chlorotrityl chloride resin was used (Scheme 3.7). Purchasing preoaded resin often gave cleaner peptide, but the first amino acid could be loaded on the resin immediately prior to synthesis as well. Synthesis of the fragment was carried out using standard solid phase peptide synthesis with Fmoc protected amino acids. After capping the N-terminus with acetic anhydride, the protected peptide fragment could be cleaved from the resin using 1% TFA, and the C-terminal acid could be coupled with a thiol to form a C-terminal thioester for ligation. To avoid premature deprotection of side chains, 1 equivalent DIPEA (relative to amount of TFA in cleavage mixture) was added to the flask into which the peptide was drained. This immediately quenched the TFA so the peptide was not exposed to acid any longer than necessary for cleavage. After installing the thioester, the peptide could be fully deprotected and purified as normal. Care was taken to insure the peptide was not left in solution for any extended periods of time and that any solution was relatively acidic to prevent hydrolysis of the thioester.



Scheme 3.7. Synthesis of the thioester fragment.

3.3.2.2 Aspartamide Formation

One of the main side reactions encountered in the synthesis of HP1 was aspartamide formation (Scheme 3.8). During the deprotection steps the side chain of aspartic acid can cyclize with the backbone amide.



Scheme 3.8. Aspartimide side reaction occurring during deprotection.

To mitigate this side reaction, deporotections were carried out in piperazine instead of piperidine, and 0.1M HOBt was added to the deprotection solution. The weaker base and addition of a weakly acidic proton donor reduces the amount of amide backbone deprotonation

by buffering the deprotection solution. Using this method of deprotection, both fragments could be consistently synthesized and purified in large quantities.

3.3.2.3 Ligation

Ligation reactions were performed under commonly used conditions with mercaptophenylacetic acid (MPAA).as a catalyst.[20] This catalyst has been shown to form a more reactive thioester for the ligation reactions. The use of this catalyst initially caused difficulty in monitoring the reaction by reverse phase HPLC in TFA buffered solvents as the peak overlapped with the peptide fragments. By switching to ammonium acetate buffer for HPLC monitoring the peptide products could be separated from the MPAA catalyst and monitored (Figure 3.4).

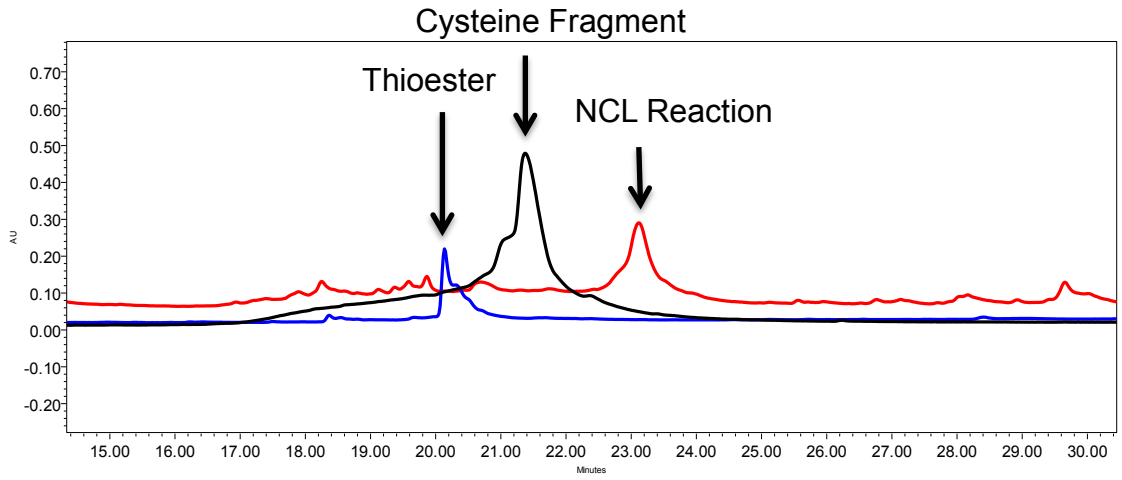


Figure 3.4. Analytical HPLC trace of peptide fragments in ligation buffer: 100mM phosphate, 6 M guanidinium HCl, 5 0mM MPAA, 20 mM TCEP at 280 nm.

Under these conditions, a new peak grew in and complete conversion was observed after 12 hours at room temperature. In addition to HPLC, the crude reaction mixture was monitored by MALDI mass spec, and the mass corresponding to the ligated product was also observed after 12 hours (Figure 3.5).

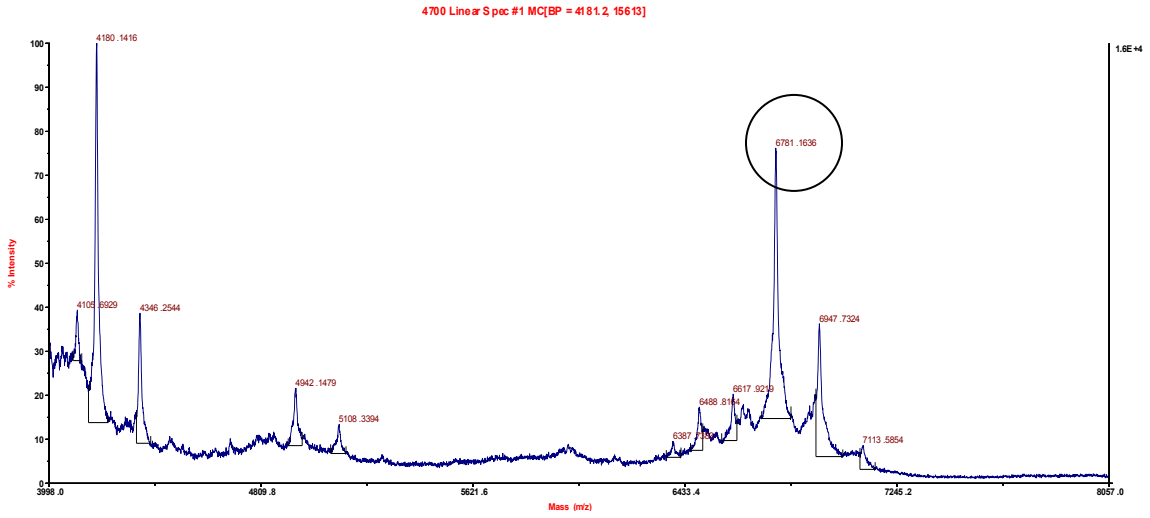


Figure 3.5. MALDI mass spec of crude ligation reaction with mass corresponding to ligated produce (circled).

The new peak in the HPLC trace of the ligation reactions was analyzed by LC-MS to confirm the identity of the ligated product (Figure 3.6).

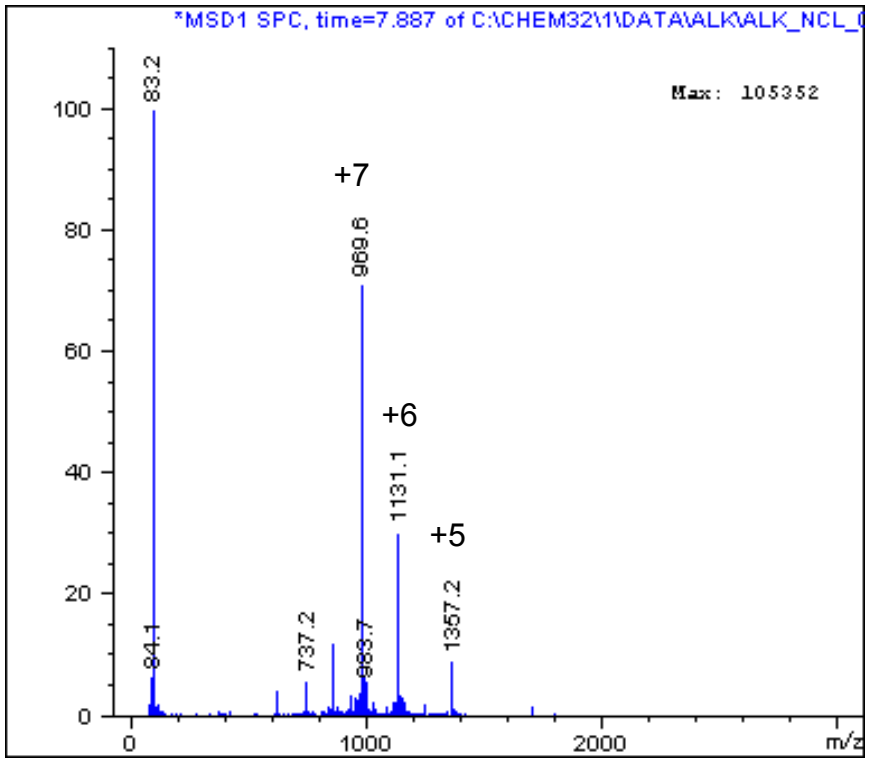


Figure 3.6. LC-MS of peak collected from ligation reactions showing the M+5, M+6, and M+7 peaks.

With the conditions for ligation and purification worked out for HP1, ligation reactions on larger scales were needed in order to synthesize enough protein for binding measurements. Unfortunately, upon injection of the reactions onto a preparatory HPLC, the ligation reactions appeared to behave differently than the small scale injections on the analytical HPLC, potentially due to aggregation issues (Figure 3.7). Using this method, isolation of enough protein from the ligation reaction was not possible.

