STRUCTURAL AND FUNCTIONAL CHARACTERIZATION OF THE TIP60 CHROMODOMAIN

A Thesis Submitted to the College of

Graduate Studies and Research

In Partial Fulfillment of the Requirements

For the Degree of Master of Science

In the Department of Biochemistry

University of Saskatchewan

Saskatoon

By

Tubagus Sifa Arrafi

PERMISSION TO USE

In presenting this thesis in partial fulfillment of the requirements for a Postgraduate degree from the University of Saskatchewan, I agree that the Libraries of this University may make it freely available for inspection. I further agree that permission for copying of this thesis in any manner, in whole or in part, for scholarly purposes may be granted by the professors who supervised my thesis work, or in their absence, by the Head of the Department or the Dean of the College in which my thesis work was done. It is understood that any copying or publication or use of this thesis or parts thereof for financial gain shall not be allowed without my written permission. It is also understood that due recognition shall be given to me and to the University of Saskatchewan in any scholarly use which may be made of any materials in my thesis.

Requests for permission to copy or make other use of material in this thesis in whole or in part should be addressed to:

Head of the Department of Biochemistry

University of Saskatchewan

Saskatoon, Saskatchewan S7N 5E5

ABSTRACT

The Tat-interactive protein of 60 kDa (Tip60) is a histone acetyltransferase enzyme that appears to have a wide range of acetylation targets, which include core histone proteins H2A and H4, transcription factors Myc and p53, the androgen receptor, and ATM kinase. Additionally, Tip60 appears to play a role in several cellular processes such as transcriptional regulation, DNA damage repair, chromatin remodeling, and apoptosis. Due to its diverse roles, the deregulation of Tip60 has been implicated in several human diseases, including Alzheimer's disease and some cancers. Several studies have been conducted in an attempt to elucidate the regulation of Tip60's acetyltransferase activity. Studies have suggested that the binding of Tip60's chromodomain to methylated lysine residues found on histone tails is important for targeting substrates and allosterically regulating the enzyme. This research aimed to determine the structure of the chromodomain, identify its binding partners, and elucidate the mechanism of binding. Ultimately, the research aimed to clarify how binding of the chromodomain to its partners would affect acetyltransferase activity. Through x-ray crystallography, the crystal structure of the *Drosophila melanogaster* Tip60 chromodomain was solved to a resolution of 1.59 Å. The binding partners of the chromodomain were revealed through the use of surface plasmon resonance and confirmed by isothermal titration calorimetry. The binding studies found that the chromodomain preferentially bound peptides, which corresponded to modifications found on the histone H4 N-terminal tail.

ACKNOWLEDGEMENTS

First, I would like to thank my supervisor, Dr. Stanley Moore for giving me the opportunity to pursue graduate studies in his lab. He was a fantastic mentor who knew when to offer words of encouragement and use his signature humour to keep the spirits high. His knowledge, guidance, and support have helped me become a better scientist and for that I will be forever grateful.

I would also like to thank the members of my advisory committee Drs. Jeremy Lee, Oleg Dmitriev, and Yuliang Wu for their knowledge and advice they have provided to me throughout my studies. Their suggestions and criticisms have helped me improve my knowledge and skills as a researcher, and for that I am grateful.

My research was made possible thanks to funding from CIHR and the Department of Biochemistry's devolved scholarship. I want to thank members of the Moore lab, including Yunhua Jia, Dakshata Ambilwade and Poonam Dhindwal for their assistance, support, advice, and friendship throughout my time in the lab. I would also like to thank the members of the Molecular Design cluster for their assistance, advice, and for making the lab a great place to work everyday. I would like to thank Jason Maley from the Saskatchewan Structural Science Centre for his assistance and advice with our surface plasmon resonance study.

Throughout my time here at the University of Saskatchewan I have met many great people and made some great friends. While I can't mention everybody, I would like to thank a few who were always there to help me decompress after a long week in the lab, so cheers to Devon Stumborg, Mathew Corrin, Jordan Bolt, and Jay Anderson. I would also like to thank the great friends I've made in my graduate studies including Brenden van Wyk, Jeremy Marshall, Kevin Voth, and Elisabet Jakova who all shared the joys (and pain) of research with me.

Finally, I would like to thank my parents, Shinta and Machdum, for their love and support throughout my education. I will be eternally thankful that you encouraged me to pursue my dreams. I would also like to thank my brother Hamada for his support and mild curiosity in my research. Last but not least I want to thank the love of my life, my wife, Paige, for all her love, support and words of encouragement. You provided the ray of sunshine that kept me smiling and laughing as I wrote this thesis, love you lots!

Cheers everyone!

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

18-crown-6	1,4,7,10,13,16-hexaoxacyclooctadecane
AcCoA	Acetyl Coenzyme A
APS	Ammonium persulfate
ATM	Ataxia-telangiectasia mutated kinase
CBP	CREB-binding protein
CREB	cAMP response element binding protein
CHES	*N*-Cyclohexyl-2-aminoethanesulfonic acid
CLS	Canadian Light Source
DSB	Double-strand break
DTT	Dithiothreitol
EDC	1-ethyl-3-(3-dimethylaminopropyl) carbodiimide
EDTA	Ethylenediaminetetraacetic acid
Esa1	Essential Sas2-related acetyltransferase 1
FPLC	Fast protein liquid chromatography
GNAT	Gcn5 related N-acetyltransferase
GST	Glutathione S-Transferase
$H3K4Me_1$	Monomethylated lysine 4 on histone H3
$H3K4Me_3$	Trimethylated lysine 4 on histone H3
$H3K9Me_3$	Trimethylated lysine 9 on histone H3
H3S10Ph	Phosphorylated serine 10 on histone H3
$H3K27Me_3$	Trimethylated lysine 27 on histone H3
$H3K36Me_1$	Monomethylated lysine 36 on histone H3

$H3K36Me_3$	Trimethylated lysine 36 on histone H3
H4K12Ac	Acetylated lysine 12 on histone H4
H4K16Ac	Acetylated lysine 16 on histone H4
$H4K20Me_1$	Monomethylated lysine 20 on histone H4
$H4K20Me_3$	Trimethylated lysine 20 on histone H4
HAT	Histone acetyltransferase
HDAC	Histone deacetylase
HP1	Heterochromatin protein 1
IPTG	Isopropyl β-D-1-thiogalactopyranoside
ITC	Isothermal titration calorimetry
MBT	Malignant brain tumour domain
MCS	Multiple cloning sites
MOF	Males-absent on the first
MPD	2-Methyl-2,4-pentanediol
MSL3	Male-specific lethal 3
MYST	MOZ, Ybf2/Sas3, Sas2, and Tip60 family
NMR	Nuclear magnetic resonance
OD_{600}	Optical density at 600 nm
p300	Protein of 300 kDa
PBS	Phosphate buffered saline
PEG	Polyethylene glycol
PHD	Plant homeodomain
PTM	Post-translational modification

R.M.S.D.	Root-mean-square deviation
RU	Response units
SEC	Size-exclusion chromatography
SDS-PAGE	Sodium dodecyl sulphate – polyacrylamide gel electrophoresis
SPR	Surface plasmon resonance
sulfo-NHS	N-hydroxysulfosuccinimide
TCEP	Tris(2-carboxyethyl)phosphine
TEMED	N,N,N’,N’ – tetramethylethylenediamine
Tip60	Tat-interactive protein of 60 kDa

1 INTRODUCTION

The DNA molecule is the building block of all living organisms. It provides instructions for the growth, development, and function of everything within an organism. Within eukaryotes, DNA is often condensed and coiled around histone proteins to form the fundamental unit of chromatin: the nucleosome. The organization and compaction of chromatin often poses a challenge to cellular replication, transcription, and repair machinery, which requires access to DNA. This problem is often solved through the use of post-translational modifications on histone proteins to alter their affinity for DNA. The addition of post-translational modifications is typically carried out by histone modifying enzymes, such as histone acetyltransferases. Additionally, the loss or deregulation of histone modifying enzymes is often implicated in human disease. Therefore, understanding the roles and mechanisms of a histone modifying enzyme's function and activity is an important step in identifying their role in human disease.

The histone acetyltransferase Tat-interactive protein of 60 kDa (Tip60) was found to possess a wide range of acetylation targets, which include core histone proteins, transcription factors, the androgen receptor, and ATM kinase. The diverse amount of acetylation targets also indicated that Tip60 is involved in numerous cellular processes, such as transcriptional regulation, DNA damage repair, chromatin remodeling, and apoptosis. It comes as no surprise that, due to Tip60's diverse roles, the deregulation of the protein is often implicated in several human diseases. Therefore, understanding how Tip60 is regulated and how it functions is an important step in understanding its role in human disease.

There have already been several studies conducted, which attempt to elucidate the regulation of Tip60's acetyltransferase activity. These studies suggest that the binding of the Tip60 chromodomain to methylated lysine residues found on histone tails is important for targeting substrates and allosterically regulating the enzyme. The structural basis of binding of the Tip60 chromodomain to methylated lysine residues and, subsequently, the allosteric regulation of its acetyltransferase are currently unknown. Previous studies on the Tip60 chromodomain have suggested that, based on a sequence alignment with the canonical methyllysine binding chromodomain of HP1α, the mechanism for binding methyllysine is achieved through the use of an aromatic cage. However, the Tip60 chromodomain shows poor sequence identity with the chromodomain of HP1α. Additionally, other closely related

chromodomains to Tip60, such as essential Sas2-related acetyltransferase 1 (Esa1) and males-absent on the first (MOF), show an absence of an aromatic cage. This casts doubt onto the data published in these prior studies of the Tip60 chromodomain. Unpublished data from the Moore lab has shown that the MOF chromodomain interacts with histone H4 tail peptides. Based on the sequence conservation between the Tip60 chromodomain and the closely related MOF chromodomain, and the unpublished data regarding the MOF chromodomain, we hypothesize that the Tip60 chromodomain will bind methyllysine found on histone H4 tails.

The research objectives aim to determine the structure of the Tip60 chromodomain through X-ray crystallography and elucidate the binding partners of the chromodomain through *in vitro* binding studies. Structure determination of the Tip60 chromodomain will utilize the chromodomains from the *Homo sapiens* Tip60 splice variant Tip60β and the *Drosophila melanogaster* Tip60 enzymes. This study also aims to identify and characterize the binding partners of the Tip60 chromodomain through *in vitro* binding studies, which include nuclear magnetic resonance (NMR), surface plasmon resonance (SPR), and isothermal titration calorimetry (ITC). Once the structure and binding partners of the chromodomain have been identified, we hope to determine the structural basis of binding through X-ray crystallography, NMR, and mutagenesis studies. Ultimately, this study aims to elucidate the mechanism for the allosteric regulation of the histone acetyltransferase (HAT) domain by the binding of the chromodomain to its targets.

2 LITERATURE REVIEW

2.1 Chromatin Packaging and Remodeling

2.1.1 Nucleosome Structure and Regulation

The genomic DNA of eukaryotes was found to be wrapped around an octamer of core histone proteins H2A, H2B, H3, and H4 to form the fundamental unit of chromatin: the nucleosome (Soria *et al.*, 2012). Approximately 145-147 base pairs of DNA are wrapped around the histone core 1.65 times in a left-handed superhelix (Figure 2.1) (Luger *et al.*, 1997). Each of the four core histone proteins are composed of two distinct domains: the histone tail and the histone fold, each of which have different functions (Andrews and Luger, 2011). A detailed look at the nucleosome core particle revealed that each histone protein possess a flexible N-terminal tail and that histone H2A also possess a flexible C-terminal tail (McGinty and Tan, 2015). The histone fold region of all core histones was also found to contain three α helices connected by two loops in an α1-L1-α2-L2-α3 configuration (Figure 2.2) (McGinty and Tan, 2015). The fold region in each of the core histones was found to contain a high level of structural similarity (Andrews and Luger, 2011).

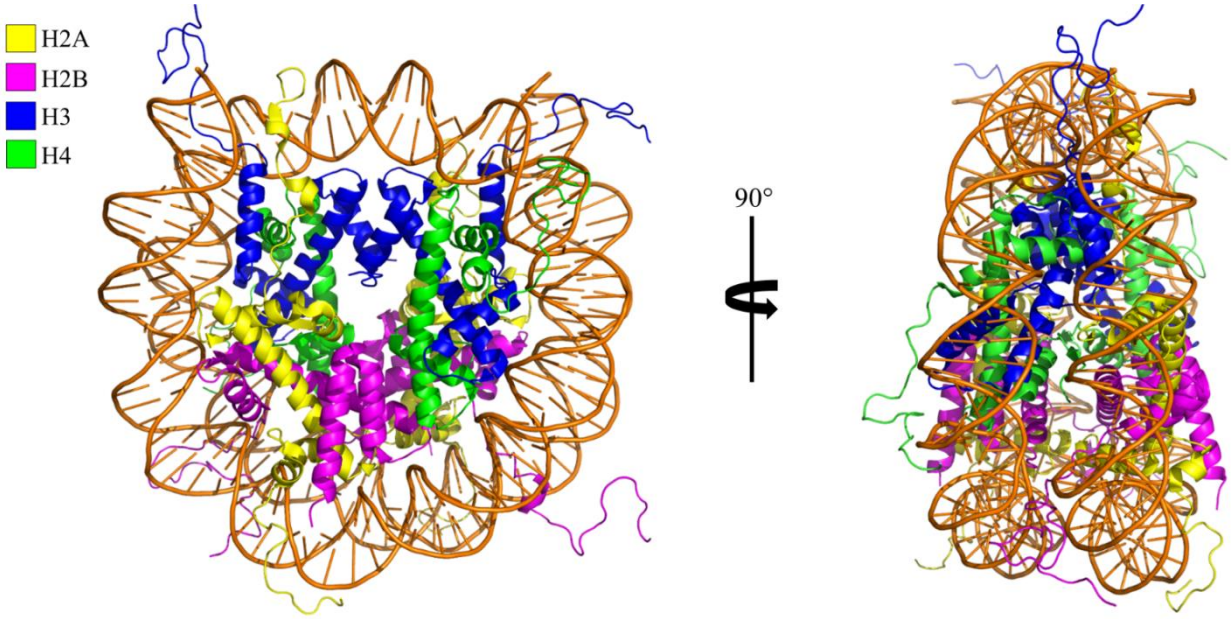

Figure 2.1 Overview of the nucleosome core particle crystal structure. The nucleosome core particle from *Xenpus laevis* (PDB ID 1KX5) solved to 1.9 Å by Davey *et al.*, 2002. The double strand DNA (orange) is shown wrapped around an octamer of core histone proteins H2A (yellow), H2B (magenta), H3 (blue), and H4 (green).

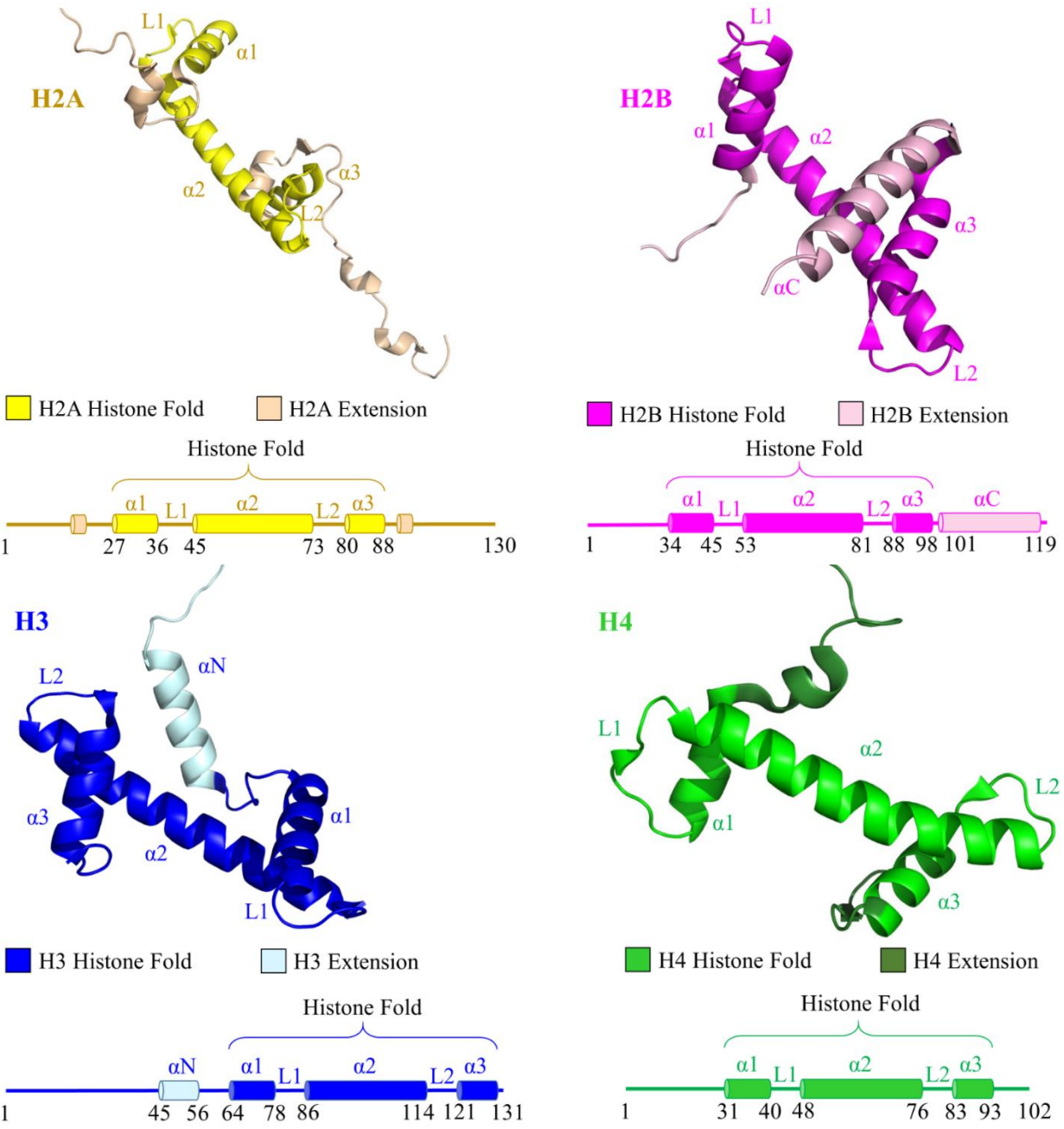

Figure 2.2 Structure and schematics of the core histone proteins. The crystal structure of the histone proteins from *Xenopus laevis* (PDB ID: 1KX5) was solved to 1.9 Å by Davey *et al.*, 2002. All four core histones are composed of a histone fold region, which contains an α1-L1-α2-L2-α3 motif. The histone extensions are shown for each histone protein in a different shade as compared to the histone fold motif. Each histone has an N-terminal tail extending from the histone fold and histones H2A and H2B contain a C-terminal tail as well. Figure adapted from McGinty and Tan, 2015.

Within the nucleosome core particle, the histone octamer is subdivided into four dimer pairs consisting of two H3-H4 pairs and two H2A-H2B pairs (Figure 2.3). These heterodimers are formed by complimentary histone folds interacting with one another. The antiparallel configuration of the heterodimer results in the L1 loop of one histone packing against the L2 loop of the complimentary histone (McGinty and Tan, 2015). The heterodimers have a crescent shaped appearance, which results in a convex surface that contains the L1L2 loops and the α1 helices, and creates a concave surface which contains the α2 and α3 helices (McGinty and Tan, 2015). This convex surface was found to carry a strong positive charge, which allows it to function as the primary DNA binding surface for each heterodimer (McGinty and Tan, 2015). To form the nucleosome, the two pairs of H3-H4 dimers form a tetramer through the interaction of a 4-helix bundle, formed between the α2 and α3 helices from the histone folds of H3 and H3' (Figure 2.4A) (Luger *et al.*, 1997). The H2A-H2B dimers were found to interact with the H3-H4 tetramer through a similar 4-helix bundle, formed between the α2 and α3 helices from the histone folds of H2B and H4 (Figure 2.4B) (Luger *et al.*, 1997). The two H2A-H2B dimers were found to only have limited interactions with one another (Andrews and Luger, 2011).

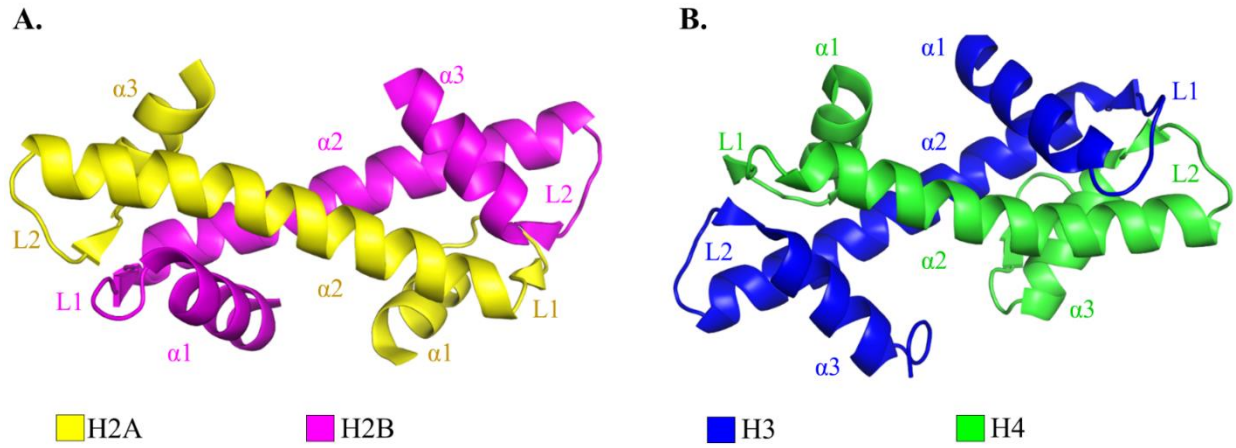

Figure 2.3 Crystal structure of the core histone heterodimers. The crystal structure of the histone fold region of the histone proteins from *Xenopus laevis* (PDB ID: 1KX5). The heterodimers are formed through the interaction of complimentary histone folds. The antiparallel configuration results in the L1 loop of one histone packing against the L2 loop of the complimentary histone. **A.** H2A-H2B dimer. **B.** H3-H4 dimer. Figure adapted from McGinty and Tan, 2015.

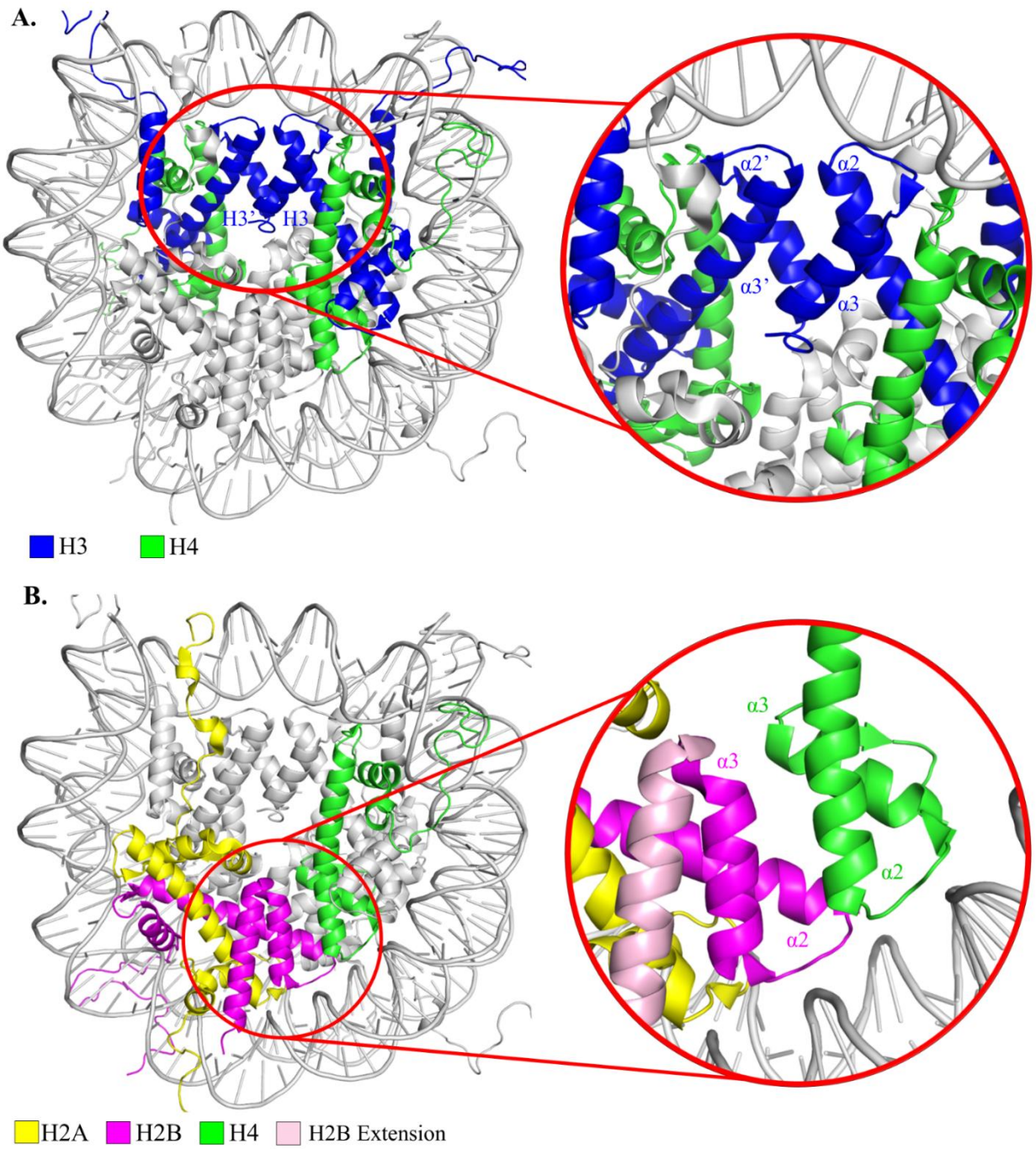

Figure 2.4 Four helix bundles found in the histone octamer. An overview of where these bundles are located in the nucleosome is shown on the left and a magnified view of the bundles is shown on the right. All structures are from *Xenopus laevis* (PDB ID: 1XK5). **A.** The H3-H4 dimers form a tetramer through a four helix bundle, which is formed by the α2 and α3 helices of the histone folds of histone H3 and H3' (shown in blue). **B.** The H2A-H2B dimers complete the octamer by forming a similar four helix bundle composed of the α2 and α3 helices of histones H2B (magenta) and H4 (green). To differentiate the αC helix from the α3 helix in histone H2B, the αC helix has been coloured light pink.

In addition to the four canonical core histone proteins, it was found that core histones H2A, H2B, and H3 possess different isoforms (referred to as histone variants) than those previously described (Venkatesh and Workman, 2015). Currently, no additional isoforms have been identified for histone H4 (Maze *et al.*, 2014). Histone variants have been found to be coded by different genes than those which code for the canonical histone proteins; it was found that the variants were highly conserved between different species (Luger *et al.*, 2012). The histone variants were found to differ from the canonical histones, either by a few amino acids or through the presence of additional domains (Venkatesh and Workman, 2015). Histone variants were found to be involved in the replacement of missing histones or transcribed specifically during the S-phase of the cell cycle to ensure high expression levels (Luger *et al.*, 2012; Venkatesh and Workman, 2015). The exchange of canonical histones with variants was found to result in the alteration of chromatin structure and dynamics, by affecting protein-protein interactions and post-translational modifications of histones (Maze *et al.*, 2014; Venkatesh and Workman, 2015).

The DNA in the nucleosome core particle was found to interact directly with the histone octamer. Approximately 121 base pairs of DNA interact with the histone fold regions of each of the four dimers (Davey *et al.*, 2002). This equates to approximately 27-28 base pairs of DNA interacting with each dimer and 4 base pairs of DNA to link each region (Luger *et al.*, 1997). The remaining 13 base pairs, located at each end of the DNA, were found to bind the αN helices at the N-terminus of each of the two H3 histones (Luger *et al.*, 1997). The binding of the DNA to the histone folds is predominantly facilitated by hydrogen bonding between the histone folds and the phosphodiester backbone of DNA (Luger *et al.*, 1997). Furthermore, it was also found that the interactions between the histones and DNA were also facilitated through water mediated hydrogen bonds (Davey *et al.*, 2002). While the majority of interactions occur through direct or water mediated hydrogen bonding between the histone and the phosphodiester backbone of DNA, it was found that approximately 20 side chains bind the DNA in the minor groove (Davey *et al.*, 2002). Additionally, it was found that the histone-DNA interaction was also mediated by ionic and non polar interactions (McGinty and Tan, 2015).

As the fundamental unit of chromatin, each nucleosome particle was found to be linked together by short segments of DNA, termed linker DNA, approximately 20-50 base pairs in length (Segal and Widom, 2009). The length of the linker region between each nucleosome core

particle was found to vary, depending on the cell type and species (Struhl and Segal, 2013). The linked nucleosome core particles form the 10 nm nucleosome fibre, which is often considered the first level of chromatin compaction and organization (Li and Reinberg, 2011). The primary mechanism for the formation of the 30 nm chromatin fibre was thought to be the interaction between the N-terminal tail of histone H4 and the acidic patch found on the canonical histone H2A (Tremethick, 2007). This interaction was found to be integral to forming the 30 nm fibre, as the deletion of the H4 tail or the replacement of canonical histone H2A with its variant H2A.Bbd, which features a truncated acidic patch, resulted in the inability to form the 30 nm fibre (Tremethick, 2007). Additionally, the presence of linker histones, such as H1 or H5, are thought to aid in the formation of the 30 nm chromatin fibres (Luger and Hansen, 2005; Segal and Widom, 2009). Finally, the 30 nm chromatin fibres are thought to undergo further condensation to form the highly ordered chromatin fibres (Luger *et al.*, 2012). An overview of chromatin structure and compaction is shown in Figure 2.5.

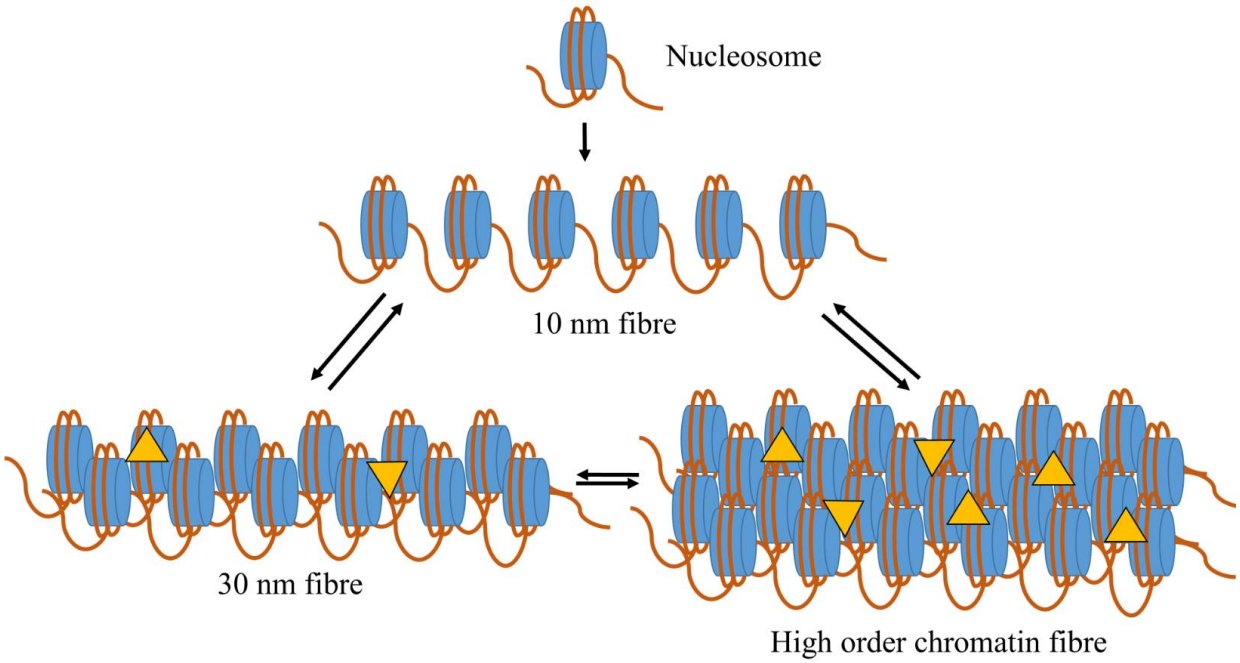

Figure 2.5 Overview of chromatin structure and compaction. The nucleosome is formed by the DNA (orange) wrapping around the core histone proteins (blue discs). Nucleosomes are then linked together by linker DNA to form the 10 nm fibre. The addition of linker histones (e.g. H1 or H5) and other scaffold proteins (yellow triangles), which interact with the nucleosomes, leads to the formation of the 30 nm fibre and high order chromatin fibres. Figure adapted from Luger *et al.*, 2012.

The highly ordered nature of chromatin organization and compaction can act as a barrier for the replication, transcription, and repair of DNA (Marmorstein and Trievel, 2009). There have been studies which have sought to determine the mechanism by which chromatin condenses and compacts. These studies have found that the mechanism by which the 30 nm chromatin fibre was constructed involved an acidic patch on canonical histone H2A and the N-terminal tail of histone H4, and that these components were integral to the fibre's formation (Tremethick, 2007). Therefore, alterations to these histones in the form of histone H2A variants and/or post-translational modifications to the H4 tail aid in the relaxation of the chromatin fibres (Luger *et al.*, 2012). The post-translational modification of histones and chromatin remodeling allows for the cellular machinery involved with replication, transcription, and repair to access the chromosomal DNA (Fischle, 2009).

2.1.2 Post-Translational Modifications of Histone Tails

The ability for cellular machinery to access DNA for replication, transcription, and repair is a vital cellular process. However, the highly ordered nature of chromatin organization and compaction serves as a barrier to the machinery to access the DNA. It was found that the structure and dynamics of chromatin could be regulated by the post-translational modifications (PTMs) of histone proteins (Musselman *et al.*, 2012). These modifications allow for the cellular machinery to access the DNA to perform necessary functions, such as replication, transcription, and DNA repair (Fischle, 2009). The process by which PTMs exert regulatory effects on chromatin structure and dynamics was found to be through two distinct methods. The first method of regulation was found to be through the direct disruption of histone interactions with other histones or DNA (Kouzarides, 2007). The second method of regulation by PTMs was through the recruitment of effector proteins or complexes (Musselman *et al.*, 2012). Additionally, it was found that a wide variety of histone PTMs exist to regulate different functions; these PTMs include, but are not limited to, acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP ribosylation, deimination, and proline isomerization (Table 2.1) (Kouzarides, 2007). It should be noted that intracellular signals govern the time, location, and type of PTMs observed, and thus not all modifications will be present at once (Kouzarides, 2007).

Table 2.1 Overview of histone modifications and the functions regulated. Table adapted from Kouzarides, 2007.

Modification Type	Residues Modified	Regulated Functions
Acetylation	Lysine	Transcription, repair, replication, condensation
ADP ribosylation	Glutamate	Transcription
Deimination	Arginine	Transcription
Methylation	Lysine Arginine	Transcription, repair Transcription
Phosphorylation	Serine Threonine	Transcription, repair, condensation Transcription, repair, condensation
Proline Isomerization	Proline	Transcription
Sumoylation	Lysine	Transcription
Ubiquitination	Lysine	Transcription, repair

The post-translational modifications, which possess the ability to directly disrupt interactions with other histones or DNA, must occur on the histones at locations that are relatively accessible for the modifying enzymes. As a result, the majority of PTMs identified on histones were found on the N-terminal tails (Kouzarides, 2007). Additionally, studies which removed the N-terminal histone tails found that the nucleosomes were unable to compact into the 30 nm chromatin fibres (Suganuma and Workman, 2011; Tremethick, 2007). This indicated that modifications made to histone tails could directly affect inter-nucleosome contacts (Suganuma and Workman, 2011). In addition to modifications on the histone tails, it was identified by mass spectrometry that PTMs were also present on the globular domains of the core histone proteins (Freitas *et al.*, 2004). Several of these PTMs were found to occur along the lateral surface of the core histones and, subsequently, many of these modified residues were found to be involved with interacting DNA (Cosgrove *et al.*, 2004). The lateral surface was found to be positively charged to facilitate binding to the negatively charged DNA backbone and, therefore, it was proposed that PTMs, which altered the charge of the lateral surface (e.g. acetylation and phosphorylation), could cause significant disruption to the interaction with DNA (Tropberger and Schneider, 2013).

In addition to the direct disruption of interactions, PTMs have been found to exert regulatory effects through the recruitment of effector proteins and complexes (Musselman *et al.*, 2012). These effector proteins and complexes typically possess histone binding domains, which recognize and bind only specific histone modifications (Campos and Reinberg, 2009). The recruitment of these effector proteins may result in the crosslinking of nucleosomes, the increased occupancy of the RNA polymerase complex, the recruitment of chromatin remodelers, or the recruitment of other chromatin modifying enzymes (Ruthenburg *et al.*, 2007). There are many examples of PTMs recruiting effector proteins and complexes to exert a regulatory effect.

As previously mentioned, there are a wide variety of histone modifications, which include, but are not limited to: acetylation, methylation, phosphorylation, ubiquitination, sumoylation, ADP ribosylation, citrullination, and proline isomerization (Kouzarides, 2007). Each modification possesses different properties, such as net charge and size, which can influence how it exerts its regulatory effects. For example, the acetylation of a lysine residue located on the histone H3 or H4 tail was proposed to neutralize the positive charge exhibited by the lysine residue and, subsequently, weaken its interaction with DNA (Zentner and Henikoff, 2013). Another example of PTMs that utilize charge manipulation to exert regulatory effects are the phosphorylation of serines, threonines, and tyrosines found on the N-terminal tails of histones (Bannister and Kouzarides, 2011). It has been proposed that the phosphorylation of histones could impart a negative charge on its modified residue, resulting in charge repulsion between the negatively charged phosphodiester backbone of DNA and the phosphorylated histone residue (Zentner and Henikoff, 2013). Additionally, the phosphorylation of histones has been implicated in the alteration of chromatin binding domains affinity for their targets. This has been observed with the phosphorylation of serine 10 on histone H3 (H3S10Ph), causing the release of the chromatin binding protein HP1 from the adjacent tri-methylated lysine nine on histone H3 ($H3K9Me_3$) (Zentner and Henikoff, 2013).

One of the more complex modifications found on histones is the methylation of lysine and arginine residues. The methylation of histones does not impart a charge to its modified residue; therefore, it exerts its regulatory effects in much less direct manner than charge manipulation (Zentner and Henikoff, 2013). Furthermore, histone methylation has been found to be much more complex, in that lysine residues can be mono-, di-, or tri-methylated on the ε-

amino group, and the arginine residues can be mono-methylated, symmetrically di-methylated, or asymmetrically di-methylated on the Ω-nitrogen, where each methylation state can correspond to a different regulatory outcome (Musselman *et al.*, 2012). For example, the mono-methylation of lysine 20 on histone H4 is associated with active gene transcription, whereas the tri-methylation of lysine 20 was associated with heterochromatin and gene repression (Barski *et al.*, 2007). A brief summary of the common modifications to the lysine residues located on the histone H3 and H4 N-terminal tails relevant to this research is shown in Figure 2.6.

Due to the diverse amount of histone modifications and the ability of the modifications to regulate different functions, it was required that there be an equally diverse amount of histone modifying enzymes and readers of their subsequent PTMs. These readers of PTMs are integral to the recruitment of the appropriate cellular machinery to chromatin, thus allowing a variety of cellular processes to be performed (Musselman *et al.*, 2012). Additionally, the misreading and/or deregulation of PTMs have been implicated in human disease (Musselman *et al.*, 2012). The readers of PTMs can be grouped into categories, depending on the modification with which they recognize. Some examples of PTM readers include methyllysine, methylarginine, acetyllysine, phosphoserine, and phosphothreonine readers (Musselman *et al.*, 2012; Taverna *et al.*, 2007). Table 2.2 outlines a few examples of specific PTM readers and their target modifications. This literature review will examine the chromodomain methyllysine reader in more detail in section 2.2.

As previously mentioned, to accommodate and regulate the wide variety of PTMs, there is an equally diverse amount of histone modifying enzymes. These enzymes serve to deposit or remove modifications on the histone residues. Examples of histone modifying enzymes include histone acetyltransferases (HATs), histone deacetylases (HDACs), lysine or arginine methyltransferases, serine/threonine kinases, ubiquitinases, and lysine demethylases (Kouzarides, 2007). It was reported that methyltransferases and kinases were the most specific of the modifying enzymes, as members of these two classes often only target one specific histone residue (Kouzarides, 2007). This literature review will only examine the histone acetyltransferase class of histone modifying enzymes.

