THE PENNSYLVANIA STATE UNIVERSITY
SCHREYER HONORS COLLEGE

DEPARTMENT OF BIOCHEMISTRY AND MOLECULAR BIOLOGY

POLYCISTRONIC COEXPRESSION AND COPURIFICATION OF THE HUMAN MLL MWRA HISTONE METHYLTRANSFERASE COMPLEX

WIL LEONARD SANTIVASI

Spring 2011

A thesis
submitted in partial fulfillment
of the requirements
for a baccalaureate degree
in Premedicine
with honors in Biochemistry and Molecular Biology

Reviewed and approved* by the following:

Song Tan, Ph. D.
Associate Professor of Biochemistry and Molecular Biology
Thesis Supervisor

C. P. David Tu, Ph. D.
Professor of Biochemistry and Molecular Biology
Honors Adviser

Wendy Hanna-Rose, Ph. D.
Associate Department Head for Undergraduate Studies
Department of Biochemistry and Molecular Biology

*Signatures are on file in the Schreyer Honors College

Abstract

Eukaryotic organisms have highly-structured and complex gene regulatory pathways. Histone methyltransferases such as the MLL complex play a prominent role in the transcriptional activation of genes via mono-, di-, and tri-methylation of lysine 4 on the N-terminal tail of histone H3. These modifications allow for the regulation of genes downstream of the methylated nucleosome. The MWRA complex is the smallest catalytically active dimethylase complex of the MLL complex, and is composed of MLL1, WDR5, RbBP5, and Ash2L. Previously, the MWRA complex was expressed as discrete subunits and reconstituted *in vitro*. Using a system developed by Dr. Song Tan, the four subunits of the MWRA subcomplex were successfully subcloned into a polycistronic expression vector and expressed as a complex in *E. coli*. The complex was copurified using Talon® metal affinity purification and Source™ S cation-exchange chromatography, yielding milligram quantities. Single-turnover kinetic experiments performed by Dr. Michael Cosgrove and Dr. Anamika Patel at Syracuse University demonstrated that bacterially coexpressed MWRA subcomplex displays the same specificity and activity as MWRA subcomplex that was reconstituted *in vitro.* The bacterially coexpressed MLL MWRA complex will facilitate biophysical characterizations including crystallization trials of this chromatin complex.

Table of Contents

Table of Figures

Table of Tables

Acknowledgements

I would like to thank all of the people that made this thesis possible. I extend my gratitude to past Tan Lab postdoctoral fellows Dr. Ravindra Makde, Dr. Sung-Hoon Jun, and Dr. Melanie Adams-Cioabă, the extraordinary Tan Lab graduate students Matt Jennings, Tom Koerber, and Jiehuan Huang, and technologist Bryan Thurston for graciously sharing their expertise, as well as past undergraduates Omkar Bhat and Adam Seitz for training me in the many techniques necessary to conduct this research. I would also like to thank the other undergraduate members of the Tan Lab, Keiran Aurori, Aaron Nogan, Zach Hostetler, Viktor Tollemar, Josh McAnulty, Ben Himes, Jen Bayly, and Jon Saperstein for fostering an environment of collaboration. I also express my gratitude to Slater Nair and Allen Minns for their invaluable contributions to the experiments outlined in this thesis. I would also like to thank Dr. Michael Cosgrove and Dr. Anamika Patel at Syracuse University for collaborating with us on aspects of this project. Most of all, I would like to thank Dr. Song Tan for his continued support, advice, and patience over the past four years. My time as a member of his research group has been the defining experience of my undergraduate career. It would be impossible to articulate exactly how grateful I am to have been a member of the Tan Lab. I also thank Dr. David Tu and Dr. Wendy Hanna-Rose for their assistance in reviewing and preparing this thesis.

Chapter 1: Introduction

1.1 Chromatin

Inside every human cell reside the nearly 3 billion base pairs of DNA that comprise the human genome. Astonishingly, this DNA, which measures nearly 2 meters in length, can fit inside a nucleus only 10 μm across. This is due in large part to the many tiers of organization applied to DNA in eukaryotes (Figure 1.1, Alberts et al, 2002). At the smallest scale, DNA is combined with histone proteins into functional units called nucleosomes.

The nucleosome core particle is composed of a 147 bp-long strand of DNA wrapped 1.65 times around a histone octamer, which in turn consists of eight histone proteins (two copies each of histones H2A, H2B, H3, and H4). Nucleosome core particles are assembled by the binding of two copies each of histone H3 and histone H4 into a tetramer, followed by the tetramer binding DNA and the recruitment of two histone H2A/H2B dimers. Repeated nucleosomes linked by linker strands of DNA give rise to the "beads on a string" structure seen in electron micrographs of euchromatin (Figure 1.2, Olins and Olins, 2003). The crystal structure of the nucleosome core particle was resolved to 2.8 Å in 1997 (Figure 1.3, Luger et al, 1997) and was further resolved to 1.9 Å in 2002 (Davey et al, 2002).

The four histone proteins that compose the histone octamer share similar structural traits. Each histone contains a globular central domain characterized by the "histone fold", a motif composed of a central α-helix flanked by shorter helices connected with DNA-

binding loops (Figure 1.4, Harp et al, 2000). Each histone protein also includes a loosely structured N-terminal tail region of 15-30 amino acids. Histone protein sequences are highly conserved. Post-translational modification has been shown to occur in both the central and N-terminal tail domains. These traits exist in all eukaryotic organisms' nucleosomes.

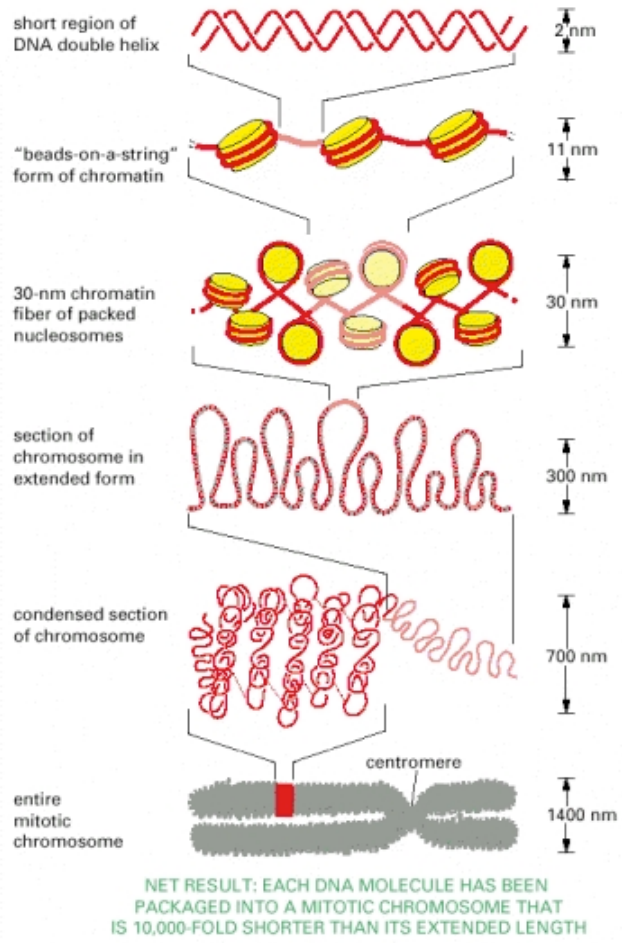

Figure 1.1: Hierarchical structure of chromatin packaging in the cell (Alberts et al, 2002)