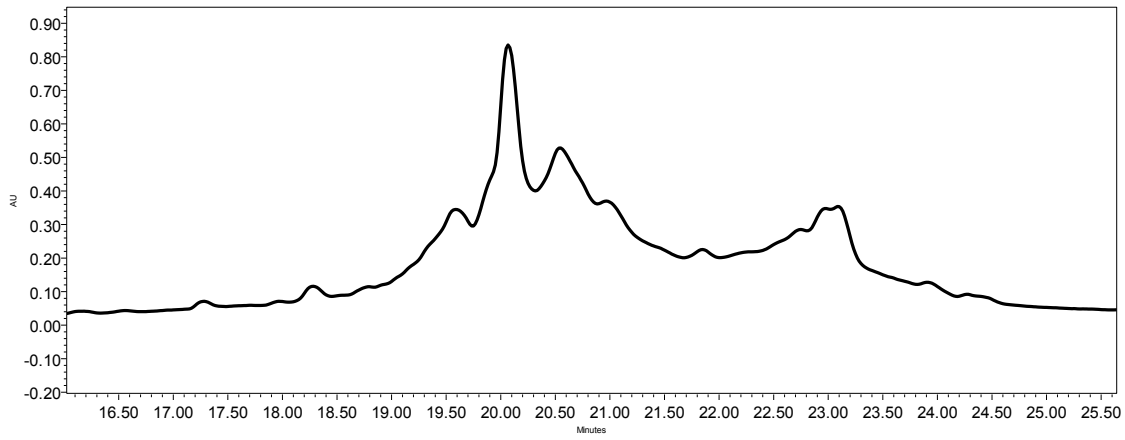


Figure 3.7. HPLC trace of large scale (5 mL reaction volume) ligation reaction.

3.4 SPPS of HP1

In addition to synthesizing HP1 by native chemical ligation, efforts to synthesize the protein linearly were pursued. Small amounts of protein were observed by mass spectrometry during linear synthesis at room temperature on a synthesizer using standard HBTU/HOBt coupling conditions, suggesting with some optimization, linear synthesis of HP1 might be feasible. To improve and simplify purification, an approach for easily purifying protein from deletion products was developed (Figure 3.8). Common side products are single amino acid deletions were coupling steps were incomplete. A common method for simplifying purification

is to acetylate after each coupling steps, so peptide chains that have not been coupled to the next amino acid are truncated. Truncation products have a vastly different size and charge than fully synthesized chains, and are often much easier to purify by HPLC. To take advantage of this method further, a tag can be coupled as the last step of the synthesis, in this case, a series of 6 histidine residues. This allows for affinity chromatography (in this case a Ni-NTA column) to be used to purify out the fully synthesized protein. Because incomplete coupling steps were truncated, only protein that has been successfully coupled with each amino acid will have the poly-histidine tag. The Ni-NTA column can be used to pull out only the full length HP1, greatly simplifying isolation of the synthetic product.

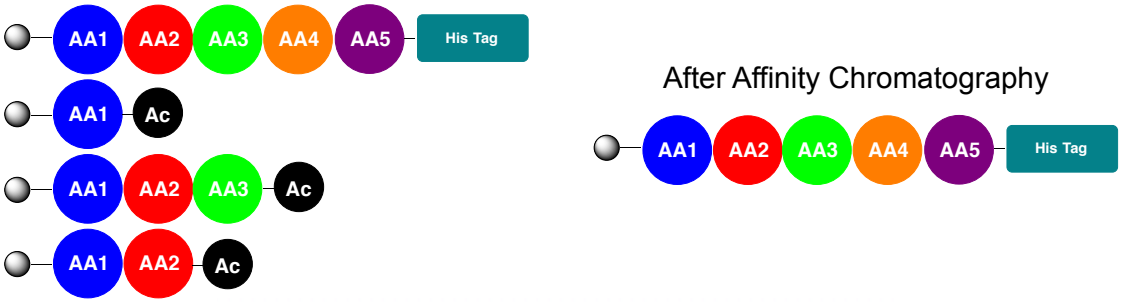


Figure 3.8. Cartoon of example of types of products in crude peptide mixture (left) that can easily be purified out (right).

The synthesis itself was also further optimized by the CEM corportation based on previously published high-efficiency peptide synthesis.[21] This method uses activation with diisopropylcarbodiimide (DIC) and ethyl(hydroxyimino)cyanoacetate (Oxyma). These conditions allow for higher temperatures for more efficient couplings and reduced waste and coupling times. Using this method, CEM was able to achieve approximately 70% crude purity by LCMS analysis on a LibertyBLUE Microwave peptide synthesizer (Figure 3.9). On a Liberty1 peptide synthesizer in our lab, the yields are not as high, but enough protein for experiments was synthesized by this method.

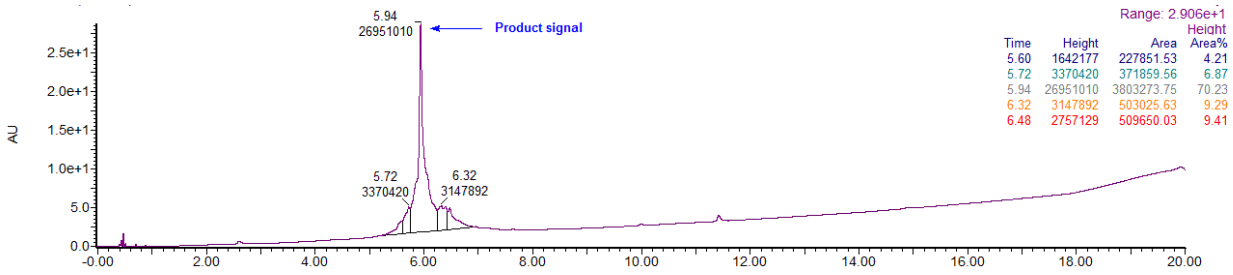


Figure 3.9. UPLC-MS trace of synthetic HP1 from CEM Corporation.

3.5 Binding measurements of HP1

With Stef Baril in the Brustad Lab, the synthetic HP1 was purified via FPLC with a Ni-NTA column followed by size exclusion chromatography. The isolated protein was then concentrated into buffer for the fluorescence anisotropy buffer that had been previously used for binding measurements with expressed HP1. The synthetic chromodomain was titrated into H3K9Me$_3$ peptide containing an N-terminal FAM tag and the binding affinity was measured by fluorescence anisotropy experiments (Figure 3.10).

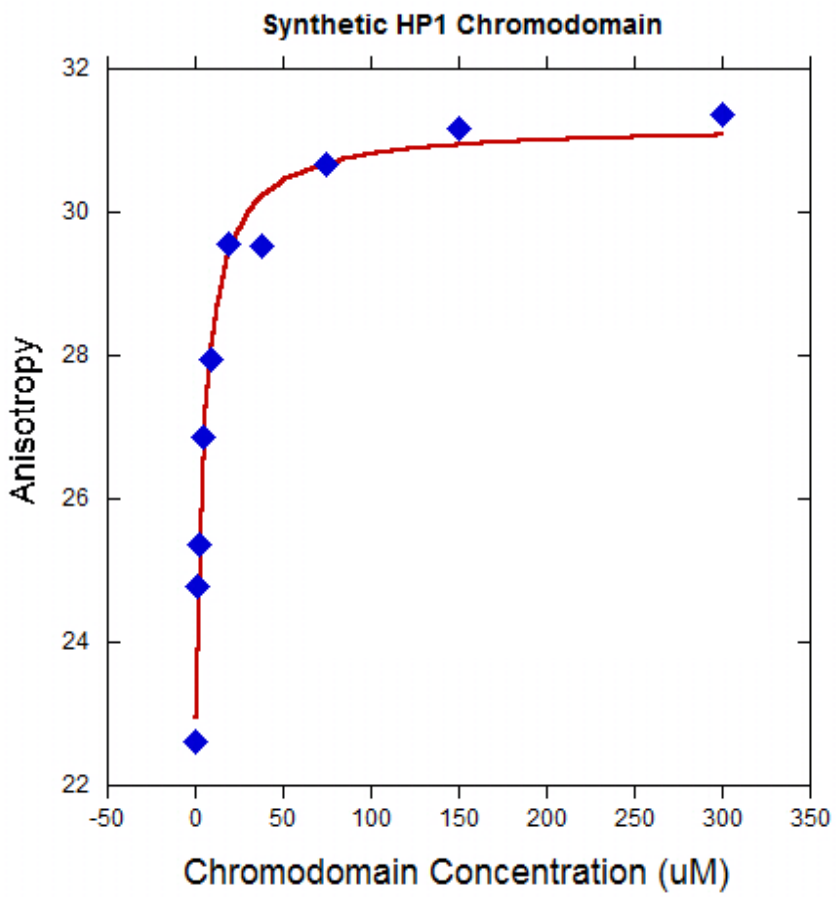


Figure 3.10. Binding curve of synthetic HP1 titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8.0, 25 nM NaCl, 2 mM DTT. K_d = 5μM.

These experiments confirmed that the synthetic protein was functional, as it bound to its natural substrate. The measure binding constant was 5 μM, which was tighter than the 20 μM measured for wild type. This difference may be attributed to the inconsistencies observed for fluorescence anisotropy experiments with HP1.

To test this method with unnatural amino acids, initial studies were done at tyrosine 48 with pentafluorophenylalanine (Y48F$_5$F). The fluorinated alanine derivatives are commercially available and provide a good test of this method before use of synthesized tryptophan derivatives

were incorporated. However, binding to this synthetic mutant was very weak (Figure 3.11), and the curve produce was could not be accurately analyzed.

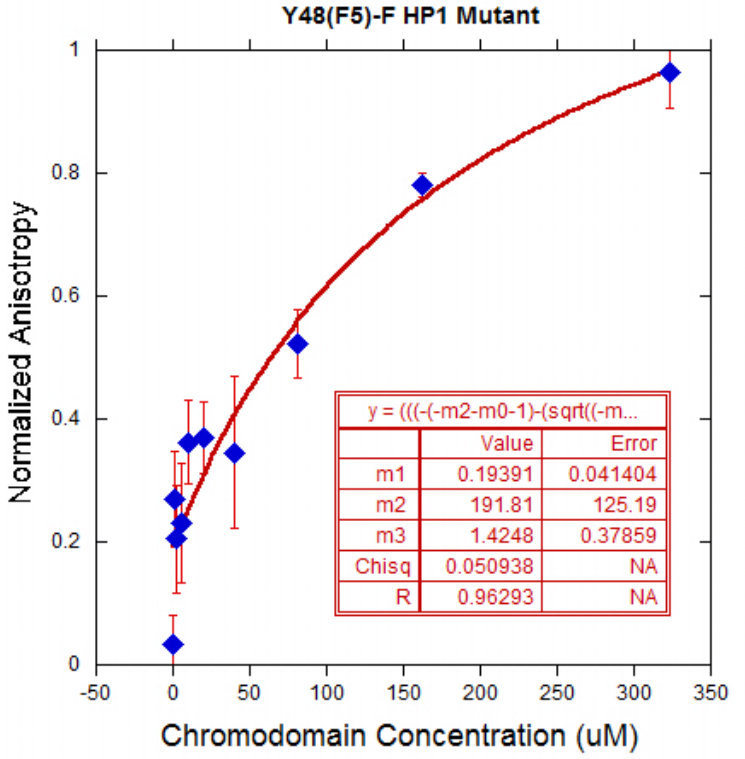


Figure 3.11. Binding curve of synthetic HP1 containing Y48F_5F mutation titrated into FAM labeled H3K9Me_3 peptide (FAM-TKQTARK$(Me)_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8.0, 25 nM NaCl, 2 mM DTT.