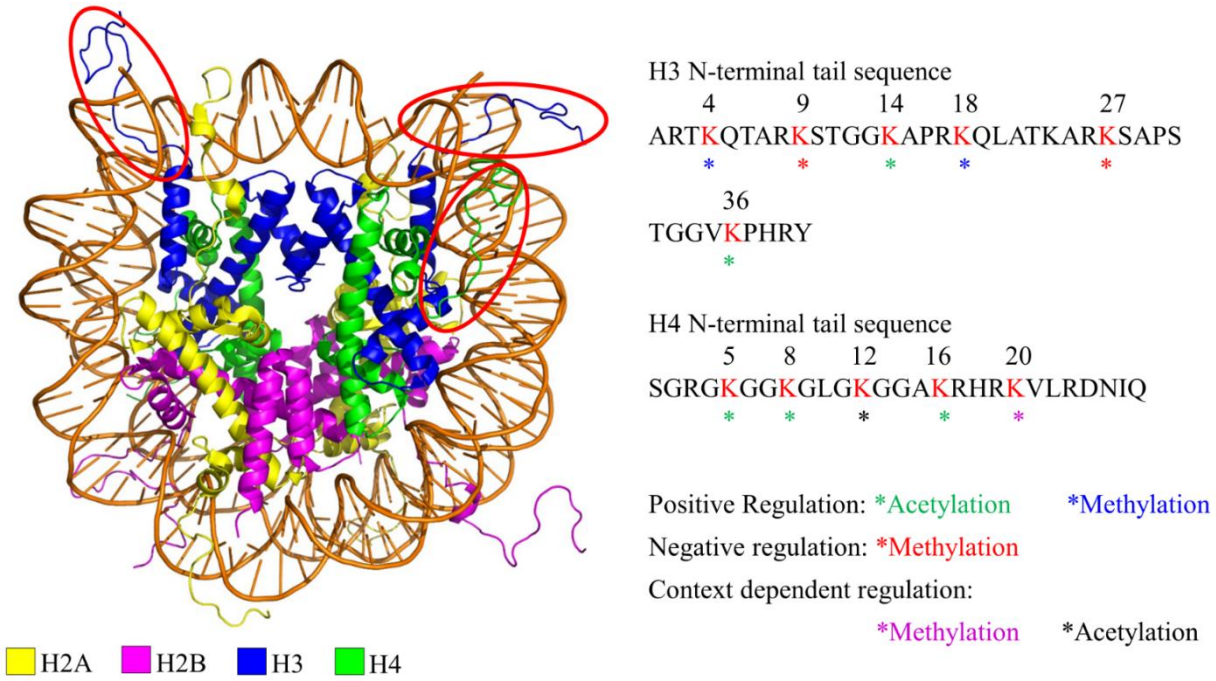

Figure 2.6 Summary of commonly modified lysine residues on the N-terminal tails of histone H3 and H4. The nucleosome core particle from *X. laevis* (PDB ID 1KX5) is shown on the left panel. The N-terminal tails of histone H3 (blue extension) and histone H4 (green extension) are circled in red. The right panel shows the N-terminal tail sequence of histones H3 and H4 from *H. sapiens*. The lysine residues are coloured red and their corresponding modification is listed below.

Table 2.2 Histone readers and the targeted modification

Modification Recognized	**PTM Reader**	**Specific Histone PTM Recognized**
Methyllysine	Chromodomain	$H3K9Me_3$, $H3K27Me_3$, $H3K4Me_1$, $H4K20Me_1$
	MBT	$H3KMe_1$, $H3KMe_2$, $H4KMe_1$, $H4KMe_2$
	PHD	$H3K9Me_3$, $H3K4Me_2$, $H3K4Me_3$
	Tudor	$H3K36Me_3$
Acetyllysine	Bromodomain	H3KAc, H4KAc, H2AKAc, H2BKAc
Methylarginine	Tudor	$H3RMe_2$, $H4RMe_2$
	WD40	$H3R2Me_2$
Phosphoserine/threonine	14-3-3	H3S10Ph, H3S28Ph
	BIR	H3T3Ph
Unmodified Histone	PHD	$H3_N$
	WD40	$H3_N$

2.2 The Chromodomain – A Reader of Post-Translational Modifications

The complexity of lysine methylation on histones (i.e. mono-, di-, or tri-methylation) results in numerous different methyllysine readers, capable of differentiating the lysine's methylation state. Some examples of methyllysine readers include chromodomains, malignant brain tumour (MBT) domains, plant homeodomain (PHD) fingers, and Tudor domains (Musselman *et al.*, 2012). In general, methyllysine readers were found to utilize two to four aromatic residues positioned in a cage-like orientation to bind methylated lysine through a cation-π interaction (Musselman *et al.*, 2012). This interaction involves the positively charged methylated lysine residue interacting with the partial negative charge from the quadrupole moment of an aromatic ring (Ma and Dougherty, 1997). The cation-π interaction was found to be the predominant driving force in the interaction between methyllysine and the aromatic cage of the methyllysine reader (Hughes *et al.*, 2007). Additionally, it was found that hydrophobic and van der Waals interactions also contribute to the binding of methyllysine to its reader protein (Hughes *et al.*, 2007).

The methyllysine readers are often divided into two classes based on structural similarity: the Royal superfamily and the PHD finger family (Taverna *et al.*, 2007). The Royal superfamily consists of the chromodomain, chromo-barrel domain, MBT domain, the PWWP domain, and the Tudor domain (Musselman *et al.*, 2012). Some members of the Royal superfamily possess a β-sheet composed of three or four curved β-strands and an adjacent α-helix (Musselman *et al.*, 2012; Yap and Zhou, 2011). The binding of the histone tail to the aforementioned members is thought to complete the β-barrel fold through the insertion of the histone tail between two of the existing β-strands (Figure 2.7A) (Taverna *et al.*, 2007). Other members of the Royal superfamily, such as the Tudor domains, possess a β-barrel composed of five β-strands (Musselman *et al.*, 2012). In this case, the histone tails are unable to insert between the β-strands and instead lie along the edge of the β-barrel with the methyllysine situated in the aromatic cage (Figure 2.7B) (Musselman *et al.*, 2012).

As previously stated, chromodomains are a member of the Royal superfamily of methyllysine readers. Typically, chromodomains are often divided into two major subgroups, canonical chromodomains and non-canonical chromodomains (Blus *et al.*, 2011). The canonical chromodomains are those which share high sequence identity and structural homology with the

HP1 chromodomain, whereas some non-canonical chromodomains adopt a similar fold to Tudor domains (Blus *et al.*, 2011). It should be noted, that other groups have further subdivided the non-canonical chromodomains into two additional groups: the chromodomain helicase-DNA binding subgroup and the chromo-barrel subgroup (Yap and Zhou, 2011). Furthermore, it has been established that chromodomains typically bind methyllysine, however, it has been found that some chromodomains also interact with non-histone proteins and nucleic acids (Yap and Zhou, 2011). Regardless of whether a chromodomain is classified as canonical or non-canonical, chromodomains are often found to be part of larger proteins or complexes with a variety of different functions (Blus *et al.*, 2011). Typically, the function of the chromodomain module is to target its associated protein or complex to methyllysine sites to facilitate regulatory effects (Yap and Zhou, 2011). Chromodomains have been found in proteins such as methyltransferases, DNA helicases, and histone acetyltransferases (Blus *et al.*, 2011).

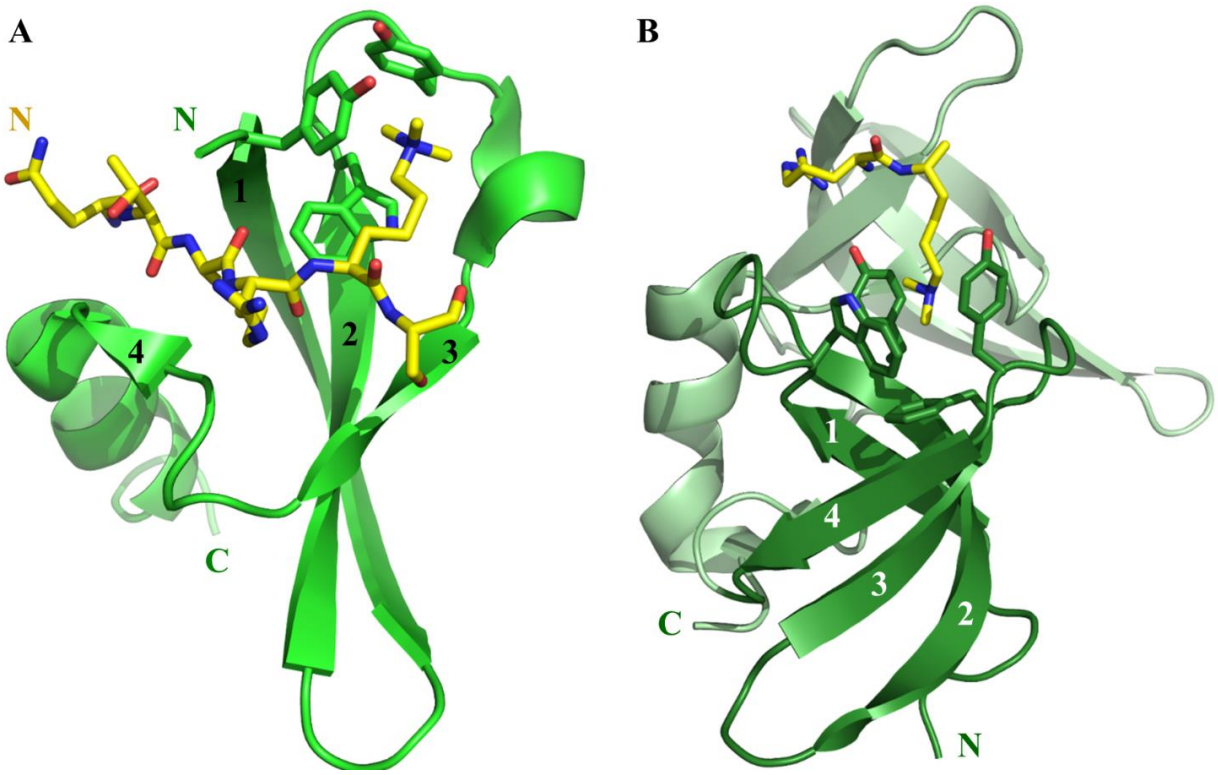

Figure 2.7 Overview of binding for members of the Royal Superfamily. **A.** The crystal structure of the HP1 chromodomain (PDB ID: 1KNE) bound to H3K9Me$_3$. The histone tail peptide corresponding to H3K9Me$_3$ (yellow) is shown inserted between β1 and β4 strands. The tri-methylated lysine is shown occupying the aromatic cage. **B.** The crystal structure of the tandem Tudor domain 53BP1 (PDB ID: 2IG0) bound to H4K20Me$_2$. The histone tail peptide corresponding to H4K20Me$_2$ (yellow) is shown to lie along the outer edge of the chromo-barrel (dark green) with the di-methylated lysine occupying the aromatic cage.

2.3 Histone Acetyltransferases

2.3.1 Histone Acetyltransferase Families & Complexes

One of the most well known and studied classes of histone modifying enzymes are the histone acetyltransferases (HATs). These enzymes have been found to catalyze the acetylation of the core histone proteins to exert regulatory effects on chromatin structure and gene transcription (Lee and Workman, 2007). There have been a large number of HATs identified across many different organisms, which are often organized into at least three distinct families based on the sequence conservation of the catalytic HAT domain (Marmorstein and Trievel, 2009). Within the catalytic region of the HAT domain it was found that only the acetyl coenzyme A (AcCoA) binding site motif was highly conserved across all HAT families (Roth *et al.*, 2001). The regions, which flank the highly conserved AcCoA binding motif, were found to be very diverse and, therefore, it is the primary basis for assigning a specific HAT to one of the three distinct HAT families (Marmorstein and Trievel, 2009). The three most well characterized HAT families include the GNAT (Gcn5 related N-acetyltransferase) family, the MYST (named after members **M**OZ, **Y**bf2/Sas3, **S**as2, and **T**ip60) family, and the p300/CBP family (Carrozza *et al.*, 2003; Marmorstein and Trievel, 2009). In this section, a brief overview of the multi-subunit HAT complexes and the three major HAT families are discussed. The catalytic mechanism and substrate binding of HATs is discussed in detail in section 2.3.2.

The majority of HATs have been found to be incorporated into larger multi-subunit complexes, which allow for the ability to acetylate free and nucleosomal histones (Carrozza *et al.*, 2003). These large multi-subunit complexes typically contained both enzymatic and non-enzymatic subunits (Table 2.3) (Marmorstein, 2001). It was found that these multi-subunit complexes were often more active than an individual catalytic subunit operating alone (Yang, 2004). Additionally, it was found that the non-enzymatic subunits could recruit substrates to the multi-subunit complexes (Yang, 2004). This suggested that the composition of the different multi-subunit complexes regulated the enzymatic activity of the catalytic subunit, determined the substrate specificity, and subsequently allowed for the targeting of specific genes (Carrozza *et al.*, 2003; Yang, 2004). Furthermore, in regards to HATs, it was found that a HAT can be present in several distinct multi-subunit complexes (Yang, 2004).

The GNAT family of HATs were originally identified as aminoglycoside acetyltransferases (Vetting *et al.*, 2005). HATs have subsequently been found to be involved in the regulation of cellular growth and development, transcriptional activation, and DNA repair (Carrozza *et al.*, 2003). Some of the most well known members of the GNAT family include Gcn5, PCAF, Elp3, Hat1, Hpa2 and Nut1 (Lee and Workman, 2007). The members of the GNAT family all share sequence and structural similarity to the Gcn5 acetyltransferase (Carrozza *et al.*, 2003). Structural and sequence data of GNAT family members have revealed that members contain a GNAT fold, which is composed of six or seven anti-parallel β-strands and four α-helices arranged as shown in Figure 2.8A (Vetting *et al.*, 2005). Within the GNAT fold region, four conserved motifs have been identified and labeled A-D (appearing in the order C, D, A, B) (Wolf *et al.*, 1998). It was found that motif A is highly conserved and contains residues from the β4 strand to the α3 helix, which are integral for AcCoA binding, motif B contains residues in the α4 helix, which also contributes to AcCoA binding, motif D, which contains residues from β2 and β3 strands provides stability for the protein, and motif C was found to be located at the N-terminus of some GNAT family members (Srivastava *et al.*, 2014). Additionally, it was found that the GNAT fold possess a V-shaped opening between the β4 and β5 strands, which is important for AcCoA binding (Figure 2.8B) (Srivastava *et al.*, 2014).

Table 2.3 Subunits found in multi-subunit complexes

Subunit	**Function**
Enzymatic Subunit	
Acetyltransferase	Lysine acetylation
Deacetylase	Lysine deacetylation
Methyltransferase	Arginine and lysine methylation
ATPase	ATP-dependent nucleosome remodeling
Kinase	Serine phosphorylation
Non-enzymatic Subunit	
Bromodomain	Acetyllysine recognition
Chromodomain	Methyllysine recognition
Zinc binding domain	Protein recognition
SANT domain	DNA binding

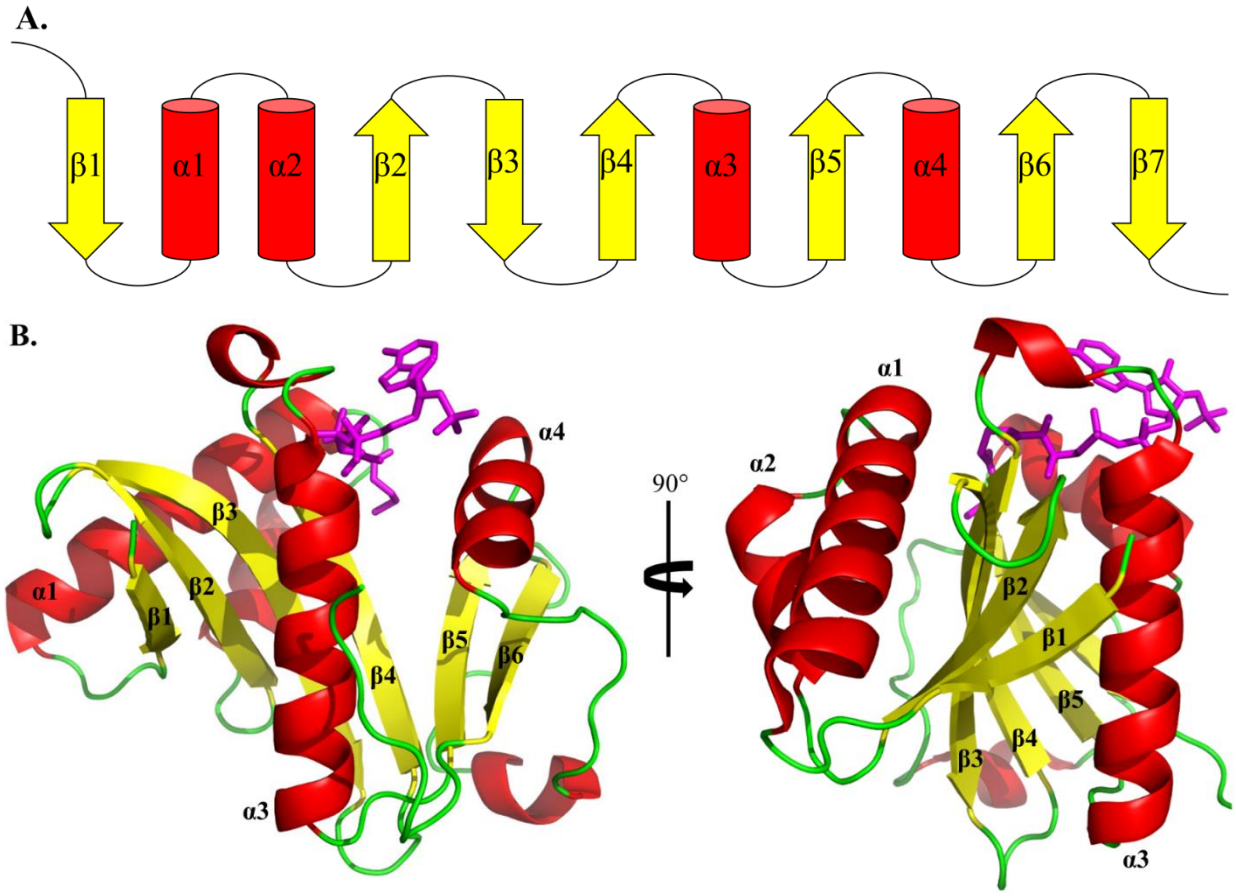

Figure 2.8 Structural overview of the GNAT family. **A.** Schematic of the GNAT fold, which is composed of 6 or 7 β-strands (yellow) and 4 α-helices (red). **B.** Crystal structure of Gcn5 from *Tetrahymena thermophila* bound with acetyl-CoA (magenta) solved to 2.3 Å (PDB ID: 1PUA). AcCoA sits between the β4 and β5 strands and contacts the α3 and α4 helices. Figure adapted from Srivastava *et al.*, 2014.

As one of the largest families of HATs, the members of the MYST family have been found to be well conserved from yeast to humans and involved in a wide range of functions, which include DNA repair, gene regulation, and cell cycle regulation (Yuan *et al.*, 2012). Some of the most prominent members of the MYST family include Esa1, Sas2, Sas3, MOF, Tip60, MOZ, and HBO1 (Sapountzi and Cote, 2011). All members of the MYST family share a highly conserved MYST domain, which is composed of an AcCoA binding motif, similar to the one found in the GNAT family of acetyltransferases, and a C_2HC zinc finger (Utley and Cote, 2003). The MYST domain was found to be composed of 13 β-strands and seven α-helices (Figure 2.9A) (Yan *et al.*, 2000). Coenzyme A was found to bind between the α3 and α4 helices; a glutamate and cysteine residue in the active site were found to mediate substrate acetylation (Yan *et al.*, 2000; Yan *et al.*, 2002). Furthermore, it was found that the autoacetylation of an active site lysine

was required for acetyltransferase activity (Yuan *et al.*, 2012). Outside of the highly conserved MYST domain, members of the MYST family have been found to possess additional conserved domains, such as a chromodomain, plant homeodomain (PHD) finger, or an additional zinc finger (Figure 2.9B) (Carrozza *et al.*, 2003; Yang, 2004). The MYST family can be divided into additional subgroups based on which of the aforementioned conserved domains is present (Carrozza *et al.*, 2003). These conserved domains serve specific functions, such as binding substrates or potentially regulating the catalytic activity of the HAT domain (Yang, 2004).

The p300/CBP family of HATs was found to be present in many multicellular organisms ranging from worms to humans (Kalkhoven, 2004). The two primary members p300 (protein of 300 kDa) and CBP (CREB-binding protein) were originally found to bind adenoviral E1A and the cAMP response element binding protein (CREB) respectively (Chan and La Thangue, 2001). Since the initial discovery, it was found that p300 and CBP possessed HAT activity and that the HAT domains possessed greater than 90% sequence identity (Wang *et al.*, 2008a). Additionally, p300 and CBP possess many conserved functional domains, including three cysteine-histidine (CH) rich domains, a bromodomain, a KIX domain, and a nuclear coactivator binding domain (NCBD) (Figure 2.10) (Delvecchio *et al.*, 2013). These functional domains have been found to be involved in the interaction between p300/CBP with transcription factors, transcription machinery, and other coactivators (Delvecchio *et al.*, 2013). Additionally, studies have found that p300 and CBP are able to acetylate the four core histone proteins as well as 70-plus proteins (Wang *et al.*, 2008a).

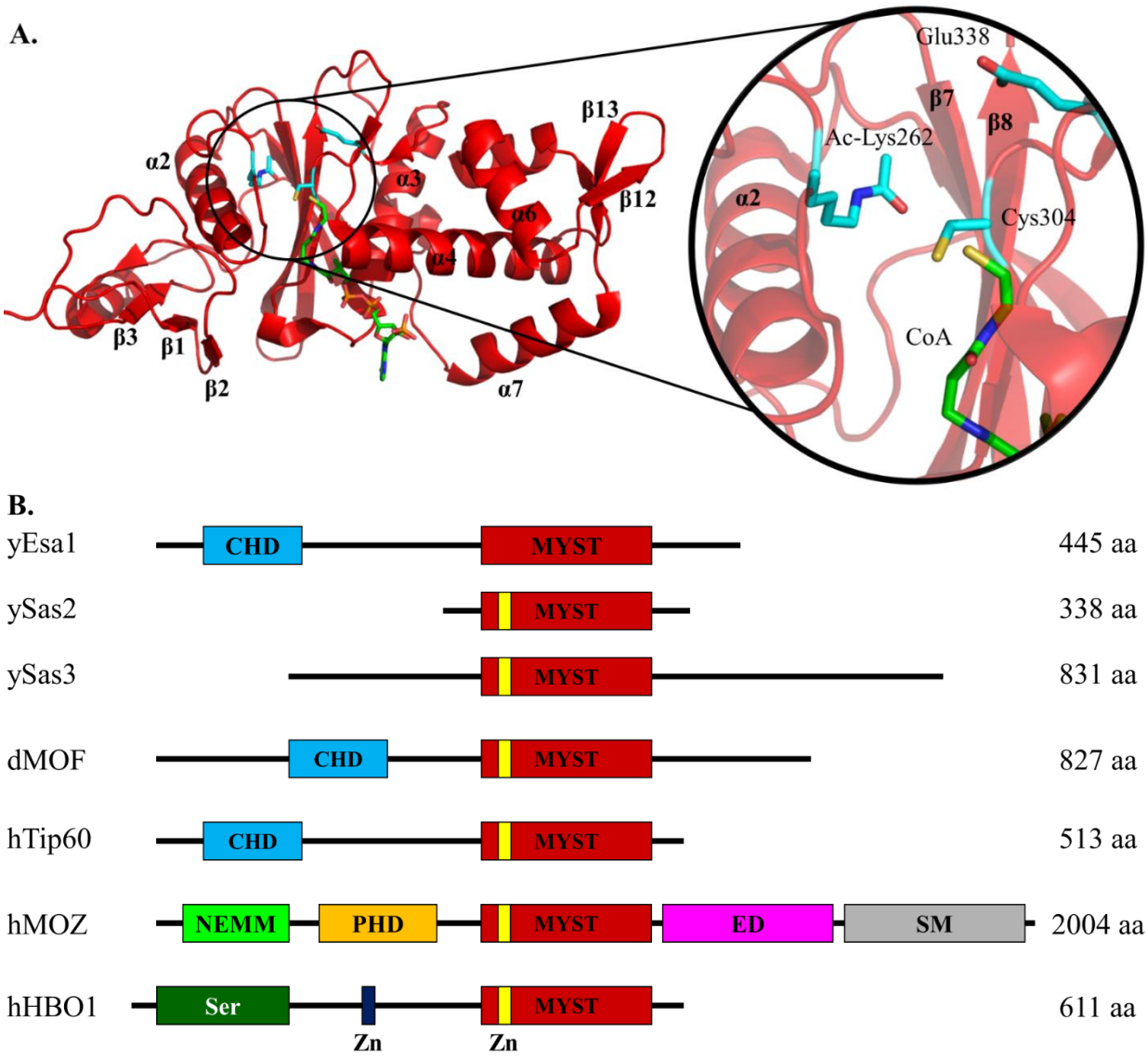

Figure 2.9 Overview of the MYST HAT family. **A.** Structural overview of the conserved MYST domain from *S. cerevisiae* Esa1 (PDB ID: 3TO7). A close up of the active site is shown in the left panel. The residues required for acetyltransferase activity, which include an autoacetylated lysine, a glutamate, and a cysteine are shown in ball-and-stick representation (cyan). CoA (green) is shown bound to the HAT domain. **B.** Schematic overview of the MYST family. The conserved MYST domain (red) contains the catalytic HAT domain and in most members also includes a Zinc finger (yellow). Many members possess additional functional domains, such as a chromodomain (blue), PHD fingers (orange), or an additional zinc finger (dark blue). Figure adapted from Carrozza *et al.*, 2003.

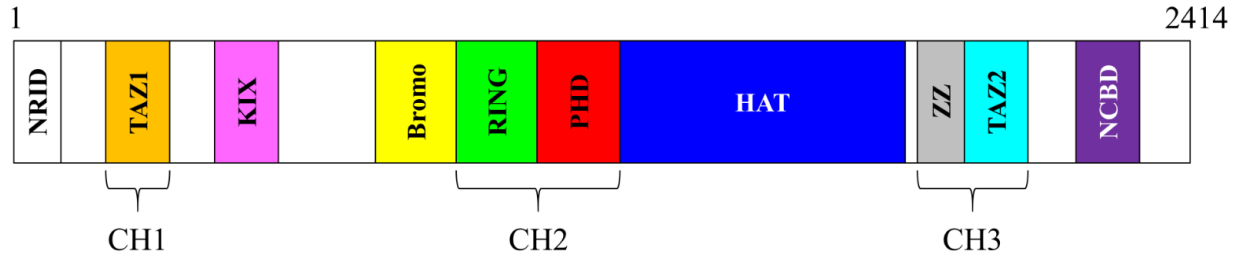

Figure 2.10 Schematic overview of the p300/CBP family. Three cysteine-histidine rich regions (CH1, CH2, CH3) are found in p300. There are four transactivation domains found in p300, which include: TAZ1 domain (orange), the CREB interacting KIX domain (pink), the CH3 region which incorporates the ZZ-type zinc finger (grey) and TAZ2 domain (cyan), and the NCBD domain (purple). These transactivation domains serve to mediate interactions between p300 and other transcription factors. The core of p300 is composed of a bromodomain (yellow), a RING domain (green), a PHD finger (red), and the catalytic HAT domain (blue). Figure adapted from Delvecchio *et al.*, 2013.

2.3.2 Substrate Binding and Catalytic Mechanism

Histone acetyltransferases are known to catalyze the acetylation of core histones by transferring the acetyl group from AcCoA to the ε-amino group of lysine residues found on the histone proteins (Roth *et al.*, 2001). The transfer of the acetyl group to the lysine residue neutralizes the positive charge of the lysine and disrupts the interaction with DNA and other chromatin binding proteins (Berndsen and Denu, 2008). As previously discussed, HATs can be grouped into at least three distinct families, each of which have a preference for different substrates and potentially different catalytic mechanism (Marmorstein and Roth, 2001). An overview of the how each family of HATs binds its substrate and the catalytic mechanism involved is discussed here.

The members of GNAT family were found to utilize a ternary complex mechanism during catalysis (Hodawadekar and Marmorstein, 2007). It was proposed that initially, before the start of catalysis, both the substrate and AcCoA bind the HAT domain (Berndsen and Denu, 2008). Following binding, a conserved glutamate residue deprotonates the ε-amino group of the target lysine. This is immediately followed by the direct nucleophilic attack on the acetyl group of the bound AcCoA, which results in the acetylation of the target lysine (Figure 2.11) (Hodawadekar and Marmorstein, 2007).

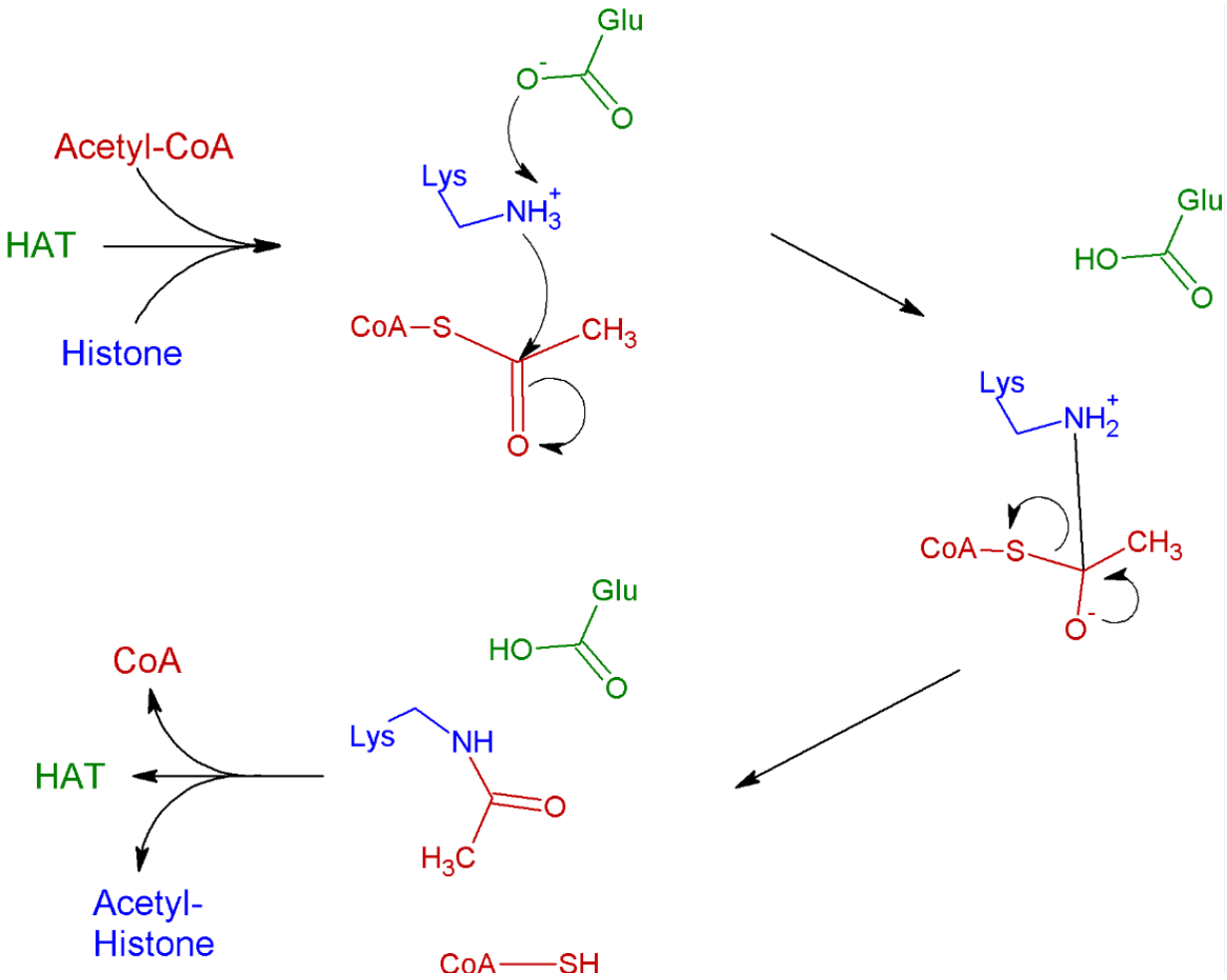

Figure 2.11 Proposed catalytic mechanism for the GNAT family. Initially, AcCoA (red) and the lysine substrate typically from a histone protein (blue) bind the HAT domain at the active site (green). The Glu residue from the HAT deprotonates the Lys substrate, which then attacks the carbonyl carbon of AcCoA and results in the acetylation of the Lys residue. Figure adapted from Berndsen and Denu, 2008.

It was proposed that the MYST family of HATs utilized a different catalytic mechanism than the GNAT family. Studies on the MYST family member essential Sas2-related acetyltransferase 1 (Esa1) from *S. cerevisiae* revealed that a ping-pong catalytic mechanism was employed (Yan *et al.*, 2002). However, a second study found that Esa1 utilized the ternary complex mechanism similar to the one employed by the GNAT family (Berndsen *et al.*, 2007). One possible reason for this discrepancy is that both studies utilized different Esa1 constructs, where Yan *et al.* utilized a truncated version and Berndsen *et al.* utilized full length Esa1 and Esa1 complexed with two additional proteins, Epl1 and Yng2. In the proposed ping-pong mechanism the HAT domain of the MYST protein binds AcCoA and the sulfhydryl group of a cysteine residue attacks the acetyl group of AcCoA, which yields an acetylated cysteine

intermediate (Berndsen and Denu, 2008). Once CoA was released, a lysine residue from the target substrate could bind the active site. The lysine residue is then deprotonated by a glutamate residue, which in turn causes the lysine nucleophile to attack the acetylated cysteine (Figure 2.12) (Berndsen and Denu, 2008).

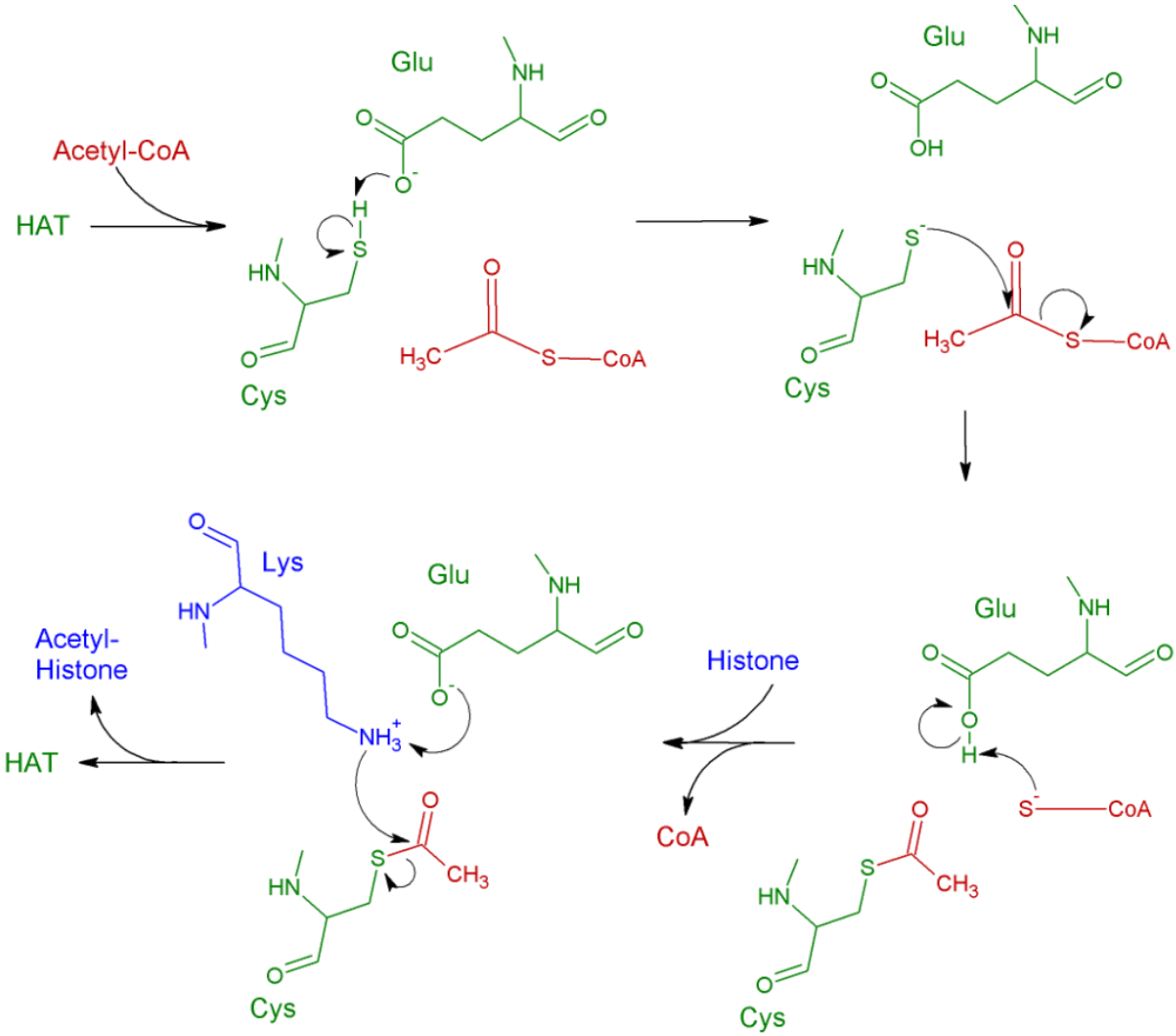

Figure 2.12 Proposed ping-pong catalytic mechanism for the MYST HAT family. Initially, the AcCoA (red) binds the HAT domain at the active site. The sulfhydryl group of a Cys residue (green) within the active site attacks the acetyl group of AcCoA, which yields and acetylated-Cys intermediary. CoA leaves the active site and the lysine substrate, typically from a histone protein, enters the active site. The Glu residue (green) in the active site deprotonates the Lys residue (blue), which subsequently attacks the acetylated Cys residue. This results in the acetylation of the lysine substrate. Figure adapted from Yan *et al.*, 2002.

The p300/CBP family was proposed to possess another distinct mechanism for catalysis. It was proposed that based on the structure of the active site, the p300/CBP family utilized the unconventional Theorell-Chance catalytic mechanism (often nicknamed 'hit-and-run' mechanism), which does not produce a ternary complex (Liu *et al.*, 2008). In this mechanism it was proposed that AcCoA binds to the HAT domain first. The positively charged lysine substrate then associates with the negatively charged P1 pocket of p300 and upon transfer of the acetyl group the substrate dissociates from the HAT domain (Wang *et al.*, 2008a). Furthermore, this proposed mechanism suggested that the indole side chain of a tryptophan residue guides the lysine substrate to attack AcCoA through van der Waals forces (Figure 2.13) (Liu *et al.*, 2008).

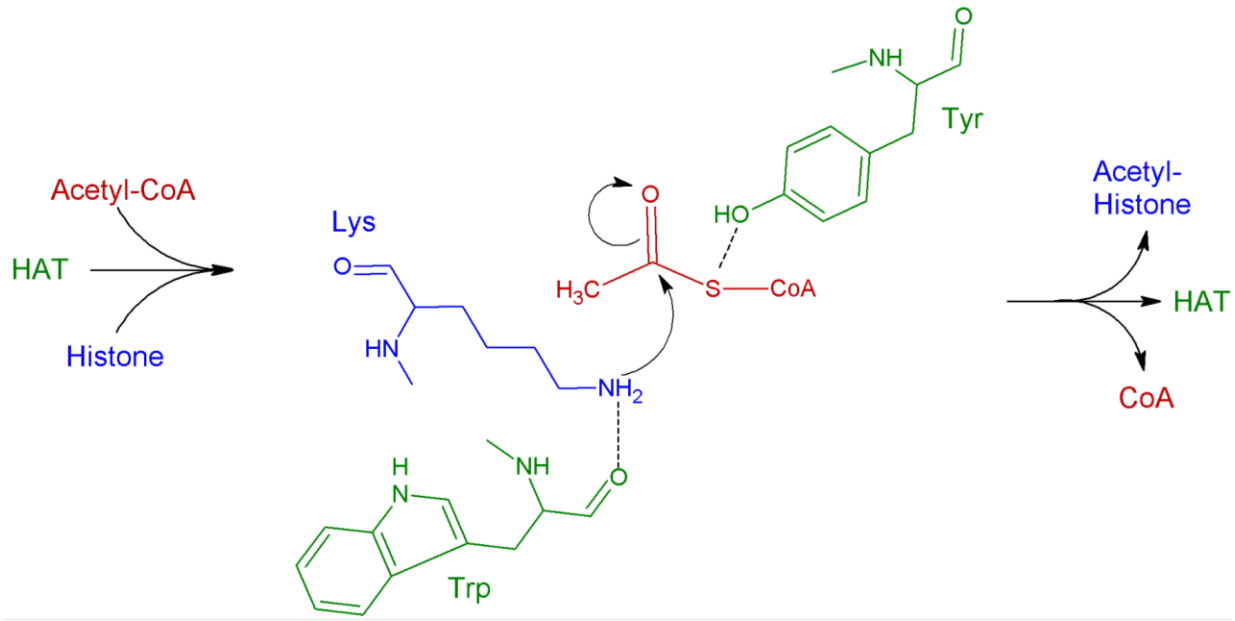

Figure 2.13 Proposed Theorell-Chance catalytic mechanism for the p300/CBP family. Once AcCoA (red) binds to the active site, the positively charged substrate Lys (blue) associates with the HAT domain and attacks the acetyl group AcCoA. Once the acetyl group has been transferred, the acetylated substrate dissociates from the active site. The backbone carbonyl group of a Trp residue (green) in the active site was proposed to hydrogen bond with the Lys residue and the Trp indole side chain was expected to help guide the Lys residue towards attacking AcCoA. Figure adapted from Liu *et al.*, 2008.