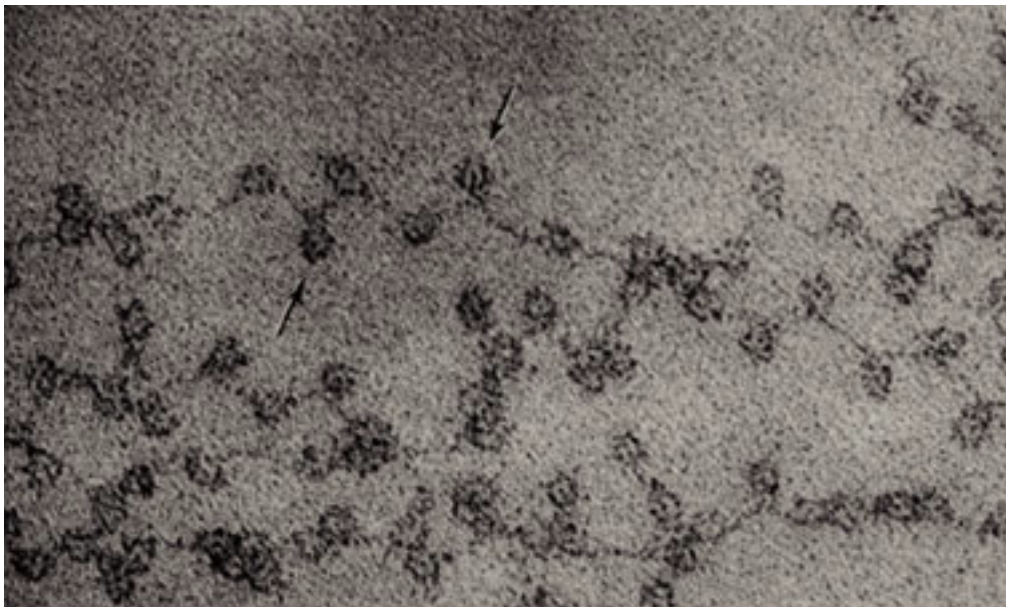

Figure 1.2: Electron micrograph showing "beads on a string" structure of euchromatin (Olins and Olins, 2003)

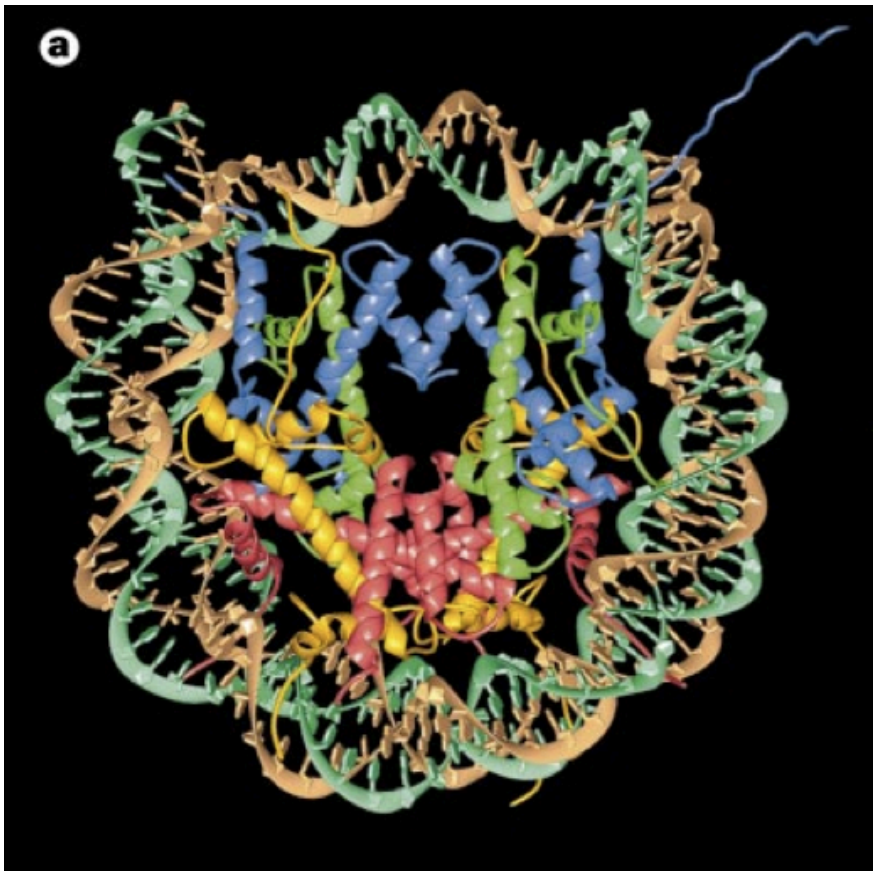

Figure 1.3: Structure of the nucleosome core particle (Luger et al, 1997)

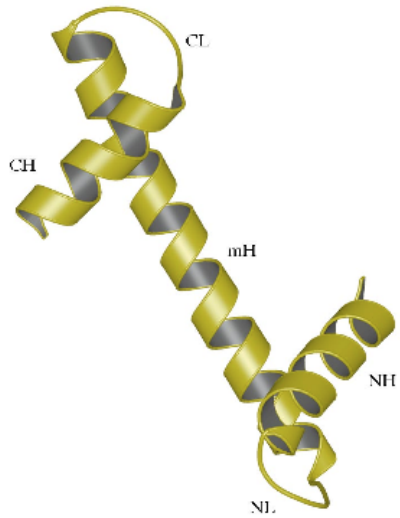

Figure 1.4: Histone fold structural motif (Harp et al, 2000)

1.2 Eukaryotic Transcription

Expression in eukaryotic cells requires the recruitment of many distinct molecules to the gene. The promoter, a region located upstream of the gene, controls the level of transcription by recruiting these components. While highly variable, most promoters contain the "TATA box," which is characterized by a TATAWAW (where "W" stands for "weak," meaning either A or T) consensus sequence 27 bp upstream of the gene. The TATA box recruits transcription factor proteins, which bind DNA and recruit RNA polymerase II to the gene (Alberts et al, 2002).