3.6 Discussion

Promising synthetic methods have been developed for accessing HP1 *via* solid phase peptide synthesis. The synthetic protein was demonstrated to be functional but measured binding to the trimethyllysine containing H3 peptide. This approach will allow for the site specific incorporation of unnatural amino acids. The method for simplifying purification also has the potential for the use of smaller amounts of unnatural amino acid in coupling reactions. If the amino acid is extremely expensive, or difficult to synthesize, the use of less than 5 equivalents is

still useful for synthesis, even if the coupling is difficult. Instead of deletion products if coupling is not complete, the peptide will truncate and only peptide chains that have incorporated the amino acid will continue through the synthesis and have the his-tag attached. This not only simplifies purification, but also provides a preliminary screen for coupling of the unnatural amino acid – some of which can be more difficult to couple than natural amino acids.

Incorporation of F_5F amino acid appeared to be successful based on isolation of material from his-tag purification, however the protein did not produce a binding curve that could be accurately analyzed. It is possible that the mutation weakened electrostatics significantly and disrupted binding significantly.

Further work to obtain CD spectra and mass spec data of synthesized mutants will be needed to ensure the proper protein is being synthesized and that the protein is well folded. Other methods can be used for determination of binding affinities that have proven to be more reliable for HP1 such as isothermal titration calorimetry (ITC) or surface plasmon resonance (SPR).

Studying the contribution of tryptophan will give a complete analysis of the cation-π interactions of all positions in the binding pocket of HP1. This detailed understanding will provide information about how the protein recognizes it's native substrates, which will aid in developing probes for proteins of this type. Further, this work provides a general approach for synthetically mutating proteins and potentially expanding the scope of proteins that can be synthesized linearly by solid phase peptide synthesis. With the incorporation of acetyl capping after coupling steps and an N-terminal affinity tag, proteins with small yields and in complex mixtures of side products from the synthesis should be easily to isolate and purify by affinity

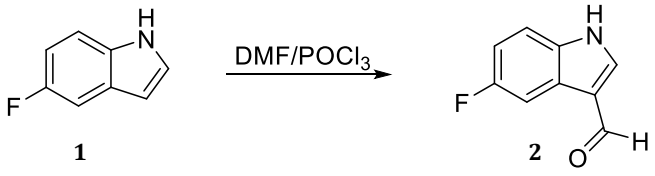
chromatography. This will allow for simpler mutation studies of proteins that are not easily amenable to expression technologies.

3.7 Experimental

3.7.1 Procedure for Synthesis of Fmoc Fluorotryptophan

Synthesis based on literature procedure.[6]

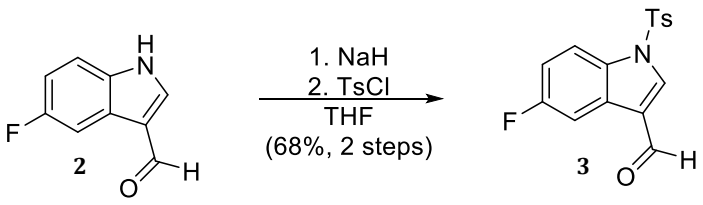
3.7.1.1 Synthesis of Compound 2



1 2

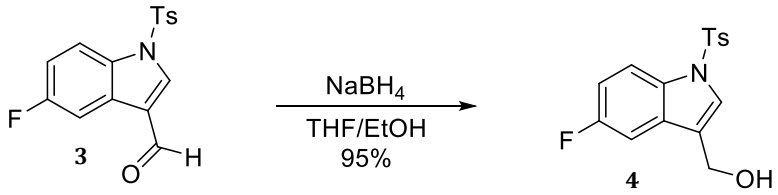
Phosphoryl chloride (1.4 mL, 14.8. mmol, 2 eq) was dissolved in 3 mL dimethyl formamide and cooled in an ice bath for 15 minutes. A solution of the indole 1 (1 g, 7.4 mmol) in 77 mL dimethyl formamide was added dropwise to the phosphoryl chloride solution and stirred at 0° C for 15 minutes. The reaction was heated to 40° C for 1.5 hours, until an opaque solid formed. Ice chips were added to the mixture until the solution became homogenous. Sodium hydroxide (5%) was added until the pH of the mixture was greater than 10. The reaction was refluxed for 5 minutes. After cooling the mixture to room temperature, the mixture was extracted with ethyl acetate, dried with anhydrous magnesium sulfate and concentrated in vacuo to a brown liquid. The crude product was used without further purification. ^{1}H NMR (400 MHz, $CDCl_3$) δ 7.125 (m, 1H), 7.54 (m, 1H), 7.77 (dd, 1H), 8.36 (d, 1H), 9.93 (s, 1H, CHO).

3.7.1.2 Synthesis of Compound 3



Aldehyde 2 (7.4 mmol; crude) was dissolved in 70 mL tetrahydrofuran and cooled to 0° C in an ice water bath. Sodium hydride (60% in mineral oil, 11.1 mmol, 1.5 eq) and tosyl chloride (8.1 mmol, 1.1 eq) was added to the reaction. The mixture was allowed to warm to room temperature and stirred overnight. The reaction was quenched with saturated ammonium chloride solution and concentrated. The residue was partitioned between ethyl acetate and water, and the aqueous layer was further extracted with ethyl acetate. The organic layers were combined, dried with magnesium sulfate and concentrated to a light yellow solid. The crude product was purified by column chromatography on silica gel with 2:1 hexanes:ethyl acetate. The purified product was isolated as a white solid. (68% over 2 steps). ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.42 (s, 3H), 7.16 (m, 1H), 7.34 (d, 2H), 7.86 (d, 2H), 7.94 (m, 1H), 8.27 (s, 1H), 10.09 (s, 1H, CHO).

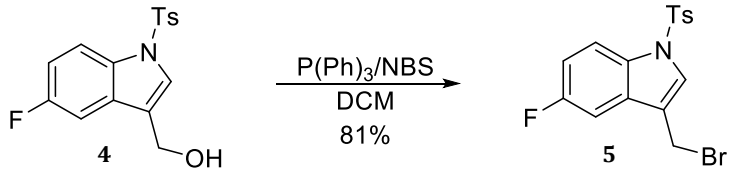
3.7.1.3 Synthesis of Compound 4



Protected aldehyde 3 (1.58 g, 5.0 mmol) was dissolved in 2:1 tetrahydrofuran:ethanol (32 mL:16 mL) and cooled to 0° C in an ice water bath. Sodium borohydride (7.5 mmol, 1.5 eq) was added. The mixture was allowed to warm to room temperature and was stirred overnight. The reaction was quenched with saturated aqueous ammonium chloride and concentrated *in vacuo*. The

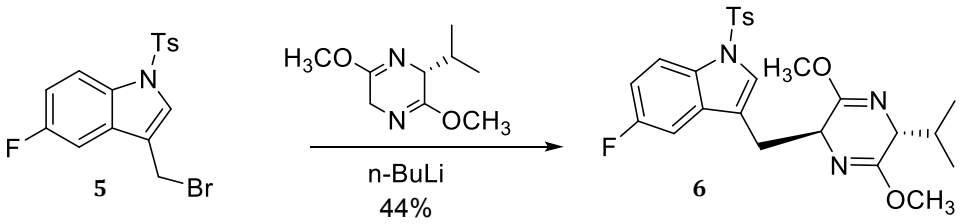
residue was dissolved in brine and extracted with dichloromethane. The organic layers were combined, dried with $MgSO_4$ and concentrated *in vacuo*. The crude product was purified by silica-gel column chromatography with 2:1 hexanes:ethyl acetate and isolated as a white solid (95%). ^{1}H NMR (400 MHz, acetone) δ 2.37 (s, 3H), 4.76 (s, 2H), 7.16 (m, 1H), 7.39 (m, 1H), 7.40 (d, 2H), 7.70 (s, 1H), 7.88 (d, 2H), 8.03 (m, 1H).

3.7.1.4 Synthesis of Compound 5



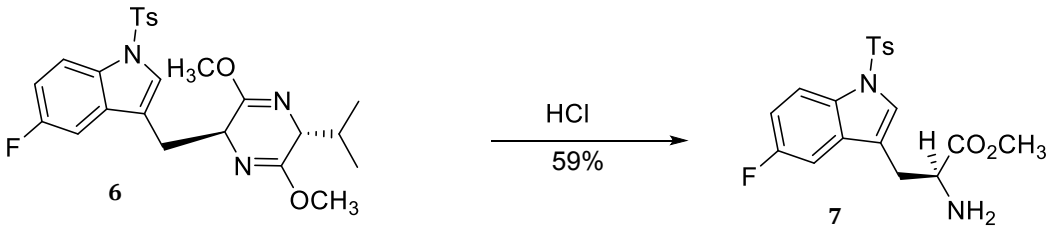
Alcohol 4 (500 mg, 1.57 mmol) was dissolved in 20 mL dichloromethane and cooled to 0° C in an ice water bath. Triphenylphosphine (1.72 mmol, 1.1 eq) was added. After 10 minutes n-bromosuccinimide (1.72 mmol, 1.1 eq.) was added to the mixture and maintained at 0° C for 1 hour. The ice water bath was then allowed to warm to room temperature and the reaction stirred overnight. The reaction was concentrated *in vacuo* and the residue was run through a short plug of silica-gel using 4:1 hexanes:ethyl acetate. The solution was concentrated in vacuo and used without further purification (81%). ^{1}H NMR (400 MHz, acetone) δ 2.38 (s, 3H), 4.86 (s, 2H), 7.22 (m, 1H), 7.42 (d, 2H), 7.47 (m, 1H), 7.91 (d, 2H), 8.01 (s, 1H), 8.05 (m, 1H).