2.4 Tip60, a Catalytic Subunit of the Tip60 HAT Complex

2.4.1 Overview of Tip60's Structure, Function, and Regulation

The histone acetyltransferase human Tat-interactive protein of 60 kDa (Tip60) belongs to the MYST family of HATs (Sapountzi *et al.*, 2006). The HTATIP gene encodes the Tip60 protein and was found to contain 14 exons and express three splice variants of the protein: Tip60 isoform 1, Tip60 isoform 2 (Tip60α), and Tip60 isoform 3 (Tip60β or phospholipase associated protein 2 [PLA2] interacting protein, PLIP) (Sapountzi *et al.*, 2006). Each of the three isoforms of Tip60 contained an N-terminal chromodomain, a C-terminal MYST domain, and a nuclear receptor-interaction (NR) box at the extreme C-terminal end (Figure 2.14) (Sapountzi *et al.*, 2006). One study found that isoform 1 was the largest of the Tip60 splice variants as it incorporated intron 1, which resulted in an extra 33 amino acids at the N-terminus of the protein (Legube and Trouche, 2003). It has been speculated that this larger splice variant could exhibit functions distinct from the other two splice variants (Legube and Trouche, 2003). The second isoform, Tip60α was found to be a 513 amino acid protein of 58 kDa and was the best characterized of the three isoforms (Sapountzi *et al.*, 2006). The third isoform, Tip60β excluded exon 5 which resulted in a 52 amino acid deletion between the chromodomain and the MYST domain (Sapountzi *et al.*, 2006).

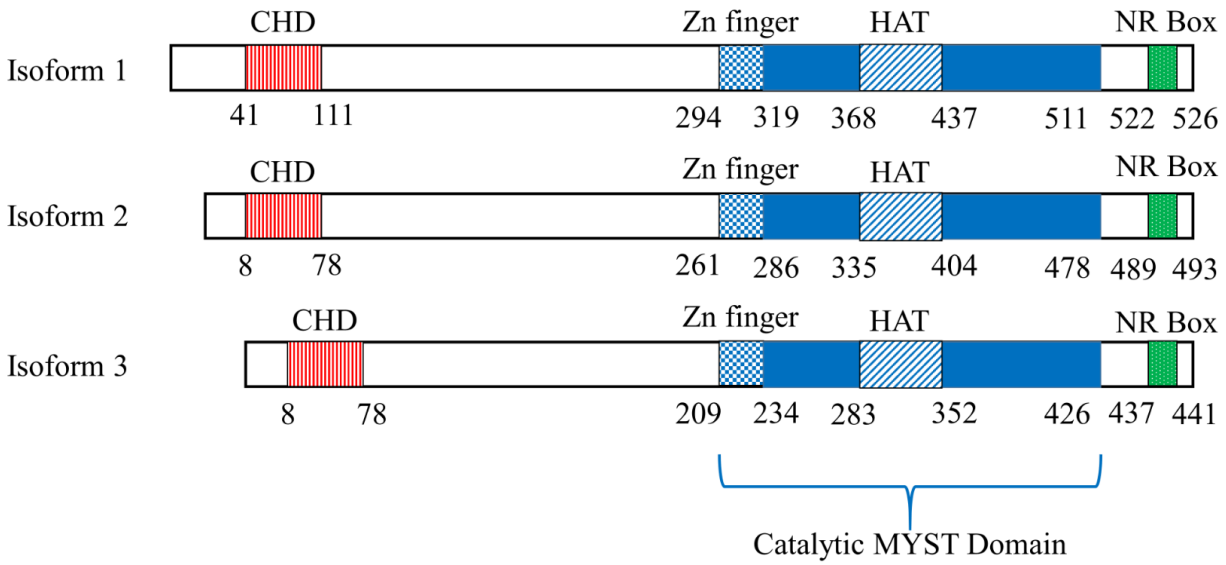

Figure 2.14 Schematic representation of the three isoforms of Tip60 found in *H. sapiens*. Each isoform contained a N-terminus chromodomain (CHD), a conserved catalytic MYST domain which was composed of a zinc finger and histone acetyltransferase domain (HAT), and a C-terminus nuclear receptor-interacting (NR) box. Figure adapted from Sapountzi *et al.*, 2006.

As previously discussed, most HATs were found to be part of larger multi-subunit complexes. Tip60, for example, was found to be a part of the NuA4 HAT complex, which has been found to be evolutionarily conserved from yeast to humans (Squatrito *et al.*, 2006). The NuA4 complex in *H. sapiens* appeared to have the functionality of both HAT and chromatin remodeling complexes, whereas in yeast these functions were performed by two separate complexes, the NuA4 and SWR1 complexes respectively (Doyon and Cote, 2004). The human NuA4 complex was composed of several subunits (Table 2.4), two of which contribute to its acetyltransferase and chromatin remodeling functions, Tip60 and p400/Domino respectively (Sapountzi *et al.*, 2006; Squatrito *et al.*, 2006). Other subunits contained within the complex are important for NuA4's function and stability; these subunits include: the transformation/transcription domain-associated protein (TRAAP), which acts as a scaffold and was found to be essential to the stability of the complex; a 53 kDa BRG-1/human BRM-associated factor (BAF53), which has an actin-related domain that may be involved in histone chaperone activity; the inhibitor of growth 3 (ING3) subunit contained plant homeodomain (PHD) fingers and was also found to be involved in the transcription of p53 as well as playing a role in apoptosis and DNA damage signaling; the mortality factor 4 related gene 15 (MRG15) subunit contained a chromodomain and was involved in the regulation of cell viability,

proliferation, and senescence (Sapountzi *et al.*, 2006). One study found that the NuA4 complexes of Tip60α and Tip60β were identical (Doyon *et al.*, 2004)

Table 2.4 Subunits of the NuA4 complex in *H. sapiens* and their homologs from *D. melanogaster* and *S. cerevisiae*

H. sapiens	***D. melanogaster***	***S. cerevisiae***	**Function**
Tip60	dTip60	Esa1	Acetyltransferase
TRRAP	dTra1	Tra1	PIKK domain
p400/Domino	Domino	Eaf1/Swr1[a]	SWI1/SWI2-like ATPase
RuvBL1	dPontin	Rvb1[a]	Helicase[b]
RuvBL2	dReptin	Rvb2[a]	Helicase[b]
Actin	Act87E	Act1	ATPase, cytoskeleton
BAF53	BAP55	Arp4	Actin-related domain, DNA repair
Mrg15	dMrg15	Eaf3	Chromodomain, senescence
MrgBP	dMrgBP	Eaf7	Scaffold Protein
ING3	dIng3	Yng2	PHD finger domain, growth inhibitor, apoptosis
Epc1	E(Pc)	Epl1	Transcription control, silencing
Epc-like protein			Transcription regulation
Brd8/TRCp120	dBrd8	Bdf1[a]	Bromodomain, thyroid receptor coactivator
YL-1	dYL-1	Vps72[a]	Chromatin remodelling
DMAP	dDMAP1	Eaf2	SANT domain, DNA replication
Gas41	dGas41	Yaf9	YEATS domain, cell viability
FLJ11730	dEaf6	Eaf6	
(H2A.X/H2A.Z)[c]	H2Av	H2A/H2A.Z[a]	Histones
(H2B)[c]	H2B	H2B[a]	Histones

[a] – Subunits belong to a distinct SWR1 complex in yeast. [b] – RuvBL1/2 do not appear to contribute to helicase activity in H. sapiens NuA4 complex. [c] – Histones H2A.X/H2A.Z and H2B have not yet been confirmed in the H. sapiens NuA4 complex (Sapountzi *et al.*, 2006)

Initially, Tip60 was thought to be involved in the regulation of viral gene expression by acting as a cofactor to the HIV-1 Tat protein (Yang *et al.*, 2012). It was originally proposed that Tip60 interacted with the N-terminal of the Tat protein, subsequently leading to Tat activation and the promotion of RNA polymerase phosphorylation, which led to increased HIV gene expression (Kamine *et al.*, 1996). However, it was later found that Tip60 possessed HAT activity and that it was able to acetylate the ε-amino group of lysine residues located on both histone and non-histone proteins (Sun *et al.*, 2010). This wide range of acetylation targets included core histone proteins H2A and H4, several transcription factors including Myc (Sapountzi *et al.*, 2006) and p53 (Tang *et al.*, 2006), the androgen receptor (Jeong *et al.*, 2011), and ATM kinase (Squatrito *et al.*, 2006). The involvement of Tip60 with a wide range of proteins indicates that it plays a vital role in numerous cellular processes, including transcriptional regulation, DNA damage repair (Sapountzi *et al.*, 2006), chromatin remodeling (Kusch *et al.*, 2004), histone acetylation (Jeong *et al.*, 2011), and apoptosis (Ikura *et al.*, 2000). Due to Tip60's involvement in a diverse amount of cellular processes, the deregulation of Tip60 has been implicated in several human diseases, including Alzheimer's disease (Pirooznia *et al.*, 2012) and some cancers (Avvakumov and Cote, 2007; Sun *et al.*, 2010).

2.4.2 Tip60 and Transcriptional Regulation

Through the regulation of a variety of transcription factors, Tip60 and the NuA4 complex were able to influence gene expression. Tip60's ability to function as a transcriptional regulator was dependent upon the specific promoter sites or the cellular context (Kim *et al.*, 2012). In some cases, Tip60 has been found to function as a coactivator of gene expression, whereas in other cases it serves to repress gene expression (Sapountzi *et al.*, 2006). Tip60's ability to act as a coactivator of gene expression typically involves the acetylation of core histones located at the target promoter (Kim *et al.*, 2012). Additionally, Tip60 was found to associate with other transcription factors, such as HIV-1 Tat, Type I nuclear hormone receptor, and the amyloid-β precursor protein (Kim *et al.*, 2012). In the case of p53, it was found that Tip60 directly acetylates the transcription factor itself and controls its transcriptional activity (Tang *et al.*, 2006).

One study found that Tip60 served to enhance the transcription of estrogen receptor alpha (ERα) target genes through the direct interaction with ERα (Jeong *et al.*, 2011). The proposed model (Figure 2.15) for this interaction stated that Tip60's LXXLL motif, found in its NR box,

first interacts directly with ERα. The occupancy of Tip60 at the gene was then stabilized by the chromodomain binding to mono-methylated lysine four of histone H3 ($H3K4Me_1$). This binding of the chromodomain was then proposed to result in the activation of Tip60, which resulted in the acetylation of lysine five on histone H2A (H2AK5Ac) leading to chromatin remodeling and eventual transcription of the target genes (Jeong *et al.*, 2011).

The transcription of active genes requires RNA polymerase II to transcribe a DNA template into RNA (Veloso et al., 2014). The control of the three stages of transcription, which include initiation, elongation, and termination, determines the level of gene expression (Veloso *et al.*, 2014). As previously described, nucleosomes pose a barrier to transcription and, therefore, the eviction of histones is critical to allow for transcriptional elongation to proceed (Ginsburg *et al.*, 2009). Studies on the Tip60 homolog Esa1 in *S. cerevisiae* revealed that Esa1 and the NuA4 complex play a role in transcriptional elongation (Ginsburg *et al.*, 2009). It was proposed that NuA4 complex was recruited to promoters where it acetylates histone H4 in the promoter region, which results in the assembly of the preinitiation complex and, subsequently, a NuA4 complex was thought to associate with phosphorylated RNA polymerase II within the coding sequence (Ginsburg *et al.*, 2009). The NuA4 complex within the coding region was then proposed to interact with methylated lysine residues on histone H3, which mediates the acetylation of histone H4 (Ginsburg *et al.*, 2014; Ginsburg *et al.*, 2009). The acetylation of histone H4 was found to result in the recruitment of chromatin remodeling complexes, which would result in the eviction of histones and subsequently increase the rate of transcriptional elongation (Ginsburg *et al.*, 2009). Furthermore, a recent study found that the tri-methylation of lysine 4 on histone H3 ($H3K4Me_3$) and the hyper-acetylated H2A variant, H2A.Z, were found at the promotor region of highly expressed loci (Kusch *et al.*, 2014). The presence of $H3K4Me_3$ resulted in the exchange of acetylated-H2A.Z by the Tip60 complex, which then stimulates the release of RNA polymerase II into elongation (Kusch *et al.*, 2014).

As previously stated, Tip60 also possesses the ability to function as a repressor of gene expression. Typically, its repressor function manifest itself through the recruitment of other repressor complexes, such as the recruitment of deacetylases (Sapountzi *et al.*, 2006). An example of this is shown in the repression of the transcription factor signal transducer and activator of transcription 3 (STAT3), as it was found that Tip60 recruits the histone deacetylase

HDAC 7 to repress STAT3 activity (Xiao *et al.*, 2003). In the case of cAMP response element binding protein (CREB), it was found that Tip60 bound directly to the transcription factor to repress CREB function (Gavaravarapu and Kamine, 2000). It was also found that Tip60 associated with transcriptional repressors, such as the zinc finger E box (ZEB) binding protein to activate its repressor activity (Sapountzi *et al.*, 2006).

Due to the role Tip60 plays in transcriptional regulation, Tip60 itself must be tightly regulated. One study found that the deacetylation of Tip60's MYST domain by NAD^+ - dependent histone deacetylase (HDAC) SIRT1 inhibits the acetyltransferase activity of Tip60 and promotes its ubiquitin-dependent degradation (Peng *et al.*, 2012). Another method of regulation was to target Tip60 for proteosomal degradation, through the ubiquitination of the protein by mouse double minute 2 (Mdm2) protein when no stimuli were present (Sapountzi *et al.*, 2006). It was found that when a DNA double strand break is detected, the Mdm2 ubiquitination of Tip60 is inhibited, which allows for an increase in Tip60 expression levels and its participation in the double-strand break repair pathway (Sapountzi *et al.*, 2006).

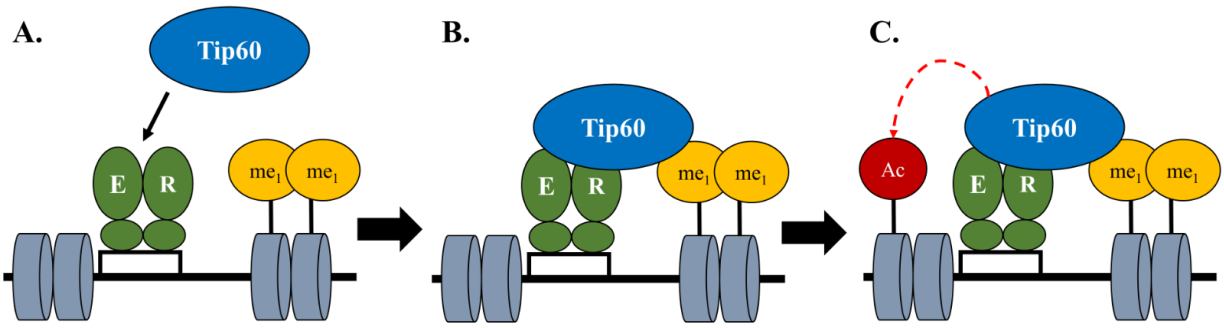

Figure 2.15 Proposed model for Tip60 regulated transcriptional activation of ERα target genes. **A.** Tip60 is initially targeted to ERα and interacts directly through its LXXLL motif found in Tip60's NR box. **B.** Tip60's occupancy at the gene was stabilized through the binding of the Tip60 chromodomain to $H3K4Me_1$. **C.** The binding of the chromodomain to $H3K4Me_1$ was proposed to activate Tip60's acetyltransferase activity. This resulted in the acetylation of H2AK5, which leads to chromatin remodeling and eventual transcription of target genes. Figure adapted from Jeong *et al.*, 2011.

2.4.3 Tip60 and Double-Strand Break Repair Pathway

Cellular exposure to endogenous and exogenous agents, such as replication errors, oxidative free radicals, or ionizing radiation, could result in damage to DNA (Sun *et al.*, 2010; van Attikum and Gasser, 2009). These sources of DNA damage could potentially result in double-strand breaks (DSBs); the inability to repair damaged coding regions of DNA could cause some mutations, which lead to apoptosis, or the mutations to tumour suppressor genes could result in uncontrolled cellular proliferation, ultimately leading to cancer (Fischle, 2009). To counter the damage caused by DNA double strand breaks cells utilize the DSB repair pathway. Tip60 and the NuA4 complex were found to play a vital role in the DSB repair pathway (Squatrito *et al.*, 2006). Upon the induction of a double strand break, the MRE11-RAD50-NBS1 (MRN) complex binds to the break site and recruits the ataxia-telangiectasia mutated (ATM) kinase (Squatrito *et al.*, 2006). One study found that ATM kinase and Tip60 exist as a complex in cells independent of DNA damage and that upon detection of a DSB event ATM was rapidly acetylated by Tip60 resulting in the activation of ATM kinase (Sun *et al.*, 2005). Another study found that the MRN complex was required to target the Tip60-ATM complex to the site of DSBs (Sun *et al.*, 2009). It was proposed that upon the detection of the of a DSB recruits MRN and CK2 to the break site. The CK2 recruitment was found to be required for the phosphorylation of HP1β allowing for its release from chromatin. Upon the release of HP1β, the Tip60-ATM complex was recruited to the break site. Tip60 was thought to interact with tri-methylated lysine nine of histone H3 ($H3K9Me_3$) through its chromodomain. It was proposed that the binding of the Tip60 chromodomain to $H3K9Me_3$ was responsible for the activation of its acetyltransferase domain (Figure 2.16) (Sun *et al.*, 2009)

Once the acetyltransferase domain of Tip60 was activated, it was able to acetylate and activate ATM kinase (Sun *et al.*, 2009). The activation of ATM kinase resulted in the phosphorylation of the mammalian histone variant H2A.X (Kusch *et al.*, 2004). Through the use of the *Drosophila melanogaster* model, a study demonstrated that when H2Av (*D. melanogaster* homolog to *H. sapiens* H2A.X) became phosphorylated during a DSB event, its exchange for unmodified H2Av was catalyzed by the NuA4 complex (Kusch *et al.*, 2004). It was found that the Tip60 subunit would acetylate the phosphorylated-H2Av, which would allow the p400/Domino subunit of the NuA4 complex to exchange the phosphorylated-H2Av for unmodified H2Av (Kusch *et al.*, 2004).

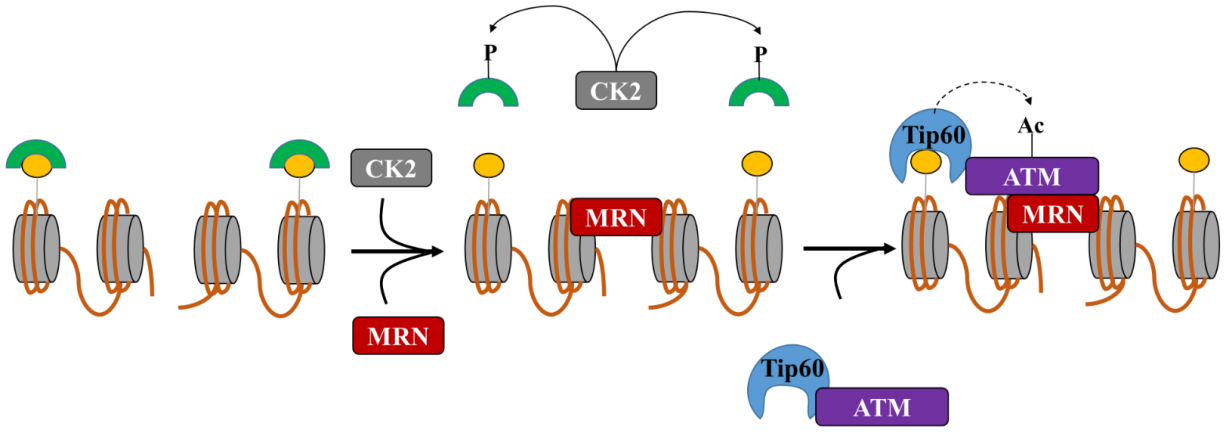

Figure 2.16 Proposed model for ATM activation. Detection of DSBs recruits MRN and CK2. Phosphorylation of HP1β by CK2 releases HP1β (green crescents) from chromatin. Tip60-ATM complex is recruited to the break site, where Tip60 chromodomain interacts with $H3K9Me_3$ (yellow spheres) stimulating HAT activity and subsequent acetylation and activation of ATM kinase. Figure adapted from Sun *et al.*, 2009.

2.5 Summary of Previous Studies on Tip60

Previous research on Tip60 has demonstrated its involvement in a variety of cellular processes, including transcriptional regulation, DNA damage repair, chromatin remodeling, and apoptosis. As a histone acetyltransferase, Tip60's role in the aforementioned cellular processes typically includes the acetylation of other proteins required in these processes. The acetylation targets were found to include core histone proteins, transcription factors, the androgen receptor, and ATM kinase. Due to Tip60's diverse role, the deregulation of Tip60 has often been implicated in human disease.

Several studies have been conducted to ascertain the mechanism by which Tip60's acetyltransferase activity is regulated. These studies suggested that the Tip60 chromodomain was responsible for targeting methylated lysine residues found on histone tails and allosterically regulating the HAT activity. It was suggested that the Tip60 chromodomain was able to bind methyllysine through an aromatic cage. However, studies conducted on the Tip60 homolog, Esa1 and Tip60 paralog, MOF, revealed that the residues required to form an aromatic cage are not present. Therefore, it is unlikely Tip60 is able to bind methyllysine through the proposed mechanism. Currently, the binding partners of the Tip60 chromodomain have not been confirmed and the mechanism with which this binding is achieved is still unknown. This research aims to determine the structure of the Tip60 chromodomain through X-ray crystallography and elucidate the binding partners of the chromodomain through *in vitro* binding studies.

3 MATERIALS AND METHODS

3.1 Reagents

Table 3.1 List of reagents and suppliers

Reagents	**Supplier**
1,4,7,10,13,16-hexaoxacyclooctadecane (18-crown-6)	Sigma Aldrich
Al's Oil	Hampton Research
Ammonium Chloride (^{15}N)	Cambridge Isotope Laboratories
Ampicillin	Bioshop
Centrifugal Filter Units	Merck Millipore
Chloramphenicol	Sigma-Aldrich
N-Cyclohexyl-2-aminoethanesulfonic acid (CHES)	Bioshop
Dialysis tubing	BioDesign Dialysis Tubing
Dithiothreitol (DTT)	Bioshop
1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC)	BioRad
Ethylenediamine	Sigma-Aldrich
Ethylenediaminetetraacetic acid (EDTA)	Bioshop
L-Glutathione reduced	Sigma-Aldrich
Glutathione Sepharose 4B	GE Healthcare Life Sciences
HydraGreen	ACTGene
Isopropyl β-D-1-thiogalactopyranoside (IPTG)	Fisher Scientific
Hen Egg White Lysozyme	Bioshop
2-Methyl-2,4-pentandediol (MPD)	Hampton Research
N-hydroxysulfosuccinimide (sulfo-NHS)	BioRad
Polyethylene glycol (PEG) 3350-8000	Hampton Research
PreScission Protease	Amersham Biosciences
Tris(2-carboxyethyl)phosphine (TCEP)	Hampton Research
Tween-20	Sigma-Aldrich

Table 3.2 List of kits and suppliers

Kit	**Supplier**
Additive Screen (HR2-428)	Hampton Research
$(NH_4)_2SO_4$ Suite	Qiagen
HR2-110 Crystal Screen Kit	Hampton Research
HR2-144 Crystal Screen Kit	Hampton Research
JCSG+ Suite	Qiagen
PACT Suite	Qiagen
QIAprep spin mini-prep columns	Qiagen
QIAquick gel extraction kit	Qiagen
QIAquick PCR purification kit	Qiagen
Plasmid Maxi kit	Qiagen
Wizard I	Rigaku Reagents
Wizard II	Rigaku Reagents

Table 3.3 Name and addresses of suppliers

Supplier	**Address**
ACTGene	Piscataway, New Jersey, USA
Amersham Biosciences	Piscataway, New Jersey, USA
Amersham PharmaciaBiotech	Uppsala, Sweden
BioDesign Dialysis Tubing	New York, USA
Bio-Rad	Hercules, California, USA
Bioshop	Burlington, Ontario, Canada
Cambridge Isotope Laboratories	Andover, Massachusetts, USA
Constant Systems	Daventry, United Kingdom
Fisher Scientific	Fair Lawn, New Jersey, USA
GE Healthcare Life Sciences	Uppsala, Sweden
Hampton Research	Aliso Viejo, California, USA
Merck Millipore	Tullagreen Carrigtwohill, Ireland
Qiagen	Toronto, Ontario, Canada

Rigaku Reagents	Bainbridge Island, Washington, USA
Sigma-Aldrich	St. Louis, Missouri, USA
Syngene	Frederick, Maryland, USA
TA Instruments	Toronto, Ontario, Canada
Thermo Scientific	Waltham, Massachusetts, USA

3.2 DNA Methods

3.2.1 Cloning

Initially four constructs were designed and cloned from an *E. coli* codon optimized synthetic DNA sequence of *Homo sapiens* Tip60 splice variant β and full length cDNA of *Drosophila melanogaster* Tip60 (provided by *Drosophila* Genetic Resource Centre). The constructs were designed to separately clone the *H. sapiens* Tip60 chromodomain (residues 3 – 72) and HAT domain (residues 68 – 459) and the *D. melanogaster* Tip60 chromodomain (residues 19 – 88) and HAT domain (residues 246 – 541). The separate cloning of each domain allowed for the separate insertion of each domain into the multiple cloning sites (MCS) of the pGEX-6P-3 expression vector (GE Healthcare Life Sciences) between the *Eco*RI and *Bam*HI restriction sites. The major features of the pGEX-6P-3 expression vector (Figure 3.1A) are the glutathione S-transferase (GST) gene upstream of the MCS, where the gene of interest would be inserted, and the antibiotic resistance gene that allowed for the selection of successful transformants. The GST gene's position upstream of the MCS allowed for the gene of interest to be expressed as a GST-fusion protein, which in turn allowed for the recombinant protein to be purified through glutathione sepharose affinity purification. Additionally, a PreScission protease recognition site is located within the linker region between the GST tag and the protein of interest, which allowed for the cleavage of the GST tag from the protein of interest. PreScission protease is a highly specific fusion protein composed of GST and human rhinovirus 3C protease (Walker *et al.*, 1994) that cleaves between Gln and Gly residues found in the recognition sequence Leu-Glu-Val-Leu-Phe-Gln/Gly-Pro (Figure 3.1B) (Cordingley *et al.*, 1990).

The targeted gene fragments were produced by the polymerase chain reaction (PCR). The cloning and amplification of the *D. melanogaster* Tip60 chromodomain and HAT domain utilized 1 ng of template cDNA in 100 µL of the PCR reaction mixture composed of 1X reaction buffer

(200 mM Tris-HCl pH 8.8, 100 mM $(NH_4)_2SO_4$, 100 mM KCl, 1 mg/mL BSA, 1% (v/v) Triton X-100, 20 mM $MgSO_4$), 200 µM dNTPs, 200 nM 5' forward primer, 200 nM 3' reverse primer, 2.5 units/µL *Pyrococcus furiosus* (*Pfu*) DNA polymerase, and H_2O. Thirty cycles of PCR amplification were performed using the MyCycler Thermal Cycler (Bio-Rad). The PCR amplification protocol used is shown in Figure 3.2.

For the cloning and amplification of the *H. sapiens* Tip60 chromodomain 1 ng, 2 ng, and 4 ng of template cDNA were used to identify which yielded the optimal level of amplification of the target gene. The template cDNAs and 1 – 4 mM $MgSO_4$ were added to 50 µL of the PCR reaction mixture previously described. The additional Mg^{2+} ions aid in the stabilization of the PCR reaction, which allowed for a greater yield of target DNA. The cloning and amplification of the *H. sapiens* Tip60 HAT domain utilized different template cDNA concentrations, which ranged from 1 – 4 ng. The template cDNAs were added to 50 µL of the PCR reaction previously described. All primers used are listed in Table 3.4.

Following PCR amplification, the PCR products were purified by ethanol precipitation. The purification was performed by adding 10% v/v of 3 M sodium acetate to the PCR products followed with a wash of 3X v/v of the PCR reaction mixture with 95% ethanol. The solution was then incubated at -20° C for 10 minutes and then the sample was pelleted by centrifugation at 16, 438 x *g* at 4° C for five minutes. The supernatant was removed and 2X of 70% ethanol was added to the mixture. The sample was then pelleted once more at 16, 438 x *g* at 4° C for five minutes. The supernatant was removed and the pellet dried by centrifugal evaporation. Once the pellet was dry, it was re-dissolved in 20 µL TE (10 mM Tris-HCl pH 7.5 and 1 mM EDTA).

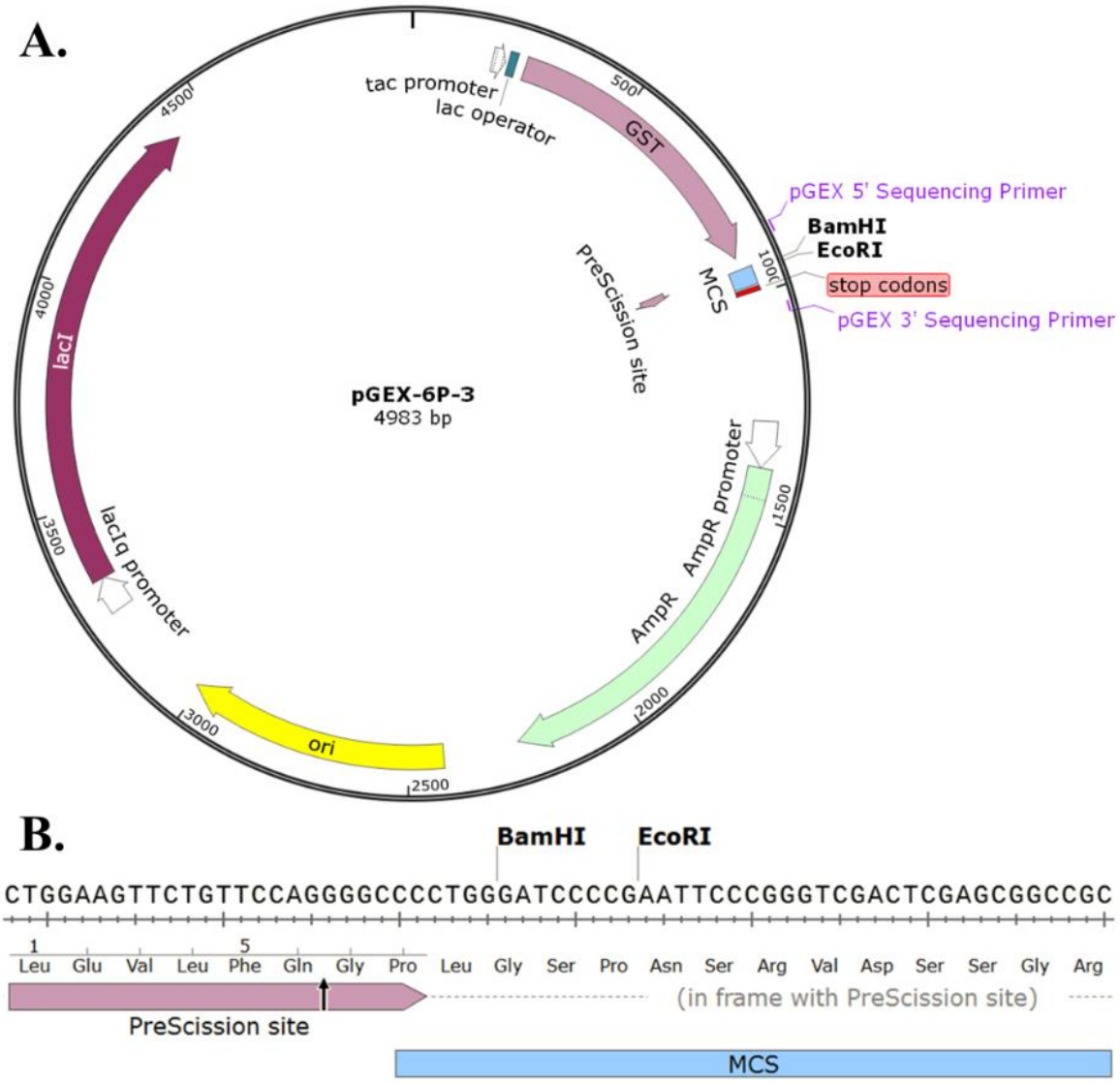

Figure 3.1 Map of the pGEX-6P-3 expression vector. **A.** Features of the pGEX-6P-3 expression vector shown include the lacI gene, the GST gene, the PreScission protease cut site, the MCS, the endonuclease cut sites for *Eco*RI and *Bam*HI, and the ampicillin resistance gene. **B.** Gene sequence and amino acid sequence of the PreScission protease cut site and MCS. The PreScission protease cut site is located between Gln and Gly as indicated by the arrow. The endonuclease cut sites for *Bam*HI and *Eco*RI are also indicated. Figure generated using SnapGene software (GSL Biotech).

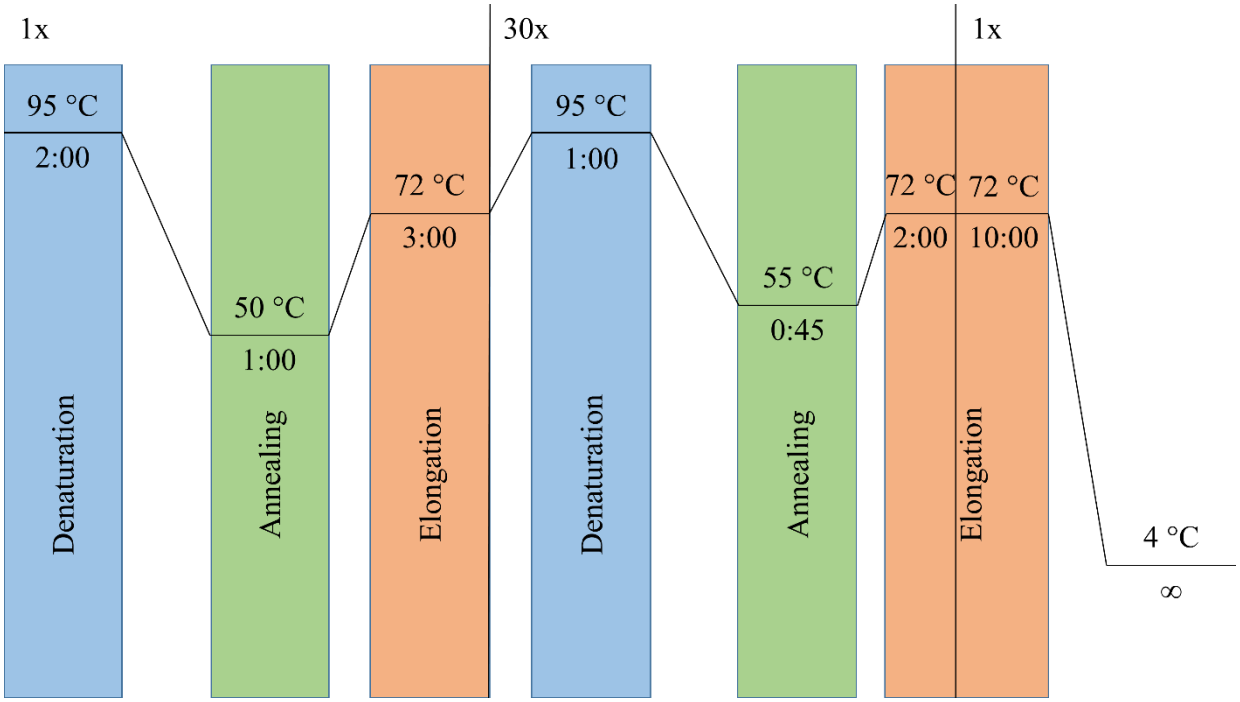

Figure 3.2 PCR amplification program.

Table 3.4 List of primers

Protein	**Primer Sequence**	**Restriction Endonuclease**
D. melanogaster Tip60 chromodomain	5'-GACCTG**GGATCC**TAACGGCCGCCCTAACC-3' 5'-GACGAG**GAATTC**CTAGCCGTCTCTCCGCGGAAA-3'	*Bam*HI *Eco*RI
D. melanogaster Tip60 HAT domain	5'-GACCTG**GGATCC**ATGGTTACGCACCAGGACGAC-3' 5'-GACGAG**GAATTC**TCATTTGGAGCGCTTGGACCA-3'	*Bam*HI *Eco*RI
H. sapiens Tip60 chromodomain	5'-GACCTG**GGATCC**GGTGAAATTATTGAAGGCTGCCGT-3' 5'-GCGCCG**GAATTC**TCACTCTTTCTTGGGGAACTGGATCTT-3'	*Bam*HI *Eco*RI
H. sapiens Tip60 HAT domain	5'-GACCTG**GGATCC**CACGACGACATCGTCACC-3' 5'-GACGAG**GAATTC**TCACCACTTCCCCCTCTTGCT-3'	*Bam*HI *Eco*RI

Once the PCR products were purified by ethanol precipitation, the gene of interest was subjected to restriction digestion with the *Eco*RI and *Bam*HI restriction endonucleases. For the restriction digest, a 50 µL mixture composed of 2 µL *Eco*RI, 2 µL *Bam*HI, 20 µL of purified gene of interest, 2X Tango buffer (66 mM Tris-acetate (pH 7.9 at 37° C), 20 mM magnesium acetate,

132 mM potassium acetate, 0.2 mg/mL BSA), and H_2O was prepared. The restriction digestion reaction was performed at 37° C for 16 hours. Afterwards, the restriction digest products were subjected to ethanol precipitation as previously described. Following ethanol precipitation, the pellet was re-dissolved in 10 µL TE buffer. The pGEX-6P-3 plasmid was also digested using this protocol however only 10 µL of plasmid DNA was required for the restriction digestion reaction.

The gene of interest and plasmid DNA from the restriction digest were then run on a 1 % agarose gel (agarose gel electrophoresis protocol described in section 3.2.5). The gel was then viewed on a UV transilluminator to locate the DNA bands of interest. The portion of the gel containing the DNA bands were then cut out and placed into separate Eppendorf tubes. The mass of the excised fragment was determined by recording the mass of the Eppendorf tube before and after the gel fragment was added. The DNA was then purified using the Gel Extraction kit from Qiagen and followed the protocol enclosed. The concentration and purity of the DNA insert was checked by using a NanoDrop 2000 spectrophotometer (Thermo Scientific) measuring absorbance at 260 nm and then by agarose gel electrophoresis.

Once the insert and plasmid had been subject to a restriction digestion and purified, a ligation reaction was performed. The ligation reaction required that the molar ratio of the insert DNA to plasmid DNA be 5 to 1. Additionally, 1 µL T4 DNA ligase, 1X (50 mM Tris-HCl (pH 7.5 at 25° C), 10 mM $MgCl_2$, 1 mM ATP, 10 mM DTT), H_2O were also added to the ligation reaction mixture. The ligation mixture was then incubated for 16 hours at 16° C or 4 hours at 22° C.

3.2.2 Preparation of Competent Cells

Three different strains of *E. coli* competent cells were used during this project, XL-1, BL21(DE3), and Rosetta 2. The XL-1 competent cells were used for cloning the gene of interest as these cells were designed for routine cloning applications and have a high transformation efficiency. By design XL-1 competent cells are deficient of endonuclease (*endA*) and recombination (*recA*), which results in the improved quality of plasmid preparation DNA and insert stability respectively. The BL21(DE3) and Rosetta 2 competent cells were used for protein expression. BL21(DE3) competent cells feature an integrated copy of T7 RNA polymerase, which is induced by IPTG, and is deficient in *lon* and *ompT* proteases to facilitate the production

of the protein of interest. The Rosetta 2 strain is a BL21 derivative, which has been designed to improve the expression of eukaryotic proteins through the inclusion of a plasmid coding for seven tRNAs that match eukaryotic codons rarely used in *E. coli*. The XL-1 and BL21(DE3) strains contain no antibiotic resistance, whereas the Rosetta 2 strain contains a chloramphenicol resistance gene located on the tRNA plasmid. To prepare the competent cells the desired strain was plated onto Luria-Bertani (LB) medium agar plates which contained no antibiotics for XL-1 and BL21(DE3) strains and 34 µg/mL chloramphenicol for the Rosetta 2 strain. The plated cells were then incubated for 16 hours at 37° C. Following incubation, a single colony was picked and used to inoculate 5 mL of LB broth media, which contained no antibiotics for XL-1 and BL21(DE3) strains and 34 µg/mL chloramphenicol for Rosetta 2 strain. The inoculated LB broth media was then incubated in a shaking incubator at 37° C for 16 hours and shaking at 250 rpm. The seed culture was then diluted 1 in 100 times with fresh LB broth media, which contained the appropriate antibiotics. The diluted culture was then incubated at 37° C in a shaking incubator shaking at 250 rpm until an optical density at 600 nm (OD_{600}) of 0.4 to 0.6 was reached. The OD_{600} was checked using a NanoDrop 2000 spectrophotometer (Thermo Scientific). Upon achieving the desired OD_{600} the cells were chilled on ice and then harvested by centrifugation with a Beckman-Coulter centrifuge. The cells were centrifuged for five minutes at 1,932 x *g* and 4° C. The cell pellet was then re-suspended in ice cold 0.1 M $MgCl_2$ in a 1 to 5 v/v ratio of the original 100 mL cell volume. The re-suspended cells were then incubated on ice for one hour. Following the incubation, the cells were then centrifuged for five minutes at 1,932 x *g* and 4° C. The cell pellet was then re-suspended in an ice cold solution composed of 0.1 M $CaCl_2$ and 14% glycerol in a 1 to 50 v/v ratio of the original 100 mL cell volume. The cells were then incubated on ice for one hour. Aliquots of the cell solution were made and flash frozen with liquid N_2 and stored at -80° C.