However, the recruitment of these proteins is dependent on the DNA being unobstructed. Genes located in nucleosome-rich regions of DNA are transcribed less than those in unbound regions. Interactions between histone tails, as well as between DNA and histone folds, are thought to repress transcription of nucleosome-bound DNA. DNA bound in a nucleosome is therefore inaccessible to RNA polymerase II and cannot be transcribed. As

such, unbound regions of DNA are often found upstream of commonly transcribed genes. Yet chromatin packaging is extremely dynamic and nucleosomes can be altered to make their DNA available for transcriptional use (Alberts et al, 2002). As such, nucleosome-rich regions are often located upstream of genes whose transcriptional activity is regulated by the cell.

Despite the high percentage of the human DNA incorporated into nucleosomes, they are not positioned randomly across the genome. In order to accommodate the histone octamer, DNA must be able to bend. Histones have been shown to bind specific sequences of DNA with increased bendability, especially those containing AA or TT dinucleotides every 6 bp (Ioshikes et al, 1996). As such, the placement of nucleosome-compatible sequences in the genome allows for transcriptional control.

1.3 Chromatin Modification and Transcriptional Regulation

As a result of their ability to be modified post-translationally, nucleosomes play a critical role in gene regulation. Certain enzymes have been shown to alter nucleosome and chromatin structure by covalently modifying certain residues on histone N-terminal tails. These modifications can lead to the relocation of the nucleosome to another DNA region (Whitehouse et al, 1999). Common modifications include the transfer of methyl, acetyl, phosphate, and ubiquitin groups to lysines, arginines and other amino acids. These covalent modifications are a key tenet of the histone code hypothesis, which suggests that such histone modifications are responsible for chromatin remodeling, both directly by altering DNA-octamer interaction and indirectly by recruiting other chromatin

remodeling enzymes (Jenuwein and Allis, 2001). Enzymes that facilitate the transfer of methyl groups to histones are known as histone methyltransferases (HMTs), and are the focus of this research.

1.4 Histone Methylation

Histone lysine methyltransferases facilitate the transfer of a methyl group from S-adenosyl methionine (SAM) to the ε-amino group of specific lysines on the histone tail (Figure 1.5, Qian and Zhou, 2006).

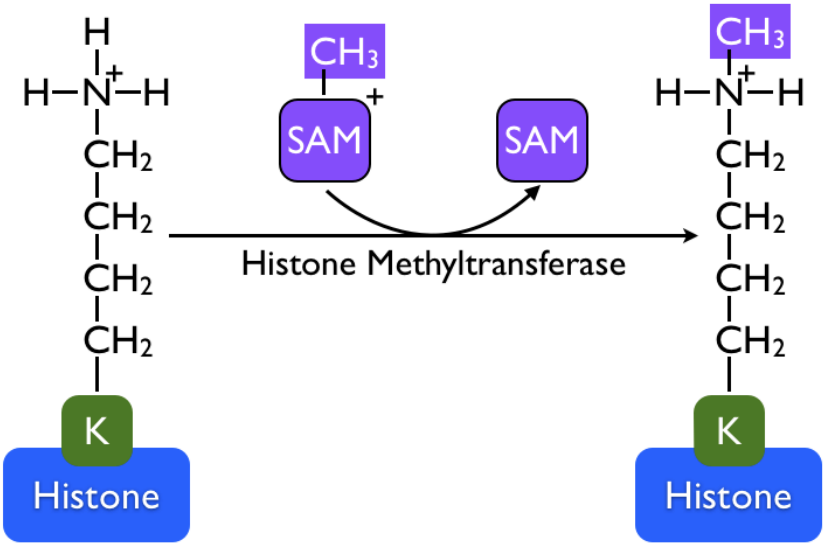

Figure 1.5: Histone methyltransferase reaction mechanism (adapted from Qian and Zhou, 2006)

Distinct enzymes with different lysine specificities modify different lysines on the histone tails. A class of histone lysine methyltransferases facilitates the transfer of methyl groups to lysine 4 of the histone H3 N-terminal tail (H3K4). Methylation of H3K4 is generally associated with transcriptional activation (Grewal and Rice, 2004). All known H3K4 methyltransferases have been shown to contain a variant of the catalytic SET

domain. The SET domain is highly conserved across proteins and across species (Guo and Guo, 2007). It contains a hydrophobic channel that allows for the positioning of an SAM molecule near the target lysine with its methyl group adjacent to the ε-amino group. While no mechanism has been confirmed, it is believed that a methyl group is transferred to the ε-amino nitrogen via an S_N2 nucleophilic attack following deprotonation of the ε-amino amino group (Qian and Zhou, 2006). Generally, methylation of H3K4 is associated with transcriptional activation due to the loosening of histone-DNA interaction (Dou et al, 2006). Monomethylation of H3K4 is indicative of transcriptionally active chromatin, while H3K4 dimethylation is observed at genes with basal transcription levels (Qian and Zhou, 2006). Trimethylation of H3K4 is commonly associated with transcriptional activation and elongation and is observed at fully active promoters (Dou et al, 2006).

1.5 Histone Demethylation

Certain enzymes can also demethylate histones. Histone demethylation is catalyzed by histone demethylases and is frequently associated with gene repression. LSD1 was the first discovered histone demethylase. It is capable of catalyzing the demethylation of mono- and dimethylated H3K4 (Shi et al, 2004). It is heavily conserved across species and serves as a transcriptional repressor. It has since been identified as a component of many complexes that facilitate gene repression and formation of heterochromatin (Noma et al, 2001).