3.7.1.5 Synthesis of Compound 6



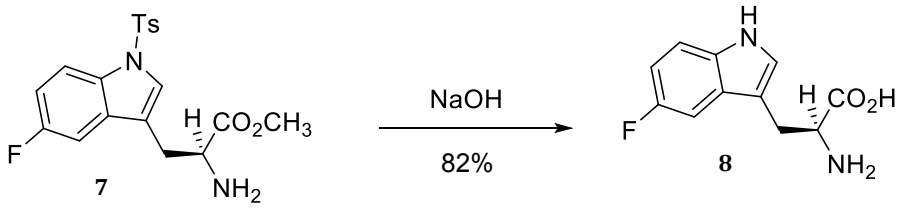
The chiral auxiliary (Schöllkof reagent) (316 mL, 1.76 mmol) was dissolved in 10 mL THF and cooled to -78° C in an isopropyl alcohol/dry ice bath. Freshly titrated n-butyllithium (2 mL, 1.623 M in hexanes, 2 eq.) was added to the reaction mixture and allowed to stir for 1 hour. The alkyl bromide was dissolved in 10 mL THF and added to the mixture dropwise. The reaction was allowed warm to room temperature and stirred overnight. The reaction was quenched with saturated aqueous ammonium chloride and solvents removed *in vacuo*. The residue was dissolved in brine and extracted with dichloromethane. The organic layers were combined and dried with $MgSO_4$ and concentrated *in vacuo*. The crude product was purified by silica-gel column chromatography with 18:1 chloroform:ethyl acetate resulting in a colorless oil (44%). ^1H NMR (400 MHz, $CDCl_3$) δ 0.615 (d, 3H), 0.877 (d, 3H), 2.08 (m, 1H), 2.35 (s, 3H) 3.11 (m, 1H) 3.12 (m, 1H), 3.18 (m, 1H), 3.66 (s, 3H), 3.69 (s, 3H), 4.32 (m, 1H), 6.98 (m, 1H), 7.19 (dd, 1H), 7.21 (d, 2H,) 7.31 (s, 1H), 7.67 (d, 2H), 7.87 (m, 1H).

3.7.1.6 Synthesis of Compound 7



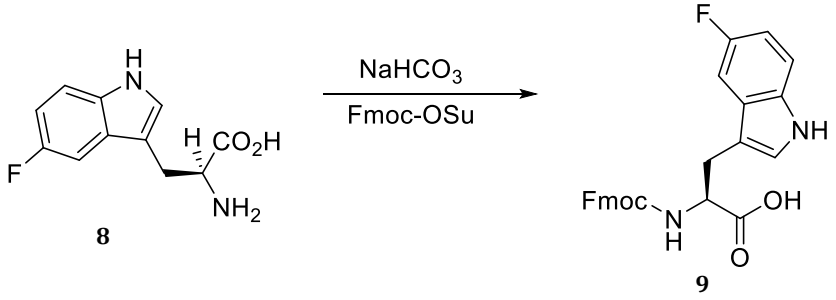
The alkylated Schöllkopf reagent (0.193 g, 0.39 mmol) was dissolved in 6 mL THF and cooled to 0° C in an ice water bath. To the mixture, 2 N HCl (3.2 mL) was added. The mixture was warmed to room temperature after 10 minutes and allowed to stir for 1 hour. The THF was removed *in vacuo* and the aqueous solution diluted with water. The product was neutralized by adding ammonium hydroxide until the pH reached 10. The mixture was extracted with ethyl acetate and dried with $MgSO_4$. The solvent was removed *in vacuo* and residue purified by silica-gel column chromatography using 5% methanol, 0.5% ammonium hydroxide in ethyl acetate. The product was isolated as a colorless oil (59%). ^{1}H NMR (400 MHz, $CDCl_3$) δ 2.078 (s, 3H), 2.69 (m, 1H) 2.84 (m, 1H), 3.19 (s, 3H), 3.87 (m, 1H), 6.78 (m, 1H), (dd, 1H), 6.97 (d, 2H), 7.25 (s, 1H), 7.47 (d, 2H), 7.65 (m, 1H).

3.7.1.7 Synthesis of Compound 8



Fluorotryptophan methyl ester (127 mg, 0.33 mmol) was dissolved in 3.6 mL ethanol and 2.2 mL 2 N sodium hydroxide added. The reaction was heated at reflux for 30 minutes. The solution was cooled to room temperature and 6N HCl added until the pH reached 2. The solution was extracted with dichloromethane. Sodium hydroxide was added until the pH reached 6 and the water was removed *in vacuo*. The residue could be either carried forward crude, or purified by recrystallization in ethanol/water to give a white to tan solid and stored as purified amino acid (82%). ^{1}H NMR (400 MHz, D_2O) δ 3.17 (m, 1H), 3.32 (m, 1H), 3.92 (m, 1H), 6.94 (m, 1H), 7.24 (s, 1H), 7.29 (m, 1H), 7.36 (m, 1H).

3.7.1.8 Synthesis of Compound 9



Fluorotryptophan (73 mg, 0.33 mmol) was dissolved in 7 mL of a 50/50 solution of sat. $NaHCO_2$ and acetonitirile. Fmoc-OSu (135 mg, 0.40 mmol) was added. The reaction mixture was stirred for 3 hours. HCl was added until the pH reached 7 and the solvents were removed under vacuum. The residue was purified by silica gel chromatography using 5% methanol in chloroform. ^{1}H NMR (400 MHz, Acetone) δ 3.22 (m, 1H), 3.38 (m, 1H), 4.22 (m, 1H), 4.45 (m, 2H), 4.58 (m, 1H), 6.90 (m, 1H), 7.38-7.28 (m, 7H, overlap) 7.64 (m, 2H), 7.84 (d, 2H).

3.7.2 Peptide Synthesis

3.7.2.1 Synthesis of the C-terminal Fragment

C-terminal NCL Fragment HP1 (residues 38-76) (CVEYYLKWKGYPETENTWEPENNLDCQDLIQQYEASRKD) was synthesized on a CEM Liberty1 Microwave Peptide Synthesizer using Fmoc protected amino acids and CLEAR Amide resin. The amino acid residues were activated with HBTU (O-benzotriazole-N, N, N', N',-tetramethyluronium hexafluorophosphate) in the presence of DIPEA (diisopropylethylamine) in DMF (N,N-dimethylformamide). Double couplings of 5 minutes each were used. Deprotections were carried out in 6% piperizine and 0.1 HOBt (hydroxy benzyl triazole) in DMF, twice for 45

seconds each. The peptide was acetylated with acetic anhydride and 2,6 lutidine in DMF after every coupling to cap any unreacted peptide chains. The N-terminus of the final cysteine residue was left unprotected. Cleavage from the resin was performed in 94:2.5:2.5:1 trifluoroacetic acid (TFA):ethanedithiol (EDT):water:triisopropylsilane (TIPS) for 3.5 hours. The TFA was blown off with nitrogen and products were precipitated with cold diethyl ether. The peptides were extracted with water and lyophilized. Peptides were purified by reversed phase HPLC using a C-18 semipreparative column and a gradient of 0 to 100% B in 60 minutes, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. The purified peptides were lyophilized and identity was confirmed by ESI mass spectrometry. Calculated: $[M+H^+]$: 4797, $[M+2H^+]$: 2399, $[M+3H^+]$: 1599.6, $[M+4H^+]$: 1200; Observed: $[M+3H^+]$: 1600.14, $[M+4H^+]$: 1200.87

3.7.2.2 Synthesis of Thioester Fragment

N-terminal Thioester Fragment of HP1 (residues 22-37) (EEYAVEKIIDRRVRKG)The thioester fragment was synthesized on an Applied Biosystems Pioneer Peptide Synthesizer using Fmoc protected amino acids and 2-chlorotrityl chloride resin. The first amino acid was preloaded on the resin using 2 equivalents amino acid, 4 equivalents DIPEA in 5 mL dichloromethane twice, for 1 hour each. Preloaded resin can also be purchased, but for NCL fragment synthesis, the residue was most often loaded by hand. Subsequent amino acid residues were coupled with HBTU (O-benzotriazole-N, N, N', N',-tetramethyluronium hexafluorophosphate) and HOBT (N-hydroxybenzotriazole) in the presence of DIPEA (diisopropylethylamine) in solvent DMF (N,N-dimethylformamide) on an Applied Biosystems automated peptide synthesizer. Double couplings of 30 minutes each were used. Deprotections were carried out in 6% piperizine and 0.1 HOBt (hydroxy benzyl triazole) in DMF, twice for 15

minutes each. The fully protected peptide fragment was cleaved from the resin using 3mL of 1% TFA in dichloromethane 3 times for three minutes each. The cleavage mixtures were drained into a flask containing 1 equivalent diisopropyl ethyl amine, and concentrated in vacuo to a yellow oil. The concentrated peptide was dissolved in 3.5 mL DMF. The thioester was formed by adding 4 equivalents DIPEA, 4 equivalents methyl thioglycolate and 4 equivalents PyBOP for 1 hour. The DMF was removed via high vacuum. Peptide deprotection was performed in 94:2.5:2.5:1 trifluoroacetic acid (TFA):ethanedithiol (EDT):water:triisopropylsilane (TIPS) for 3.5 hours. The TFA was evaporated and products were precipitated with cold diethyl ether. The peptides were extracted with water and lyophilized. Peptides were purified by reversed phase HPLC using a C-18 semipreparative column and a gradient of 0 to 100% B in 60 minutes, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. The purified peptides were lyophilized and identity was confirmed by ESI mass spectrometry. Calculated: $[M+H^+]$: 2091, $[M+2H^+]$: 1046, $[M+3H^+]$: 967.7; Observed: $[M+2H^+]$: 1046.04, $[M+3H^+]$: 967.87.