3.2.3 Transformation of Bacterial Cells

The transformation of XL-1 *E. coli* cells required that the entirety of the ligation mixture be added to 100 µL of cells. The cells were then cold shocked by incubation at 4° C for 30 – 45 minutes. Following incubation, the cells were plated onto a pre-warmed LB agar plate which contained 100 µg/mL of ampicillin. The plated cells were then incubated for 16 hours at 37° C. Afterwards the plates were stored at 4° C until needed. The BL-21 and Rosetta 2 *E. coli* strains were used for protein expression and 40 – 80 ng of plasmid DNA was used to transform these

strains. The transformation procedure used was the same as described previously. For the transformation of the Rosetta 2 *E. coli* strain 34 µg/mL of chloramphenicol was also added to the LB agar plate.

3.2.4 Plasmid Preparation

Following the transformation of XL-1 *E. coli* cells the transformant colonies were screened by colony PCR. The colony PCR was carried out to check for the presence of the gene of interest. To perform the colony PCR, 250 µL reaction mixture was prepared to screen ten colonies. The reaction mixture was composed of 1X reaction buffer (10 mM Tris-HCl pH 8.3, 50 mM KCl, 1.5 mM $MgCl_2$), 5 units/µL *Thermus aquaticus* (*Taq*) DNA polymerase, 200 µM dNTPs, 200 nM 5' forward primer complementary to the 3' end of the GST gene, 200 nM 3' reverse primer complementary to the 3' end of the insert gene, and H_2O. The reaction mixture was then split into ten 25 µL aliquots and each aliquot was inoculated with a different transformant colony. The PCR procedure used was the same as described previously. The PCR products were then run on a 1% agarose gel as described in section 3.2.5 to check for the presence of the insert.

Positive colonies were selected to inoculate 5 mL of LB broth containing 100 µg/mL of ampicillin and grown for 16 hours at 37° C shaking at 250 rpm. Following incubation, 2 mL of culture were harvested in an Eppendorf tube by centrifuging at 15,588 x *g* for one minute. The supernatant was then discarded and the pellet was subjected to the plasmid preparation process. The plasmid preparation was done by using the QIAprep Spin Miniprep Kit (Qiagen) and following the instructions provided. To confirm the presence of the gene of interest a restriction digestion, PCR purification, and DNA sequencing were performed. The restriction digestion followed the protocol previously described. Sanger DNA sequencing was performed at the Plant Biotechnology Institute (PBI) at the National Research Council (NRC) of Canada in Saskatoon, SK.

3.2.5 Agarose Gel Electrophoresis

For the quantification and purification of the gene of interest, agarose gel electrophoresis was performed. The composition of the agarose gel included 1% (w/v) agarose and 1X TAE solution (40 mM Tris-acetate, pH 8.0 and 1 mM EDTA). The agarose was dissolved in TAE buffer by heating in a microwave for 30 – 45 second intervals until the agarose was completely

dissolved. Once the agarose mixture had cooled a 1X concentration of HydraGreen (ACTGene) was added to allow for visualization of the DNA. The gel mixture was then poured into a gel casting tray containing a 10 well comb and allowed to set. The gel was then immersed in 1X TAE buffer. The DNA samples were mixed with DNA loading dye to allow for visualization during the resolving phase. The DNA samples and a DNA ladder were loaded into the wells and the gel was resolved at a constant 100 V for 45 to 60 minutes. The HdyraGreen added to the agarose gel has fluorescence excitation maxima between 295 nm and 490 nm. To visualize the DNA bands on the gel a 365 nm wavelength was used. The resolved agarose gel was then photographed by a G:Box Chemi XX9 Gel Doc Imaging System (Syngene).

3.3 Protein Expression and Purification

The purification of proteins in this project was accomplished through the use of fast protein liquid chromatography (FPLC), specifically using the ÄKTA FPLC system developed by Amersham Pharmacia Biotech (now GE Healthcare Life Sciences).

3.3.1 Protein Expression Trials

Protein expression trials were conducted to find the optimal conditions for the expression of the *H. sapiens* Tip60 chromodomain and *D. melanogaster* Tip60 chromodomain. Initially BL21(DE3) competent *E. coli* cells were transformed with 40 ng of the pGEX-6P-3 plasmid carrying the chromodomain insert and plated onto LB agar plates containing 100 µg/mL of ampicillin. The plates were left to incubate for 16-18 hours at 37° C. A single colony was then used to inoculate 5 mL of LB broth media (for clarity this will be referred to as the overnight culture) containing 100 µg/mL of ampicillin and was left to incubate for 16-18 hours at 37° C and shaking at 220 rpm. Fifty millilitres of LB broth media containing 100 µg/mL of ampicillin was inoculated with 1% v/v of the overnight culture and grown at 37° C and shaking at 220 rpm until a desired optical density at 600 nm (OD_{600}) was reached. During the expression trial three different OD_{600} were chosen for protein induction which were 0.4, 0.6, and 0.8. Once the desired OD_{600} were reached, a 5 µL aliquot was taken for sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) analysis and the expression was induced with 0.1 mM of isopropyl β-D-1-thiogalactopyronoside (IPTG) at 37° C and 25° C for at least 16 hours. Following induction, 5 mL of cell culture were pelleted by centrifugation at 15,588 x *g* for two minutes. The cells were re-suspended in 500 µL of 1X phosphate buffered saline (PBS) and were lysed

through sonication. Another 5 µL aliquot was taken of the cell lysate for SDS-PAGE analysis. The cell lysate was then centrifuged at 15, 588 x *g* for two minutes. The resultant supernatant was then moved to a clean Eppendorf tube and the remaining pellet was re-dissolved in 500 µL of 1X PBS. Five microliter aliquots of the supernatant and re-dissolved pellet were also taken for SDS-PAGE analysis. Twenty microliters of a 70% glutathione sepharose slurry was then added to the supernatant solution and incubated at 4° C for 1 hour while gently mixing. Following incubation, the glutathione sepharose beads were pelleted by centrifugation at 1,476 x *g* for 20 seconds. The beads were washed five times in 200 µL 1X PBS and then re-suspended in 20 µL of SDS-PAGE loading buffer. The aliquots prepared during the expression trials were then analysed through SDS-PAGE analysis following the protocol detailed in section 3.4.1. The optimal conditions for protein expression were selected and scaled up to allow for large scale production of the protein of interest.

The expression of the *D. melanogaster* Tip60 chromodomain was achieved by following the protocol previously described for conducting protein expression trials with a few alterations. The transformation and plating of the transformants used the same parameters as previously described. The overnight culture was prepared as previously described in the protein expression trials except that the volume was increased to 50 – 100 mL of LB broth media; the concentration of ampicillin, the incubation time and temperature were the same. One litre of LB broth media containing 100 µg/mL of ampicillin was inoculated with 1% v/v of the overnight culture and was grown at 37° C and shaking at 210 rpm until an OD_{600} of 0.8 was reached. The expression of protein was induced with 0.1 mM IPTG at 25° C and shaking at 210 rpm for at least 16 hours. The cells were harvested by centrifugation for 20 minutes at 5,393 x *g*. The cell pellet was re-suspended in 3X volume of lysis buffer (50 mM Tris-HCl (pH 7.0), 150 mM NaCl, and 1 mM EDTA) to mass of cell pellet (e.g. 3 mL lysis buffer for every gram of cell pellet) and stored at -20° C.

3.3.2 Glutathione Sepharose Affinity Purification

The purification of the GST-tagged chromodomain began with glutathione sepharose affinity purification using an ÄKTA FPLC and a HR 16/10 column packed with 22 – 25 mL of glutathione sepharose 4B (GE Healthcare Life Sciences). The column was initially equilibrated with two column volumes of 1X PBS buffer (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4,

and 1.8 mM KH_2PO_4 at pH 7.3) until conductivity was stable. The *E. coli* cells were thawed and 1 mg/mL of lysozyme was added to aid in the lysis of the cells. The cells were then lysed using the TS Series Benchtop Cell Disruptor (Constant Systems). The lysate was then centrifuged at 20,406 x *g* for one hour at 4° C. Following centrifugation, the supernatant was loaded into a 60 mL super loop on the ÄKTA FPLC and then injected onto the glutathione sepharose affinity column. The absorbance at 280 nm was monitored to detect the amount of protein passing through the column throughout the purification process. The flow-through fractions were collected and two to five column volumes of 1X PBS were used to wash out unbound protein. The GST-fusion protein bound the column and was eluted with a glutathione elution buffer composed of 50 mM Tris, 150 mM NaCl, and 10 mM reduced L-glutathione at pH 8.0.

Following the initial GST affinity purification, the eluted fractions which contained the GST-fusion protein were pooled together and dialyzed in 4 L of cleavage buffer composed of 50 mM Tris, 150 mM NaCl, and 1 mM EDTA at pH 8.0 for 12 – 16 hours at 4° C. Afterwards, the fusion protein was cleaved by incubating 30 µL of PreScission protease for 16 hours at 4° C. After cleavage of the fusion protein, GST affinity purification was repeated. The protein of interest, which no longer possesses the GST tag, was eluted in the flow through whereas the GST tag and PreScission protease remained bound to the column. Aliquots of the fractions were taken and resolved by SDS-PAGE analysis to determine the level of purity of each fraction sample.

3.3.3 Anion-exchange Purification

Following glutathione sepharose affinity purification, the fractions containing the chromodomain were pooled together and dialyzed in a low salt buffer (50 mM Tris (pH 8.5) and 10 mM NaCl) for at least 16 hours to prepare the protein for anion-exchange purification. The protein was purified using a HR 16/5 column packed with 15 – 20 mL of Source 15Q beads (GE Healthcare Life Sciences). The Source Q column was equilibrated with three column volumes of low salt buffer (50 mM Tris (pH 8.5) and 10 mM NaCl) until conductivity was stable. The protein was then injected onto the column using a super loop and five column volumes of low salt buffer was used to wash out unbound protein. The elution of the protein was achieved by using a mixture of low salt buffer and high salt buffer (50 mM Tris (pH 8.5) and 2 M NaCl). The high salt buffer was gradually increased over 5.5 column volumes until the solution mixture, which flowed through the column, contained 25% high salt buffer (500 mM NaCl). Aliquots of

the fractions were taken and resolved by SDS-PAGE analysis to determine the purity of each fraction sample.

3.3.4 Size-Exclusion Chromatography

Size-exclusion chromatography was performed to predict the oligomerization state of the *D. melanogaster* Tip60 chromodomain. Following expression of the chromodomain, the protein was purified by glutathione sepharose affinity and anion-exchange as previously outlined. Following anion-exchange the protein was concentrated to 2 mg/mL and 500 µL of this concentrated protein solution was used for size-exclusion chromatography. The Superdex 75 10/300 GL column (GE Healthcare Life Sciences), which has a 23 mL bed volume and a molecular weight separation range between 3,000 and 70,000 Da, was used. The running buffer used was composed of 50 mM Tris, pH 8.5, and 150 mM NaCl. Additionally, the same size-exclusion chromatography procedure was performed using a running buffer composed of 30 mM CHES, pH 9.5, and 200 mM NaCl to determine if the storage buffer caused the oligomerization state of the chromodomain to change.

Initially, the Superdex 75 10/300 GL column was equilibrated with three column volumes of running buffer. Following equilibration, the protein was injected into the column and the fractions were collected. The volume at which the protein was eluted was compared to a molecular weight standard which showed the eluted volume for proteins of various sizes.

3.3.5 Buffer Exchange and Protein Concentration

Following purification, the pure protein was first buffer exchanged into a storage buffer where the composition was dependent on what study the protein was to be used for. The buffer exchange process was the same regardless of the storage buffer. Initially, the purified protein solution volume was reduced to approximately 7 mL by centrifugation at 2,218 x *g* for 20 minutes at 4° C using a 15 mL Amicon Ultra Centrifugal filter (EMD Millipore) with a 3,000 Da molecular weight cut off. Following this reduction in protein solution volume, an equal amount of storage buffer is added to the centrifugal filter device and centrifugation is repeated. The process of volume reduction and storage buffer addition is repeated five times to complete buffer exchange into the storage buffer. Once buffer exchange was completed, the protein solution was concentrated as desired through the use of the centrifugal filter and centrifugation, in a process similar to buffer exchange, except that no additional buffers were added as the volume was

reduced. The concentration of the purified chromodomain was checked via a NanoDrop 2000 spectrophotometer (Thermo Scientific) measuring the absorbance at 280 nm.

For crystallization trials and ITC, the storage buffer was composed of 30 mM CHES pH 9.5, 200 mM NaCl, 2 mM TCEP, and 10 % MPD. The chromodomain was concentrated to 2.5 mg/mL for use in crystallography trials. For NMR studies the storage buffer was composed of 40 mM phosphate buffer pH 7.5, 50 mM NaCl, and 5 % MPD. The chromodomain was concentrated to 0.15 mM. Initially, the SPR studies used a storage buffer composed of 30 mM phosphate buffer pH 7.4, 100 mM NaCl, 2 mM TCEP, and 10 % MPD. The chromodomain was concentrated to 2.5 mg/mL for SPR studies.

3.4 Protein Visualization Techniques

3.4.1 Sodium Dodecyl Sulphate – Polyacrylamide Gel Electrophoresis

Sodium dodecyl sulphate – polyacrylamide gel electrophoresis (SDS-PAGE) was utilized to visualize the protein following overexpression and purification. All SDS-PAGE gels were cast with a 4% acrylamide stacking gel composed of 4% acrylamide, 124 mM Tris-HCl (pH 6.8), 0.1% SDS, 0.1% ammonium persulfate (APS), 0.1 % N,N,N',N' – tetramethylethylenediamine (TEMED), and H_2O. The resolving gel was composed of 15% acrylamide, 373 mM Tris-HCl [pH 8.8], 0.1% SDS, 0.05% APS, 0.5% TEMED, and H_2O. The protein samples were prepared by mixing with 2X SDS gel loading buffer (100 mM Tris-HCl [pH 6.8], 4% w/v SDS, 0.2% w/v bromophenol blue, 20% v/v glycerol, and 200 mM β – mercaptoethanol) and boiled for 2 minutes. Ten microliters of sample was loaded into the wells of the stacking gel and the samples were resolved by applying a constant 120 V for approximately 2 hours or until the dye front reached the bottom of the resolving gel. The running buffer used was composed of 24.8 mM Tris-HCl [pH8.2], 192 mM glycine, 0.1% w/v SDS, and H_2O.

The gel was then stained with coomassie blue stain (40% v/v ethanol, 10% v/v acetic acid, 0.1% w/v Coomassie R250, and H_2O) for 20 – 30 minutes and then destained with a destaining solution (40% v/v ethanol, 10% v/v acetic acid, and H_2O) for 20 minutes and H_2O for at least 4 hours. The gel images were captured using a G:Box Chemi XX9 Gel Doc Imaging System (Syngene).

3.5 Protein Crystallization

3.5.1 Sparse Matrix Crystal Screening

The commercially available sparse matrix crystal screen kit, Wizard Classic I (Emerald Biosciences now Rigaku Reagents) was initially used to identify potential crystallization conditions. Crystal screens were setup by using the hanging drop vapour diffusion method, in which a droplet composed of 1 µL protein solution and 1 µL precipitant solution are suspended over a reservoir containing the same precipitant solution and allowed to equilibrate (Figure 3.3). The crystallization conditions were setup in 24 well plates and allowed to equilibrate at 20° C. Each condition was checked periodically for the presence of crystals. The conditions which produced crystals were noted and selected for optimization. Additionally, other sparse matrix screening kits such as the Wizard Classic II (Emerald Biosciences now Rigaku Reagents), JCSG+ Suite, PACT Suite, $(NH_4)_2SO_4$ Suite (Qiagen), HR2-110 kit, and HR2-144 kit (Hampton Research) were used as well.

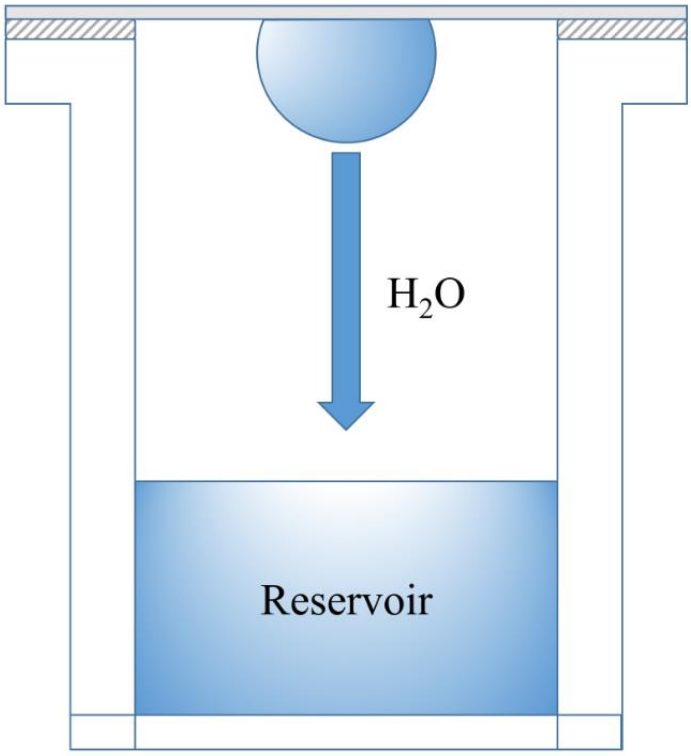

Figure 3.3 Hanging drop vapour diffusion. Droplet composed of a protein, buffer, and precipitant mixture was suspended above a reservoir containing the same buffer and precipitant. The precipitant and protein concentration gradually increase in the drop as the system equilibrates and the water evaporates from the droplet.

3.5.2 Optimization of Crystal Screens

Promising crystallization conditions from the initial sparse matrix screens were selected for optimization. The optimization process was carried out in iterative steps in an attempt to identify the best condition for crystallization of the *D. melanogaster* Tip60 chromodomain. Typically, the precipitant concentration and pH of the buffer were varied first to probe around the initial condition identified by the sparse matrix screen. An illustration of how the optimization procedure was performed is shown in Figure 3.4. As the crystals improved in size and appearance, the concentration and buffer range of the solution was narrowed. Conditions which yielded large single crystals were harvested for x-ray diffraction and data collection.

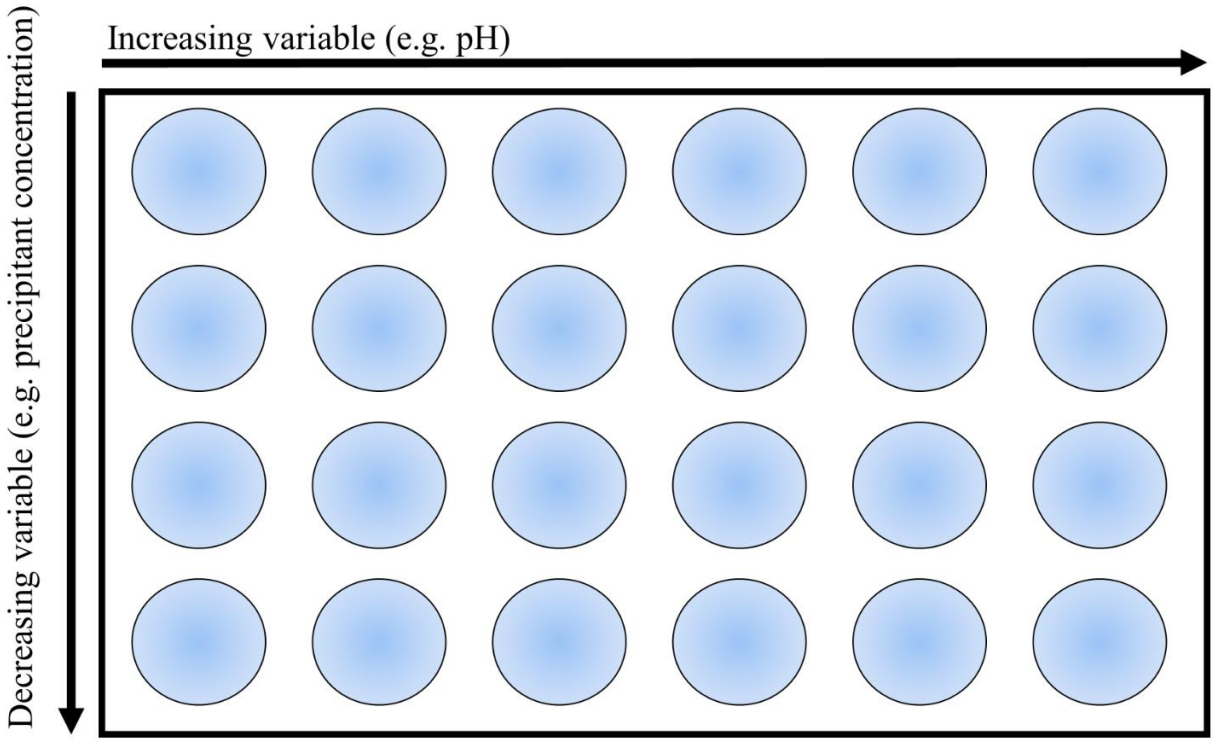

Figure 3.4 Optimization of crystal screens. Illustrated is a 24 well plate where one variable is increased moving from left to right and the other variable is decreased moving from top to bottom. The variables that could be changed include the pH, precipitant concentration, additive concentrations, salt concentrations, or buffer types at a constant pH.

3.6 Structure Determination and Refinement

3.6.1 Collecting Diffraction Data

The protein crystals were taken to the Canadian Light Source for x-ray diffraction and data collection. To prepare the crystals for diffraction, cryo-protectants were prepared for each condition that generated crystals suitable for diffraction. The cryo-protectants were composed of the same conditions which the crystals were grown in with glycerol and/or MPD added. The cryo-protectant was added to the drop, which contained the crystals that were selected for diffraction. The crystals were mounted using nylon loops (Hampton Research) and flash frozen with liquid nitrogen. The data was collected at either the CMCF 08B1-1 Bending Magnet beamline or the CMCF 08ID-1 20 mm hybrid small gap in-vacuum undulator beamline.

3.6.2 Molecular Replacement

The diffraction data was processed with HKL2000 (Otwinowski and Minor, 1997). The structure was solved using the PHASER molecular replacement program (McCoy *et al.*, 2007) from the PHENIX software suite (Adams *et al.*, 2010). The search model used for molecular replacement was a poly-alanine model generated from the unpublished NMR structure of the *H. sapiens* Tip60 chromodomain (PDB ID: 2EKO) deposited in the Protein Data Bank (PDB). The *H. sapiens* Tip60 chromodomain and the *D. melanogaster* Tip60 chromodomain share 70% sequence identity; therefore, it was expected the two proteins adopted similar structures.

3.6.3 Model Building and Refinement

The model building was achieved by using the molecular modeling software COOT (Emsley *et al.*, 2010). The first step of building the *D. melanogaster* Tip60 chromodomain model was to add side chain and backbone residues into the unoccupied areas of the electron density map. Once the first series of changes were completed the updated model was run through the PHENIX refinement program. The process of updating the model to fit side chain residues and backbone residues into the improved electron density map was repeated several times until the R_{free} and R_{work} no longer decreased in value.

The R values were used to monitor the progress of the model building and refinement of the *D. melanogaster* Tip60 chromodomain was. The R values were calculated using Equation 3.1. The R_{work} value measured the agreement between the model and the experimental data, the lower the number the higher the degree of agreement between the model and experimental data.

The R_{free} value randomly selected 5% of the experimental data to not be used during the model building or refinement process. This ensured that no biases were introduced when model building. The lower the R_{free}, the better the model predicts the unused data.

$$R = \frac{\sum_{hkl} ||F_{obs}| - |F_{calc}||}{\sum_{hkl} |F_{obs}|}$$

Equation 3.1 Calculation for the R factor values. Where F is the structure factor amplitudes observed (F_{obs}) or calculated (F_{calc}) for each reflection.

3.7 Peptide Binding Experiments

3.7.1 Preparation of Protein for Recording of $^{1}H–^{15}N$ HSQC Spectra

The *D. melanogaster* Tip60 chromodomain was isotopically labelled with ^{15}N for the NMR study. To isotopically label the *D. melanogaster* Tip60 chromodomain, the initial protein expression conditions are altered to allow for the protein to be expressed as an isotopically labelled protein. This was done by following the methods outlined in section 3.3.1 with the following changes; once an OD_{600} of 0.8 is obtained the *E. coli* cells are pelleted by centrifugation at 5,393 x *g* for 20 minutes and washed in M9 minimal media (22 mM KH_2PO_4, 90 mM Na_2HPO_4, 8 mM NaCl, 22 mM D-glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$). The pelleting and washing of the *E. coli* cells was repeated three times. After the final wash step the cells were pelleted once more by centrifugation at 5,393 x *g* for 20 minutes and then re-suspended in 1 L of M9 minimal media (22 mM KH_2PO_4, 90 mM Na_2HPO_4, 8 mM NaCl, 22 mM D-glucose, 2 mM $MgSO_4$, 0.1 mM $CaCl_2$) containing 1 g of $^{15}NH_4Cl$. The cells were allowed a 20–30 minute recovery period at 37° C before induction. The protein expression was induced with 0.1 mM IPTG for 12 – 16 hours at 25° C and shaking at 210 rpm. The cells were then harvested and the protein purified as outlined in section 3.3. The protein was concentrated and stored as outlined in section 3.3.5.

3.7.2 Surface Plasmon Resonance to Identify Potential Binding Partners of Tip60 Chromodomain

The surface plasmon resonance study was conducted at the Saskatchewan Structural Sciences Centre (University of Saskatchewan) and used the ProteOn XPR36 protein interaction array system from Bio-Rad. This system provided a 6 x 6 interaction array, which allowed for up to six separate ligands to be immobilized onto the sensor chip surface and six separate analytes to be flowed over the sensor chip at once, which then allowed for the detection of 36 separate

interactions (Figure 3.5). The *D. melanogaster* Tip60 chromodomain was covalently linked to the carboxylated alginate polymer surface of the sensor chip (GLM, Bio-Rad) through amine coupling. The carboxyl groups were activated for ligand immobilization by injecting an activation solution containing a 1:1 mixture of 20 mM 1-ethyl-3-(3-dimethylaminopropyl) carbodiimide (EDC) and 5 mM N-hydroxysulfosuccinimide (sulfo-NHS) in the vertical direction for 60 seconds. This resulted in the formation of highly reactive sulfo-NHS esters that reacted with free amines exposed on the ligand surface, which in turn caused the ligand to immobilize onto the sensor chip surface. Due to the overall negative charge of the sensor chip surface following activation, the ligand must have an overall positive charge which was achieved using immobilization buffers with a pH lower than the pI of the protein of interest. For the *D. melanogaster* Tip60 chromodomain, 25 µg/mL was incubated with an immobilization buffer composed of 10 mM sodium malonate at pH 4.5. Immobilization was accomplished by injecting the chromodomain in the vertical direction at a flow rate of 30 µL/min for five minutes. It should be noted that, with SPR, it is assumed that the protein of interest will be immobilized in the same orientation and with all binding sites available.

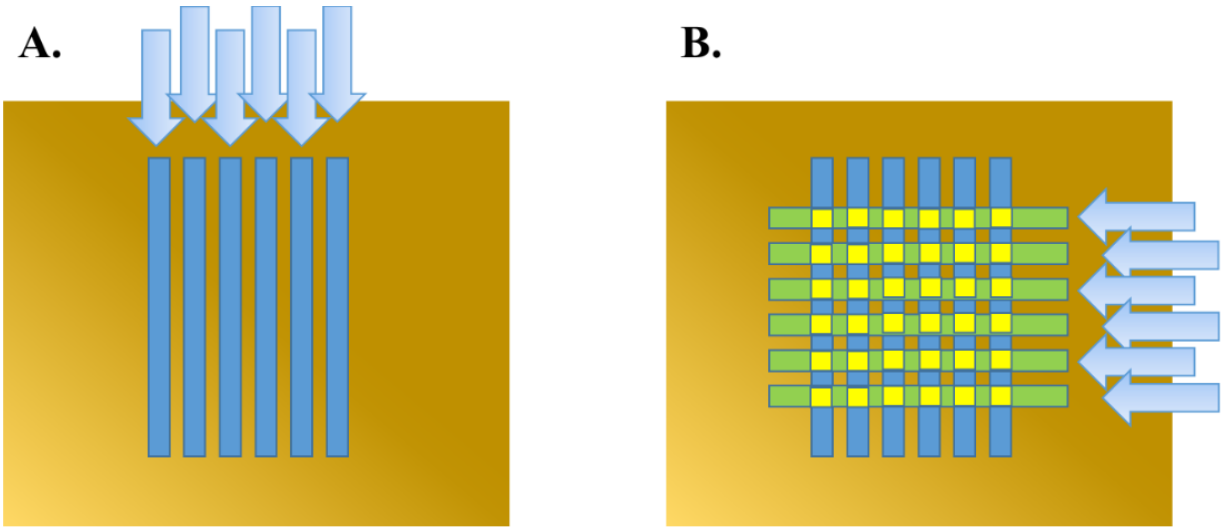

Figure 3.5 Schematic of the ProteOn XPR36 protein interaction array from Bio-Rad. **A.** Initially the protein is immobilized to the sensor chip in the vertical direction (blue rectangles). **B.** The sensor chip is then rotated to allow for analytes to be flowed in the horizontal direction. The six horizontal channels (green rectangles) overlap the vertical channels (blue rectangles) at 36 unique locations (yellow). This allows for 36 separate interactions to be measured at once.

To prevent the interaction between the sensor chip surface and the analytes, the activated and unreacted carboxyl groups were deactivated by injecting 130 mM ethylenediamine at a flow rate of 30 µL/min for five minutes in the vertical direction. Upon the completion of the deactivation step, the response units (RU) were recorded to give an approximate value of the amount of protein immobilized. Following the deactivation of unreacted carboxyl groups, a stable baseline was established through several washes of the sensor chip surface with running buffer composed of 50 mM HEPES (pH 7.4), 150 mM NaCl, 3 mM EDTA, 0.005% Tween20 and a regeneration buffer composed of 10 mM glycine (pH 9.0), 0.5 M NaCl, 0.1% Tween20.

Once the washing of the sensor chip was complete the system was ready for analytes to be injected. The analytes used in this experiment were peptides, which correspond to sections of the H3 or H4 histone N-terminal tail sequences. Additionally, some peptide sequences contained additional modifications to a lysine residue such as acetylation or methylation, which corresponded to post-translational modifications found on the N-terminal histone tails. The peptides used are listed in Table 3.5. These peptides were dissolved in the running buffer to minimize any refractive index changes between different buffers.

To rapidly screen the peptide library for potential binding partners each peptide was injected at a concentration of 100 µM at a flow rate of 30 µL/min and the response monitored. Following each peptide injection, the sensor chip was washed with regeneration buffer to remove any bound peptides. Every injection was double referenced; the first reference required that one ligand channel was activated/deactivated but no protein was immobilized; the second reference was to inject buffer over the immobilized protein. The responses to both references were monitored and subtracted from the signal generated by the protein-peptide interaction.

There are two common methods utilized for collecting kinetic data: pre-steady state and steady state experiments. The steady state experiment allows for the dissociation constant (K_D) of a binding interaction to be approximated. The K_D values for the histone tail peptides were measured by a steady state affinity experiment, which required ten different concentrations for each peptide. The concentration ranges for the titrated peptides varied depending on the expected K_D. One group of peptides had concentrations which ranged from 1.85 µM to 300 µM; another group of peptides had concentrations which ranged from 6.17 µM to 1 mM. The RUs from each interaction were then normalized using Equation 3.2 and the normalized responses were plotted

against the concentrations. The normalization of the RUs allowed for the data obtained from each peptide to be compared. The K_D was then calculated using Equation 3.3. The R_{max}, which is the maximum theoretical response of a particular peptide interacting with the ligand, was calculated using Equation 3.4. The R_{max} value was used in the calculation of the K_D.

Table 3.5 Histone tail peptides for binding experiments

Histone tail peptide	**N/C Terminus**	**Sequence**	**Molecular Weight (Da)**
H3K4Me$_1$	Ac/Amide	ART**(KMe)**QTARKSTGGY	1579
H3K9Me$_3$	Ac/Amide	KQTAR**(KMe3)**STGGY	1281
H3K27Me$_3$	Ac/Amide	SKAAR(**KMe3**)SAPAY	1234
H3K36Me$_1$	Ac/Amide	STGGV**(KMe)**KPHRY	1284
H3K36Me$_3$	Ac/Amide	STGGV**(KMe3)**KPHRY	1314
H4$_{6-24}$ Unmodified	H/NH$_2$	GGKGLGKGGAKRHRKVLRDY	2153
H4K12Ac	NH$_2$/Amide	GGKGLG(**KAc**)GGAKRHRKVLRDY	2195
H4K16Ac	NH$_2$/OH	GGKGLGKGGA(**KAc**)RHRKVLRDY	2194
H4K20Me$_1$	NH$_2$/Amide	GGKGLGKGGAKRHR(**KMe**)VLRDY	2166
H4K20Me$_1$ short	Ac/Amide	YGAKRHR(**KMe**)V	1169
H4K20Me$_3$	Ac/Amide	GAKRHR**(KMe3)**VLRDNY	1698
H4$_{9-15}$	NH$_2$/Amide	GLGKGGAY	721
H4$_{16-20}$	NH$_2$/Amide	KRHRKY	886
H4$_{20-24}$	NH$_2$/Amide	KVLRDNY	793
H4D24K	NH$_2$/Amide	HRKVLR**K**NY	1212
H4K20 short	Ac/Amide	YGAKRHRKV	1155
H4R17A	Ac/Amide	YGAK**A**HRKV	1070
H4R17E	Ac/Amide	YGAK**E**HRKV	1128
H4H18A	Ac/Amide	YGAKR**A**RKV	1089
H4H18E	Ac/Amide	YGAKR**E**RKV	1147
H4R19A	Ac/Amide	YGAKRH**A**KV	1070
H4R19E	Ac/Amide	YGAKRH**E**KV	1128
*H4$_{13-20}$ Unmodified	Ac/Amide	YGGAKRHRKV	1212

* – used only in ITC experiments

$$RU_{\text{Normalized}} = \left(\frac{RU_{\text{Observed}}}{RU_{\text{Ligand Immobilized}}} \times \frac{\text{Molecular Weight}_{\text{Ligand}}}{\text{Molecular Weight}_{\text{Analyte}}}\right) \times \frac{\text{Molecular Weight}_{\text{Heaviest Analyte}}}{\text{Molecular Weight}_{\text{Analyte of interest}}}$$

Equation 3.2 The normalization equation for SPR. The ratio of the response units observed for a particular peptide to the response units of the immobilized ligand was multiplied by the ratio of the molecular weight of the ligand to the peptide (analyte) of interest. This value was then multiplied by the ratio of the molecular weight of the heaviest peptide used in the series to the peptide of interest.

$$R_{eq} = \frac{R_{max}[A]}{K_D + [A]}$$

Equation 3.3 Calculating K_D at equilibrium. R_{eq} is the response at equilibrium, R_{max} is the maximum theoretical response of particular peptide interacting with the ligand, [A] is the concentration of peptide.

$$R_{max} = n\frac{M_A}{M_L}R_L$$

Equation 3.4 The R_{max} equation calculates the maximum theoretical response of a particular peptide interacting with the ligand. *n* is the stoichiometric value of the interaction, M_A is the mass of the analyte (peptide), M_L is the mass of the immobilized ligand, and R_L is the amount of the ligand immobilized in response units (RUs).

3.7.3 Isothermal Titration Calorimetry to Confirm and Characterize Interaction between Tip60 Chromodomain and Peptides

Isothermal titration calorimetry (ITC) was performed at the Protein Characterization and Crystallization Facility (University of Saskatchewan) and used the Nano-ITC from TA Instruments. The protein and the peptides were prepared in a buffer composed of 30 mM CHES pH 9.5, 200 mM NaCl, 2 mM TCEP, and 10% MPD. The $H4_{13\text{-}20}$ unmodified peptide was used in the ITC study and the sequence is listed in Table 3.5. This peptide was selected based on the results from SPR.

The *D. melanogaster* Tip60 chromodomain was concentrated to 0.34 mM (3.0 mg/mL) and 170 µL was placed into the reaction cell. Two hundred microliters of buffer were placed into the reference cell and the 50 µL syringe was filled with 6 mM – 8.5 mM of the $H4_{13\text{-}20}$ unmodified peptide. The syringe was placed into the sample cell, the temperature was set to 25° C, and the stir rate set to 200 rpm. Once the stirring began, the system was allowed to equilibrate for 30 –

45 minutes until a stable baseline was achieved. Once the baseline was stabilized the experiment was started. The experiments were performed with either 21 or 26 total injections, using an injection volume of 2.5 µL or 2.02 µL respectively. Two small volume injections were performed at the beginning of every injection series to minimize the equilibration artifacts typically seen with the first two injections of every series; these injections were not included in the data analysis.

Once the initial peptide titration was completed, two series of reference titrations were also completed. The first series of reference titrations used 50 µL of peptide loaded into the syringe and injected into the reaction cell containing 170 µL of buffer. The second series of reference titrations used 50 µL of buffer loaded into the syringe and injected into the reaction cell containing 170 µL of protein. Both series of reference titrations followed the same parameters used in the peptide into protein titrations. Once all titrations were completed, the data could be analysed using the NanoAnalyze software from TA Instruments.

4 RESULTS

4.1 Sequence alignment of the Tip60 chromodomain to other related chromodomains

Previous studies conducted by two different research groups suggested that the Tip60 chromodomain preferentially bound to either H3K9Me$_3$ (Sun *et al.*, 2009) or H3K4Me$_1$ (Jeong *et al.*, 2011) through the use of an aromatic cage. The canonical chromodomain, HP1α was known to interact with H3K9Me$_3$ through a hydrophobic cage composed of three aromatic amino acids (Nielsen *et al.*, 2002). The aforementioned research groups performed a sequence alignment between the Tip60 and HP1α chromodomains and found that Tip60 possessed a similar arrangement of conserved aromatic residues in its primary sequence (Trp26, Phe43, and Tyr47), presuming it to be required for H3K9Me$_3$ binding (Jeong *et al.*, 2011; Sun *et al.*, 2009). As previously discussed, Tip60 belongs to the MYST family of HATs and it was found that another member of this family, the males-absent-on-the-first (MOF) chromodomain, did not possess an aromatic cage (Conrad *et al.*, 2012). This was significant, as it was found that *Homo sapiens* MOF and *H. sapiens* Tip60 share approximately 45% sequence identity over the course of the entire protein (Peng *et al.*, 2012).

The high level of sequence identity between Tip60 and MOF over the entire protein indicated that the two proteins likely adopt similar structures. To further examine this, a sequence alignment of the chromodomains belonging to Tip60, MOF, and Esa1 was performed (Figure 4.1A). The sequence alignment revealed a high level of sequence conservation between the MYST family members as expected. However, a sequence alignment between the chromodomain containing MYST family members and the HP1α chromodomain showed poor sequence conservation. Furthermore, it was found that the Tip60 chromodomain and HP1α chromodomain share only 17% sequence identity. Due to the high level of sequence conservation between the Tip60 chromodomain and its related MYST family members, it was reasonable to speculate that the Tip60 chromodomain was likely to adopt a structure similar to the MOF or Esa1 chromodomain, as opposed to the HP1α chromodomain. Furthermore, a previous study compared the chromodomains of MOF and HP1α and revealed that, not only does the MOF chromodomain lack an aromatic cage, but the residue Arg387 would actually interfere with methyllysine binding (Figure 4.1B) (Nielsen *et al.*, 2005). This information casts doubt onto the data published by Sun *et al.* (2009) and Jeong *et al.* (2011), attributing the methyllysine binding capabilities to the Tip60 chromodomain via an aromatic cage.