1.6 MLL Complex

The human MLL (mixed-lineage leukemia) complex is a 1 MDa histone 3 lysine 4 (H3K4) trimethyltransferase complex that interacts with MOF, an H4-specific acetyltransferase. It facilitates the activation of *Hox* genes, which are critical for the embryonic development of proper body structure (Carroll, 1995). The complex activates these genes by trimethylating the nucleosomes that contain *Hox* gene promoters (Qian and Zhou, 2006). Max, E2F6, and p53 transcription factors recruit MLL to these genes. The complex's recognition of chromatin is not limited to the TATA box region of *Hox* genes. In fact, MLL binds the transcriptional unit much like RNA polymerase II. It has been demonstrated that the MLL complex regulates genes due to its role in transcriptional elongation. (Slany, 2005) A generalized figure of its role in transcriptional regulation can be found below. (Figure 1.6, Slany, 2005)

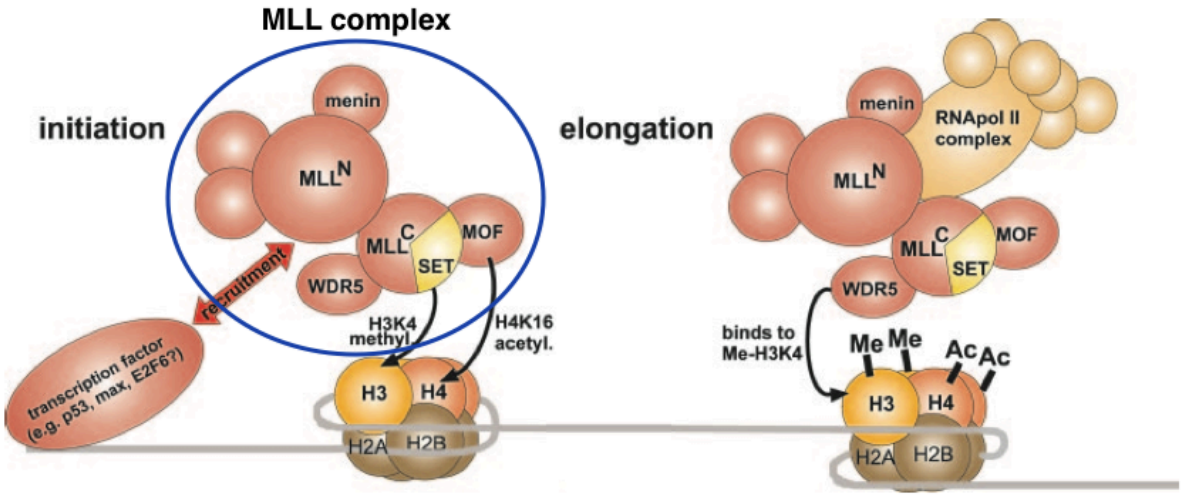

Figure 1.6: General role of the MLL complex in transcriptional control (Slany, 2005)

1.7 MLL MWRA Complex

The MLL MWRA complex is a four-subunit, 182.1 kDa H3K4 dimethylase and is the smallest catalytically active subcomplex of the MLL complex. It is comprised of a C-terminal portion of MLL1 containing the Win WDR5-binding domain and the SET domain, WDR5, RbBP5, and Ash2L. Interaction has been demonstrated between MLL1 and WDR5, between WDR5 and RbBP5, and between RbBP5 and Ash2L (Dou et al, 2006; Patel et al, 2009). While no crystal structure has been determined for the complex, previous research has produced a general figure for its inter-subunit binding, as well as its interaction with the tail of histone H3 (Figure 1.7, Cao et al, 2010).

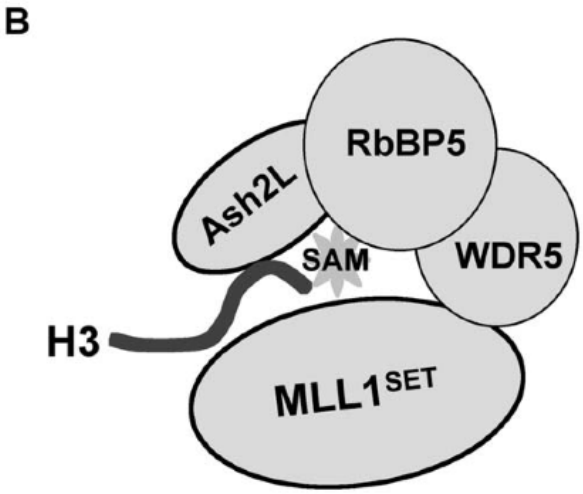

Figure 1.7: Organization of the MWRA complex with histone H3 tail and S-adenosyl methionine (Cao et al, 2010)

The MLL MWRA complex was first produced in *E. coli* using individual subunit expression, purification, and *in vitro* complex reconstitution by the Cosgrove Laboratory at Syracuse University (Patel et al, 2009). The complex has also been successfully reconstituted by infecting cells with multiple engineered baculoviruses, each containing

one of the individual subunits' genes (Dou et al, 2008). The purpose of this research was to determine if the MWRA complex could be coexpressed in *E. coli* and subsequently purified as an intact complex.

1.8 MLL1

Mixed-lineage leukemia protein 1 (MLL1) serves as the catalytic subunit of the MLL complex. Fusion proteins containing MLL1 have been implicated in some types of leukemia (Liu et al, 2009). When incorporated into the MLL complex and as a single protein, it is capable of H3K4 monomethylation. After translation, the MLL1 protein is cleaved into two peptides: MLL^N and MLL^C. MLL^N has been shown to interact with menin, while MLL^C interacts with MOF and WDR5. Both peptides are necessary for recruitment of RNA polymerase II (Slany, 2005). The MWRA subcomplex only contains a portion of MLL^C. WDR5, RbBP5, and Ash2L regulate MLL1's specific catalytic activity in the MWRA subcomplex. (Patel et al, 2008a)

1.8.1 Structural and Functional Domains

MLL1's C-terminal region contains two components critical to the assembly and function of the MLL MWRA complex. Residues 3745-3768, also known as the Win (WDR5 interaction) domain, have been identified as necessary for binding WDR5. Of particular importance is arginine 3765, without which MLL1 is unable to recognize WDR5. Residues 3769-3969 contain the SET domain and are sufficient to recognize, bind, and monomethylate H3K4 (Figure 1.7, Patel et al, 2009). The SET domain is a catalytic domain that is highly conserved across the SET superfamily of histone lysine

methyltransferases (Dillon et al, 2005). The crystal structure of the SET domain of MLL1 has been resolved to 2.0 Å (Figure 1.8, Southall et al, 2009). As a result, the Win and SET domains are sufficient, along with WDR5, RbBP5, and Ash2L for the formation of the MWRA complex. In these experiments, this 200-amino acid truncation (MLL1Δ1) was used instead of a full-length version of MLL1.