3.7.2.3 Synthesis of Full Length HP1 with His-tag.

(MKKHHHHHHAEEEEEEYAVEKIIDRRVRKGKVEYYLKWKGYPETENTWEPEN NLDCQDLIQQYEASRKD) Full length HP1 was synthesized on a CEM Liberty1 Microwave Peptide Synthesizer using Fmoc protected amino acids and Rink Amide resin. The amino acid residues (5 eq) were activated with DIC (diisopropylcarbodiimide – 5 eq) in the presence of Oxyma (Ethyl (hydroxyimino)cyanoacetate – 5 eq). Double couplings of 4 minutes each at 90° C were used. Arginine and histidine residues were coupled at RT for 4 min and 50° C for 8 min. Deprotections were carried out in 10% piperazine and 0.1 HOBt (hydroxy benzyl triazole) in 9:1 NMP:EtOH, twice for 60 seconds each at 90° C. The peptide was acetylated with acetic

anhydride and 2,6 lutidine in DMF after every coupling to truncate any unreacted peptide chains. Cleavage from the resin was performed in 92.5:2.5:2.5:2.5 trifluoroacetic acid (TFA): (ethylenedioxy)diethanethiol (DODT):water:triisopropylsilane (TIPS) for 3.5 hours. The TFA was blown off with nitrogen and products were precipitated with cold diethyl ether. The peptides were extracted with water and lyophilized. His-tagged protein was purified by FPLC on a Ni-NTA column, followed by size exclusion chromatography and concentrated in an appropriate buffer for experiments.Protein was confirmed by LC-MS. Calculated: $[M+H^+]$: 8434, $[M+5H^+]$: 1688, $[M+6H^+]$: 1407, $[M+7H^+]$: 1206; Observed: $[M+5H^+]$: 1688.44, $[M+6H^+]$: 1407.32, $[M+7H^+]$: 1206.61.

3.7.2.4 Synthesis of H3 1-15 K9Me$_3$

H3 1-15 for SPR containing KMe_3 at position 9 (ARTKQTARK(Me)$_3$STGGKAY) was synthesized by hand using Fmoc protected amino acids and Rink Amide resin on a 0.06 mmol scale. The amino acid residues were activated with HBTU (O-benzotriazole-N, N, N', N',-tetramethyluronium hexafluorophosphate) and HOBT (N-hydroxybenzotriazole) in the presence DIPEA (diisopropylethylamine) in DMF (N,N-dimethylformamide). Double couplings of 30 minutes each were used. Deprotections were carried out in 20% piperidine in DMF, twice for 15 minutes each.

Dimethyllysine was coupled using 2 equivalents of Fmoc-Lys(Me)2-OH□HCl for 5 hours. Immediately after coupling, the resin was washed with DMF and the residue was methylated to form trimethyllysine with 7-methyl- 1,5,7-triaza-bicyclo[4.4.0]dec-5-ene (MTDB, 10.8 uL, 0.075 mmol) and methyl iodide (37.4 uL, 0.6 mmol, 10 equiv.) in DMF (5 mL) for 5 hours at room temp. The resin was washed with DMF and peptide synthesis was continued as normal.

Cleavage from the resin was performed in 95:2.5:2.5 trifluoroacetic acid (TFA):water:triisopropylsilane (TIPS) for 4 hours. The TFA was evaporated and products were precipitated with cold diethyl ether. The peptides were extracted with water and lyophilized. Peptides were purified by reversed phase HPLC using a C-18 semipreparative column and a gradient of 0 to 100% B in 60 minutes, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. The purified peptides were lyophilized, and identity was confirmed by MALDI mass spectrometry. Calculated: 1765 Observed: 1765.9.

3.7.2.5 Synthesis of FAM-capped H3 3-15

(FAM-TKQTARK(Me)$_3$STGGKA) Peptide containing N-terminal 5,6-carboxyfluroscein (FAM) was synthesized for fluorescence anisotropy using the same procedure. Prior to cleavage from the resin, FAM was coupled to the N-terminus using 2 eq FAM, 2 eq PyBOP, 2 eq HOBT, and 4 eq DIPEA. The coupling was bubbled with N_2 overnight in a foil wrapped reaction flask and kept in the dark. Cleavage from the resin and purification were performed as previously described. Peptide was used for experiments without desalting, and concentration was determined using UV-Vis absorbance of FAM at 492 nm (ε=78000 $M^{-1}cm^{-1}$). Calculated: [M+H$^+$]: 1,708, [M+2H$^+$]: 854.5, [M+3H$^+$]: 570; Observed: [M+2H$^+$]: 854, [M+3H$^+$]: 570.

3.7.3 Native Chemical Ligation Reactions

Ligation reactions were performed in 100 mM phosphate buffer, 6 M guanidinium HCl, 50 mM MPAA, 20 mM TCEP at pH 7, based partly on conditions from the literature.[20] Buffer was degassed prior to use and both peptide fragments were dissolved at approximately 1 mM each. The pH was adjusted back up to 7.0 after dissolving the peptides, as the TFA salts lower the pH. The reactions were done at 25° C in a sealed vial. Aliquots were taken every hour and quenched by diluting the reaction 6 fold with 0.1% TFA. The reaction was analyzed by reversed

phase analytical HPLC using a C-18 semipreparative column and a gradient of 0 to 100% B in 60 minutes, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. Peaks were isolated and further analyzed by LC-MS and MALDI. Calculated: [M]: 6784.5, [M+5H$^+$]: 1357.9, [M+6H$^+$]: 1131.8, [M+7H$^+$]: 970.2; Observed: [M+5H$^+$]: 1357.2, [M+6H$^+$]: 1131.1, [M+7H$^+$]: 969.6.

3.7.4 Fluorescence Anisotropy Binding Measurements

Fluorescence anisotropy measurements were performed using a PolarStar omega plate reader by BMG Labtech. Chromodomain solutions of different concentrations were made by performing serial half dilutions of a concentrated stock. Concetration was determined by UV-Vis absorbance at 280 nm (ε=17780). Chromodomain (25 µL) was added into 1 µM FAM labeled peptide (25 µM) in a 96 well plate in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, and 2 mM DTT). Samples were allowed to equilibrate at 25° C for 30 minutes. The samples were analyzed using an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The data were analyzed in kaleidegraph using the equation:

$$r = \left[\left(\frac{p + [c] + K_D \pm \sqrt{(-p-[c]-K_D)^2 - 4(p \cdot [c])}}{2p}\right) \cdot (r_\infty - r_0)\right] + r_0$$

where r is fluorescence anisotropy, r_0 is the anisotropy of the free histone tail, r_∞ is the anisotropy of the fully bound histone tail, ρ is the total concentration of histone peptide, [c] is the total concentration of chromodomain, and K_d is the dissociation constant. The variables r_0, r_∞, and K_d were all treated as floating variables.

REFERENCES

(1) Dougherty, D. A. *Acc. Chem. Res.* **2013**, *46* (4), 885–893.

(2) Xiu, X.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. *Nature* **2009**, *458* (7237), 534–537.

(3) Zhong, W.; Gallivan, J. P.; Zhang, Y.; Li, L.; Lester, H. a; Dougherty, D. a. *Proc. Natl. Acad. Sci. U. S. A.* **1998**, *95* (21), 12088–12093.

(4) Tavares, X. D. S.; Blum, A. P.; Nakamura, D. T.; Puskar, N. L.; Shanata, J. A. P.; Lester, H. A.; Dougherty, D. A. *J. Am. Chem. Soc.* **2012**, *134* (28), 11474–11480.

(5) Beene, D. L.; Brandt, G. S.; Zhong, W.; Zacharias, N. M.; Lester, H. A.; Dougherty, D. A. *Biochemistry* **2002**, *41* (32), 10262–10269.

(6) Konas, D. W.; Seci, D.; Tamimi, S. *Synth. Commun.* **2012**, *42* (1), 144–152.

(7) Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11184–11188.

(8) Ueki, H.; Ellis, T. K.; Martin, C. H.; Boettiger, T. U.; Bolene, S. B.; Soloshonok, V. A. *J. Org. Chem.* **2003**, *68* (18), 7104–7107.

(9) Sorochinsky, A. E.; Aceña, J. L.; Moriwaki, H.; Sato, T.; Soloshonok, V. a. *Amino Acids* **2013**, *45* (4), 691–718.

(10) Jacobs, S. A.; Khorasanizadeh, S. *Science (80-.).* **2002**, *295* (5562), 2080–2083.

(11) Dawson, P. E.; Muir, T. W.; Clark-Lewis, I.; Kent, S. B. *Science* **1994**, *266* (5186), 776–779.

(12) Chen, J.; Wan, Q.; Yuan, Y.; Zhu, J.; Danishefsky, S. J. *Angew. Chemie* **2008**, *120* (44), 8649–8652.

(13) Yan, L. Z.; Dawson, P. E. *J. Am. Chem. Soc.* **2001**, *123* (4), 526–533.

(14) Tan, Z.; Shang, S.; Danishefsky, S. J. *Angew. Chem. Int. Ed. Engl.* **2010**, *49* (49), 9500–9503.

(15) Chen, J.; Wang, P.; Zhu, J.; Wan, Q.; Danishefsky, S. J. *Tetrahedron* **2010**, *66* (13), 2277–2283.

(16) Yang, R.; Pasunooti, K. K.; Li, F.; Liu, X. W.; Liu, C. F. *J. Am. Chem. Soc.* **2009**, *131* (38), 13592–13593.

(17) Crich, D.; Banerjee, A. *J. Am. Chem. Soc.* **2007**, *129* (33), 10064 – +.

(18) Pandey, A. K.; Naduthambi, D.; Thomas, K. M.; Zondlo, N. J. *J. Am. Chem. Soc.* **2013**, *135* (11), 4333–4363.

(19) Gagné, M. R.; Ghosh, S.; Ingerman, L. a.; Frye, A. G.; Lee, S. J.; Waters, M. L. *Org. Lett.* **2010**, *12* (5), 1860–1863.

(20) Johnson, E. C. B.; Kent, S. B. H. *J. Am. Chem. Soc.* **2006**, *128* (20), 6640–6646.

(21) Collins, J. M.; Porter, K. A.; Singh, S. K.; Vanier, G. S. **2013**, 10–13.