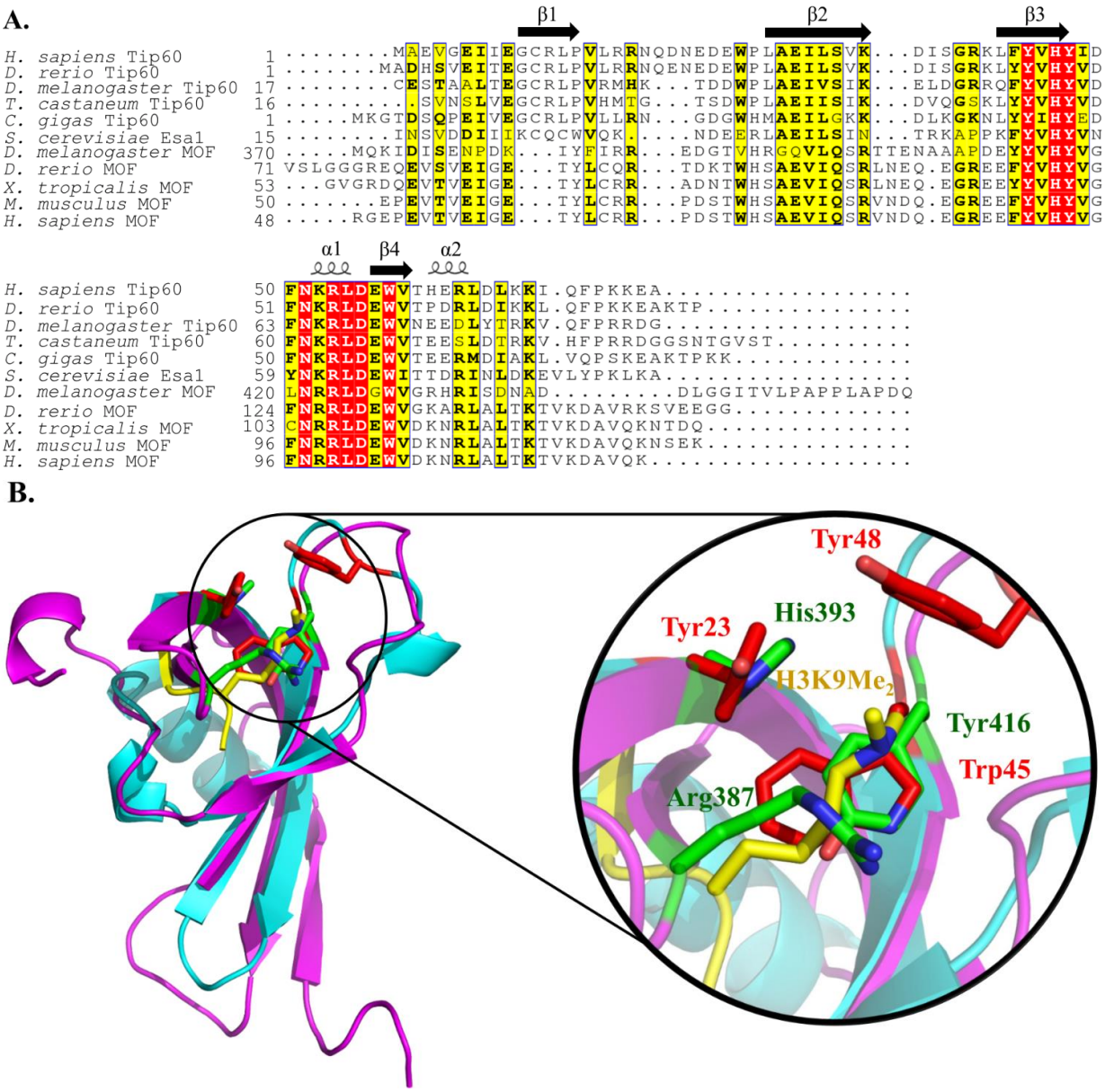

Figure 4.1 Comparison of different chromodomains. **A.** Sequence alignment of the chromodomain containing MYST family members. The residues highlighted in red show strict sequence identity whereas those residues highlighted in yellow and bolded show sequence similarity. Sequence alignment was generated using clustal omega (Goujon *et al.*, 2010; Sievers *et al.*, 2011). Secondary sequence features and conserved sequences identified using ESPript (http://espript.ibcp.fr) (Robert and Gouet, 2014). **B.** Superimposed close-up view of the methyllysine binding pocket of MOF (magenta) and HP1α (cyan) chromodomains from *D, melanogaster*. By superimposing the MOF chromodomain onto the HP1α chromodomain bound to H3K9Me$_2$ (yellow) it was revealed that an arginine residue in the MOF chromodomain would interfere with methyllysine binding. The key residues found in the HP1α binding pocket are coloured red and the equivalent residues in MOF are coloured green. Figure 4.1B right panel adapted from Nielsen *et al.*, 2005. Overlay generated using CCP4 software suite (Sievers *et al.*, 2011; Winn *et al.*, 2011) secondary structure matching program (Krissinel and Henrick, 2004).

4.2 Cloning of *Homo sapiens* and *Drosophila melanogaster* Tip60 domains

The chromodomain and HAT domain of *H. sapiens* were cloned separately from *E. coli* codon optimized synthetic DNA and the *D. melanogaster* Tip60 chromodomain and HAT domain were cloned individually from full length cDNA sequences using PCR. The domains were cloned separately to simplify the crystallization process. The results from cloning the *D. melanogaster* Tip60 chromodomain and HAT domain are shown in Figure 4.2A. The cloning of the *H. sapiens* Tip60 HAT domain are shown in Figure 4.2B. The cloning of the *H. sapiens* Tip60 HAT domain yielded low concentrations of PCR products; therefore, different conditions were attempted to increase the yield of PCR products. The template concentrations were varied from 1 ng/µL to 2 ng/µL in combination with using either *Taq* or *Pfu* DNA polymerases. To further optimize the yield, the template concentration was increased to 4 ng/µL and 2 mM $MgSO_4$ was added, however, this proved to be unsuccessful (Figure 4.2C). It was found that the use of *Taq* polymerase was unsuccessful for the production of any PCR products as seen in Figure 4.2B.

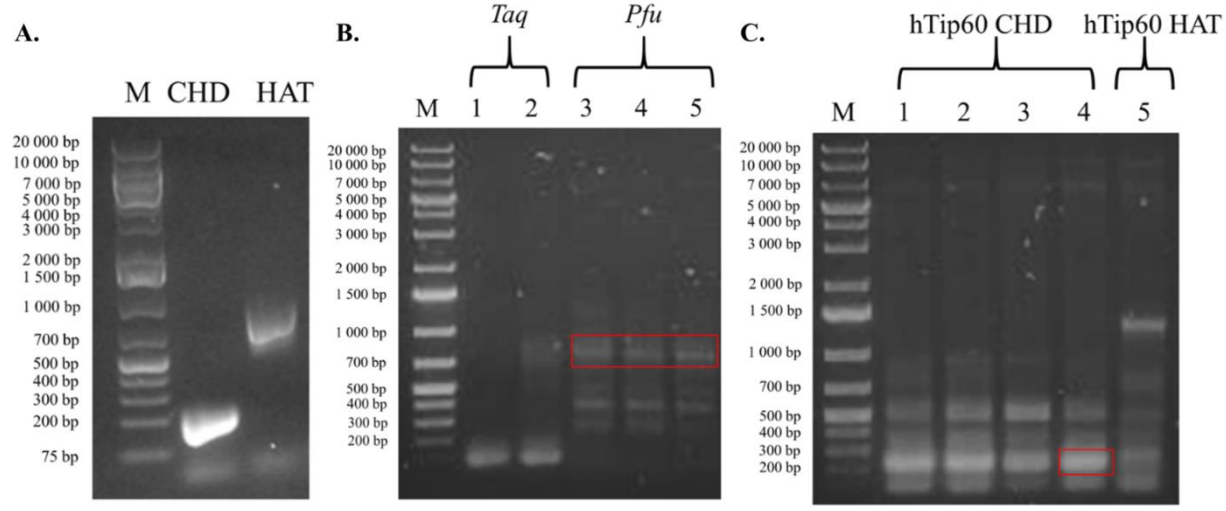

Figure 4.2 Cloning of *Drosophila melanogaster* and *Homo sapiens* Tip60 chromodomains and HAT domains. All DNA samples were run on 1 % agarose gel and HydraGreen was used to visualize the DNA. **A.** Cloning of the *D. melanogaster* Tip60 chromodomain (CHD) and HAT domain. 1 ng/µL of template DNA was used. **B.** Cloning of the *H. sapiens* Tip60 HAT domain used 1 – 2 ng/µL of template DNA and *Taq* or *Pfu* DNA polymerase. Each lane contained different amount of template DNA as follows: Lane 1 - 1 ng/µL, Lane 2 - 2 ng/µL, Lane 3 – 1 ng/µL, Lane 4 – 1.5 ng/µL, Lane 5 - ng/µL. The reactions highlighted by the red box were combined together and used for transformation. **C.** Cloning of the *H. sapiens* Tip60 chromodomain with varying amounts of $MgSO_4$. Lanes 1 – 3 contained 2 ng/µL template DNA and varying amounts of $MgSO_4$. The amount of $MgSO_4$ in lanes 1 – 3 is as follows: Lane 1 – 1 mM, Lane 2 – 2 mM, Lane 3 – 4 mM. Lane 4 contained 4 ng/µL of DNA template and 2 mM $MgSO_4$. The attempted optimization of the cloning of the HAT domain (lane 5) used 4 ng/µL of template DNA and an additional 2 mM $MgSO_4$.

The inefficient cloning of the *H. sapiens* Tip60 domains could be attributed to the template DNA originating from an *E. coli* codon optimized synthetic DNA sequence.

The cloned domains were subjected to a restriction digest using *Bam*HI and *Eco*RI to cut the 5' and 3' ends of the insert respectively. The inserts were then ligated into the pGEX-6P-3 expression vector and used to transform XL-1 *E. coli* competent cells. The cells were plated and the resultant colonies were screened by colony PCR (Figure 4.3). Figure 4.3A shows the colony PCR results of the *E. coli* cells transformed with the pGEX-6P-3 expression vector carrying the *D. melanogaster* Tip60 chromodomain insert. Of the eight colonies picked, only colonies 1 and 8 do not appear to contain the chromodomain insert. Colonies 2 – 7 clearly show the presence

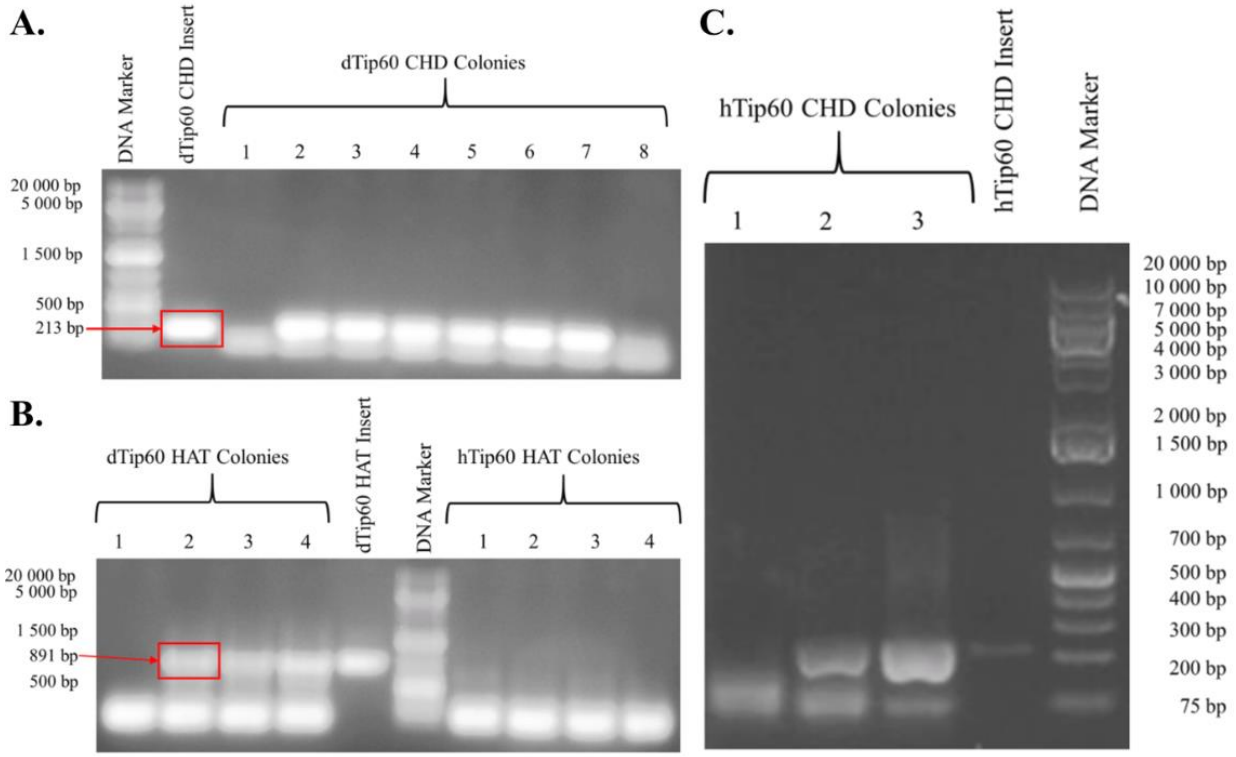

Figure 4.3 Colony PCR of *Drosophila melanogaster* and *Homo sapiens* Tip60 domains. All DNA samples were run on 1 % agarose gel and HydraGreen was used to visualize the DNA. **A.** Shows colony PCR of XL-1 *E. coli* cells transformed with the pGEX-6P-3 vector carrying the *D. melanogaster* Tip60 chromodomain insert. The *D. melanogaster* Tip60 chromodomain insert was run alongside as a positive control. **B.** Colony PCR of XL-1 *E. coli* cells transformed with pGEX-6P-3 vector carrying either the *D. melanogaster* Tip60 HAT domain or the *H. sapiens* Tip60 HAT domain. The *D. melanogaster* Tip60 HAT domain insert was run alongside as a positive control. **C.** Colony PCR of XL-1 *E. coli* cells transformed with the pGEX-6P-3 vector carrying the *H. sapiens* Tip60 chromodomain. The *H. sapiens* Tip60 chromodomain insert was run alongside as a positive control.

of the *D. melanogaster* Tip60 chromodomain. Colonies 6 and 7 were selected for sequencing as these were the most intense bands in the colony PCR gel which indicated that a higher concentration of the target DNA insert was present in the cells. Figure 4.3B shows the results of colony PCR of the cells transformed with the pGEX-6P-3 vector carrying either the *D. melanogaster* Tip60 HAT domain or *H. sapiens* Tip60 HAT domain insert. Of the colonies transformed with the *D. melanogaster* Tip60 HAT domain, only colony 1 did not have the insert present and colonies 2 and 4 were selected for sequencing as these appeared to have the most intense bands in the agarose gel. The colonies transformed with the pGEX-6P-3 vector carrying the *H. sapiens* Tip60 HAT domain appeared to only have one positive result; colony 1. Figure 4.3C shows the result of colony PCR of the colonies transformed with the pGEX-6P-3 vector carrying the *H. sapiens* Tip60 chromodomain. The screen revealed that colony 1 does not contain the insert, whereas colonies 2 and 3 show presence of the chromodomain insert. Colonies 2 and 3 were selected for DNA sequencing.

To confirm the presence of the gene of interest in each of the positive colonies, a plasmid preparation, double restriction digest, and DNA sequencing were performed. The sequencing data revealed that only two sequenced colonies did not contain the correct sequence for the gene of interest. Two colonies, one of which contained the *H. sapiens* Tip60 chromodomain and one of which contained the *H. sapiens* Tip60 HAT domain, were unable to integrate the gene of interest into the expression vector successfully. The results of the DNA sequencing are summarized in Table 4.1.

Table 4.1 Colony PCR sequencing results

Colony Number	**Gene of Interest Present**	**Sequencing Result**
6	*D. melanogaster* Tip60 chromodomain	Correct sequence
7	*D. melanogaster* Tip60 chromodomain	Correct sequence
2	*D. melanogaster* Tip60 HAT domain	Correct sequence
4	*D. melanogaster* Tip60 HAT domain	Correct sequence
2	*H. sapiens* Tip60 chromodomain	Incorrect sequence
3	*H. sapiens* Tip60 chromodomain	Correct sequence
1	*H. sapiens* Tip60 HAT domain	Incorrect sequence

4.3 Protein expression and purification

Following the insertion of the chromodomains and the *D. melanogaster* HAT domain into the pGEX expression vectors, trials were conducted to determine the ideal growth and expression conditions of the constructs. Initially, plasmids which contained the *D. melanogaster* Tip60 chromodomain, HAT domain, and *H. sapiens* Tip60 chromodomain inserts were selected for the expression trials. Due to the expression of the proteins of interest as GST fusion proteins, a GST pull down was utilized. The expected size of the fusion proteins are as follows: *H. sapiens* Tip60 chromodomain is 35.3 kDa, *D. melanogaster* Tip60 chromodomain is 35.2 kDa, and the *D. melanogaster* Tip60 HAT domain is 62 kDa.

From the GST-pull downs it was found that the *H. sapiens* Tip60 chromodomain was optimally expressed in Rosetta 2 *E. coli* competent cells as opposed to BL21(DE3) *E. coli* competent cells (Figure 4.4). When comparing the two bands in the GST-pull down lanes highlighted in Figure 4.4, it could be seen that the band corresponding to the fusion protein was much darker in the Rosetta 2 gel (Figure 4.4B), which indicated the presence of a higher concentration protein. The most likely reason for this difference could be attributed to the Rosetta 2 *E. coli* strain possessing the additional tRNAs for improved eukaryotic protein expression. Additionally, it should be noted that the presence of free GST in the GST-pull down lanes was likely due to incomplete protein synthesis by *E. coli*, resulting in the early termination of the recombinant protein's synthesis or the presence of contaminants resulting in proteolytic cleavage (Figure 4.4).

The expression trials for the *D. melanogaster* Tip60 chromodomain and HAT domain revealed the optimal conditions for the expression of both proteins. For the HAT domain it was found that the protein was better expressed within the BL21(DE3) strain than the Rosetta 2 strain (Figure 4.5). This could be seen when the bands (highlighted in red in Figure 4.5A and B) in the GST pull down columns were compared. The band corresponding to the fusion protein in the BL21(DE3) gel appeared to be thicker than that in the Rosetta 2 gel, which indicated a higher concentration of protein.

The *D. melanogaster* Tip60 chromodomain was found to be optimally expressed in the BL21(DE3) strain (Figure 4.6). The amount of the GST-chromodomain fusion protein derived from *D. melanogaster* was significantly greater than the amount of GST-chromodomain fusion

protein derived from *H. sapiens* (Figure 4.6A and B). For this reason, all subsequent experiments were carried out using the *D. melanogaster* Tip60 chromodomain.

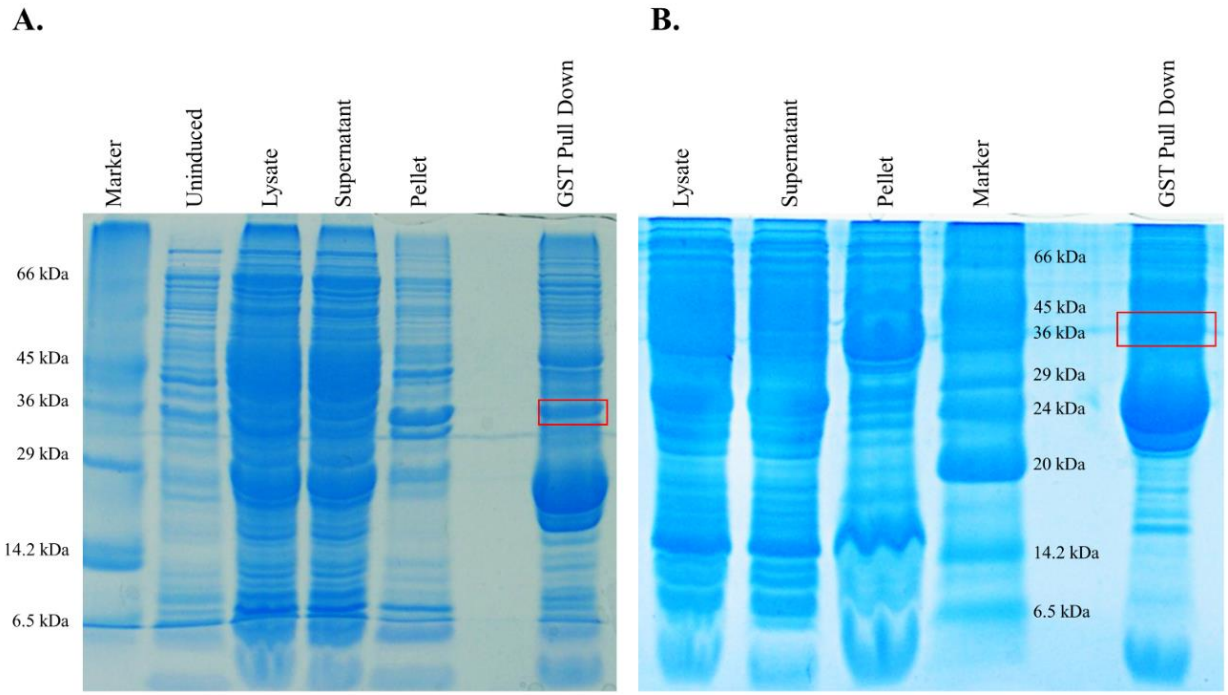

Figure 4.4 Results of protein expression trials for *H. sapiens* Tip60 chromodomain. **A.** *H. sapiens* Tip60 chromodomain expressed in BL21(DE3) competent cells. The chromodomain was expressed by induction with 0.1 mM IPTG at 25° C for 16 hours. The GST-chromodomain fusion protein is highlighted in red. **B.** *H. sapiens* Tip60 chromodomain expressed in Rosetta 2 competent cells. The chromodomain was expressed by induction with 0.1 mM IPTG at 25° C for 16 hours. The GST-chromodomain fusion protein is highlighted in red.

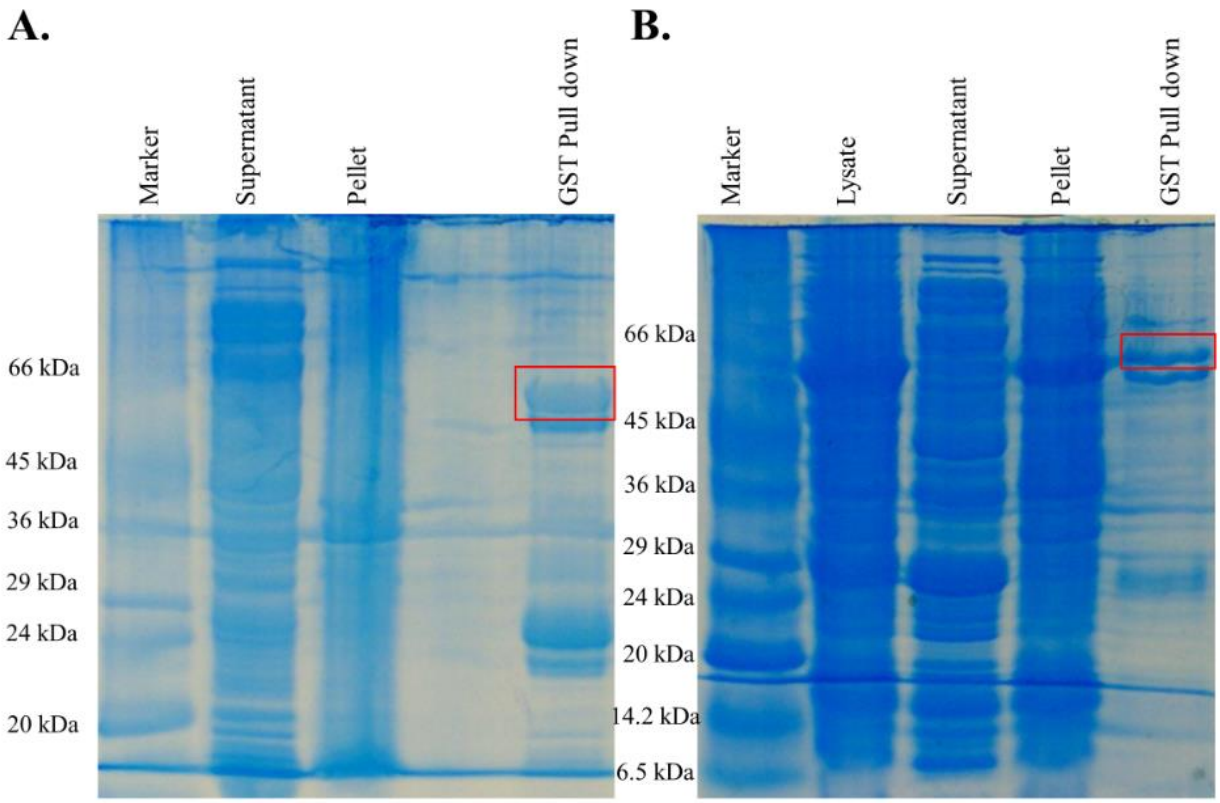

Figure 4.5 Results of protein expression trials for *D. melanogaster* Tip60 HAT domain. **A.** *D. melanogaster* Tip60 HAT domain expressed in BL21(DE3) competent cells. Expression was induced with 0.1 mM IPTG at 25° C for 16 hours. The GST-HAT fusion protein is highlighted in red. **B.** *D. melanogaster* Tip60 HAT domain expressed in Rosetta 2 competent cells. Expression was induced with 0.1 mM IPTG at 25° C for 16 hours. The GST-HAT fusion protein is highlighted in red.

Following the expression trials, large scale purification of the *D. melanogaster* Tip60 chromodomain was carried out. As described in section 3.3.2, the protein was initially purified with two rounds of glutathione sepharose affinity purification. The first round of glutathione sepharose affinity purification was designed to separate the GST-tagged chromodomain from the cellular lysate (shown in the lane labeled "1st GST Eluate" in Figure 4.7A). As outlined in section 3.3.2, the GST-tagged chromodomain was subjected to cleavage by PreScission Protease (shown in the lane labeled "GST-DrosoCHD + PreScission Protease" in Figure 4.7A), followed by a second round of glutathione sepharose affinity purification. This additional round of glutathione sepharose affinity purification resulted in the chromodomain being eluted free of the GST tag (highlighted in red in Figure 4.7A). To further purify the protein, anion-exchange was performed using the method that was previously described in section 3.3.3. The results of anion exchange

showed that any impurities which had remained from the glutathione sepharose affinity purification had been removed; only the pure chromodomain remained (Figure 4.7B).

Size-exclusion chromatography was also performed on the *D. melanogaster* Tip60 chromodomain to determine its oligomerization state and follows the method outlined in section 3.3.4. It was important to elucidate the oligomerization state of the chromodomain, as it could have potential implications on how the protein packs in a crystal and what surfaces could be available for binding. It was found that the chromodomain, which has a molecular weight of 8.411 kDa, eluted at a volume of 14.97 mL (Figure 4.8). A comparison of this volume to those of the proteins used in the molecular weight standard revealed that the chromodomain eluted between the volumes of the 6.5 kDa and 13.5 kDa standards. This indicated that the chromodomain was a monomer in solution.

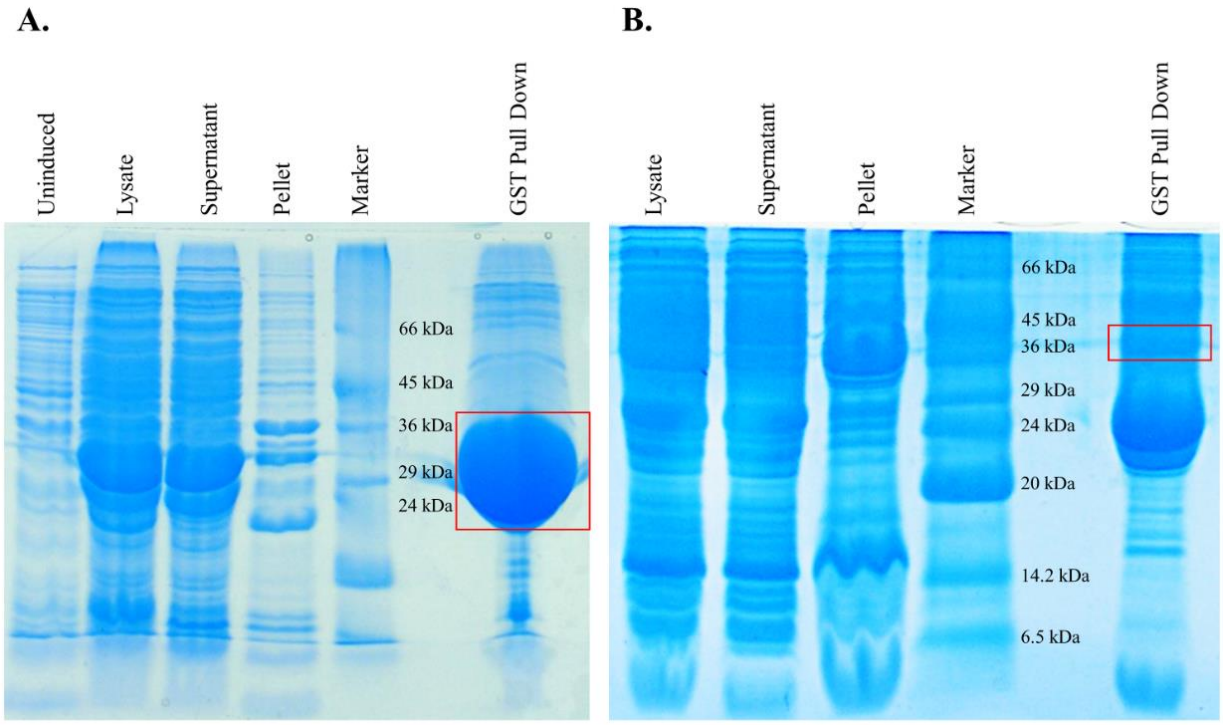

Figure 4.6 Comparison between *D. melanogaster* Tip60 chromodomain and *H. sapiens* Tip60 chromodomain. **A.** *D. melanogaster* Tip60 chromodomain expressed in BL21(DE3) competent cells. Expression was achieved through induction with 0.1 mM IPTG at 25° C for 16 hours. The GST-chromodomain fusion protein is highlighted in red. **B.** *H. sapiens* Tip60 chromodomain expressed in Rosetta 2 competent cells. Expression conditions were identical to those listed in Figure 4.4B. The GST-chromodomain fusion protein is highlighted in red.

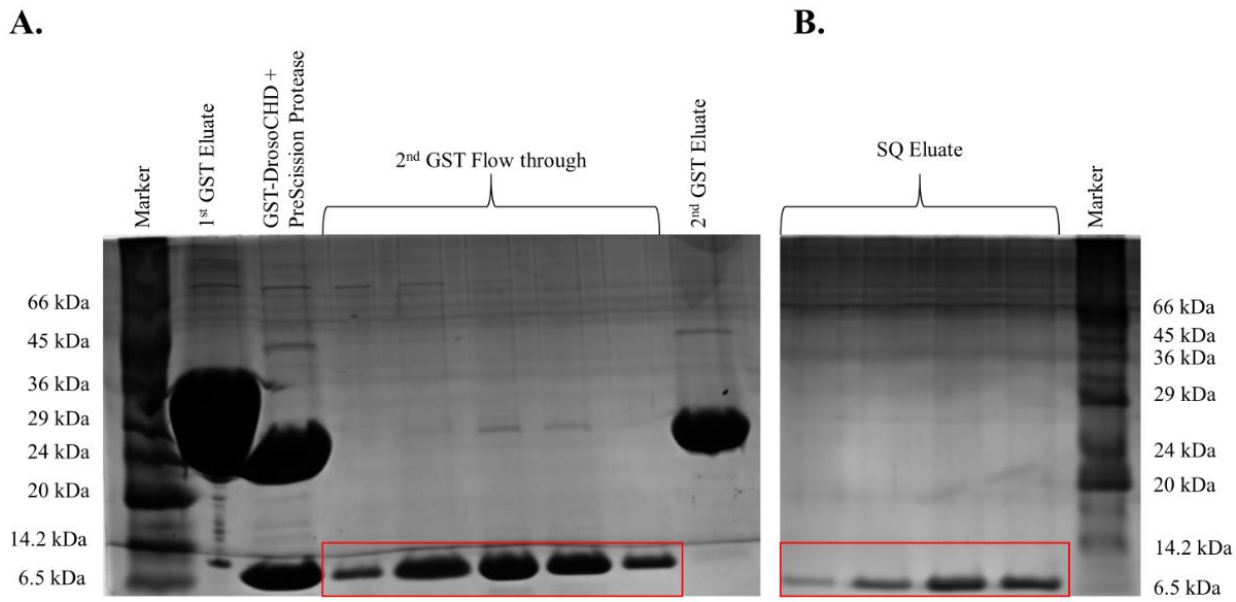

Figure 4.7 Results of the protein purification of *D. melanogaster* Tip60 chromodomain. **A.** The results of two rounds of glutathione sepharose affinity purification were shown. The eluted GST-chromodomain fusion protein could be seen in the lane labeled "1st GST Eluate". The cleaved products from the addition of PreScission Protease were shown in the lane labeled "GST-DrosoCHD + PreScission Protease." The cleaved GST tag is seen between the 24 kDa and 29 kDa markers and the chromodomain is seen close to the 6.5 kDa marker. The results of second round of glutathione sepharose affinity purification was shown in the lanes labeled "2nd GST Flow through" and "2nd GST Eluate." During this round of purification, the chromodomain appeared in the flow through fractions (highlighted in red) and the GST tag remained bound to the column until eluted by the addition of excess glutathione. The GST tag could be seen in the "2nd GST Eluate" lane. **B.** The results of anion-exchange (SQ) purification of the chromodomain. The purified chromodomain is highlighted in red.

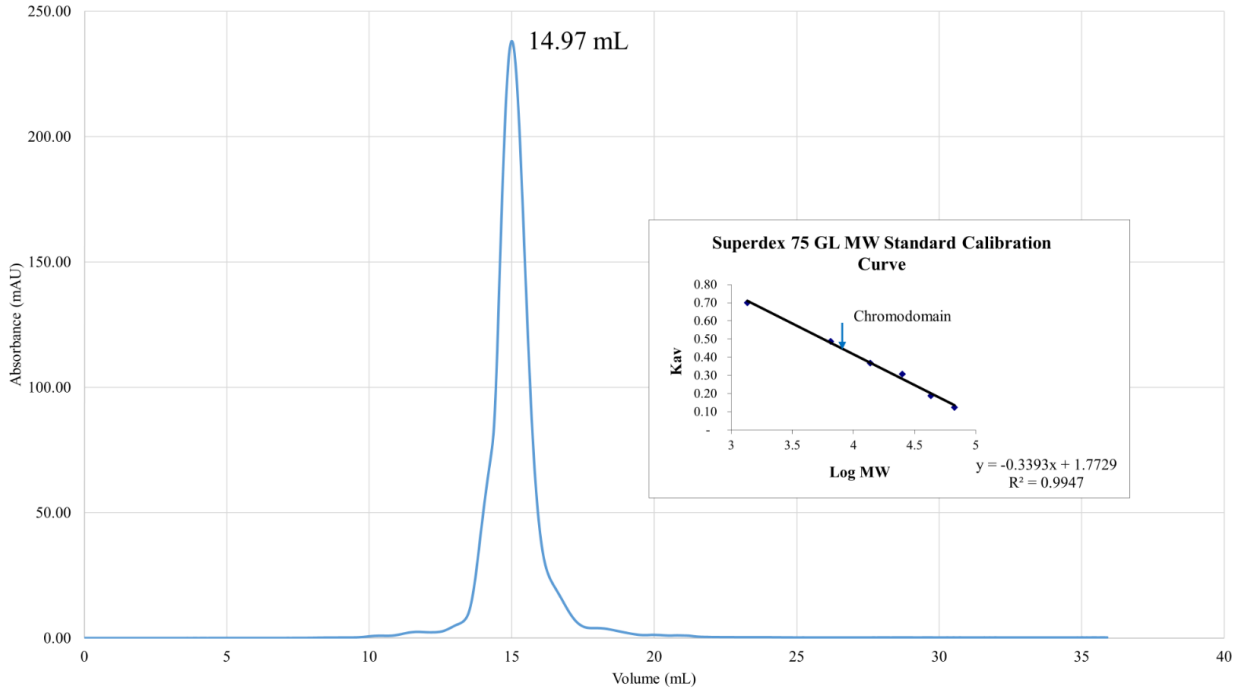

Figure 4.8 Size-exclusion chromatography of the *D. melanogaster* Tip60 chromodomain. Chromatogram of the size-exclusion chromatography of the chromodomain. Five hundred microliters of a 2 mg/mL protein solution was loaded onto a Superdex 70 10/300 column. The absorbance at 280 nm was monitored during the elution of the protein. The chromodomain eluted at 14.97 mL.

4.4 Crystal screening and optimization of crystallization conditions for *Drosophila melanogaster* Tip60 chromodomain

Initial crystallization trials utilized the Wizard I sparse matrix screening kit from Emerald Biosciences (now Rigaku Reagents) to identify potential crystallization conditions for the chromodomain. It was found from the sparse matrix screen that there were many conditions which produced crystals of the chromodomain. The sparse matrix screen also revealed that many of these conditions were typically above or below the isoelectric point (pI) of the chromodomain (the pI was 6.06) by two pH units. Additionally, many of the conditions used either a polyethylene glycol (PEG) of varying molecular weights (e.g. PEG3000 or PEG8000) or ammonium sulphate as the main precipitant reagent. From the 48 different conditions screened in the kit, the most promising conditions were selected for further optimization (Figure 4.9). These conditions were selected on the basis that the crystals produced were fairly large and singular.

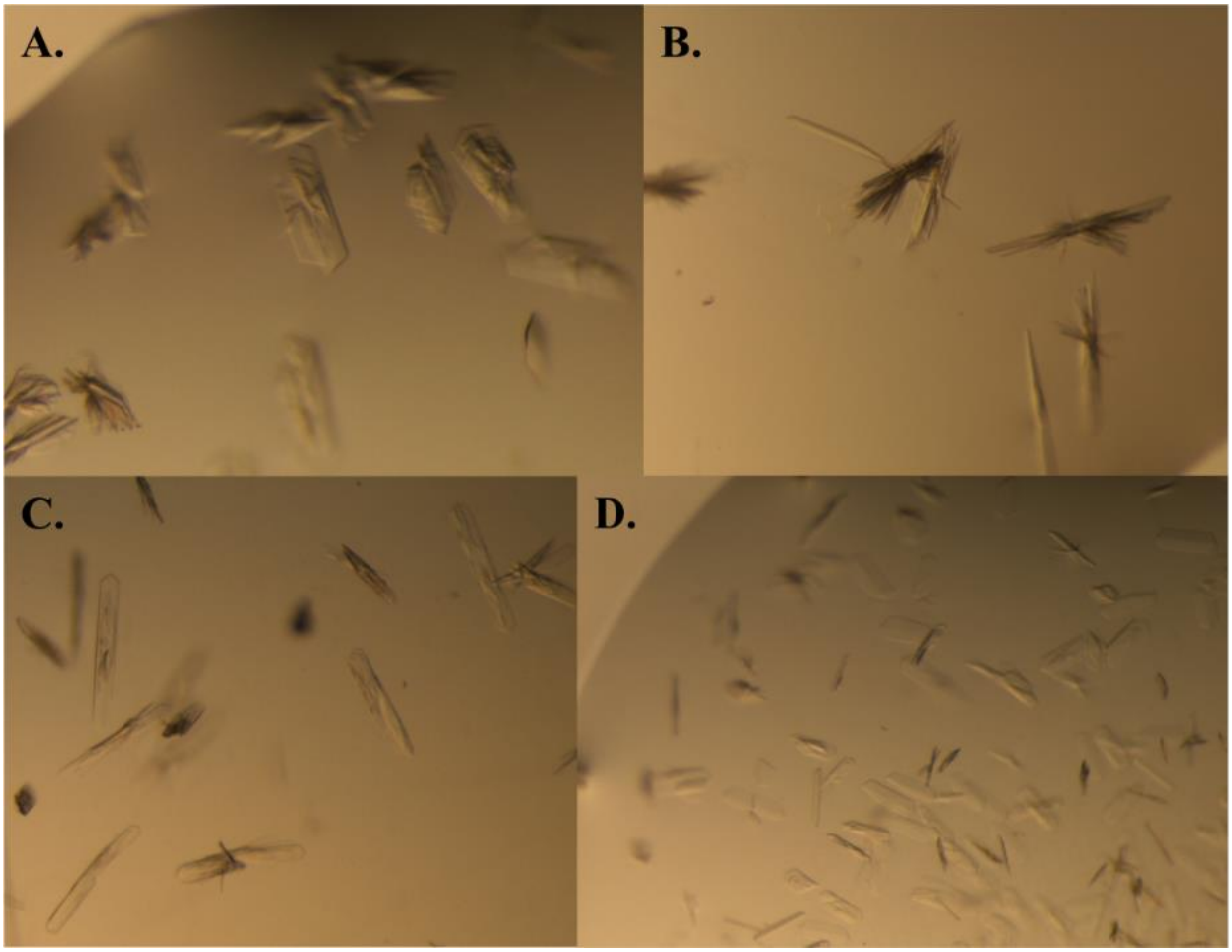

Figure 4.9 Crystals of the *D. melanogaster* Tip60 chromodomain produced from the Wizard I Sparse Matrix screening kit. All images shown were captured at 4X magnification. The chromodomain was concentrated to 2.5 mg/mL and the hanging drop vapour diffusion method was used. Each droplet was composed of 1 µL protein solution and 1 µL precipitant solution. **A.** The precipitant solution was composed of 0.1 M citrate (pH 5.5) and 20 % PEG3000. **B.** The precipitant solution was composed of 0.1 M cacodylate (pH 6.5) and 1.26 M $(NH_4)_2SO_4$. **C.** The precipitant solution was composed of 0.1 M CAPS (pH 10.5), 0.2 M Li_2SO_4, and 1.2 M NaH_2PO_4/0.8 M K_2HPO_4. **D.** The precipitant solution was composed of 0.1 M phosphate-citrate (pH 4.2), 0.2 M NaCl, and 20 % PEG8000.