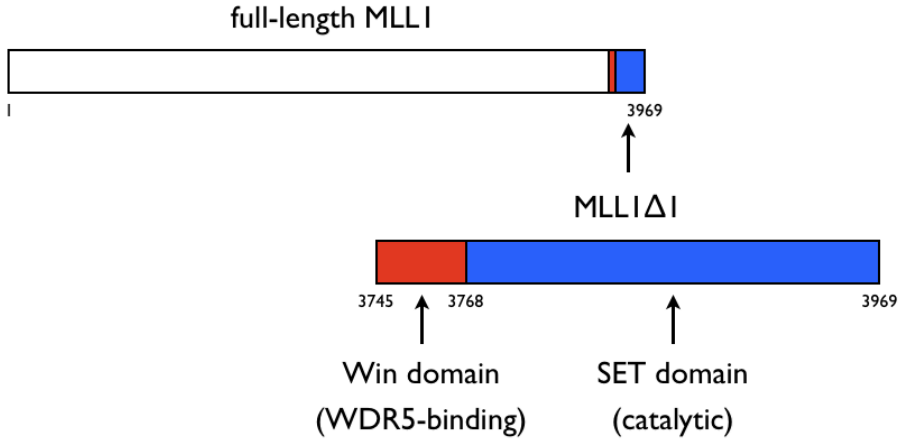

Figure 1.8: C-terminus domains of MLL1 (adapted from Patel et al, 2009)

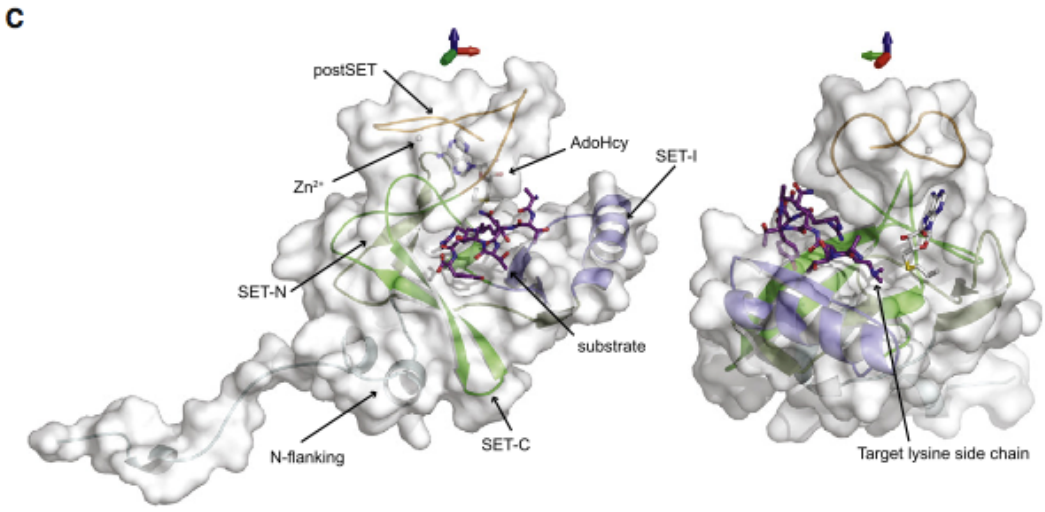

Figure 1.9: Structure of MLL1 SET domain (Southall et al, 2009)

1.8.2 Role in Human Disease

MLL1 and fusion proteins containing MLL1 have been implicated in a number of forms of acute leukemia. The MLL1 gene normally resides at locus q23 of chromosome 11. This locus is especially prone to chromosomal translocation. Translocations of the MLL1 gene result in the expression of MLL1-fusion proteins, which ultimately causes acute myeloid leukemia, acute lymphoblastic leukemia, and mixed-lineage leukemia (Krivtsov and Armstrong, 2007). It is believed that MLL1 fusion proteins interfere with PPP1R15A-mediated apoptosis (Liu et al, 2009). Under normal circumstances, regulation of cell homeostasis by PPP1R15A allows cells whose DNA sustains damage from ionizing radiation to self-destruct, stopping potential mutations from being propagated to later cell generations. When this apoptotic pathway is compromised, cells with damaged DNA cannot detect the damage, allowing for proliferation of their potentially harmful mutations, especially in rapidly dividing cells such as leukocytes (Krivtsov and Armstrong, 2007).

1.9 WDR5

WD repeat-containing protein 5 (WDR5) serves as a scaffold protein between MLL1, RbBP5 and histone H3. It recognizes arginine 2 of histone H3 and helps to stabilize lysine 4 for methylation by other subunits of the MWRA complex. It has been shown to interact with other SET-family methyltransferase complexes in a similar fashion. (Patel et al, 2008a) The crystal structure of WDR5 bound to residues 371-380 of RbBP5 and the Win domain of MLL1 has been resolved to 2.35 Å (Figure 1.9, Avdic et al, 2011). While

it is known that WDR5 is necessary for proper H3K4 methylation, its specific enzymatic role in the WMRA complex has not been identified (Odho et al, 2010).

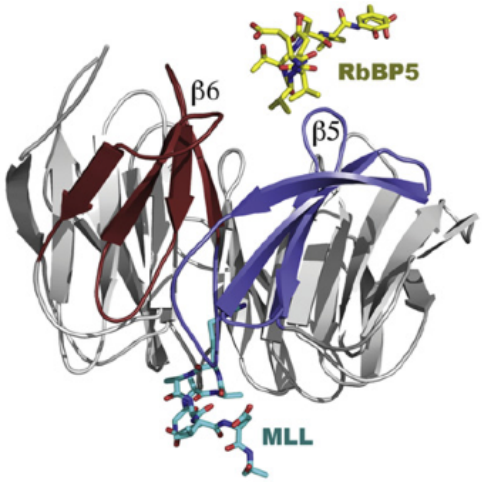

Figure 1.10: Structure of WDR5 bound to residues 371-380 of RbBP5 and Win domain of MLL1 (Patel et al, 2008a)

1.10 RbBP5

Retinoblastoma-binding protein 5 (RbBP5) serves as a scaffold protein between WDR5 and Ash2L. It is known to contain a 7-bladed WD repeat β-propeller domain, a structurally variable "hinge" domain, and a C-terminal region, the function of which remains unknown (Avdic et al, 2011). The presence of an RbBP5/Ash2L heterodimer has been shown to substantially increase the catalytic activity of the MLL1 SET domain. While RbBP5 does not exhibit methyltransferase activity, this heterodimer has been shown to successfully methylate monomethylated H3K4 (Cao et al, 2010).