CHAPTER 4: PROBING CATION-π INTERACTIONS WITH DIFFERENT HISTONE SEQUENCES AND ADDITONAL NONCOVALENT INTERACTIONS

4.1 Significance

A broader understanding of the mechanism employed by reader proteins in the recognition of their substrates can be gained via a comparative analysis of reader proteins with different binding pockets, as well as the impact that other noncovalent interactions have on both binding and the contribution of cation-π interactions to the overall recognition. Reader proteins that recognize lower methylation states contain salt bridges as part of their recognition motif (Figure 4.1). Others have different aromatic residues (both in number and type) in the binding pockets, and some have an incomplete pocket in the unbound state. Additionally, the contribution cation-π interactions to inhibitor binding is unknown.

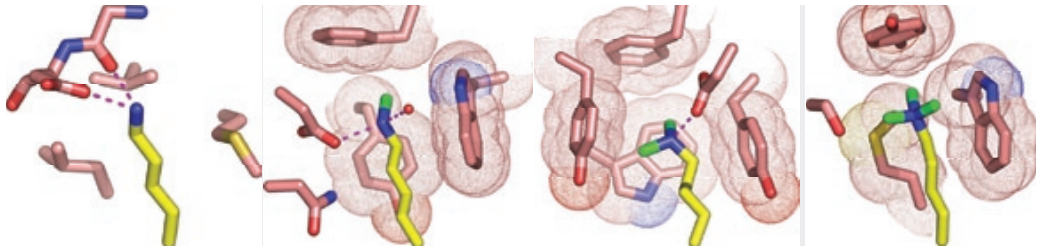


Figure 4.1. Binding pockets of various proteins that recognize different methylation states of lysine. A) PHD domain protein BHC80 binding unmethylated lysine B) MBT domain protein L3MBTL1 binding monomethyllysine C) Tudor domain protein 53BP1 binding dimethyllysine D) PHD domain protein ING2 binding Trimethyllysine. Reproduced with permission from *Nat. Struct. Mol. Biol.* **2007**, *14* (11), 1025–1040. Copyright 2007 Nature Publishing Group.[1]

By expanding these studies to include other reader proteins and the influence of other noncovalent interactions on this recognition, we can gain a broader understanding of this

mechanism. Understanding the similarities and differences between different reader proteins will aid in the development of probes with specificities for their desired targets.[2]

4.2 Reader Protein Inhibitors

Several inhibitors have been identified for different reader proteins, however it is unclear if cation-π interactions play the same role in binding inhibitors as it does in binding the natural substrates. An inhibitor for the chromodomain protein CBX7 contains a pyrrolidinelysine while some inhibitors for KMe2 binding proteins also contain this pyrrolidine ring, or a tert-butyl amine (Figure 4.2). These types of compounds, while still possessing a positive charge, place methyl groups or methylene groups with a smaller partial positive into the binding pocket than the methyl groups of methylated lysine (Figure 4.3). It is possible then, that the impact of cation-π interactions of these binding events are less than they are with the native substrate. Other noncovalent interactions may provide more a driving force for this recognition.

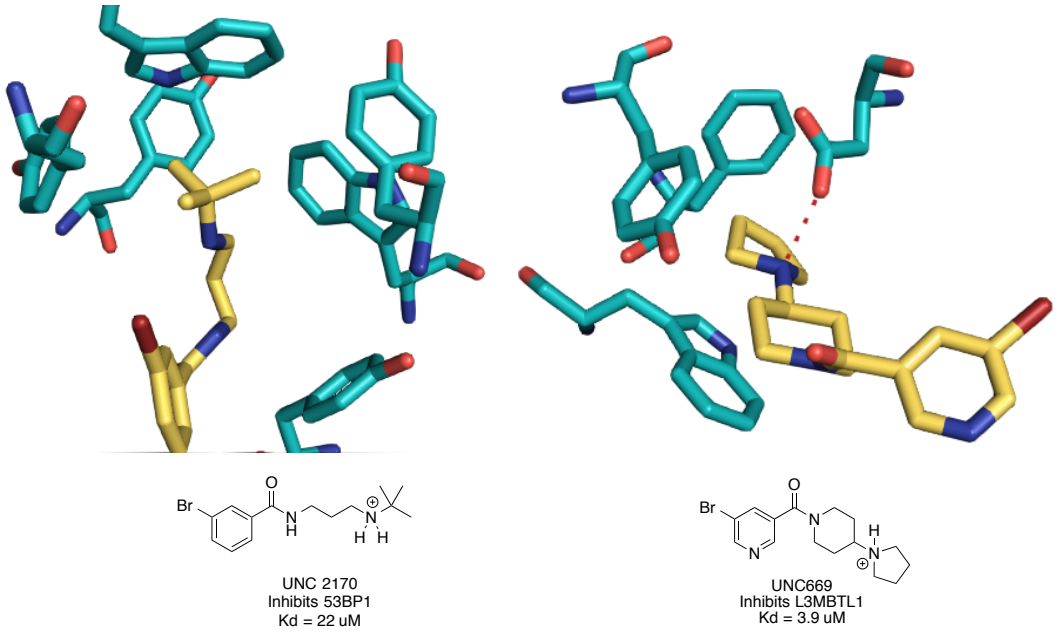


Figure 4.2. Inhibitors for KMe2 reader proteins (PDB:4RG2, 3P8H).[3,4]

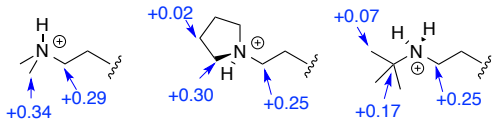


Figure 4.3. Partial charges on side chain of KMe2 (left) compared to the carbons of the amines present in various inhibitors (center, right).

4.2.1 Studying Inhibitor Interactions

Performing mutation studies to probe the cation-π interactions of the aromatic residues of binding proteins with their inhibitors in addition to the studies with their natural peptide substrates would provide a useful comparison of their binding mechanism. Additionally, these types of motifs can be incorporated in the peptide sequence to study their effect on the cation-π interactions with the HP1 protein, which we have already studied.

Efforts are currently underway to incorporate these side chains into peptides. This can be done either by synthesizing unnatural residues to incorporate via solid phase peptide synthesis, or by performing reactions to install the side chain after peptide synthesis.

One approach allowing for rapid synthesis of many different side chain mimics involves the use of the thiol nucleophile by incorporating a cysteine at the desired location. This cysteine can then be reacted with a variety of electrophiles to install a mimic of some of these side chains (Figure 4.4).

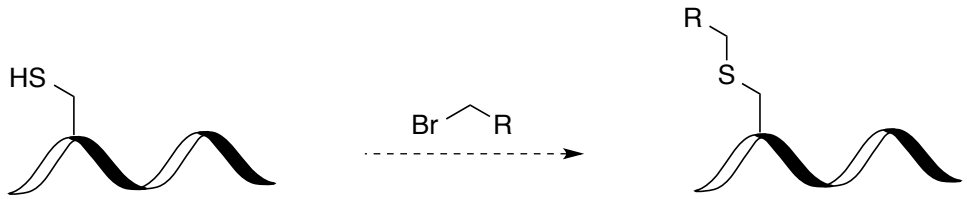


Figure 4.4. Approach for rapidly synthesizing peptides with a variety of side chain mimics, using cysteine mutation.

This method has been shown to be successful in mimicking various post translational modifications in proteins, such as acetylated lysine, as well as mono- di- and trimethyllysine.[5–7] By using pH control, alkylation can be done at cysteine, while avoiding reaction with amine nucleophiles that may also be present, providing a good way to achieve site selective incorporation of the modification.[5] This approach provides a rapid way to mimic some of the structural motifs found in inhibitors, and can provide an initial screen for their impact on the cation-π interactions with reader proteins.

4.3 Salt Bridges in Reader Proteins

While reader proteins of lower methylation states still have an aromatic cage to recognize the cation, a carboxylic acid is generally found in the binding pocket to form a salt-bridge with the NH groups in lower methylation states of Lys. For example, HP1 also recognizes dimethyllysine and contains a glutamic acid that forms a water mediated salt bridge with KMe2 in the binding pocket in addition to the cation-π interactions with the aromatic residues (Figure 4.5).

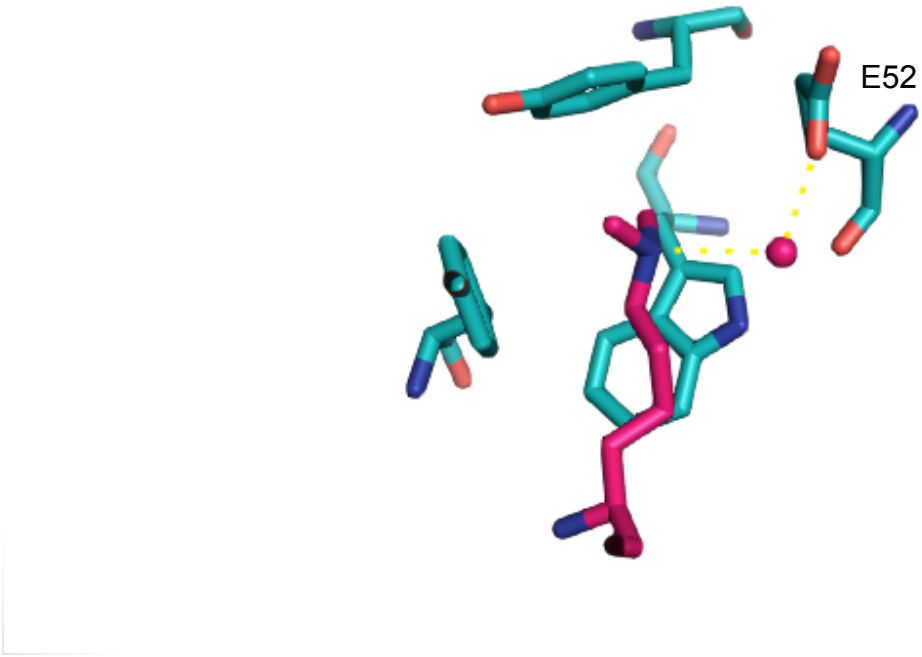


Figure 4.5. HP1 bound with KMe2 peptide containing a water-mediated salt bridge (shown in yellow) between E52 and KMe2. (PDB: 1KNA). [8]

4.3.1 E52Q Mutation of HP1

Previous studies by our lab have shown how important this salt bridge is for the recognition of dimethyllysine by the HP1 chromodomain. By replacing the glutamic acid with a glutamine, which contains an amide, the ability for the residue to hydrogen bond is reduced, and the charge needed for an electrostatic interaction is removed (Figure 4.6).