The crystallization conditions from the sparse matrix screen served as the basis for which to further investigate the crystallization space of the chromodomain. From each condition selected there were two variables that were varied: the pH and the precipitant concentration. By varying these conditions around the base condition found in the sparse matrix screen, it was possible to narrow down the optimal conditions for the production of large singular crystals. The optimization process revealed that the conditions which produced the best crystals were composed of 0.1 M phosphate-citrate at a pH range of 4.6 – 5.4, 0.2 M NaCl, 10 % MPD, and

PEG8000 at a concentration range of 18 – 24 % (Figure 4.10). These crystals were found to be large and singular and suitable for x-ray diffraction and data collection. The crystals were diffracted at the Canadian Light Source (CLS), where it was found that the condition composed of 0.1 M phosphate-citrate (pH 4.6), 0.2 M NaCl, 10 % MPD, and 22 % PEG8000 (Figure 4.10B) diffracted the best. These results are discussed in more detail in section 4.5.

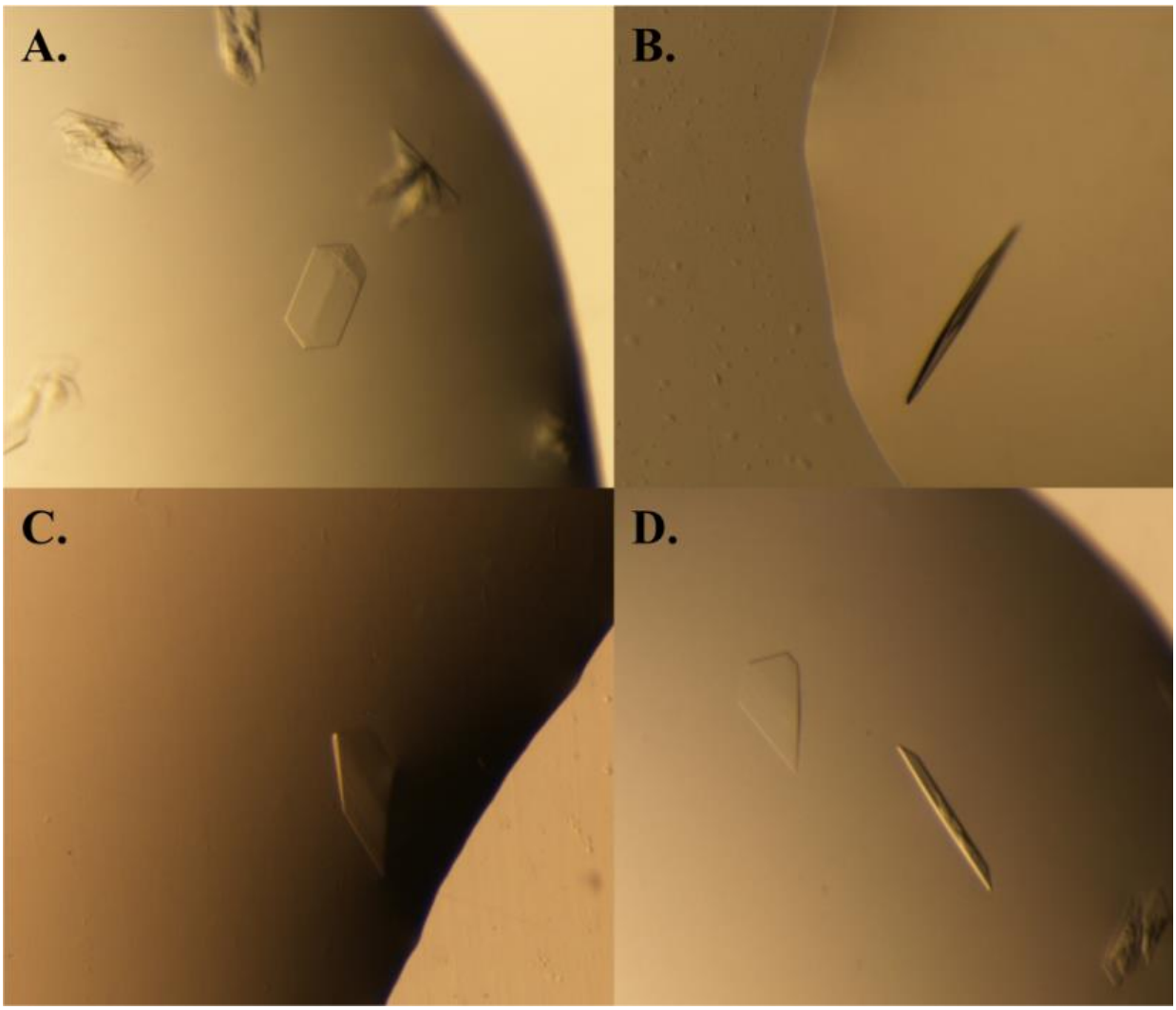

Figure 4.10 *D. melanogaster* Tip60 chromodomain crystals produced from optimization. All images shown were captured at 4X magnification. The chromodomain was concentrated to 2.5 mg/mL and the hanging drop vapour diffusion method was used. Each droplet was composed of 1 µL protein solution and 1 µL precipitant solution. The precipitant solution for all droplets pictured was composed of 0.1 M phosphate-citrate at varying pH's, 0.2 M NaCl, 10 % MPD, and PEG8000 at varying concentrations. **A.** The droplet condition used pH 5.0 and 24 % PEG8000. **B.** The droplet condition used pH 4.6 and 22 % PEG8000. **C.** The droplet condition used pH 5.0 and 20 % PEG8000. **D.** The droplet condition used pH 5.4 and 18 % PEG8000.

The co-crystallization of the *D. melanogaster* Tip60 chromodomain with a peptide derived from the N-terminal tail sequence of the histone H3 was also attempted. Previously published research stated that the Tip60 chromodomain was able to bind to the N-terminal tail sequence of histone H3 at a mono-methylated lysine 4 (H3K4Me$_1$) (Jeong *et al.*, 2011). Based on this study, the co-crystallization was attempted to elucidate the structural basis of binding. As outlined in section 3.5.1, several sparse matrix screening kits were used to probe for the ideal co-crystallization conditions. However, the criteria for the selection of crystals for optimization required the crystals have either a different morphology than those found to produce crystals of the apo-form of the protein or conditions which produced co-crystals but not apo-protein crystals. The rationale for this was that, due to the way the apo-protein packs within the crystal (described in detail in section 4.5), it was likely that if the H3K4Me$_1$ peptide bound the protein it would alter the crystal packing.

The sparse matrix screens revealed that the majority of conditions produced crystals similar in morphology to the apo-protein. This indicated there was a high likelihood that those crystals would not contain the bound H3K4Me$_1$. However, there was one condition which produced crystals with a different morphology when compared to the apo-protein crystals. This condition was composed of 4.0 M sodium formate and was selected for optimization. To probe the crystallization space around this condition, the concentration of sodium formate was varied. Additionally, other formates such as magnesium formate and ammonium formate, were also screened. What was found was that the initial sodium formate condition could not reproduce crystals. However, it was found that the ammonium formate conditions were able to form very long and thin needle-like crystals (Figure 4.11). The production of large singular co-crystals has proven to be very difficult to date, therefore further optimization would be required. Furthermore, based on the results of the binding studies discussed in section 4.6, co-crystallization trials with H4 peptides will have be done as well.

Figure 4.11 Potential co-crystal of the *D. melanogaster* Tip60 chromodomain with the H3K4Me$_1$ peptide. The image was captured with 4X magnification. The chromodomain was concentrated to 2.5 mg/mL and the peptide was incubated with the protein in a 2:1 peptide:protein ratio. Hanging drop vapour diffusion was used where each droplet was composed of 1 µL protein-peptide solution and 1 µL precipitant solution. This condition was composed of 4 M NH_4 formate.

4.5 Crystal structure of the *Drosophila melanogaster* Tip60 chromodomain

The crystallization condition composed of 0.1 M phosphate-citrate (pH 4.6), 0.2 M NaCl, 10 % MPD, and 22 % PEG8000 was found to give the best diffracting crystal. The crystal was diffracted at the Canadian Light Source (CLS) with the bending magnet beamline and data collected to 1.59 Å resolution. The crystal was found to have a space group of $P2_12_12$ and unit cell parameters a = 64.812 Å, b = 68.207 Å, c = 38.313 Å, $\alpha = \beta = \gamma = 90°$. The data collection statistics are reported in Table 4.2.

Table 4.2 Data collection statistics.
Values in parentheses are for highest resolution shell.
*R_{merge}= $\Sigma_{hkl}\Sigma_i|I_i(hkl) - \langle I(hkl)\rangle|/ \Sigma_{hkl}\Sigma_i I_i(hkl)$

X-ray source	Bending Magnet
Wavelength (Å)	1.03320
Space group	$P2_12_12$
Unit-cell parameters (Å, °)	*a*=64.812, *b*=68.207, *c*=38.313, α=β=γ=90
Resolution range (Å)	46.98 – 1.596 (1.653 – 1.596)
Observed reflections	141096
Unique reflections	23143 (2176)
Multiplicity	7.1 (5.6)
Completeness (%)	99.59 (96.20)
R_{merge}^* (%)	4.7 (23.2)
R_{work}	0.1835 (0.2292)
R_{free}	0.2084 (0.3040)
$\langle I/\sigma(I)\rangle$	41.78 (6.72)
Number of non-hydrogen atoms	1309
macromolecules	1182
water	127
Protein residues	141
RMS (bonds)	0.013
RMS (angles)	1.32
Ramachandran favored (%)	100
Ramachandran outliers (%)	0
Clashscore	1.99
Average B-factor	17.70
macromolecules	17.10
solvent	23.70

The crystal structure was solved by molecular replacement as described in section 3.6.2. The solution found by molecular replacement revealed that the asymmetric unit contained two copies of the chromodomain. Previously, it was determined by size-exclusion chromatography that the chromodomain was a monomer, which indicated that the two copies found in the asymmetric unit associate with one another at least partly due to crystal packing. Therefore, it was possible to examine each copy of the chromodomain in the asymmetric unit individually to determine its features and potential binding site, because it was expected that the protein existed in solution as a monomer.

The final model was determined through iterative refinements using the PHENIX refinement program (Adams *et al.*, 2010). The final model of the chromodomain crystal structure is shown in Figure 4.12. Within the final model of the crystal structure, two distinct copies of

the chromodomain are found. When the two copies of the chromodomain were overlaid upon each other, it was found the two copies had a root-mean-square deviation (r.m.s.d.) of 0.49, with the only significant differences between the two being the conformations of some of the side chains. It was revealed that the chromodomain was composed of four β-strands, two truncated α-helices, and loops connecting the β1-β2 strands and β3-β4 strands (Figure 4.12). The four curved β-strands were found to form a β-barrel, the truncated α1 helix was found to be located within the β3-β4 loop, and the C-terminal α2 helix was found to form after the β4 strand. Further examination of the final model also revealed that the α2 helix of one of the chromodomain copies packs against the curved surface formed by β-strands 2 – 4.

Within the final model, it was found that the electron density was missing for several side chains and terminal backbone residues. This resulted in those amino acids missing side chain density to be left as alanine residues, while residues missing backbone density were omitted from

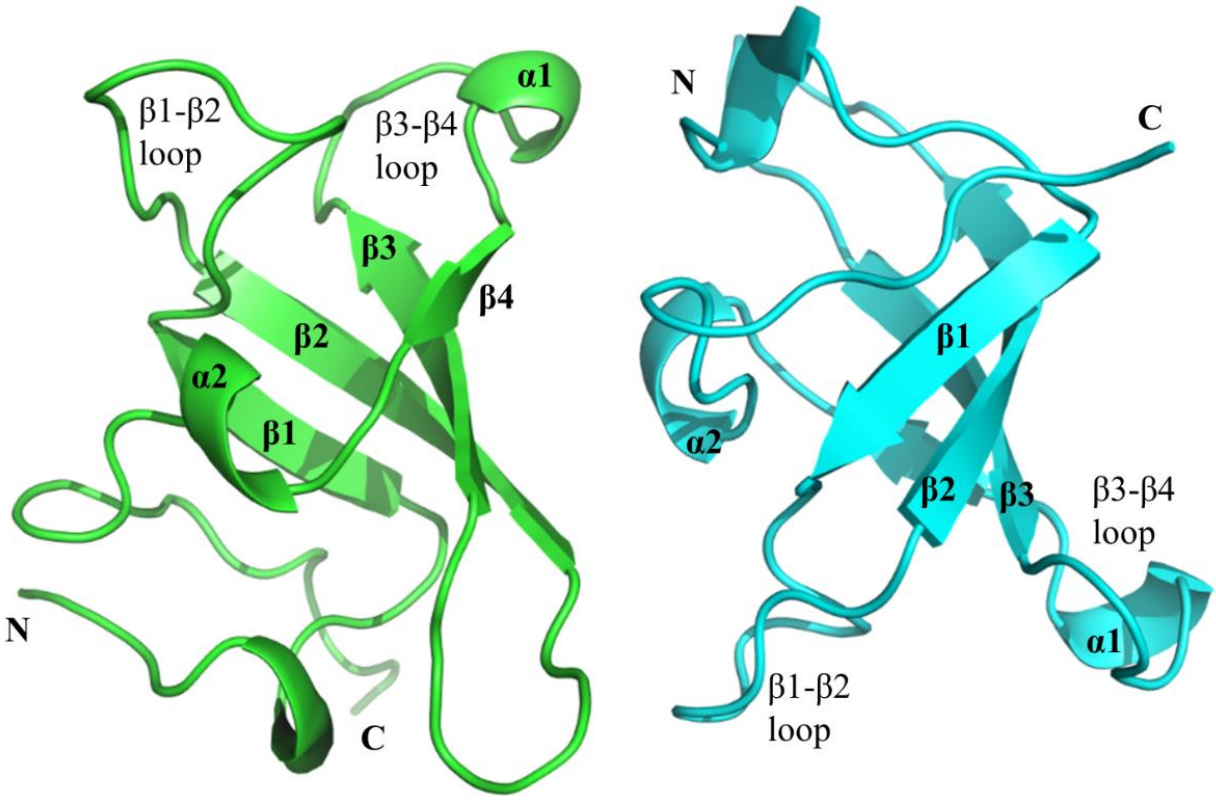

Figure 4.12 Cartoon representation of the *D. melanogaster* Tip60 chromodomain crystal structure. The structure was solved to a resolution of 1.59 Å. Two copies of the chromodomain were found in the asymmetric unit and are shown here. Both Chain A (green) spans residues 19-88 and Chain B (cyan) spans residues 19-85. The secondary structure elements and the N and C terminus for both copies of the chromodomain are labeled.

the final model. The rationale behind these omissions was that without electron density it was impossible to indicate with confidence the orientation of side chains or terminal residues. Additionally, the reason that electron density for these side chain and terminal backbone residues were missing was likely due to a high level of mobility within the crystal. The differences between the known *D. melanogaster* Tip60 chromodomain sequence and the model are summarized in Table 4.3.

As previously mentioned, the Tip60 chromodomain shares 45% sequence identity with the MOF chromodomain. This high level of sequence conservation results in the chromodomains adopting a similar structure, as it was found that the MOF chromodomain was composed of five β-strands and one C-terminal α-helix (Figure 4.13A). The structural similarities are easily visualized when the two structures are superimposed upon one another (Figure 4.13B). Furthermore, based on unpublished data from the Moore lab of the MOF chromodomain, it was

Table 4.3 Differences between dTip60 chromodomain sequence and final model

Chain ID	**Residue in dTip60 sequence**	**Residue in model**	**Comments**
A	N/A	Leu3	Part of fusion protein linker
A	N/A	Gly4	Part of fusion protein linker
A	N/A	Ser5	Part of fusion protein linker
A	Glu49	Glu49	Alternate conformation
A	Lys80	Ala80	No side chain density
A	Arg86	Ala86	No side chain density
A	Asp87	Ala87	No side chain density
A	Gly88	N/A	No backbone density
B	N/A	Ser5	Part of fusion protein linker
B	Glu49	Glu49	Alternate conformation
B	Lys80	Ala80	No side chain density
B	Arg85	Ala85	No side chain density
B	Arg86	N/A	No backbone density
B	Asp87	N/A	No backbone density
B	Gly88	N/A	No backbone density

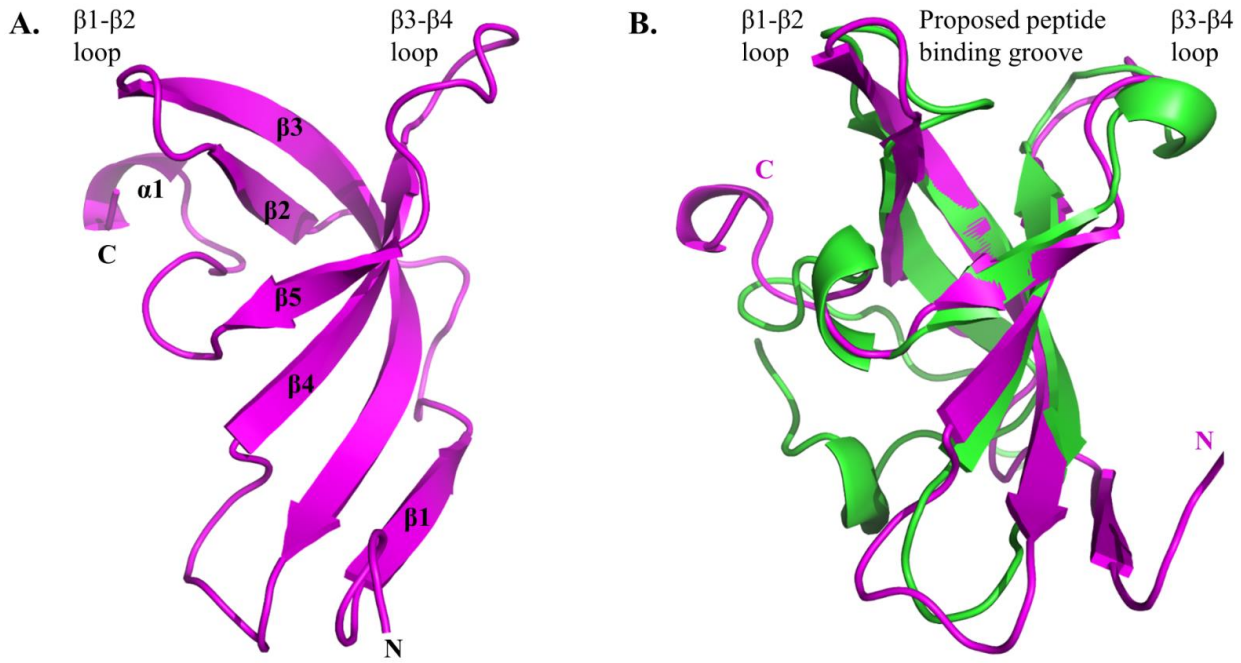

Figure 4.13 Structure of MOF chromodomain. **A.** The unpublished crystal structure of the *D. melanogaster* MOF chromodomain. The chromodomain was composed of five β-strands and one C-terminal α-helix. **B.** An overlay of the MOF chromodomain (magenta) with the *D. melanogaster* Tip60 chromodomain (green). The proposed peptide binding groove in MOF was thought to be located between the β1-β2 loop and the β3-β4 loop.

proposed that the peptide binding groove was located between the β1-β2 loop and the β3-β4 loop (Figure 4.13B). Due to the structural similarities between Tip60 and MOF, it is likely that this location in Tip60 could bind substrates as well.

Previously published data suggested that the Tip60 chromodomain residues Trp39, Phe56, and Tyr60 (equivalent to residues Trp26, Phe43, Tyr47 in the *H. sapiens* Tip60 chromodomain) were involved in the formation of an aromatic cage proposed to be a methyllysine binding site (Jeong *et al.*, 2011; Sun *et al.*, 2009). However, the crystal structure revealed that the formation of this specific aromatic cage was not possible (Figure 4.14). The reason for this was that it was found that Phe56 was actually buried and not solvent accessible, rendering it unable to contribute to binding. Furthermore, Trp39 and Tyr60 were found to be positioned away from each other and unlikely to contribute together towards the formation of a binding site. As previously mentioned, the peptide binding groove was thought to be located between the β1-β2 loop and the β3-β4 loop. The crystal structure revealed that, of the three residues originally proposed to form an aromatic binding cage (Trp39, Phe56, Tyr60), only Tyr60 was located within the proposed peptide binding groove (Figure 4.14). Additionally, it was found

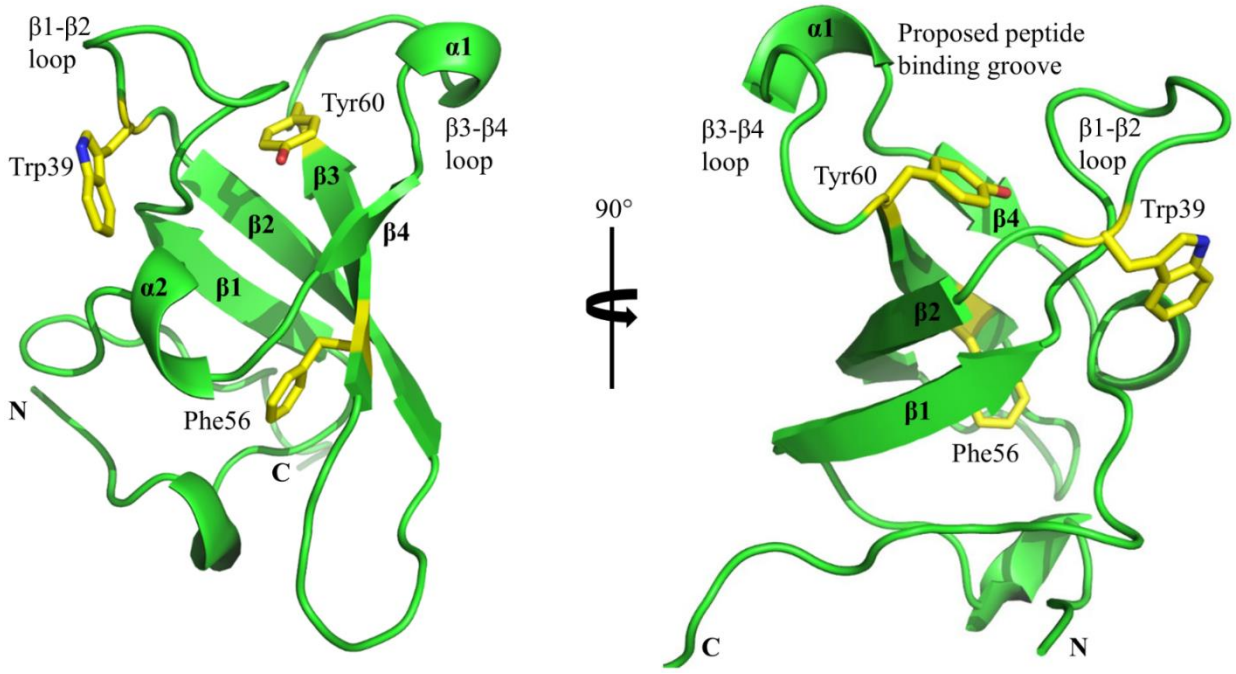

Figure 4.14 Previously proposed residues for methyllysine binding. The three aromatic residues, Trp39, Phe56, and Tyr60, previously proposed to form the methyllysine binding pocket are shown in yellow. The chromodomain was rotated 90° counterclockwise to give a better view of how the three residues were oriented. The three residues do not form the proposed methyllysine binding pocket. It was found that Trp39 was located on the β1-β2 loop and its side chain was positioned on the outer surface of the chromodomain. The Tyr60 residue was found to be located at the base of the proposed peptide binding groove. Finally, Phe56 was found to be buried within the core of the protein and unlikely to be involved with binding.

that Trp39 was located on the β1-β2 loop and that its side chain was positioned on the outer surface of the chromodomain. Although the residues Trp39, Phe56, and Tyr60 do not bind methyllysine through the aforementioned mechanism, it was found that Trp39 and Tyr60 may be located at separate potential binding sites (Figure 4.15).

The first potential binding site found in the chromodomain crystal structure was composed of residues Trp39 and Tyr77. The crystal structure revealed that Trp39 was located in the β1-β2 loop and Tyr77 was located following the α2-helix. Additionally, it was found that these two residues were located on one of the crystal contact surfaces, which indicated that these residues were buried in the crystal structure. It was found that due to crystal packing the Lys48 residue of one chromodomain molecule would pack between the Trp39 and Tyr77 of another chromodomain molecule (Figure 4.15A). The presence of a Lys residue between these two aromatic residues suggested this site could serve as a potential location for the binding of methyllysine. Although this surface was buried in the crystal structure, the data obtained from

the size-exclusion chromatography experiments indicated that the chromodomain exists as a monomer in solution and, therefore, indicated that this surface was potentially available for peptide binding.

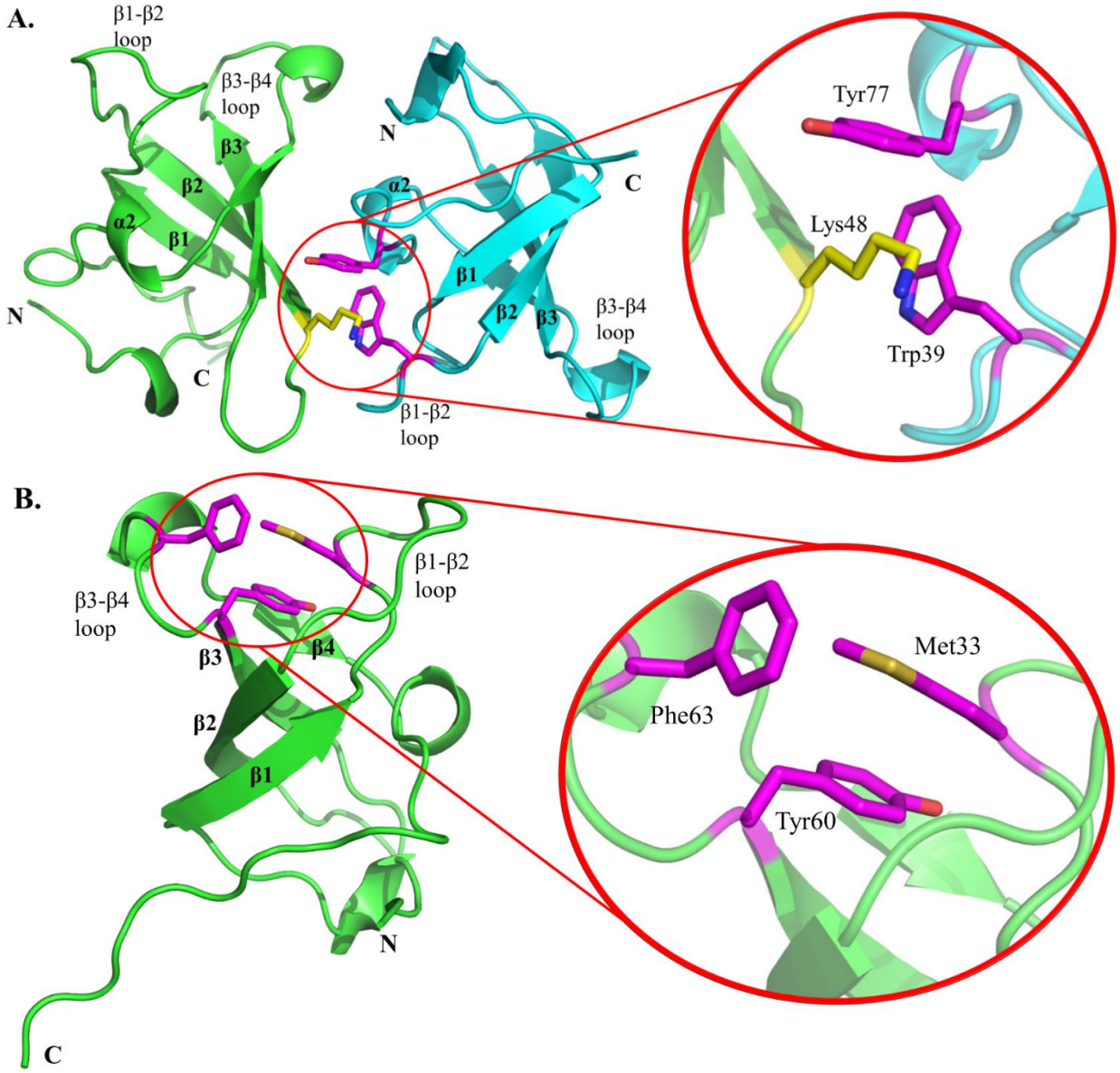

Figure 4.15 Potential methyllysine binding sites. **A.** Close up view of a potential binding pocket formed by Trp39 and Tyr77. Circled here are a ball and stick representation of Lys48 (yellow) located on Chain A (green), Trp39 and Tyr77 (magenta) located on Chain B (cyan). Due to crystal packing Lys 48 of Chain A sits between Trp39 and Tyr77 of Chain B, which indicates a potential methyllysine binding site. **B.** Close up view of the potential binding pocket formed by Tyr60 and Phe63. The ball and stick representation of Tyr60, Phe63, and Met33 (magenta) are circled. The two aromatic residues were found to be located within the proposed peptide binding groove. However, the position of Met33 blocks access to the substrate binding groove.

However, when examining a sequence alignment between the *D. melanogaster* Tip60 chromodomain with other MYST chromodomain homologs it was found while Trp39 was highly conserved, the tyrosine residue was only found in the *D. melanogaster* Tip60 chromodomain (Figure 4.16). The lack of sequence conservation at the tyrosine residue could be seen as an indicator that this site does not bind methyllysine. It should be noted, however, that, although this site was unlikely to bind methyllysine in the *D. melanogaster* Tip60 homologs, it was possible that this location could serve as a unique binding site specific only to *D. melanogaster*.

The second potential binding site revealed by the crystal structure contained residues Tyr60, Phe63, and Met33 (Figure 4.15B). As previously discussed, it was proposed that the peptide binding groove was positioned between the β1-β2 loop and β3-β4 loop. The crystal structure showed that the Met33 residue was located on the β1-β2 loop, the Phe63 residue was located on the β3-β4 loop, and the Tyr60 residue was located at the floor of the proposed peptide binding groove. The presence of these two aromatic residues in the proposed binding groove indicated this location was a possible methyllysine binding site. However, the position of Met33 in the *D. melanogaster* Tip60 chromodomain would block substrate binding unless it was flexible enough to undergo a conformational change. An alignment of the Tip60 chromodomain with other chromodomain containing MYST proteins revealed that the Tyr60 residue was strictly conserved across the MYST family (Figure 4.16). Furthermore, the sequence alignment revealed that other members of the MYST family contained an aromatic residue at the equivalent position as Phe63 and that the Met33 residue was poorly conserved (Figure 4.16).

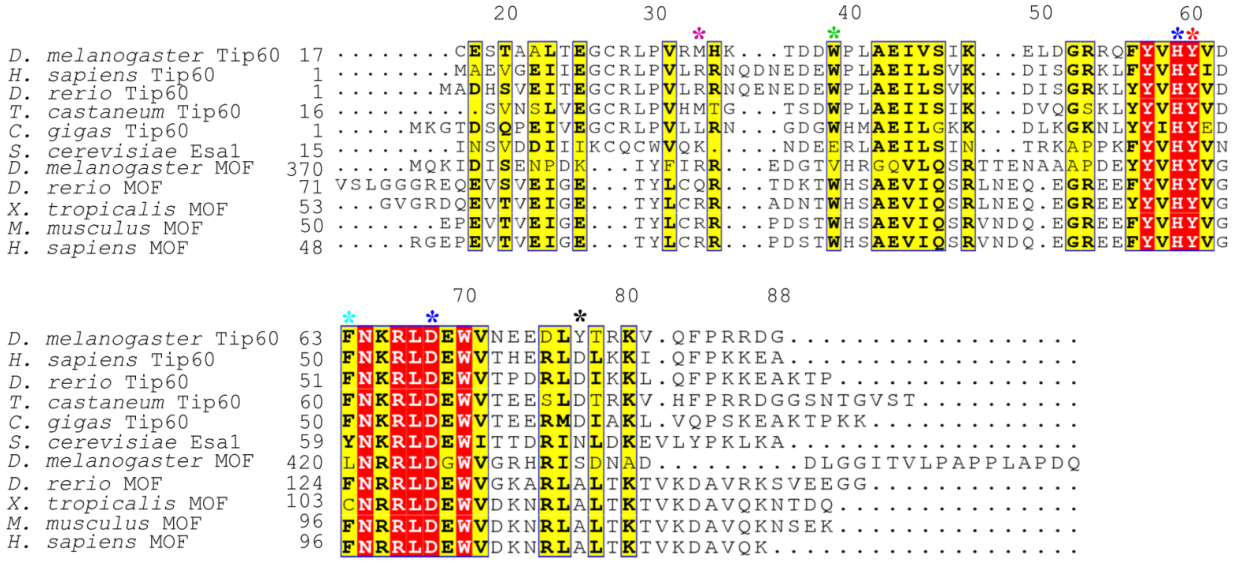

Figure 4.16 Sequence alignment of the chromodomain containing MYST proteins. The residues highlighted in red show strict sequence identity whereas those residues highlighted in yellow and bolded show sequence similarity. Sequence alignment was generated using clustal omega (Goujon *et al.*, 2010; Sievers *et al.*, 2011). Asterisks correspond to residues found in *D. melanogaster* Tip60 chromodomain discussed in this section. * - Met33, * - Trp39, * - His59 and Asp68, * - Tyr60, * - Phe63, * - Tyr77.

As previously stated, the position of Met33 would block substrate binding, unless it was able to undergo a conformational change. To asses the flexibility of Met33 within the crystal structure, it was necessary to examine the B-factors (also known as the temperature factors), which described the displacement from the average atomic position in square angstroms (Å^2). This value was assigned to each atom in the crystal structure. Therefore, if the Met33 side chain was flexible the B-factor values of the side chain atoms were expected to be high. The B-factors for the Met33 side chain ranged from 12-17 Å^2. This indicated that the side chain atoms were inflexible within the crystal structure. This apparent rigidity in Met33 indicated two possibilities: in solution, the chromodomain exhibited this rigidity and, therefore, substrate binding at this site could only be achieved if a major conformational change occurred, or the second possibility was that this rigidity, seen in the crystal structure, was simply an artefact of crystal packing. If the second possibility were found to be true and Met33 and its surrounding residues were flexible, it would be likely that in the presence of substrate Met33 could move away from the binding pocket.

The final site of interest in the crystal structure of the chromodomain contained residues His59 and Asp68. The crystal structure revealed that the His59 residue was located on the C-terminal end of the β3 strand and the Asp68 residue was located on the β3-β4 loop (Figure 4.17).

Examining the crystal structure revealed that His59 and Asp68 formed a hydrogen bond with a distance of 2.7 Å. Based on this distance, it was likely that the two residues were able to form a salt bridge as well. The salt bridge interaction is defined as the combination of two non-covalent interactions: the hydrogen bond and an electrostatic interaction (Bosshard *et al.*, 2004). Therefore, at physiological pH, it would be possible for His59 and Asp68 to form a salt bridge, due to the presence of a partial positive charge on His59 and the negative charge on Asp68. Due to the position of both residues, it could be possible that the salt bridge formed aids in the stabilization of the β3-β4 loop structure. Additionally, a sequence alignment amongst the chromodomain containing MYST family members revealed that both the histidine and aspartate residues were strictly conserved among all MYST chromodomains (Figure 4.16).

The conservation of this His-Asp pair were found to extend into other chromodomains, such as those found in MSL3 (Figure 4.18A). Previously, it had been reported that chromodomains belonging to MSL3 (Kim *et al.*, 2010) and members of the MYST family such as Esa1 (Shimojo *et al.*, 2008) and MOF (Conrad *et al.*, 2012) were able to contact nucleic acids. From the crystal structure of MSL3, it could be seen that the His-Asp pair formed a salt bridge with the phosphate backbone of duplex DNA (Figure 4.18B). An overlay of MSL3 with the Tip60 chromodomain revealed structural similarity between the two chromodomains and highlighted conservation and positions of the His-Asp pair relative to the DNA backbone (Figure 4.18B). Due to the strict conservation of this His-Asp pair across chromodomains and its ability to contact the phosphodiester backbone of duplex DNA in MSL3, it is possible that these amino acids play a role in the interaction and/or recognition of nucleosomal DNA.

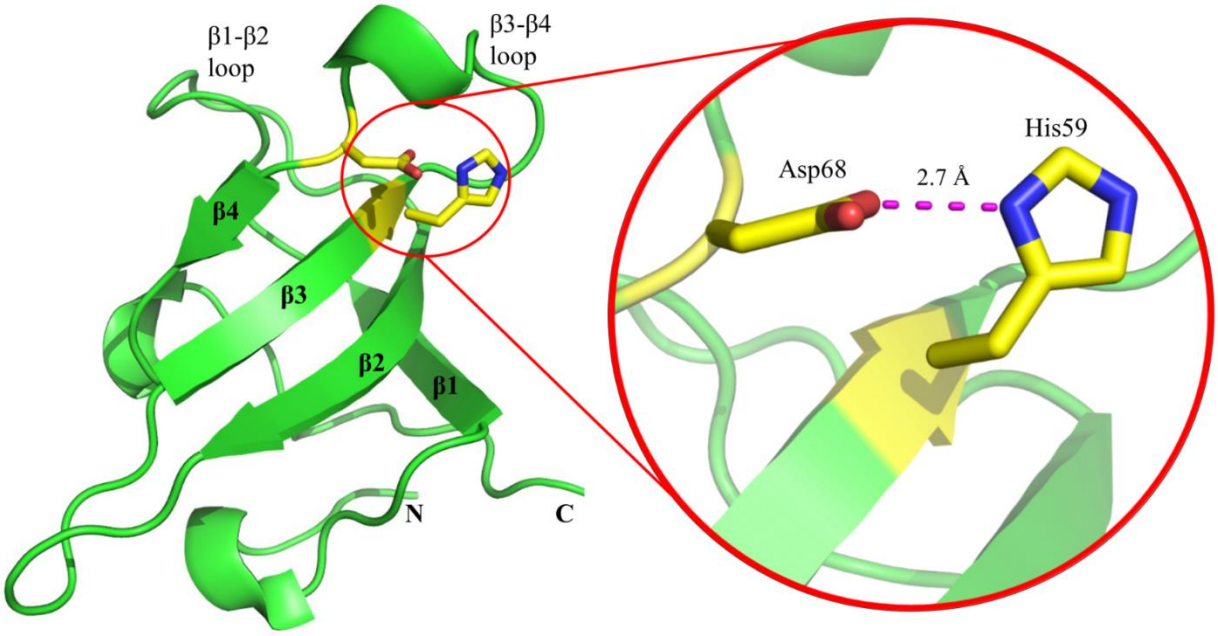

Figure 4.17 Structural view of the strictly conserved His-Asp pair. Circled here are a ball and stick representation of Asp68 and His59 (yellow) in relation to the entire chromodomain (shown as a cartoon coloured green). An enlarged view of the two residues showed that a hydrogen bond of 2.7 Å is formed between histidine and aspartate (magenta dashes). At physiological pH, it would be possible for these residues to form a salt bridge, as in addition to the hydrogen bond, the partial positive charge found on the His residue would interact with the negative charge of the Asp residue as well. The His residue could potentially be involved in contacting the phosphate backbone of duplex DNA through a salt bridge.

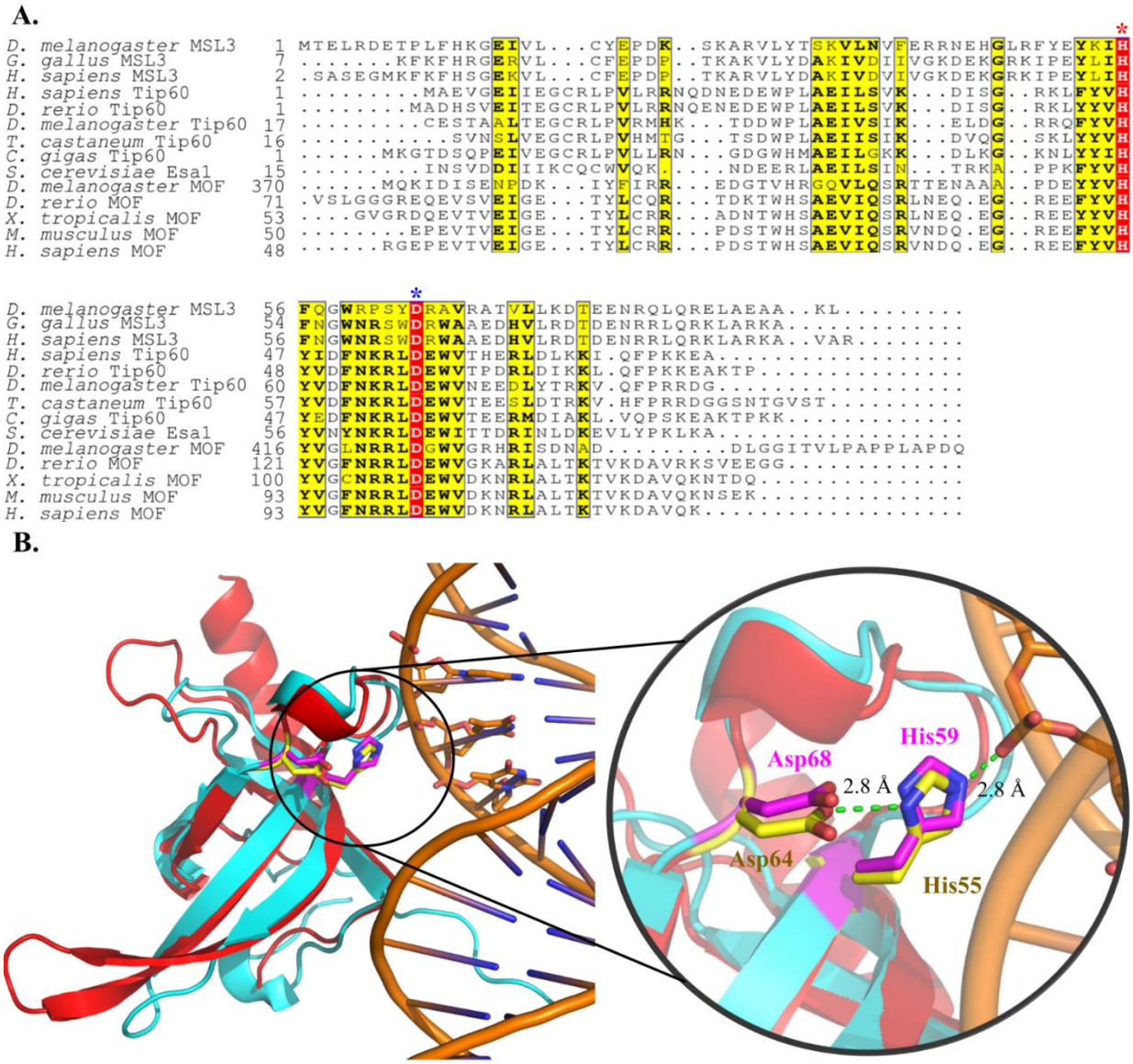

Figure 4.18 Comparison of MYST chromodomains with the MSL3 chromodomains. **A.** Sequence alignment of the MYST chromodomains with MSL3 chromodomains. The alignment found that the *D. melanogaster* Tip60 residues His59 and Asp68 were strictly conserved amongst all MYST and MSL3 chromodomains (both residues highlighted in red). **B.** An overlay of the *D. melanogaster* Tip60 chromodomain (cyan) with *H. sapiens* MSL3 (red) (PDB ID: 3OA6). The duplex DNA (orange) is also shown interacting with MSL3. An enlarged view of the conserved His-Asp pair showed that MSL3's Asp64 (yellow) was within 2.8 Å of His55 (yellow) to form a hydrogen bond or salt bridge. It was also found that MSL3's His55 residue was positioned 2.8 Å from the phosphodiester backbone of DNA, which allowed it to form a hydrogen bond or salt bridge with DNA. The overlay also showed that Tip60's His 59 (magenta) would be in a position to contact duplex DNA as well. Note that in the MSL3 crystal structure, it contacts a second DNA molecule, only one is shown here.