1.11 Ash2L

Absent, small, or homeotic 2-like protein (Ash2L) is the fourth component of the MWRA complex. It contains an N-terminal C4 zinc finger domain and a C-terminal SPIa

Ryanodine Receptor domain (Avdic et al, 2011). Ash2L has been shown to recognize S-adenosyl methionine, and it has been suggested that it also helps form a stable channel to contain H3K4 during dimethylation (Cao et al, 2010). Experiments have confirmed that Ash2L is necessary for trimethylation of H3K4 (Steward et al, 2006). However, Ash2L will not catalyze trimethylation of H3K4 if H3K9 is trimethylated (Wysocka et al, 2003). Since H3K9 trimethylation indicates gene silencing, hAsh2L and the MLL complex are incapable of transcriptionally activating genes silenced by H3K9 trimethyltransferases.

1.12 Polycistronic Expression

A polycistronic expression system was employed to express the MLL MWRA complex. A polycistronic plasmid serves as the template for a single mRNA transcript that contains coding information for multiple genes. The pST44 plasmid suite is a T7 RNA polymerase-based expression system derived from the pET3a system (Figure 1.11, Tan et al, 2005). In the Tan Laboratory, it has been used to successfully coexpress the subunits of other chromatin-modifying enzymes such as the yeast Piccolo NuA4 histone acetyltransferase complex. pST69 is a polycistronic expression vector similar to pST44, but it includes rare restriction sites (SgrAI and RsrII) in Translation Cassette 1 (Song Tan, personal communication). These sites are extremely uncommon in genomic DNA (Veselkov et al, 1996). The pST50Tr series of transfer plasmids facilitate the incorporation of four genes into the polycistronic system. pST66Tr is a transfer plasmid allowing for the incorporation of a gene into the first cassette of pST69. pST73 is a pET3a-based expression plasmid. All plasmids can be transcribed by T7 RNA polymerase, as they contain the T7 promoter and terminator. The pST50Tr transfer

plasmids contain translational enhancers, Shine-Dalgarno sequences, initiation, and termination codons, and can be used to express individual genes. All plasmids contain the pMB1 ColE1 origin of replication and confer ampicillin resistance via the β-lactamase gene. (Tan et al, 2005)

pST73 is a pK 187x4-based polycistronic expression plasmid which contains the p15A origin of replication and confers kanamycin resistance. It employs the same transfer plasmids as pST69 and was developed by Song Tan.

Use of a single plasmid for coexpression allows for compatibility with expression strains, including BL21(DE3)pLysS and BL21-CodonPlus(DE3)-RIL, which contain plasmids to improve expression (Tan et al, 2005). Construction of a four-cassette polycistronic expression vector can be quite complex. A schematic overview of the subcloning scheme used to incorporate the MWRA subunits into pST69 can be found in Section 1.13 below.

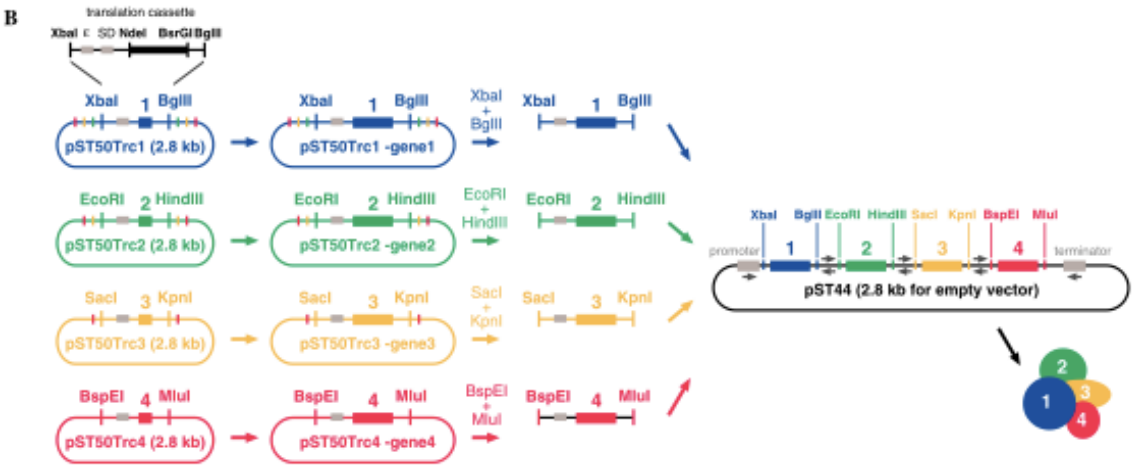

Figure 1.11: Schematic overview of the pST44 polycistronic expression system (Tan et al, 2005)

1.13 Schematic Overviews of MLL MWRA Polycistronic Expression Vector Preparation

Diagrams outlining the subcloning schemes used to generate the expression plasmids used in these experiments can be found in Figure 1.12 and Figure 1.13.