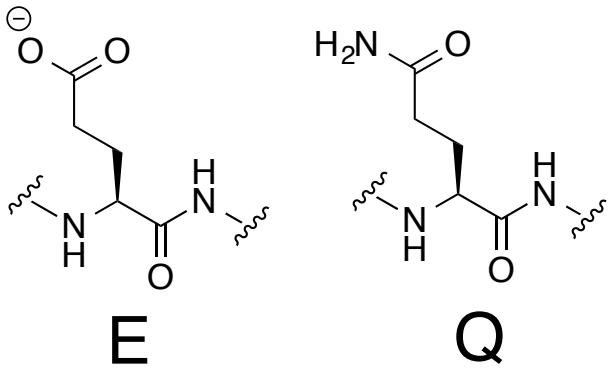


Figure 4.6. Glutamic acid (left) and glutamine (right).

Mutation of the carboxylic acid to the amide in the E52Q mutation does not reduce the ability of the protein to recognize trimethyllysine. However, the affinity for dimethyllysine is reduced approximately two-fold (Table 4.1).[9]

Table 4.1. Dissociation constants of E25Q mutant with di- and trimethyllysine

Protein	$K(Me)_3$ Kd (uM)	$K(Me)_2$ Kd (uM)
Wild Type	17	20
E52Q	15	52

Electrostatics do not play a role in the binding of KMe3 as there is no chance from WT HP1. However, the carboxylic acid is important for the water mediated hydrogen bond in the recognition of KMe2. With this added interaction, it is unknown whether cation-π interactions still play a similar role in the recognition of the lower methylation state. These interactions can be probed by mutating the binding pocket aromatics with electron withdrawing residues and measuring the effect on binding to dimethyllysine and compared to the trimethyllysine studies already done (Chapter 2). By using surface plasmon resonance (SPR) to measure these binding interactions, multiple peptides can be studied within a short period of time, as protein is loaded onto a chip with peptide flowed over to measure binding.

Binding studies with trimethyllysine containing peptide can be followed up with dimethyllysine containing peptide with the same proteins to further insure the results can be compared directly do each other. Initial work is underway to synthesize analogous peptides containing dimethyllysine for these studies.

4.4 Binding Neutral Analogs

An approach that has been used numerous times to probe cation-π interactions is the replacement of the cation with a neutral analog. It has already been shown by work done in our lab that binding of HP1 to histone peptide containing a *tert*-butyl analog of trimethyllysine significantly reduces the binding affinity (Figure 4.7).[10]

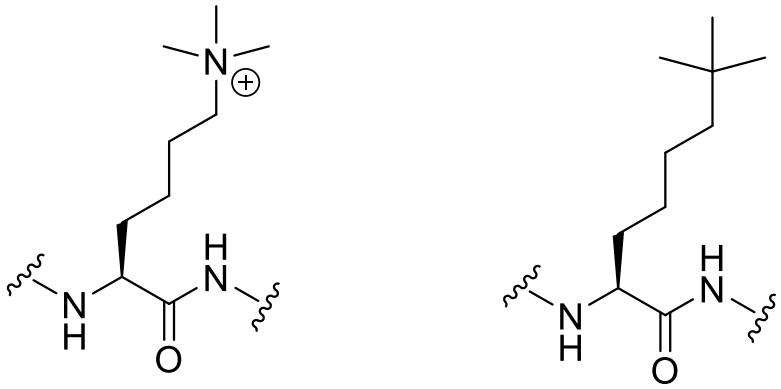


Figure 4.7. Neutral analog of trimethyllyinse used to probe cation-π interactions.

4.4.1 Neutral Analog and E52Q Mutant

This neutral analog can be further used to study the role of the carboxylic acid in the binding of trimethyllysine. As the E52Q mutant does not show a reduced binding to trimethyllyine, and electrostatics are not thought to play a role in the recognition of KMe_3, however, the electrostatics may have contributed to the weakened binding observed between HP1 and the neutral *tert*-butyl group. In the E52Q mutatnt, this repulsion is relieved, and it is possible the *tert*-butyl group will bind more strongly to the mutant than the WT protein. This

will provide more insight into the contribution of reduced cation-π interactions as apposed to potential added negative electrostatic interactions upon binding to the neutral side chain.

Stef Baril in the Brustad lab successfully expressed this mutant for binding studies. Initally, binding measurements were done by fluorescence anisotropy of wild type HP1 chromodomain (Figure 4.8) and the E52Q mutant (Figure 4.9) with H3K9Me3 peptide to verify we could repeat the results that had been previously reported.[9] These experiments, however, measured a binding affinity of 38 μM for WT and 76 μM for E52Q, which were not only different from each other, but also different from those reported in the literature and measured previously by us. We again determined that anisotropy was giving inconsistent results, and further experiments will be carried out using isothermal titration calorimetry (ITC) or SPR, which has been shown to be more reliable for measuring binding affinities with our HP1 mutants.

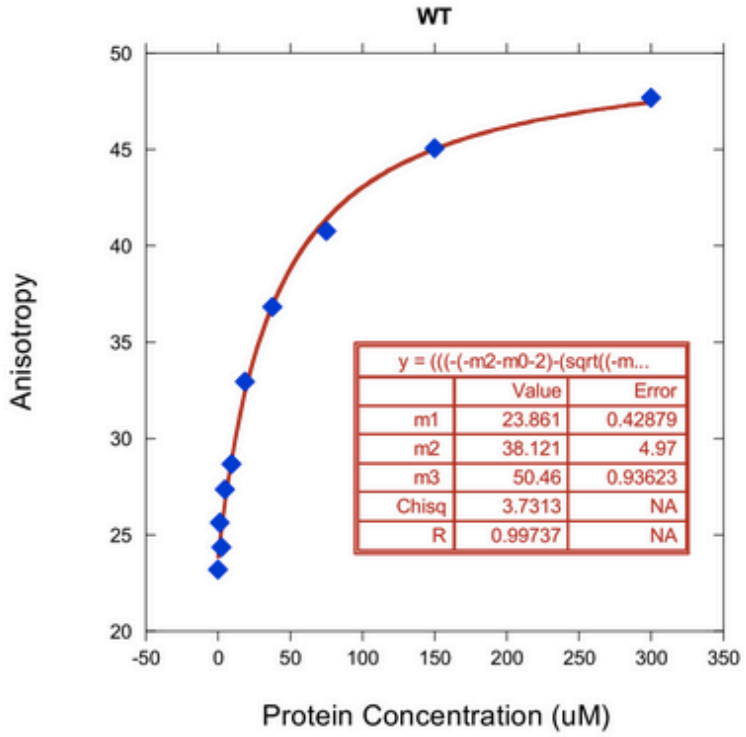


Figure 4.8. Binding curve of Wild Type HP1 titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8, 25 nM NaCl, 2 mM DTT.

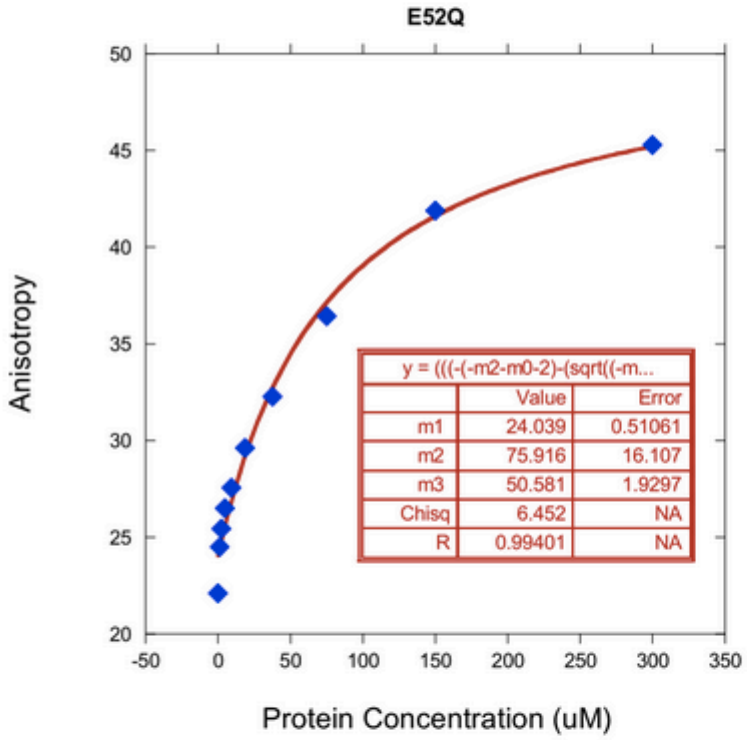


Figure 4.9. Binding curve of E52Q HP1 titrated into FAM labeled H3K9Me$_3$ peptide (FAM-TKQTARK(Me)$_3$STGGKA – 1 μM) in 50 mM phosphate buffer, pH 8, 25 nM NaCl, 2 mM DTT.

4.4.2 Neutral Analog and HP1 Mutants

Binding of the neutral analog with the mutated binding pockets of HP1 can also be analyzed. As SPR allows for easily studying the proteins loaded on the chip with multiple peptide ligands, this method will be useful for these experiments as well. Because there is no cation, it is not expected that this residue will exhibit the same linear free energy relationships that are observed upon binding of the mutants to trimethyllysine. It is possible that the mutations have no effect on binding. Alternatively, the binding may actually increase with the neutralization of the partial negative on the face of the ring. We will determine whether any trend exists with hydrophobicity or polarizability of the mutants as well.

4.4.3 Neutral Analog and Other Reader Proteins

Lastly, the neutral analog can be used to probe a variety of reader proteins. In addition to the work with HP1, this approach has been used with several other reader proteins, with some of them resulting in a similar drop in affinity to that observed with HP1, but not all, indicating a potential different recognition mechanism.[11] The Frye lab at UNC has a variety of proteins that can be screened with various peptides for binding (Table 4.2)

Table 4.2. Reader proteins the sequence of peptides they bind to.