The high resolution structure of the *D. melanogaster* Tip60 chromodomain allowed for the identification of two potential methyllysine binding sites. The first potential binding site involved residues Trp39 and Tyr77, which were found to be located on the outer surface of the chromodomain. However, due to poor sequence conservation of the Tyr77 across MYST chromodomains, it was unlikely that this was the primary location for methyllysine binding. It could be possible that the Trp39 and Tyr77 site is unique to the *D. melanogaster* Tip60 chromodomain. The second potential methyllysine binding site was composed of residues Tyr60 and Phe63, which were found to be located in the proposed peptide binding groove located between the β1-β2 loop and the β3-β4 loop. Furthermore, a sequence alignment of MYST chromodomains showed the strict conservation of residue Tyr60 and that an aromatic residue was present at the equivalent position of Phe63. Finally, it was found that His59 and Asp68 were located on the β3 strand and β3-β4 loop respectively. These two residues were found to interact through a hydrogen bond and were likely to form a salt bridge at physiological pH. Additionally, a sequence alignment revealed that this His-Asp pair was strictly conserved across all MYST and MSL3 chromodomains. Furthermore, an overlay with an MSL3 crystal structure complexed with duplex DNA indicated that this His-Asp pair was likely involved in contacting the phosphodiester backbone of DNA. These insights formed the foundation for identifying the structural basis of binding of the Tip60 chromodomain to its partners. Before attempting to determine the structural basis of binding, several experiments were done to elucidate and confirm the binding partners of the Tip60 chromodomain.

4.6 Binding Studies between *Drosophila melanogaster* Tip60 chromodomain and histone peptides

4.6.1 Aggregation of *Drosophila melanogaster* Tip60 chromodomain during recording of ^{1}H–^{15}N HSQC spectra

The NMR study of the *D. melanogaster* Tip60 chromodomain was unable to produce results as the protein aggregated during data collection (Figure 4.19). This was likely due to the buffer the protein was stored in during data collection. Previously, it was found that the *D. melanogaster* Tip60 chromodomain was most soluble at pH 9.5, several pH units away from its pI of 6.06, and that the protein was insoluble in most buffers with a pH range of 6.0 – 8.5. To collect good HSQC spectra, the protein was required to be solubilized in a buffer composed of 40 mM phosphate buffer pH 7.5, 50 mM NaCl, and 5 % MPD. Due to the pH of the phosphate

buffer, it was likely that the long term storage of the protein was not possible, which resulted in the precipitation and aggregation of the chromodomain.

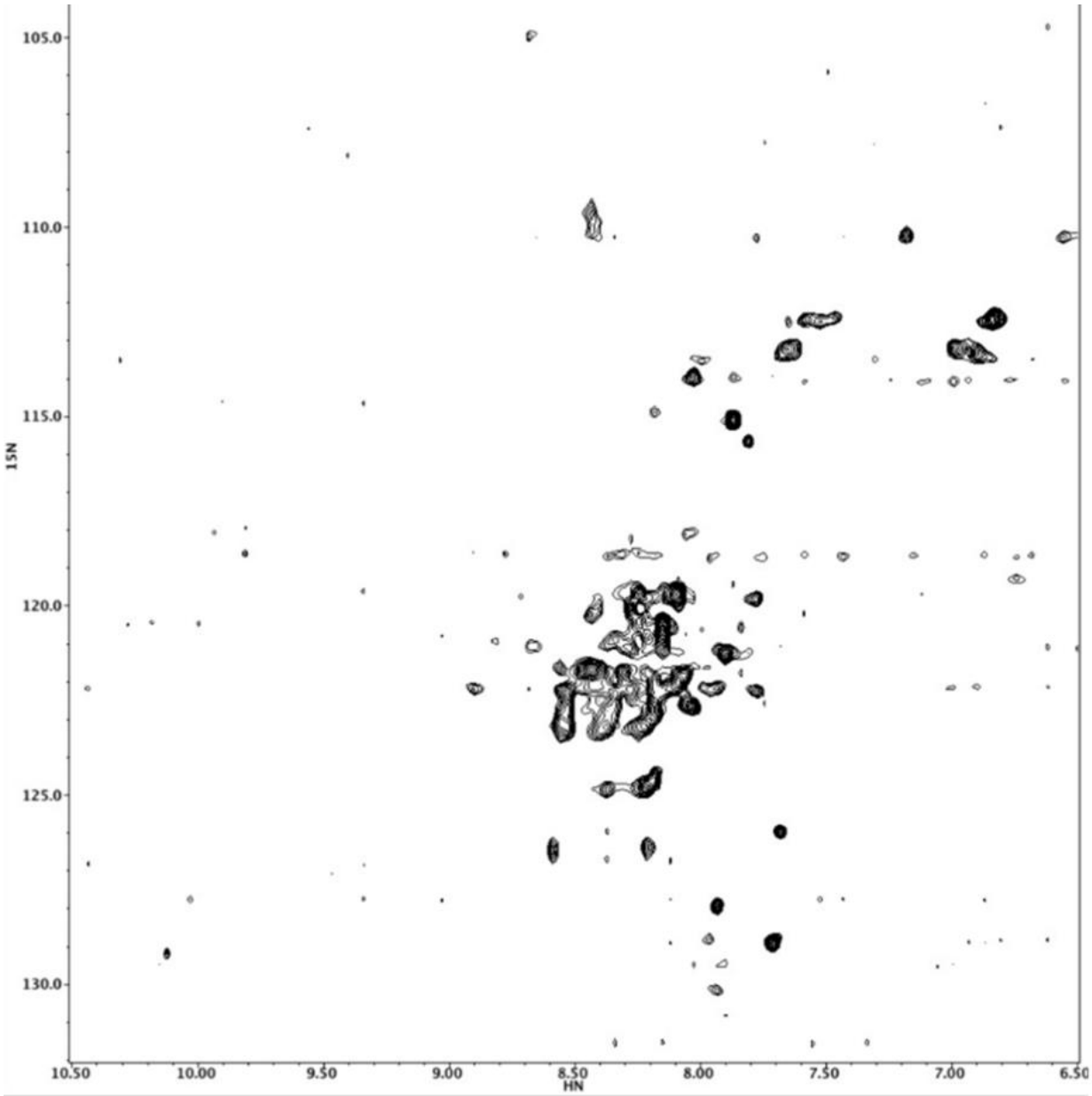

Figure 4.19 HSQC spectra for *D. melanogaster* Tip60 chromodomain. The HSQC spectra for the Tip60 chromodomain showed that many of the peaks had clustered. This indicated that the protein aggregated during the measurement of the HSQC spectra. The y-axis shows ^{15}N measurement and the x-axis shows the proton chemical shifts in parts per million (ppm).

4.6.2 Surface plasmon resonance revealed interaction between histone H4 peptides and *Drosophila melanogaster* Tip60 chromodomain

There were several candidates which were identified as potential substrates for the *D. melanogaster* Tip60 chromodomain. These candidates were peptide sequences derived from either the N-terminal tail sequence of histone H3 or H4 and typically contained a distinctly modified lysine residue, which corresponded to a post-translational modification involved with gene regulation. To rapidly screen this peptide library, surface plasmon resonance (SPR) was utilized. The SPR experiments were conducted at the Saskatchewan Structural Sciences Centre and utilized the ProteOn XPR36 protein interaction array system from BioRad.

As described in section 3.7.2, the chromodomain was covalently linked to the carboxylated alginate polymer surface of a sensor chip (GLM, Bio-Rad) through amine coupling. Based on this immobilization procedure, 25 µg/mL of the *D. melanogaster* Tip60 chromodomain resulted in approximately 2200 RU of the chromodomain immobilized onto the sensor chip. Additionally, for the initial screening of the peptide library, the *D.* melanogaster MOF chromodomain was used as a control due to it sharing 45% sequence identity with Tip60 and because the Moore lab had previously elucidated its binding partners. For the immobilization of the MOF chromodomain, 20 µg/mL was incubated with an immobilization buffer composed of 10 mM sodium acetate at pH 4.5. This resulted in 3000 RU of the MOF chromodomain to be immobilized onto the sensor chip. This level of protein density on the surface of the sensor chip was found to give the optimal response for the detection of protein-peptide interactions for this specific system.

The 6 x 6 interaction array system allowed for the Tip60 chromodomain and MOF chromodomain to be immobilized in parallel and the peptide library to be screened against both chromodomains simultaneously. As described in section 3.7.2, the initial screening procedure used a fixed concentration of 100 µM for each peptide to quickly assess which of the peptides interacted with the chromodomains. This initial group of peptides screened reflected those which corresponded to the sequences found on the N-terminal tails of either histone H3 or H4 and are listed in Table 4.4.

Table 4.4 Peptides used in SPR experiments

Histone Tail Peptide	N/C Terminus	Sequence	Molecular Weight (Da)
*^$H4_{6-24}$ Unmodified	H/NH_2	GGKGLGKGGAKRHRKVLRDY	2153
*^H4K12Ac	NH_2/Amide	GGKGLG(**KAc**)GGAKRHRKVLRDY	2195
*^H4K16Ac	NH_2/OH	GGKGLGKGGA(**KAc**)RHRKVLRDY	2195
*^H4K20Me_1	NH_2/Amide	GGKGLGKGGAKRHR(**KMe**)VLRDY	2166
*^H4K20Me_3	Ac/Amide	GAKRHR**(KMe$_3$)**VLRDNY	1698
^H3K4Me_1	Ac/Amide	ART**(KMe)**QTARKSTGGY	1579
^H3K9Me_3	Ac/Amide	KQTAR**(KMe3)**STGGY	1281
H3K27Me_3	Ac/Amide	SKAAR(**KMe3**)SAPAY	1234
H3K36Me_1	Ac/Amide	STGGV**(KMe)**KPHRY	1284
^H3K36Me_3	Ac/Amide	STGGV**(KMe3)**KPHRY	1314
$H4_{9-15}$	NH_2/Amide	GLGKGGAY	721
*^$H4_{16-20}$	NH_2/Amide	KRHRKY	886
$H4_{20-24}$	NH_2/Amide	KVLRDNY	793
*H4D24K	NH_2/Amide	HRKVLR**K**NY	1212
#H4K20 short	Ac/Amide	YGAKRHRKV	1155
#H4R17A	Ac/Amide	YGAK**A**HRKV	1070
#H4R17E	Ac/Amide	YGAK**E**HRKV	1128
#H4H18A	Ac/Amide	YGAKR**A**RKV	1089
#H4H18E	Ac/Amide	YGAKR**E**RKV	1147
#H4R19A	Ac/Amide	YGAKRH**A**KV	1070
#H4R19E	Ac/Amide	YGAKRH**E**KV	1128

* - denotes positive hits from initial SPR screening
^ - denotes peptides used for steady state affinity experiments
# - denotes $H4_{14-21}$ tail peptides used for identifying key residues for binding

The screening process found several peptide sequences derived from the histone H4 N-terminal tail that interacted with the Tip60 chromodomain (Table 4.4). It was also found that those peptides which the Moore lab had previously determined to interact with the MOF

chromodomain also exhibited an interaction during the screening process. This was important for two reasons: the first reason was that it confirmed that the ProteOn XPR36 protein interaction array system worked, the second reason was that, due to sharing 45% sequence identity with the MOF chromodomain, it was expected the Tip60 chromodomain would have similar binding partners. The screening process also revealed that many of the peptides derived from the N-terminal tail sequence of histone H3 did not show any interaction or showed a relatively weak interaction. This data allowed for the selection of peptides, which would be used in titration experiments to determine the binding affinity. Additionally, the binding affinity of the Tip60 chromodomain for the H3 peptides was also determined as these peptides had previously been reported (Jeong *et al.*, 2011; Sun *et al.*, 2009) to bind the Tip60 chromodomain. The titration of these peptides was outlined in section 3.7.2.

The steady state affinity experiments utilized ten different peptide concentrations, which allowed for the dissociation constant (K_D) to be determined. These concentrations were dependent upon the results of the initial screening process. For the H4 peptides, which appeared to show an interaction with the Tip60 chromodomain at 100 µM, the concentrations used were as follows: 300 µM, 150 µM, 100 µM, 50 µM, 33.3 µM, 16.7 µM, 11.1 µM, 5.55 µM, 3.7 µM, and 1.85 µM. The H3 peptides, which showed minimal to no interaction at 100 µM, required that the steady state affinity experiment be performed with a higher concentration range (1 mM, 500 µM, 333 µM, 167 µM, 111 µM, 55.6 µM, 37 µM, 18.5 µM, 12.3 µM, and 6.17 µM). The peptides used for the steady state experiment are listed in Table 4.4.

The RUs for each interaction were recorded and the sensorgrams for each interaction with the H4 peptides are shown in Figure 4.20; the interactions with H3 peptides are shown in Figure 4.21. However, close examination of the raw sensorgrams revealed a high background signal for all peptides and a signal-to-noise of 1.3 (Figure 4.22). It was found that the reference channel, which does not have protein bound but underwent the same activation/deactivation process, showed a high signal in response to the peptide flowing over the channel. This indicated the presence of non-specific binding of the peptides to the sensor chip surface. The reason for this non-specific binding could be attributed to the positive charge the peptides carried and the negative charge found on the sensor chip surface. However, even with the presence of non-specific binding, several of the peptides (specifically $H4_{6\text{-}24}$ unmodified, H4K12Ac, H4K16Ac,

and H4K20Me$_1$) possessed the same sequence and charge, yet exhibited different binding profiles. This indicated that the interaction observed between the peptides and the Tip60 chromodomain were likely to be real, but further binding studies would need to be conducted to confirm and quantify the interaction.

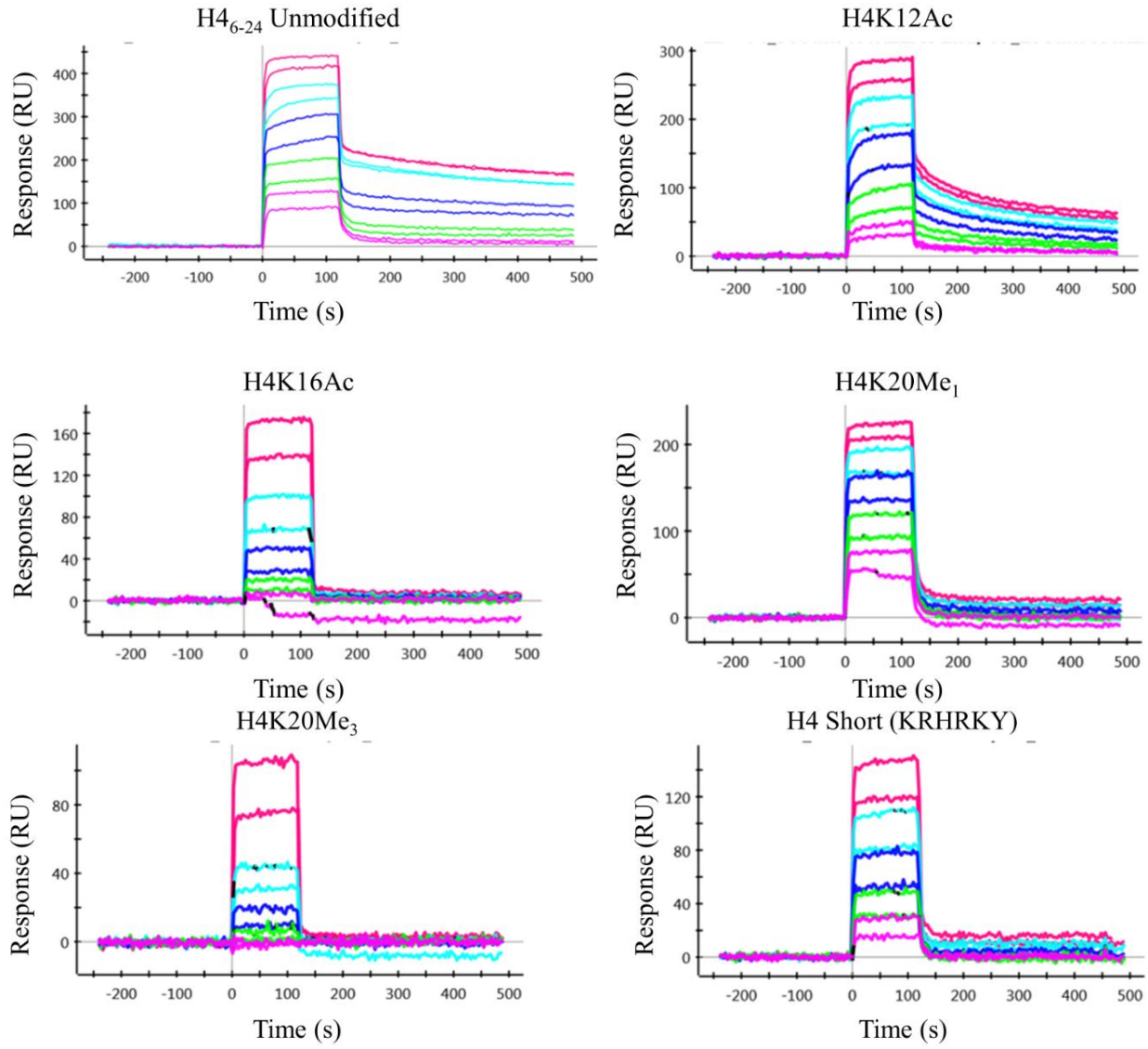

Figure 4.20 SPR sensorgrams for each H4 peptide interaction. The sensorgrams measured the response detected over time for a given concentration of peptide. The signal detected for each concentration is shown in the graph as coloured peaks. The peaks in each graph, in descending order, correspond to the following concentration: 300 µM, 150 µM, 100 µM, 50 µM, 33.3 µM, 16.7 µM, 11.1 µM, 5.55 µM, 3.7 µM, and 1.85 µM. Additionally, the reference channel signals have been subtracted from the sensorgrams shown here. Note that the y-axis scale for each graph is different.

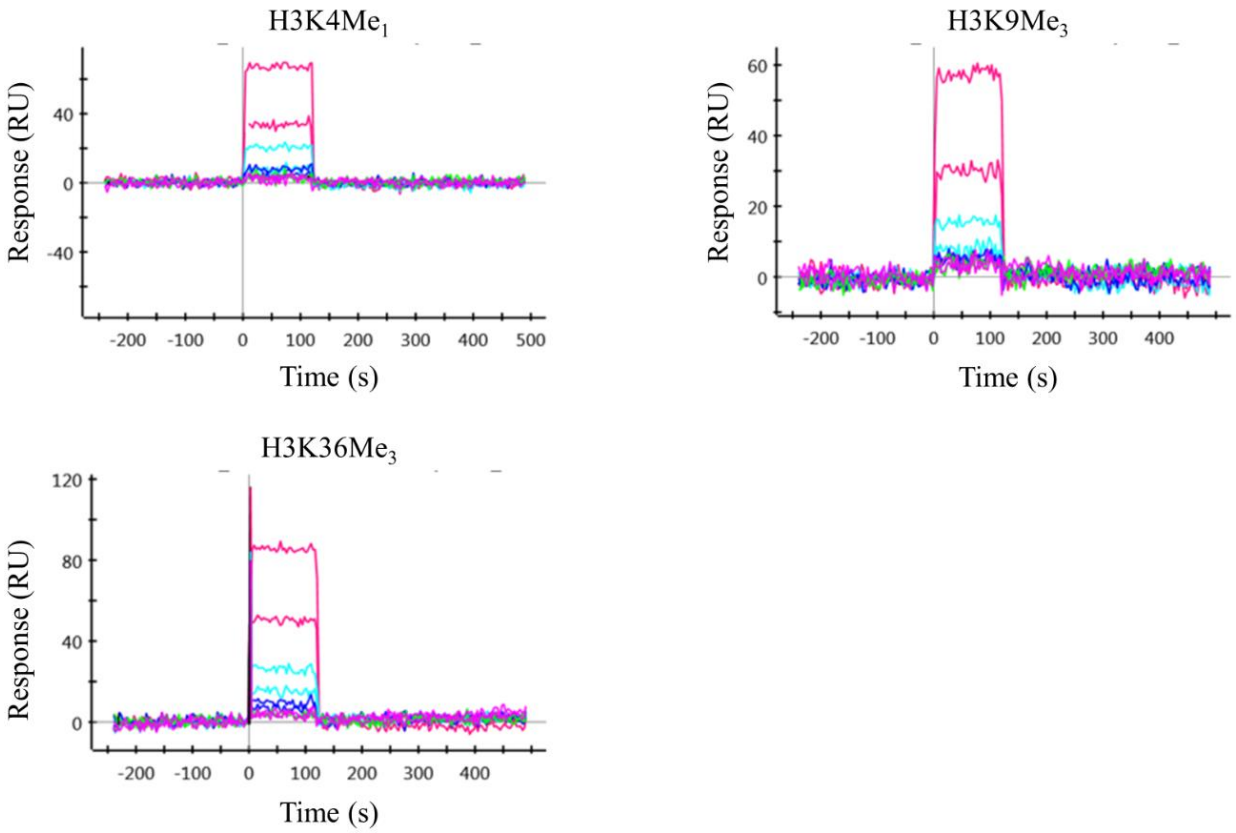

Figure 4.21 SPR sensorgrams for each H3 peptide interaction. The sensorgrams measured the response detected over time for a given concentration of peptide. The signal detected for each concentration is shown in the graph as coloured peaks. The peaks in each graph, in descending order, correspond to the following concentration: 1 mM, 500 µM, 333 µM, 167 µM, 111 µM, 55.6 µM, 37 µM, 18.5 µM, 12.3 µM, and 6.17 µM. Additionally, the reference channel signals have been subtracted from the sensorgrams shown here. Note that the y-axis scale for each graph is different.

From the steady state affinity experiments, it was possible to determine an approximate K_D value for each peptide. First, the values were normalized according to Equation 3.2, which was previously described in section 3.7.2 and these normalized values were plotted against the concentration of the peptide. The purpose of the normalization procedure was to allow for the accurate comparison of the different peptides (Davis and Wilson, 2000). The normalized data for the H4 and H3 peptides was fitted independently in Origin (from OriginLab) to produce the R_{max} and K_D values. It should be noted that for the H3 peptides, a reasonable R_{max} or K_D value could not be calculated, as the binding site could not be saturated. The reason that the H3 peptides were not able to saturate the binding site was likely due to the very weak nature of the binding, therefore the peptide concentration required to saturate the binding site would have to be much higher. This revealed that the Tip60 chromodomain showed a minimal interaction with the H3

peptides and instead, preferentially bound to the H4 peptides (Figure 4.23, Table 4.5). The difference in R_{max} between the H4 peptides could be attributed to several possibilities, such as the difference in molecular weight between the peptides or the degradation and/or unfolding of the immobilized chromodomain. The Tip60 chromodomain appeared to prefer binding the H4K20Me$_1$ (K_D of 8.8 µM), H4$_{6-24}$ Unmodified (K_D of 12 µM), and H4K12Ac (K_D of 21 µM) peptides. Additionally, a weaker interaction was observed with H4K16Ac (K_D of 140 µM). The sequence length and charge of these four peptides were found to be very similar, the major difference between each was the type and location of the modified lysine residue, with each exhibiting a different level of interaction with the chromodomain. Furthermore, the H4-short sequence (KRHRKY) which corresponded to the residues between H4K16 and H4K20, exhibited an interaction with the Tip60 chromodomain. This indicated that it was possible that at least one of the three residues between H4K16 and H4K20 was integral for binding to the Tip60 chromodomain.

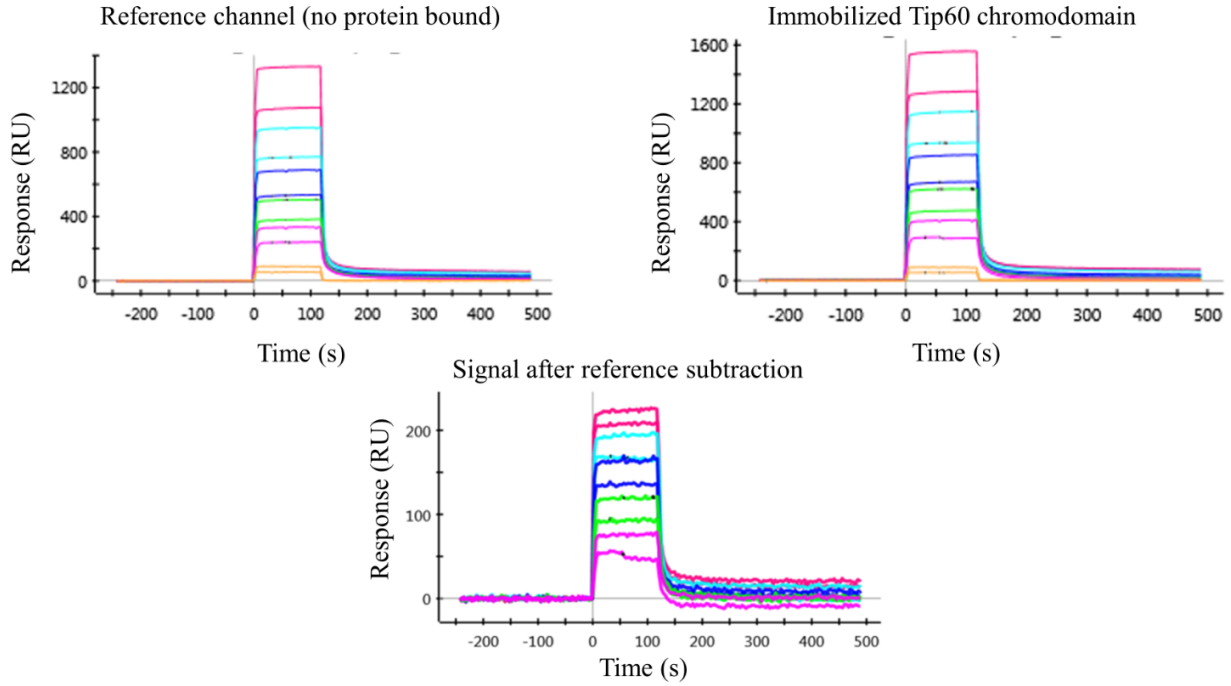

Figure 4.22 Raw SPR sensorgrams for the histone peptides. The sensorgrams corresponding to the reference channel (top left) and the channel with the immobilized Tip60 chromodomain (top right) are shown. The signal recorded by the reference channel indicated that background noise of SPR experiments were very high. The bottom panel shows the signal of the interaction between the protein and peptide after the background signal was subtracted. Note that the y-axis scale for each graph is different.

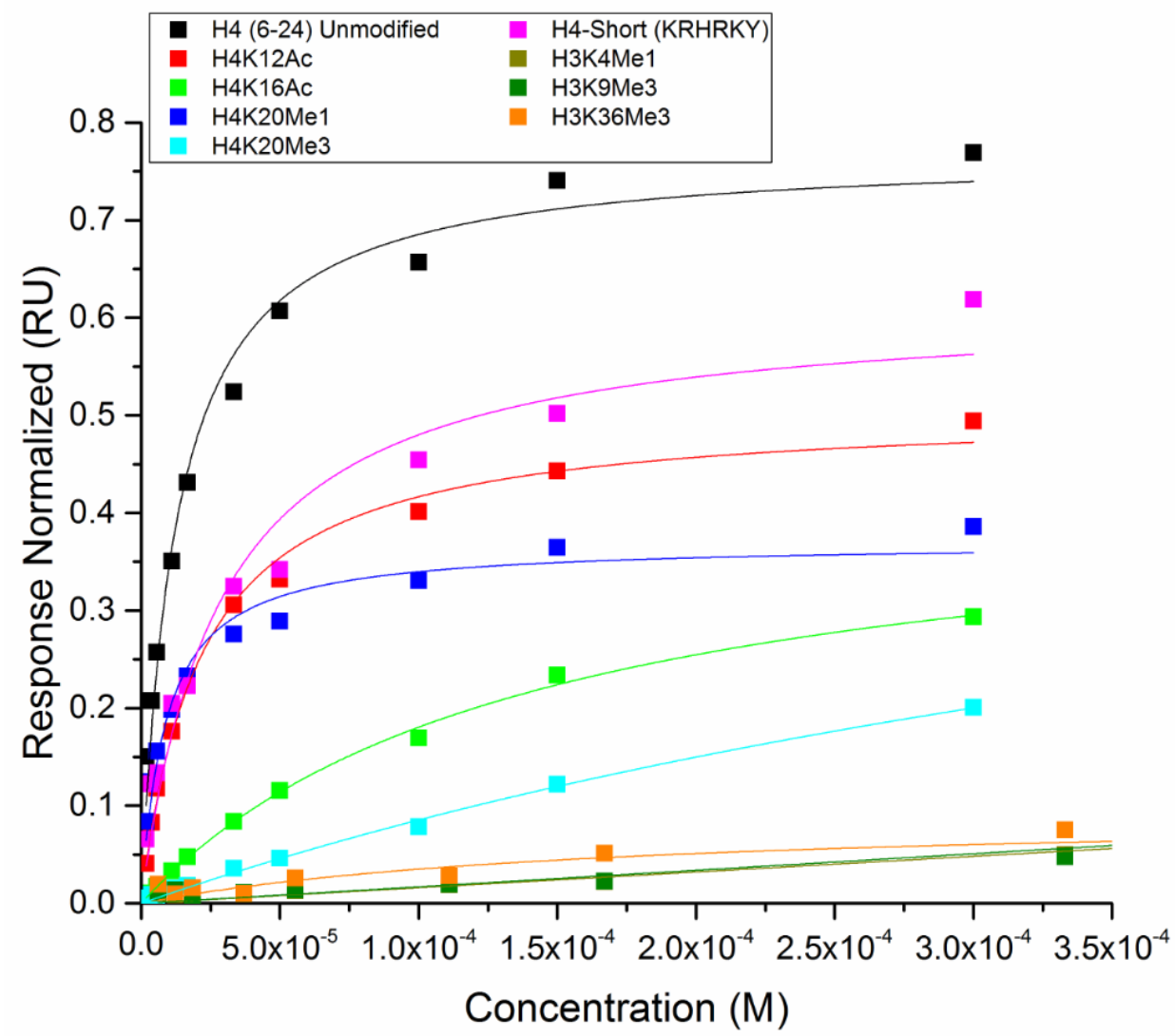

Figure 4.23 SPR steady state titrations of histone tail peptides with *D. melanogaster* Tip60 chromodomain. The H4 and H3 peptides were fitted independently to determine the R_{max} and K_D for each peptide. As the binding of the H3 peptides does not saturate the binding site, it was not possible to calculate the R_{max} or K_D for these peptides. The inability of the H3 peptides to saturate the binding site could be attributed to the very weak nature of the binding interaction.

Table 4.5 Calculated R_{max} and K_D for H4 and H3 peptides

Peptide	**R_{max}**	**Standard Error for R_{max}**	**K_D (M)**	**Standard Error for K_D (M)**
$H4_{6-24}$ Unmodified	0.76	0.02	1.2×10^{-5}	1.4×10^{-6}
H4K12Ac	0.51	0.01	2.1×10^{-5}	1.8×10^{-6}
H4K16Ac	0.43	0.02	1.4×10^{-4}	1.1×10^{-5}
$H4K20Me_1$	0.37	0.01	8.8×10^{-6}	1.2×10^{-6}
$H4K20Me_3$	0.61	0.09	6.1×10^{-4}	1.2×10^{-4}
H4-Short (KRHRKY)	0.62	0.04	2.8×10^{-5}	5.6×10^{-6}
$^{\#}H3K4Me_1$	N/A	N/A	N/A	N/A
$^{\#}H3K9Me_3$	N/A	N/A	N/A	N/A
$^{\#}H3K36Me_3$	N/A	N/A	N/A	N/A

* - denotes estimated value

- denotes peptides which did not saturate binding site therefore R_{max} and K_D could not be calculated

To determine which of the three residues between H4K16 and H4K20 was integral for binding, short H4 peptides with either an alanine or glutamate mutation to one of the three residues were examined (Table 4.4). The glutamate mutations were expected to cause a greater disruption to the binding than the alanine mutations, due to the reduction of the overall positive charge of the peptide caused by the replacement with the glutamate residue. It was expected that, if one the three residues between H4K16 and H4K20 was integral to binding, then a mutation to that residue would cause a decrease in the K_D as determined by a steady state affinity experiment.

The steady state affinity experiment confirmed that the glutamate mutations, H4R17E, H4H18E, and H4R19E caused a severe disruption to the binding, as the response observed was near to the baseline. The glutamate mutations were also found to reduce the overall positive charge of the peptide, which could disrupt any potential ionic interactions with the chromodomain or the sensor chip surface, resulting in a near baseline response. It was also found that the three alanine mutations, H4R17A, H4H18A, and H4R19A were able to disrupt the binding of the peptides to the Tip60 chromodomain (Figure 4.24). It was not possible to determine the R_{max} or K_D for any of the mutant peptides, as none of the peptides were able to saturate the binding site (Figure 4.24, Table 4.6). The inability to saturate the binding site could be attributed to the

concentrations utilized in the titrations being significantly lower than the expected K_D. To reasonably estimate the R_{max} and K_D of the mutant peptides, higher concentrations will have to be used in the titrations. When compared to the wild type sequence, the mutations to Arg17, His 18, and Arg19 clearly showed that the binding of the peptide to the chromodomain was disrupted, which indicated that at least one of these residues could be involved in interacting with the Tip60 chromodomain.

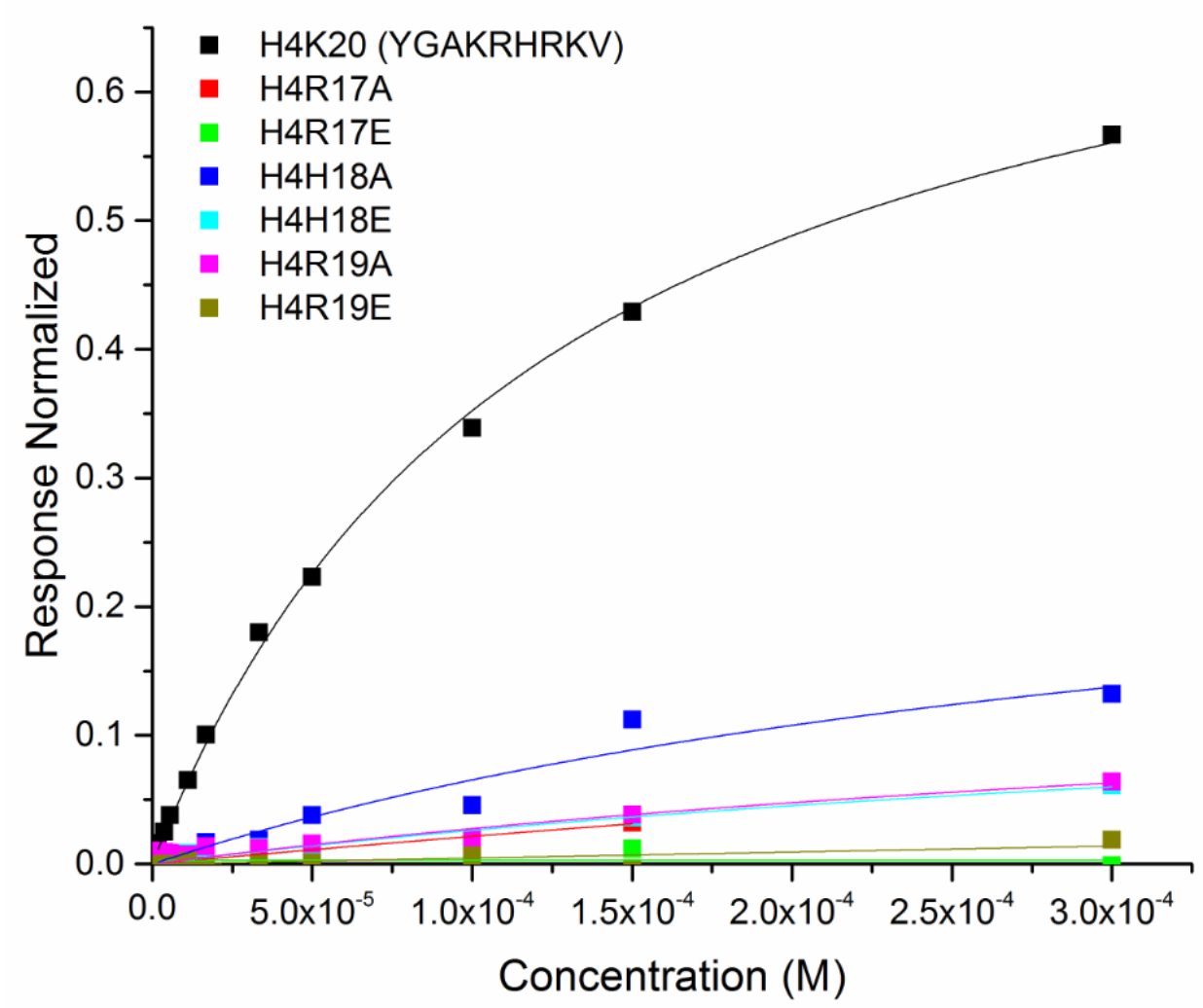

Figure 4.24 SPR steady state titrations of the mutant $H4_{14-21}$ tail peptides. Titrations were performed using peptide concentration ranges between 300 µM and 1.85 µM. The WT H4K20 (YGAKRHRKV) peptide was fitted independently to approximate the R_{max} and K_D values. It should be noted that it was not possible to reasonably estimate the R_{max} and K_D values for the mutant H4 peptides as the binding site could not be saturated.

Table 4.6 Calculated K_D for H4 mutant peptides

Peptide	**K_D (M)**	**Standard Error for K_D (M)**
H4K20 (YGAKRHRKV)	1.3×10^{-4}	7.5×10^{-6}
H4R17A	N/A	N/A
H4R17E	N/A	N/A
H4H18A	N/A	N/A
H4H18E	N/A	N/A
H4R19A	N/A	N/A
H4R19E	N/A	N/A

The SPR data revealed that the *D. melanogaster* Tip60 chromodomain preferentially interacted with peptides corresponding to residues 6 – 24 found in the N-terminal tail sequence of histone H4. Additionally, it appeared as though the chromodomain specifically interacted with a specific region within the H4 tail sequence. The SPR binding data suggested that the Tip60 chromodomain interacted with the H4 tail sequence between residues 16 – 20. Through the use of point mutations to residues Arg17, His18, and Arg19, it was determined that at least one of the three residues were integral for binding to the Tip60 chromodomain; as mutations to these residues showed a significant decrease in binding when compared to the wild type sequence. The interaction between the Tip60 chromodomain and the H4 peptides was then validated by isothermal titration calorimetry.

4.6.3 Interaction between *Drosophila melanogaster* Tip60 chromodomain and histone H4 peptides confirmed by isothermal titration calorimetry

To confirm the binding data results of the *D. melanogaster* Tip60 chromodomain and the H4 peptides obtained from SPR, isothermal titration calorimetry was used. These experiments were conducted at Protein Characterization and Crystallization Facility at the University of Saskatchewan with the Nano-ITC instrument from TA Instruments. The standard control experiment that utilized 1,4,7,10,13,16-hexaoxacyclooctadecane (18-crown-6) and KCl (Turnbull and Daranas, 2003) was performed, which ensured proper operation of the ITC instrument (Figure 4.25).