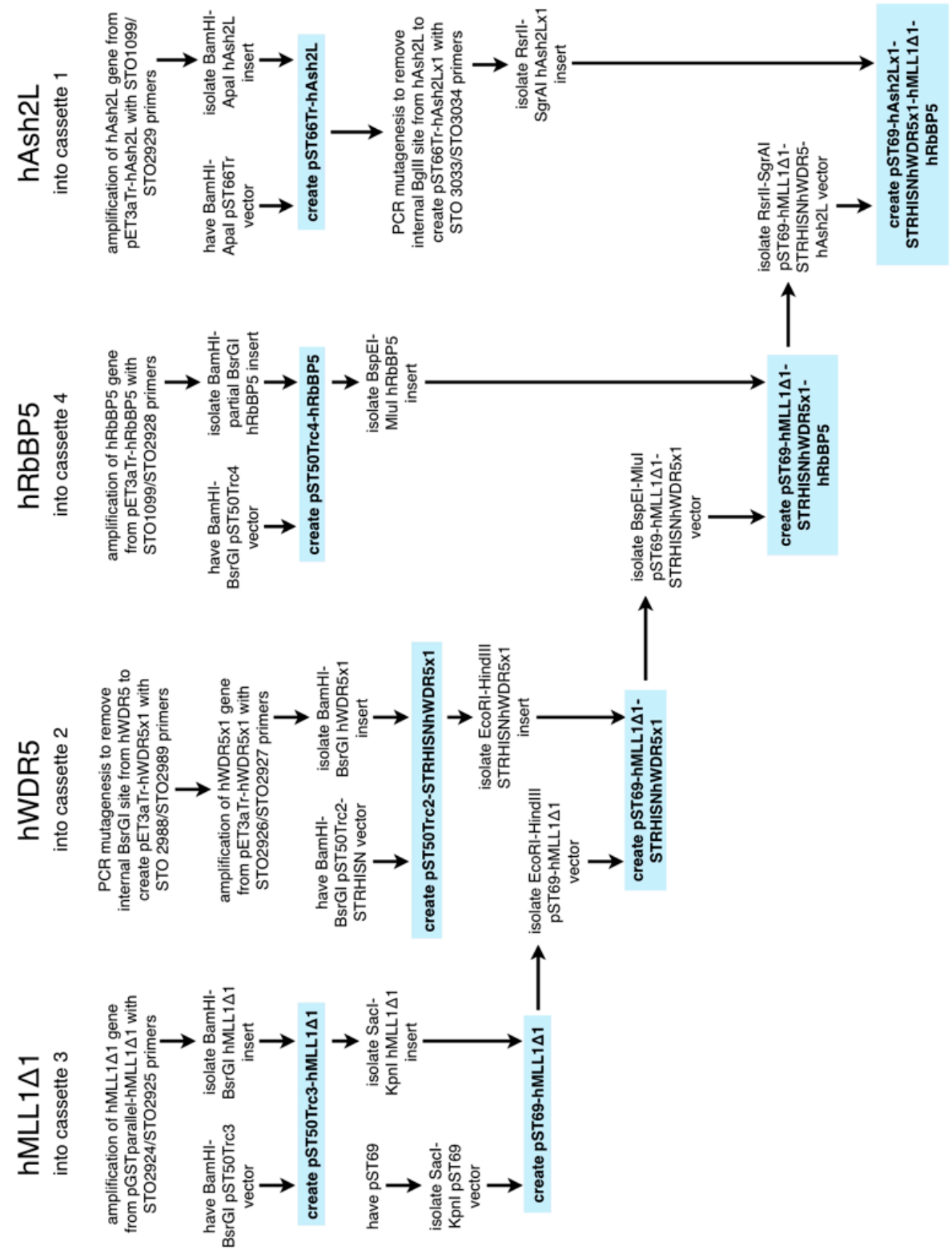

Figure 1.12: Cloning schematic to coexpress the MLL MWRA subcomplex by polycistronic expression

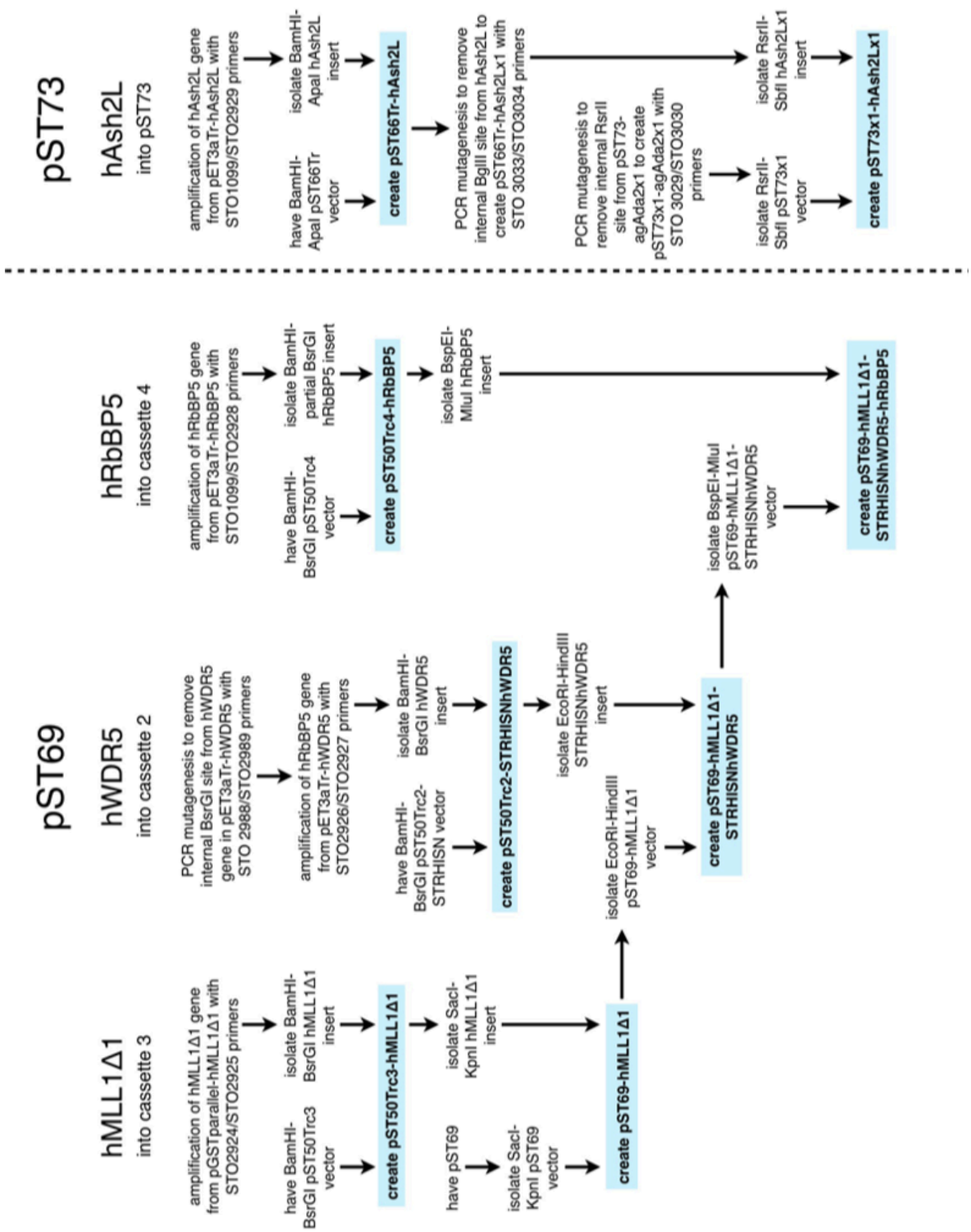

Figure 1.13: Cloning schematic to coexpress the MLL MWRA subcomplex using two plasmids

Chapter 2: Materials and Methods

2.1 Bacteriological Methods

2.1.1 Bacteriological Media

For all transformation, plasmid preparation, and protein expression, 2x TY liquid media was used. 2x TY contained 1.6% BactoTryptone (w/v), 1% yeast extract (w/v), and 0.5% NaCl (w/v) dissolved in deionized water. All media was sterilized via autoclave prior to use. Ampicillin (50 µg/mL), kanamycin (50 µg/mL), and chloramphenicol (25 µg/mL) were added before use as necessary. 2x TY was stored at room temperature until use.