Protein	Peptide	Peptide sequence
53BP1	H4K20Me2	KGGAKRHRK(Me2)VLRDNIQ-OH
L3MBTL1	H4K20Me1	KGGAKRHRK(Me1)VLRDNIQ-OH
L3MBTL3	H4K20Me2	KGGAKRHRK(Me2)VLRDNIQ-OH
MBTD1	H4K20Me1	KGGAKRHRK(Me1)VLRDNIQ-OH
CBX7	H3K9Me3	ARTKQTARK(Me3)STGGKAPRKQL-NH2
UHRF1	H3K9Me3	ARTKQTARK(Me3)STGGKAPRKQL- NH2
PHF23	H3K4Me3	ARTK(Me3)QTARKSTGGKAPRKQYT- NH2
JARID1A	H3K4Me3	ARTK(Me3)QTARKSTGGKAPRKQYT- NH2
PHF1	H3K36Me3	KSAPSTGGVK(Me3)KPHRYRPGTV- NH2
PHF19	H3K36Me3	KSAPSTGGVK(Me3)KPHRYRPGTV- NH2

The binding of these proteins can be studied with the replacement of the natural sequence with one containing the neutral analog. This would give a broader comparison across reader proteins, and reveal any potential different binding mechanisms. Potential differences between readers of trimethyllysine and lower methylation states can also be revealed. Initial peptide synthesis has been unsuccessful, as complex mixtures that are difficult to purify are often formed during synthesis of these longer histone sequences. While optimizing conditions for histone sequence synthesis, automated methylation of dimethyllysine was developed for a microwave peptide synthesizer.

4.5 Experimental

4.5.1 Peptide Synthesis

4.5.1.1 Synthesis of H3 1-15 K9Me2

H3 1-15 containing KMe3 at position 9 (ARTKQTARK$(Me)_2$STGGKAY) was synthesized by hand using Fmoc protected amino acids and Rink Amide resin on a 0.06 mmol scale. The amino acid residues were activated with HBTU (O-benzotriazole-N, N, N', N',-tetramethyluronium hexafluorophosphate) and HOBT (N-hydroxybenzotriazole) in the presence DIPEA (diisopropylethylamine) in DMF (N,N-dimethylformamide). Double couplings of 30 minutes each were used. Deprotections were carried out in 20% piperidine in DMF, twice for 15 minutes each.

Dimethyllysine was coupled using 2 equivalents of Fmoc-Lys(Me)2-OH□HCl for 5 hours. The resin was washed with DMF and peptide synthesis was continued as normal.

Cleavage from the resin was performed in 95:2.5:2.5 trifluoroacetic acid (TFA):water:triisopropylsilane (TIPS) for 4 hours. The TFA was evaporated and products were precipitated with cold diethyl ether. The peptides were extracted with water and lyophilized. The peptide was purified by reversed phase HPLC using a C-18 semipreparative column and a gradient of 0 to 100% B in 60 minutes, where solvent A was 95:5 water:acetonitrile, 0.1% TFA and solvent B was 95:5 acetonitrile:water, 0.1% TFA. HPLC traces were complex, and complete purification was difficult. After 3 purifications, impurities remained, and the peptide was not fully purified. The peptide was lyophilized, and identity was confirmed by MALDI mass spectrometry. Calculated: 1751 Observed: 1752.6

4.5.1.2 Synthesis of FAM-capped Peptides

H3K9Me3 peptide (residues 3-15) containing N-terminal 5,6-carboxyfluroscein (FAM) (FAM-TKQTARK$(Me)_3$STGGKA) was synthesized for fluorescence anisotropy using the same procedure. After coupling of dimethyllysine, the resin was washed with DMF and the residue was methylated to form trimethyllysine with 7-methyl- 1,5,7-triaza-bicyclo[4.4.0]dec-5-ene (MTDB, 10.8 uL, 0.075 mmol) and methyl iodide (37.4 uL, 0.6 mmol) in DMF (5mL) for 5 hours at room temperature. Peptide synthesis was then continued to finish the sequence. Prior to cleavage from the resin, FAM was coupled to the N-terminus using 2 eq FAM, 2 eq PyBOP, 2 eq HOBT, and 4 eq DIPEA. The coupling was bubbled with N_2 overnight in a foil wrapped reaction flask and kept in the dark. Cleavage from the resin and purification were performed as previously described. Peptide was confirmed by ESI-MS. Calculated: $[M^+]$ 1,733, $[M^+ + H^+]$: 867, $[M^+ + 2H^+]$: 578.33 Observed: $[M^+ + H^+]$: 866.9, $[M^+ + 2H^+]$: 578.39.

4.5.1.3 Synthesis of KMe3 and Microwave Methylation

H3K9Me3 peptide (ARTKQTARKSTGGKAY) was synthesized on a CEM Liberty1 Microwave Peptide Synthesizer using Fmoc protected amino acids and Rink Amide resin. The amino acid residues (5 eq) were activated with DIC (diisopropylcarbodiimide, 5 eq)) in the presence of Oxyma (Ethyl (hydroxyimino)cyanoacetate, 5 eq). Double couplings of 4 minutes each at 90 °C were used. Arginine and histidine residues were coupled at RT for 4 min and 50 °C for 8 min. Deprotections were carried out in 10% piperazine and 0.1 HOBt (hydroxy benzyl triazole) in 9:1 NMP:EtOH, twice for 60 seconds each at 90 °C. Coupling of Fmoc-Lys(Me)2-OH HCl was performed for 5 hours at room temperature and injected using a plastic insert containing the unnatural amino acid in place of an unused amino acid

bottle. Methylation immediately followed coupling, and was performed by placing the methylation solution (7-methyl- 1,5,7-triaza-bicyclo[4.4.0]dec-5-ene (MTDB, 10.8 uL, 0.075 mmol) and methyl iodide (37.4 uL, 0.6 mmol) in 2.5 mL DMF) at an amino acid position and running a coupling method without the injection of coupling reagents. The methylation was performed for 5 hours at room temperature. Coupling steps as normal continued after methylation. Cleavage and purification were performed as previously described. Peptide characterized by MALDI. Calculated: 1765 Observed: 1765.9.

4.5.2 Fluorescence Anisotropy Binding Experiments

Fluorescence anisotropy measurements were performed using a PolarStar omega plate reader from BMG Labtech. Chromodomain solutions of different concentrations were made by performing serial half dilutions of a concentrated stock. Concentrations were determined by UV-Vis absorbance at 280 nm (ε=17780). Chromodomain (25 µL) was added into 1 µM FAM labeled peptide (25 µM) in a 96 well plate in 50 mM phosphate buffer, pH 7.4, 25 mM NaCl, and 2 mM DTT). Samples were allowed to equilibrate at 25 °C for 30 minutes. The samples were analyzed using an excitation wavelength of 485 nm and an emission wavelength of 520 nm. The data were analyzed in Kaleidegraph using the equation:

$$r = \left[\left(\frac{p + [c] + K_D \pm \sqrt{(-p-[c]-K_D)^2 - 4(p \cdot [c])}}{2p}\right) \cdot (r_\infty - r_0)\right] + r_0$$

where r is fluorescence anisotropy, r_0 is the anisotropy of the free histone tail, r_∞ is the anisotropy of the fully bound histone tail, ρ is the total concentration of histone peptide, [c] is the total concentration of chromodomain, and K_d is the dissociation constant. The variables r_0, r_∞, and K_d were all treated as floating variables.

REFERENCES

(1) Taverna, S. D.; Li, H.; Ruthenburg, A. J.; Allis, C. D.; Patel, D. J. *Nat. Struct. Mol. Biol.* **2007**, *14* (11), 1025–1040.

(2) Herold, J. M.; Ingerman, L. A.; Gao, C.; Frye, S. V. *Curr. Chem. Genomics* **2011**, *5* (SPEC. ISSUE 1), 51–61.

(3) Perfetti, M. T.; Baughman, B. M.; Dickson, B. M.; Mu, Y.; Cui, G.; Mader, P.; Dong, A.; Norris, J. L.; Rothbart, S. B.; Strahl, B. D.; Brown, P. J.; Janzen, W. P.; Arrowsmith, C. H.; Mer, G.; McBride, K. M.; James, L. I.; Frye, S. V. *ACS Chem. Biol.* **2015**, *10* (4), 1072–1081.

(4) J. Martin Herold, Tim J. Wigle, Jacqueline L. Norris, Robert Lam, V. K. K.; Cen Gao, Lindsey A. Ingerman, Dmitri B. Kireev, Guillermo Senisterra, M.; Vedadi, Ashutosh Tripathy#, Peter J. Brown, Cheryl H. Arrowsmith, Jian Jin, W. P.; Janzen, and S. V. F. *J Med Chem.* **2011**, *54* (7), 2504–2511.

(5) Chalker, J. M.; Bernardes, G. J. L.; Lin, Y. A.; Davis, B. G. *Chem. - An Asian J.* **2009**, *4* (5), 630–640.

(6) Huang, R.; Holbert, M. A.; Tarrant, M. K.; Curtet, S.; Colquhoun, D. R.; Dancy, B. M.; Dancy, B. C.; Hwang, Y.; Tang, Y.; Meeth, K.; Marmorstein, R.; Cole, R. N.; Khochbin, S.; Cole, P. A. *J. Am. Chem. Soc.* **2010**, *132* (29), 9986–9987.

(7) Krall, N.; da Cruz, F. P.; Boutureira, O.; Bernardes, G. J. L. *Nat. Chem.* **2015**, *8* (November), 1–11.

(8) Jacobs, S. A.; Khorasanizadeh, S. *Science (80-.).* **2002**, *295* (5562), 2080–2083.

(9) Eisert, R. J.; Waters, M. L. **2011**, *27599*, 2786–2790.

(10) Hughes, R. M.; Wiggins, K. R.; Khorasanizadeh, S.; Waters, M. L. *Proc. Natl. Acad. Sci. U. S. A.* **2007**, *104* (27), 11184–11188.

(11) Kamps, J. J. a. G.; Huang, J.; Poater, J.; Xu, C.; Pieters, B. J. G. E.; Dong, A.; Min, J.; Sherman, W.; Beuming, T.; Matthias Bickelhaupt, F.; Li, H.; Mecinović, J. *Nat. Commun.* **2015**, *6*, 8911.