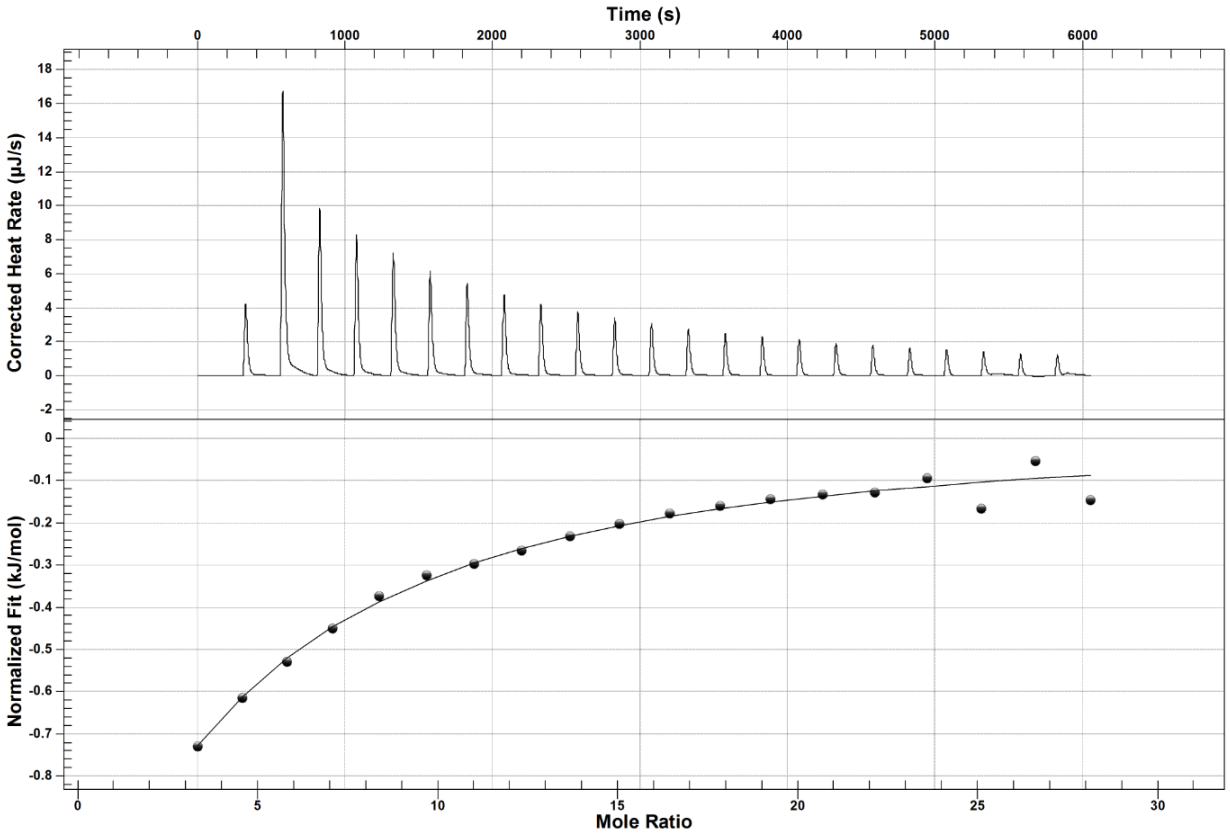

Figure 4.25 Isothermal titration calorimetry analysis of crown ether and KCl. The top panel shows the heat rates for the 21 total injections completed. The first two injections were not included in the fitting process (bottom panel) as the purpose of the first two injections was to minimize the equilibration artifacts typically seen at the start of every titration series. The bottom panel plots the integrated heat peaks against the molar ratio of KCl.

As described in section 3.7.3, the *D. melanogaster* Tip60 chromodomain was concentrated to 0.34 mM and placed into the reaction cell. The $H4_{13\text{-}20}$ unmodified peptide was prepared to a concentration of 6.5 mM and placed in the syringe. The ITC experiment was performed at 25° C, utilized a stir rate of 200 rpm, and 19 injections of 2.5 µL. The data was then analysed with the NanoAnalyze software from TA Instruments.

The ITC data confirmed the interaction between the Tip60 chromodomain and the $H4_{13\text{-}20}$ unmodified peptide (sequence shown in Table 3.5) (Figure 4.26). The data was fitted using the independent binding model, which found that the K_D was 7.3 x 10^{-5} M and that the stoichiometry of the reaction was n = 3.7; the other statistics determined by the fitting were shown in Table 4.7. It was expected that the binding of the H4 peptide to the chromodomain would exhibit 1:1 binding however, the estimated stoichiometry of the reaction was approximately 4:1 (peptide:chromodomain). The overestimation of the stoichiometry could be explained by the

weak nature of the interaction between the peptide and the Tip60 chromodomain. Other studies have found that, in order to reasonably estimate the stoichiometry of a reaction, the concentration of the protein must be at least 10 times greater than the K_D (Turnbull and Daranas, 2003). If the concentration of the chromodomain used in these experiments was too close to the actual K_D of the reaction, the stoichiometry could not be reasonably estimated. Furthermore, the data showed that the reaction was not saturated. The inability to saturate the binding site could be attributed to the dimerization or unfolding of the chromodomain throughout the reaction. To distinguish the difference between structural perturbations and those caused by binding, circular dichroism spectroscopy or NMR would have to be performed.

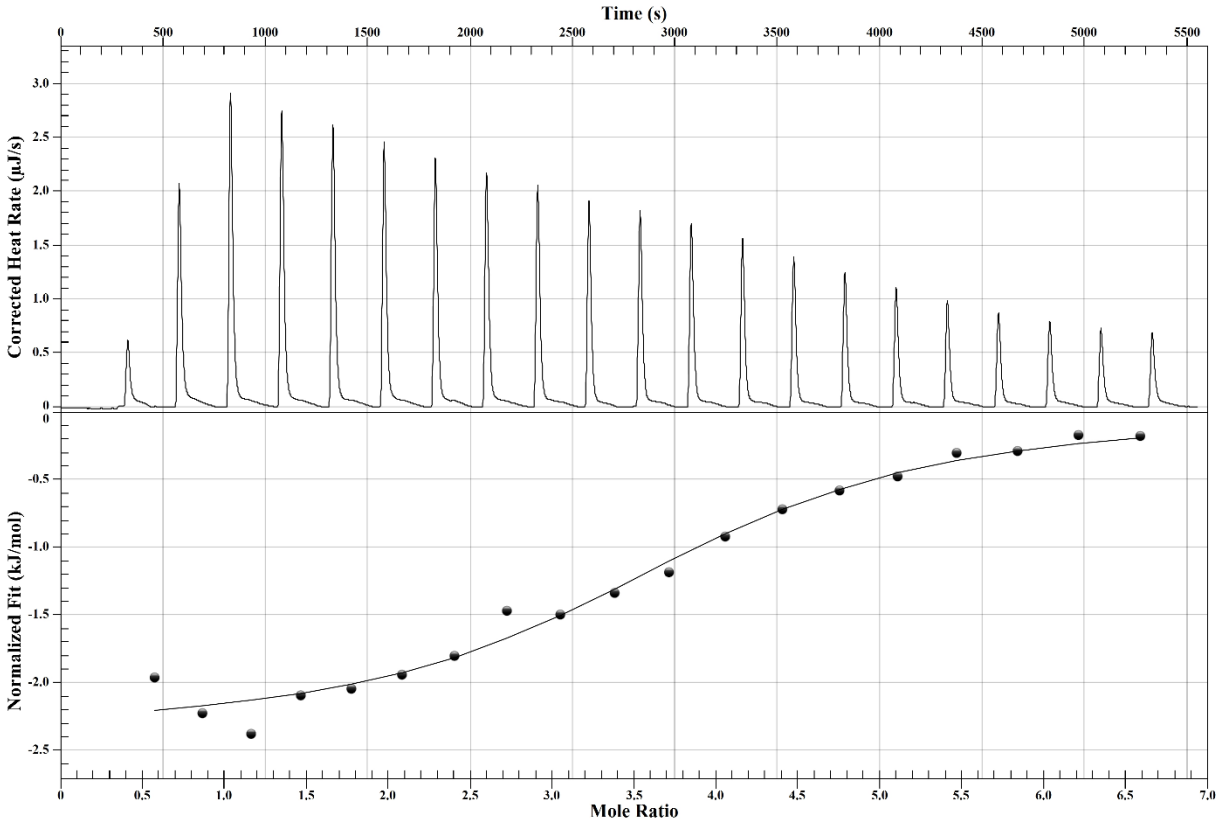

Figure 4.26 Isothermal titration calorimetry analysis of the $H4_{13-20}$ unmodified binding to *D. melanogaster* Tip60 chromodomain. The top panel shows the heat rates for the 21 total injections completed. The first two injections were not included in the fitting process (bottom panel) as the purpose of the first two injections was to minimize the equilibration artifacts typically seen at the start of every titration series. The bottom panel plots the integrated heat peaks against the molar ratio of the peptide.

Table 4.7 Statistics from ITC bindings experiment

K_D (M)	7.3×10^{-5}
n	3.7
ΔH (kJ/mol)	-2.4
ΔS (J/mol K)	71

It was assumed that if the reaction were saturated, it would be possible to elucidate both the stoichiometry and strength of the interaction. Therefore, the ITC experiment was repeated, with either more injections of peptides at the same peptide concentration or with the same number of injections but a higher concentration of peptide; in both cases, the molar ratio of peptide to protein was increased. However, it was found that these adjustments were not sufficient to result in saturation of the binding site.

The ITC data confirmed that the *D. melanogaster* Tip60 chromodomain does interact with the unmodified $H4_{13-20}$ peptide; however, it was unable to confirm the stoichiometry or K_D of the interaction due to the inability to saturate the binding site. This inability to saturate the binding site could be attributed to dimerization or unfolding of the chromodomain throughout the reaction. To elucidate these variables, further optimization of the ITC study will be required.

5 DISCUSSION

5.1 Structural comparison of *Drosophila melanogaster* Tip60 chromodomain to other related chromodomains

The canonical chromodomains, such as HP1, have been found to bind methylated lysine residues located on N-terminal histone tails through the presence of an aromatic cage (Blus *et al.*, 2011). Previous studies have suggested that the Tip60 chromodomain was able bind methylated lysine residues to allosterically regulate its HAT domain (Jeong *et al.*, 2011; Sun *et al.*, 2009). Furthermore, these studies attributed this binding to the presence of an aromatic cage and identified the residues Trp26, Phe43, and Tyr47 in *H. sapiens* Tip60 chromodomain to be responsible for the formation of this cage. Additionally, these studies linked the loss of any of these three aromatic residues to abrogate binding to methyllysine. However, studies on the Tip60 paralog MOF revealed that the chromodomain of MOF does not contain an aromatic cage for methyllysine binding (Nielsen *et al.*, 2005). This was found to be significant as Tip60 and MOF share 45% sequence identity, which indicated the two are likely to adopt similar structures.

This research was able to determine the crystal structure of the *D. melanogaster* Tip60 chromodomain, which provided insight as to whether the predicted aromatic cage for methyllysine binding was present or not. The crystal structure revealed that the aromatic residues Trp39, Phe56, and Tyr60 (equivalent residues to Trp26, Phe43, and Tyr47 in *H. sapiens* Tip60 chromodomain) do not form an aromatic cage as originally proposed by Sun *et al.* (2009) and Jeong *et al.* (2011). It was found that Trp39 was positioned on the outer surface of the chromodomain, Phe56 was buried within the core the protein, and only Tyr60 was located in the proposed peptide binding groove. Therefore, it was unlikely that the proposed aromatic cage could be formed by these three residues. Moreover, it was not possible to conclusively link the loss of binding observed in prior studies to the mutation of one of these three residues. The loss of function that was attributed to the mutation of Trp39, Phe56, and Tyr60, could be caused by the misfolding of the protein as opposed to the disruption of the aromatic binding cage.

The crystal structure of the *D. melanogaster* Tip60 chromodomain also revealed two potential methyllysine binding sites. One site was found to be formed by Trp39 and Tyr77, while the second site was formed by Tyr60 and Phe63. A sequence alignment of the *D. melanogaster* Tip60 chromodomain with other chromodomain containing MYST HATs found that, of the two

potential methyllysine binding sites, only the site composed of Tyr60 and Phe63 showed high level of conservation with other chromodomains. It was found that Tyr60 was strictly conserved amongst MYST HAT chromodomains and that an aromatic residue was present at the equivalent position of Phe63. The strict conservation of the Tyr60 was found to be of note, due to the suggestion that it was one of the residues responsible for the formation of an aromatic cage (Jeong *et al.*, 2011; Sun *et al.*, 2009). Furthermore, the sequence alignment revealed that a second aromatic residue positioned near Tyr60 was present in several MYST HAT chromodomains. An aromatic cage has been previously defined to be composed of two to four aromatic residues oriented in a cage-like structure (Musselman *et al.*, 2012). By this definition, the site composed of Tyr60 and Phe63 in the *D. melanogaster* Tip60 chromodomain was the ideal location for an aromatic cage to bind methyllysine. However, close inspection of this potential binding site revealed the Met33 residue blocks access to the aromatic residues. Therefore, for this region to bind substrates, the Met33 residue would have to move away from the potential binding site.

The sequence alignment of the Tip60 chromodomain with other MYST HAT chromodomains and the canonical chromodomain MSL3 revealed a strictly conserved His-Asp pair. Previous studies on MSL3 found that this histidine residue was involved in contacting the phosphodiester backbone of DNA (Kim *et al.*, 2010). Additionally, it has been reported that the MYST family members Esa1 and MOF were also able to interact with nucleic acids, although the mechanism by which this is achieved has not been confirmed (Conrad *et al.*, 2012; Shimojo *et al.*, 2008). It has also been reported that Tip60 plays a role in the acetylation of histones H2A and H4 (Sapountzi *et al.*, 2006) and is involved in the chromatin remodeling process (Kusch *et al.*, 2004). Due to the involvement with histones, it is likely that Tip60 is also able to interact with nucleosomal DNA. However, further experiments need to be performed to identify whether Tip60 can interact with nucleic acids.

5.2 *Drosophila melanogaster* Tip60 chromodomain and identified binding partners

It has been previously reported that the Tip60 chromodomain interacted with H3K4Me$_1$ (Jeong *et al.*, 2011) and H3K9Me$_3$ (Sun *et al.*, 2009). Both studies attributed the binding of the chromodomain to its substrate resulted in the activation of the HAT domain. It is still unknown how the binding of the chromodomain to its substrates results in acetyltransferase activity. It is

likely that a yet to be determined allosteric mechanism is responsible for HAT activity. The research performed here attempted to elucidate the binding partners for Tip60 chromodomain.

This study was able to determine that the *D. melanogaster* Tip60 chromodomain interacted predominantly with peptides corresponding to the N-terminal tail sequence of histone H4. The SPR data obtained provided an initial estimate of the strength of the interaction between the *D. melanogaster* Tip60 chromodomain and the H4 tail sequences. The ITC data obtained confirmed that an interaction between the *D. melanogaster* Tip60 chromodomain and H4 peptide occurred. However, to date, this study has been unable to ascertain, with a high level of confidence, specific aspects of the binding event, such as the strength and stoichiometry of the interaction due to complications with the chosen methods.

The SPR experiments had issues with the non-specific binding of the histone tail peptides to the surface of the sensor chip. This was likely due to the positive charge present on the histone tail peptides and the highly negative charge of the sensor chip surface, which resulted in an interaction between the two. However, even with non-specific binding issues, it was found that the Tip60 chromodomain interacted with the H4 tail peptides. The rationale behind this notion was that several of the peptides used for SPR were identical in sequence and charge, with the only difference occurring at the type and location of the modification, yet, each peptide exhibited a different signal. While this indicated the Tip60 chromodomain preferentially interacted with the H4 tail peptides, it was only able to give an approximation of the strength of this interaction. The reason K_D can only be approximated by the SPR experiment is due to the limitations of SPR itself. In an SPR experiment, it is assumed that, when the protein is immobilized to the sensor chip surface, all protein molecules are active and in the correct orientation for binding. Therefore, it was necessary to perform a secondary binding assay to confirm the interaction between the H4 tail peptides and the Tip60 chromodomain.

To verify the interaction between the Tip60 chromodomain and the H4 peptides, isothermal titration calorimetry was selected as a secondary *in vitro* binding assay. The benefits of using ITC to verify the interaction were that the reaction occurs in solution, therefore alleviating non-specific binding issues encountered in SPR or changes to the protein caused by immobilization to a sensor chip (Doyle, 1999). The ITC experiments utilized the titration of the H4 peptide into a reaction cell containing the Tip60 chromodomain in solution. Typically, the

gradual titration of the ligand into the reaction cell containing the protein is able to saturate the binding sites found on the protein. However, careful consideration must be made to the volume of ligand being injected throughout the experiment, as the saturation of the binding site is dependent on the molar ratio of the ligand to protein. Additionally, to control the molar ratio the concentrations of both the ligand and protein must be precisely known. Even with these considerations, it was found that the Tip60 chromodomain could not be saturated. Due to this inability to saturate the protein, an accurate measurement of the K_D and stoichiometry could not be taken. To solve these issues, the ITC experiment requires further optimization which will include varying the temperature of the reaction, the stir speed, and the volume and number of injections used. If these optimization steps fail to resolve the saturation issue, a homolog of the *D. melanogaster* Tip60 chromodomain may need to be utilized instead. To date the ITC study has only examined the interaction between the *D. melanogaster* Tip60 chromodomain and the unmodified $H4_{13\text{-}20}$ peptide. An interaction between the Tip60 chromodomain and the unmodified $H4_{13\text{-}20}$ peptide was confirmed to occur, however, the strength and stoichiometry of the reaction is yet to be determined.

It is known that the full length Tip60 protein acetylates histones H2AK5, H3K14, H4K5, H4K8, H4K12, and H4K16 (Sapountzi *et al.*, 2006). Based on the data from the binding experiments, it was found that the Tip60 chromodomain interacted with the H4 tail peptides. The binding data suggested that the Tip60 chromodomain preferentially bound $H4K20Me_1$ and exhibited weaker interactions with the unmodified H4 peptide, H4K12Ac, H4K16Ac, and $H4K20Me_3$. The $H4K20Me_1$ mark has been found to be associated with transcriptional activation (Wang *et al.*, 2008b). Furthermore, it has been found that specific histone modifications serve to recruit and activate the enzymatic activity of specific proteins (Kouzarides, 2007). The preferential binding of the Tip60 chromodomain to $H4K20Me_1$ would be consistent with this notion, as it is possible that the binding of the chromodomain could activate the HAT activity of the full length Tip60 protein. However, from the crystal structure, it is still unknown which potential binding site is responsible for the targeting and binding of substrates. One possibility that has yet to be explored, is the potential for the Tip60 chromodomain and HAT domain to cooperatively bind methyllysine.

5.3 Future Directions

The research to date has revealed a high resolution crystal structure of the *D. melanogaster* Tip60 chromodomain, which has been solved to 1.59 Å. The crystal structure revealed the mechanism previously thought to bind histone tails to be incorrect. The residues thought to be involved in the formation of an aromatic cage for methyllysine binding were found to not be close enough to one another to form this structure. Instead, it was found that the chromodomain possessed two distinct potential binding sites for methyllysine. The first potential binding site was composed of residues Trp39 and Tyr 77 and the second potential site was composed of residues Tyr60 and Phe63. The crystal structure also identified a highly conserved His-Asp pair which may play a role in contacting the phosphodiester backbone of duplex DNA. The research had also identified the binding partners of the *D. melanogaster* Tip60 chromodomain. Through SPR and ITC, it was found that the chromodomain preferentially bound the histone H4 tail sequence between residues 13 – 20. However, the binding data to date has not been able to ascertain the strength and stoichiometry of the interaction.

In the future, this research will need to conclusively determine the strength and stoichiometry of the interaction between the H4 peptide and the Tip60 chromodomain. This will be accomplished through further optimization of the SPR and ITC studies already conducted. The future of this research still requires that the structural basis of binding be identified. To do this, co-crystallization trials will need to be restarted using short peptide sequences derived from the histone H4 N-terminal tail sequences. Additionally, mutation studies on the residues suspected to be involved with binding will need to be done. Once a residue has been mutated, the SPR and ITC experiments need to be repeated with the mutant chromodomain to assess the effect of the mutation on binding. It would also be interesting to determine whether the Tip60 chromodomain has nucleotide binding capabilities. To accomplish this, SPR and ITC experiments could be performed using short double stranded DNA and the wild type chromodomain. Finally, the research aims to elucidate the mechanism for how the chromodomain binding to its partners is able to allosterically regulate the HAT activity of the full length Tip60.

6 REFERENCES

Adams, P.D., Afonine, P.V., Bunkoczi, G., Chen, V.B., Davis, I.W., Echols, N., Headd, J.J., Hung, L.W., Kapral, G.J., Grosse-Kunstleve, R.W., *et al.* (2010). PHENIX: a comprehensive Python-based system for macromolecular structure solution. Acta Crystallogr. D Biol. Crystallogr. *66*, 213-221.

Andrews, A.J., and Luger, K. (2011). Nucleosome structure(s) and stability: variations on a theme. Annu Rev Biophys *40*, 99-117.

Avvakumov, N., and Cote, J. (2007). The MYST family of histone acetyltransferases and their intimate links to cancer. Oncogene *26*, 5395-5407.

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

Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Schones, D.E., Wang, Z., Wei, G., Chepelev, I., and Zhao, K. (2007). High-resolution profiling of histone methylations in the human genome. Cell *129*, 823-837.

Berndsen, C.E., Albaugh, B.N., Tan, S., and Denu, J.M. (2007). Catalytic Mechanism of a MYST Family Histone Acetyltransferase. Biochemistry *46*, 623-629.

Berndsen, C.E., and Denu, J.M. (2008). Catalysis and substrate selection by histone/protein lysine acetyltransferases. Curr. Opin. Struct. Biol. *18*, 682-689.

Blus, B.J., Wiggins, K., and Khorasanizadeh, S. (2011). Epigenetic virtues of chromodomains. Crit. Rev. Biochem. Mol. Biol. *46*, 507-526.

Bosshard, H.R., Marti, D.N., and Jelesarov, I. (2004). Protein stabilization by salt bridges: concepts, experimental approaches and clarification of some misunderstandings. J. Mol. Recognit. *17*, 1-16.

Campos, E.I., and Reinberg, D. (2009). Histones: annotating chromatin. Annu. Rev. Genet. *43*, 559-599.

Carrozza, M.J., Utley, R.T., Workman, J.L., and Côté, J. (2003). The diverse functions of histone acetyltransferase complexes. Trends Genet. *19*, 321-329.

Chan, H.-M., and La Thangue, N.B. (2001). p300/CBP proteins: HATs for transcriptional bridges and scaffolds. J. Cell Sci. *114*, 2363-2373.

Conrad, T., Cavalli, F.M., Holz, H., Hallacli, E., Kind, J., Ilik, I., Vaquerizas, J.M., Luscombe, N.M., and Akhtar, A. (2012). The MOF chromobarrel domain controls genome-wide H4K16 acetylation and spreading of the MSL complex. Dev. Cell *22*, 610-624.

Cordingley, M.G., Callahan, P.L., Sardana, V.V., Garsky, V.M., and Colonno, R.J. (1990). Substrate Requirements of Human Rhinovirus 3C Protease for Peptide Cleavage in vitro. J. Biol. Chem. *265*, 9062-9065.

Cosgrove, M.S., Boeke, J.D., and Wolberger, C. (2004). Regulated nucleosome mobility and the histone code. Nat. Struct. Mol. Biol. *11*, 1037-1043.

Davey, C.A., Sargent, D.F., Luger, K., Maeder, A.W., and Richmond, T.J. (2002). Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9Å Resolution. J. Mol. Biol. *319*, 1097-1113.

Davis, T.M., and Wilson, W.D. (2000). Determination of the refractive index increments of small molecules for correction of surface plasmon resonance data. Anal. Biochem. *284*, 348-353.

Delvecchio, M., Gaucher, J., Aguilar-Gurrieri, C., Ortega, E., and Panne, D. (2013). Structure of the p300 catalytic core and implications for chromatin targeting and HAT regulation. Nat. Struct. Mol. Biol. *20*, 1040-1046.

Doyle, M.L. (1999). Titration Microcalorimetry. In Current Protocols in Protein Science (John Wiley & Sons, Inc.), pp. 20.24.21-20.24.24.

Doyon, Y., and Cote, J. (2004). The highly conserved and multifunctional NuA4 HAT complex. Curr. Opin. Genet. Dev. *14*, 147-154.

Doyon, Y., Selleck, W., Lane, W.S., Tan, S., and Cote, J. (2004). Structural and Functional Conservation of the NuA4 Histone Acetyltransferase Complex from Yeast to Humans. Mol. Cell. Biol. *24*, 1884-1896.

Emsley, P., Lohkamp, B., Scott, W.G., and Cowtan, K. (2010). Features and development of Coot. Acta Crystallogr. D Biol. Crystallogr. *66*, 486-501.

Fischle, W. (2009). Tip60-ing the balance in DSB repair. Nat. Cell Biol. *11*, 1279-1281.

Freitas, M.A., Sklenar, A.R., and Parthun, M.R. (2004). Application of mass spectrometry to the identification and quantification of histone post-translational modifications. J. Cell. Biochem. *92*, 691-700.

Gavaravarapu, S., and Kamine, J. (2000). Tip60 inhibits activation of CREB protein by protein kinase A. Biochem. Biophys. Res. Commun. *269*, 758-766.

Ginsburg, D.S., Anlembom, T.E., Wang, J., Patel, S.R., Li, B., and Hinnebusch, A.G. (2014). NuA4 links methylation of histone H3 lysines 4 and 36 to acetylation of histones H4 and H3. J. Biol. Chem. *289*, 32656-32670.

Ginsburg, D.S., Govind, C.K., and Hinnebusch, A.G. (2009). NuA4 lysine acetyltransferase Esa1 is targeted to coding regions and stimulates transcription elongation with Gcn5. Mol. Cell. Biol. *29*, 6473-6487.

Goujon, M., McWilliam, H., Li, W., Valentin, F., Squizzato, S., Paern, J., and Lopez, R. (2010). A new bioinformatics analysis tools framework at EMBL-EBI. Nucleic Acids Res. *38*, W695-699.

Hodawadekar, S.C., and Marmorstein, R. (2007). Chemistry of acetyl transfer by histone modifying enzymes: structure, mechanism and implications for effector design. Oncogene *26*, 5528-5540.

Hughes, R.M., Wiggins, K.R., Khorasanizadeh, S., and Waters, M.L. (2007). Recognition of trimethyllysine by a chromodomain is not driven by the hydrophobic effect. Proc. Natl. Acad. Sci. U. S. A. *104*, 11184-11188.

Ikura, T., Ogryzko, V.V., Grigoriev, M., Groisman, R., Wang, J., Horikoshi, M., Scully, R., Qin, J., and Nakatani, Y. (2000). Involvement of the TIP60 Histone Acetylase Complex in DNA Repair and Apoptosis. Cell *102*, 463-473.

Jeong, K.W., Kim, K., Situ, A.J., Ulmer, T.S., An, W., and Stallcup, M.R. (2011). Recognition of enhancer element-specific histone methylation by TIP60 in transcriptional activation. Nat. Struct. Mol. Biol. *18*, 1358-1365.

Kalkhoven, E. (2004). CBP and p300: HATs for different occasions. Biochem. Pharmacol. *68*, 1145-1155.

Kamine, J., Elangovan, B., Subramanian, T., Coleman, D., and Chinnadurai, G. (1996). Identification of a Cellular Protein That Specifically Interacts with the Essential Cysteine Region of the HIV-1 Tat Transactivator. Virology *216*, 357-366.

Kim, C.H., Kim, J.W., Jang, S.M., An, J.H., Song, K.H., and Choi, K.H. (2012). Transcriptional activity of paired homeobox Pax6 is enhanced by histone acetyltransferase Tip60 during mouse retina development. Biochem. Biophys. Res. Commun. *424*, 427-432.

Kim, D., Blus, B.J., Chandra, V., Huang, P., Rastinejad, F., and Khorasanizadeh, S. (2010). Corecognition of DNA and a methylated histone tail by the MSL3 chromodomain. Nat. Struct. Mol. Biol. *17*, 1027-1029.

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

Krissinel, E., and Henrick, K. (2004). Secondary-structure matching (SSM), a new tool for fast protein structure alignment in three dimensions. Acta Crystallogr. D Biol. Crystallogr. *60*, 2256-2268.

Kusch, T., Florens, L., Macdonald, W.H., Swanson, S.K., Glaser, R.L., Yates, J.R., Abmayr, S.M., Washburn, M.P., and Workman, J.L. (2004). Acetylation by Tip60 is required for selective histone variant exchange at DNA lesions. Science *306*, 2084-2087.

Kusch, T., Mei, A., and Nguyen, C. (2014). Histone H3 lysine 4 trimethylation regulates cotranscriptional H2A variant exchange by Tip60 complexes to maximize gene expression. Proc. Natl. Acad. Sci. U. S. A. *111*, 4850-4855.

Lee, K.K., and Workman, J.L. (2007). Histone acetyltransferase complexes: one size doesn't fit all. Nat. Rev. Mol. Cell Biol. *8*, 284-295.

Legube, G., and Trouche, D. (2003). Identification of a larger form of the histone acetyl transferase Tip60. Gene *310*, 161-168.

Li, G., and Reinberg, D. (2011). Chromatin higher-order structures and gene regulation. Curr. Opin. Genet. Dev. *21*, 175-186.

Liu, X., Wang, L., Zhao, K., Thompson, P.R., Hwang, Y., Marmorstein, R., and Cole, P.A. (2008). The structural basis of protein acetylation by the p300/CBP transcriptional coactivator. Nature *451*, 846-850.

Luger, K., Dechassa, M.L., and Tremethick, D.J. (2012). New insights into nucleosome and chromatin structure: an ordered state or a disordered affair? Nat. Rev. Mol. Cell Biol. *13*, 436-447.

Luger, K., and Hansen, J.C. (2005). Nucleosome and chromatin fiber dynamics. Curr. Opin. Struct. Biol. *15*, 188-196.

Luger, K., Mäder, A.W., Richmond, R.K., Sargent, D.F., and Richmond, T.J. (1997). Crystal structure of the nucleosome core particle at 2.8 Å resolution. Nature *389*, 251-260.

Ma, J.C., and Dougherty, D.A. (1997). The Cation−π Interaction. Chem. Rev. *97*, 1303-1324.

Marmorstein, R. (2001). Protein modules that manipulate histone tails for chromatin regulation. Nature Reviews Molecular Cell Biology *2*, 422-432.

Marmorstein, R., and Roth, S.Y. (2001). Histone acetyltransferases: function, structure, and catalysis. Curr. Opin. Genet. Dev. *11*, 155-161.

Marmorstein, R., and Trievel, R.C. (2009). Histone modifying enzymes: Structures, mechanisms, and specificities. Biochim. Biophys. Acta *1789*, 58-68.

Maze, I., Noh, K.M., Soshnev, A.A., and Allis, C.D. (2014). Every amino acid matters: essential contributions of histone variants to mammalian development and disease. Nat. Rev. Genet. *15*, 259-271.

McCoy, A.J., Grosse-Kunstleve, R.W., Adams, P.D., Winn, M.D., Storoni, L.C., and Read, R.J. (2007). Phaser crystallographic software. J Appl Crystallogr *40*, 658-674.

McGinty, R.K., and Tan, S. (2015). Nucleosome structure and function. Chem. Rev. *115*, 2255-2273.

Musselman, C.A., Lalonde, M.E., Cote, J., and Kutateladze, T.G. (2012). Perceiving the epigenetic landscape through histone readers. Nat. Struct. Mol. Biol. *19*, 1218-1227.

Nielsen, P.R., Nietlispach, D., Buscaino, A., Warner, R.J., Akhtar, A., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2005). Structure of the chromo barrel domain from the MOF acetyltransferase. J. Biol. Chem. *280*, 32326-32331.

Nielsen, P.R., Nietlispach, D., Mott, H.R., Callaghan, J., Bannister, A., Kouzarides, T., Murzin, A.G., Murzina, N.V., and Laue, E.D. (2002). Structure of the HP1 chromodomain bound to histone H3 methylated lysine 9. Nature *416*, 103-107.

Otwinowski, Z., and Minor, W. (1997). Processing of X-ray Diffraction Data Collected in Oscillation Mode. In Volume 276: Macromolecular Crystallography, Part A, J. Carter, C.W., and R.M. Sweet, eds. (New York: Academic Press), pp. 307-326.

Peng, L., Ling, H., Yuan, Z., Fang, B., Bloom, G., Fukasawa, K., Koomen, J., Chen, J., Lane, W.S., and Seto, E. (2012). SIRT1 negatively regulates the activities, functions, and protein levels of hMOF and TIP60. Mol. Cell. Biol. *32*, 2823-2836.

Pirooznia, S.K., Sarthi, J., Johnson, A.A., Toth, M.S., Chiu, K., Koduri, S., and Elefant, F. (2012). Tip60 HAT activity mediates APP induced lethality and apoptotic cell death in the CNS of a Drosophila Alzheimer's disease model. PLoS One *7*, e41776.

Robert, X., and Gouet, P. (2014). Deciphering key features in protein structures with the new ENDscript server. Nucleic Acids Res. *42*, W320-324.

Roth, S.Y., Denu, J.M., and Allis, C.D. (2001). Histone Acetyltransferases. Annu. Rev. Biochem. *70*, 81-120.

Ruthenburg, A.J., Li, H., Patel, D.J., and Allis, C.D. (2007). Multivalent engagement of chromatin modifications by linked binding modules. Nat. Rev. Mol. Cell Biol. *8*, 983-994.

Sapountzi, V., and Cote, J. (2011). MYST-family histone acetyltransferases: beyond chromatin. Cell. Mol. Life Sci. *68*, 1147-1156.

Sapountzi, V., Logan, I.R., and Robson, C.N. (2006). Cellular functions of TIP60. Int. J. Biochem. Cell Biol. *38*, 1496-1509.

Segal, E., and Widom, J. (2009). What controls nucleosome positions? Trends Genet. *25*, 335-343.

Shimojo, H., Sano, N., Moriwaki, Y., Okuda, M., Horikoshi, M., and Nishimura, Y. (2008). Novel structural and functional mode of a knot essential for RNA binding activity of the Esa1 presumed chromodomain. J. Mol. Biol. *378*, 987-1001.

Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Soding, J., *et al.* (2011). Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal Omega. Mol. Syst. Biol. *7*, 539.

Soria, G., Polo, S.E., and Almouzni, G. (2012). Prime, repair, restore: the active role of chromatin in the DNA damage response. Mol. Cell *46*, 722-734.

Squatrito, M., Gorrini, C., and Amati, B. (2006). Tip60 in DNA damage response and growth control: many tricks in one HAT. Trends Cell Biol. *16*, 433-442.

Srivastava, P., Khandokar, Y.B., Swarbrick, C.M., Roman, N., Himiari, Z., Sarker, S., Raidal, S.R., and Forwood, J.K. (2014). Structural characterization of a Gcn5-related N-acetyltransferase from Staphylococcus aureus. PLoS One *9*, e102348.

Struhl, K., and Segal, E. (2013). Determinants of nucleosome positioning. Nat. Struct. Mol. Biol. *20*, 267-273.

Suganuma, T., and Workman, J.L. (2011). Signals and Combinatorial Functions of Histone Modifications. Annu. Rev. Biochem. *80*, 473-499.

Sun, Y., Jiang, X., Chen, S., Fernandes, N., and Price, B.D. (2005). A role for the Tip60 histone acetyltransferase in the acetylation and activation of ATM. Proc. Natl. Acad. Sci. U. S. A. *102*, 13182-13187.

Sun, Y., Jiang, X., and Price, B.D. (2010). Tip60: connecting chromatin to DNA damage signaling. Cell Cycle *9*, 930-936.

Sun, Y., Jiang, X., Xu, Y., Ayrapetov, M.K., Moreau, L.A., Whetstine, J.R., and Price, B.D. (2009). Histone H3 methylation links DNA damage detection to activation of the tumour suppressor Tip60. Nat. Cell Biol. *11*, 1376-1382.

Tang, Y., Luo, J., Zhang, W., and Gu, W. (2006). Tip60-dependent acetylation of p53 modulates the decision between cell-cycle arrest and apoptosis. Mol. Cell *24*, 827-839.

Taverna, S.D., Li, H., Ruthenburg, A.J., Allis, C.D., and Patel, D.J. (2007). How chromatin-binding modules interpret histone modifications: lessons from professional pocket pickers. Nat. Struct. Mol. Biol. *14*, 1025-1040.

Tremethick, D.J. (2007). Higher-order structures of chromatin: the elusive 30 nm fiber. Cell *128*, 651-654.

Tropberger, P., and Schneider, R. (2013). Scratching the (lateral) surface of chromatin regulation by histone modifications. Nat. Struct. Mol. Biol. *20*, 657-661.

Turnbull, W.B., and Daranas, A.H. (2003). On the value of c: can low affinity systems be studied by isothermal titration calorimetry? J. Am. Chem. Soc. *125*, 14859-14866.

Utley, R.T., and Cote, J. (2003). The MYST Family of Histone Acetyltransferases. In Protein Complexes that Modify Chromatin, J.L. Workman, ed. (Springer Berlin Heidelberg), pp. 203-236.

van Attikum, H., and Gasser, S.M. (2009). Crosstalk between histone modifications during the DNA damage response. Trends Cell Biol. *19*, 207-217.

Veloso, A., Kirkconnell, K.S., Magnuson, B., Biewen, B., Paulsen, M.T., Wilson, T.E., and Ljungman, M. (2014). Rate of elongation by RNA polymerase II is associated with specific gene features and epigenetic modifications. Genome Res. *24*, 896-905.

Venkatesh, S., and Workman, J.L. (2015). Histone exchange, chromatin structure and the regulation of transcription. Nat. Rev. Mol. Cell Biol. *16*, 178-189.

Vetting, M.W., LP, S.d.C., Yu, M., Hegde, S.S., Magnet, S., Roderick, S.L., and Blanchard, J.S. (2005). Structure and functions of the GNAT superfamily of acetyltransferases. Arch. Biochem. Biophys. *433*, 212-226.

Walker, P.A., Leong, L.E.-C., Ng, P.W.P., Tan, S.H., Waller, S., Murphy, D., and Porter, A.G. (1994). Efficient and Rapid Affinity Purification of Proteins Using Recombinant Fusion Proteases. Nat. Biotechnol. *12*, 601-605.

Wang, L., Tang, Y., Cole, P.A., and Marmorstein, R. (2008a). Structure and chemistry of the p300/CBP and Rtt109 histone acetyltransferases: implications for histone acetyltransferase evolution and function. Curr. Opin. Struct. Biol. *18*, 741-747.

Wang, Z., Zang, C., Rosenfeld, J.A., Schones, D.E., Barski, A., Cuddapah, S., Cui, K., Roh, T.Y., Peng, W., Zhang, M.Q., *et al.* (2008b). Combinatorial patterns of histone acetylations and methylations in the human genome. Nat. Genet. *40*, 897-903.

Winn, M.D., Ballard, C.C., Cowtan, K.D., Dodson, E.J., Emsley, P., Evans, P.R., Keegan, R.M., Krissinel, E.B., Leslie, A.G., McCoy, A., *et al.* (2011). Overview of the CCP4 suite and current developments. Acta Crystallogr. D Biol. Crystallogr. *67*, 235-242.

Wolf, E., Vassilev, A., Makino, Y., Sali, A., Nakatani, Y., and Burley, S.K. (1998). Crystal Structure of a GCN5-Related N-acetyltransferase: Serratia marcescens Aminoglycoside 3-N-acetyltransferase. Cell Cycle *94*, 439-449.

Xiao, H., Chung, J., Kao, H.Y., and Yang, Y.C. (2003). Tip60 is a co-repressor for STAT3. J. Biol. Chem. *278*, 11197-11204.

Yan, Y., Barlev, N.A., Haley, R.H., Berger, S.L., and Marmorstein, R. (2000). Crystal Structure of Yeast Esa1 Suggests a Unified Mechanism for Catalysis and Substrate Binding by Histone Acetyltransferases. Mol. Cell *6*, 1195-1205.

Yan, Y., Harper, S., Speicher, D.W., and Marmorstein, R. (2002). The catalytic mechanism of the ESA1 histone acetyltransferase involves a self-acetylated intermediate. Nat. Struct. Biol. *9*, 862-869.

Yang, C., Wu, J., and Zheng, Y.G. (2012). Function of the active site lysine autoacetylation in Tip60 catalysis. PLoS One *7*, e32886.

Yang, X. (2004). The diverse superfamily of lysine acetyltransferases and their roles in leukemia and other diseases. Nucleic Acids Res. *32*, 959-976.

Yap, K.L., and Zhou, M.M. (2011). Structure and mechanisms of lysine methylation recognition by the chromodomain in gene transcription. Biochemistry *50*, 1966-1980.

Yuan, H., Rossetto, D., Mellert, H., Dang, W., Srinivasan, M., Johnson, J., Hodawadekar, S., Ding, E.C., Speicher, K., Abshiru, N., *et al.* (2012). MYST protein acetyltransferase activity requires active site lysine autoacetylation. EMBO J. *31*, 58-70.

Zentner, G.E., and Henikoff, S. (2013). Regulation of nucleosome dynamics by histone modifications. Nat. Struct. Mol. Biol. *20*, 259-266.