For all experiments requiring solid media, TYE was used. TYE contained 1.5% agar (w/v), 1% BactoTryptone (w/v), 0.5% yeast extract (w/v), and 0.8% NaCl (w/v) dissolved in deionized water. All media was sterilized via autoclave prior to use. After cooling to room temperature, ampicillin (100 µg/mL), kanamycin (50 µg/mL), and chloramphenicol (25 µg/mL) were added. The media was then poured into Petri dishes and allowed to solidify at room temperature. TYE plates were stored at 4ºC until use.

2.1.2 Bacteriological Strains

Bacteriological strains used for plasmid preparation and protein expression are outlined in Table 2.1. TG1 cells were used for subcloning. BL21(DE3)pLysS and BL21-CodonPlus(DE3)-RIL cells were used for expression of recombinant proteins.

Table 2.1: Bacteriological Strains

Strain	Source	Genotype	Antibiotic Resistance
TG1	Toby Gibson	*Δ(lac-pro), supE, thi, hsdD5/F', traD36, proA+B+, $lacI_q$, lacZΔ, M15*	None
BL21(DE3)pLysS	Stratagene	*B F-, dcm, ompT, hsdS (r_B- m_B-) gal λ(DE3) [pLysS Cam^r]*	Chloramphenicol
BL21-CodonPlus(DE3)-RIL	Stratagene	*B F-, ompT, hsdS (r_B- m_B-) gal λ(DE3) endA Hte [argU ileY leuW Cam^r]*	Chloramphenicol

2.1.3 Competent Cell Preparation

Some expression experiments required two plasmids to be present. Competent cells containing a first plasmid conferring antibiotic resistance were prepared prior to transformation with the second plasmid, which conferred an additional form of antibiotic resistance. As such, resistance to both antibiotics indicated the successful transformation of both plasmids.

Competent cells were prepared by transforming the first plasmid into appropriate cells and streaking the cells onto a TYE plate with appropriate antibiotics. The plate was incubated at 37ºC overnight. A flask containing 100mL 2xTY and 2.2 mL of salt solution (0.44 M $MgCl_2$, 0.44 M $MgSO_4$, 0.11 M KCl) was inoculated with three colonies from the plate and grown in a shaking incubator at 200 rpm at 37ºC for 2-4 hours. Once the cells had reached an OD_{600} of 0.4-0.6, the culture was chilled on ice for 10 minutes and spun at 2000 rpm (Heraeus Megafuge 1.0R, #7570G rotor) at 4ºC for 7 minutes. The supernatant was discarded and the cell pellet was resuspended gently in 30mL of ice cold FB solution (10 mM PIPES, 250 mM KCl, 15 mM $CaCl_2$) and incubated on ice for 10 minutes. The cells were then spun again at 2000 rpm (Heraeus Megafuge 1.0R, #7570G

rotor) at 4ºC for 7 minutes. The supernatant was discarded, the cell pellet was resuspended in 8mL of cold FB solution, and 0.6 mL of 100% DMSO was added. The cells were incubated on ice for 10 minutes prior to being aliquoted into sterile Eppendorf tubes, flash frozen in liquid nitrogen, and stored at -80ºC.

2.2 DNA Methods

2.2.1 Processing of Custom Oligonucleotide Primers

Primers for polymerase chain reaction (PCR) experiments were obtained from IDT (Coralville, IA). The primers, which arrived lyophilized and desalted, were diluted to 50 μM in TE (10, 0.1) (10 M Tris-HCl pH 8.0, 0.1 mM EDTA) and incubated at room temperature for 5 minutes. The primer solution was then vortexed for 15 seconds and diluted to 10 μM in TE (10, 0.1) for use in PCR experiments.

2.2.2 Ethanol Precipitation

Ethanol precipitation was used to concentrate DNA and remove small organic molecules from DNA samples. 0.1 volumes of 3 M NaAc pH 5.2 were added to the sample, followed by 2.5 volumes of 95% ethanol. The sample was vortexed for 5 seconds and spun at full speed for 10 minutes in a microcentrifuge. The supernatant carefully removed via aspiration, and the pellet was allowed to air-dry for 3 minutes. The pellet was then resuspended in an appropriate volume of TE (10, 0.1).

2.2.3 Phenol/CIA Extraction

Phenol/CIA extraction was used to purify DNA from samples with contaminating proteins. Phenol/CIA was prepared by combining equal volumes of TE-equilibrated phenol and CIA (24 volumes of chloroform and 1 volume of isoamyl alcohol), vortexing the mixture, and spinning it at full speed for 1 minute in a microcentrifuge. Equal volumes of phenol/CIA and sample were combined, vortexed for 5 seconds, and spun at full speed for 1 minute in a microcentrifuge. The mixture separated into an upper aqueous phase and a lower organic phase. The aqueous phase was extracted and transferred to a new tube.

2.2.4 Agarose Gel Electrophoresis

Agarose gel electrophoresis was used to separate DNA fragments by molecular weight for analysis and preparative purposes. An appropriate amount of high gelling temperature Agarose (0.7-1.5% w/v) was added to 30 mL of 0.5x TBE (45 mM Tris base, 45 mM boric acid, 1.5 mM EDTA). Agarose concentrations are dependent on the size of the expected DNA fragment(s) and are summarized in Table 2.2 below.

Table 2.2: Agarose Concentrations Necessary for Visualization of DNA Fragments

DNA Fragment Size	Agarose Concentration
< 300 bp	1.5%
300-600 bp	1.2%
600-1500 bp	1%
1500-3000 bp	0.7%

The agarose was dissolved by heating the solution in a microwave for 60-90 seconds and swirling until no agarose powder remained. The solution was allowed to cool for 5 minutes at room temperature. 1.5 µL of 10 µg/mL ethidium bromide was added and the

solution was poured into a gel casting block. An appropriately sized well comb was installed and the gel was allowed to solidify for 30 minutes. The gel, along with its casting tray, were transferred to an electrophoresis box and covered with 0.5x TBE containing approximately 0.1 mg of ethidium bromide. Samples containing 1x gel loading buffer (GLB) were prepared by adding 1 volume of 6x GLB (0.25% bromophenol blue (w/v), 0.25% xylene cyanol (w/v), 30% glycerol (w/v) 60mM EDTA) to 5 volumes of DNA-containing sample. Samples were inserted into the wells and the gel was electrophoresed at 125 V until the bromophenol blue dye migrated to the far end of the gel (typically 40 minutes). Specific gel concentrations and electrophoresis times were dependent on the size of the desired DNA fragment. Electrophoresed DNA was visualized using a UV transilluminator.

2.2.5 Polymerase Chain Reaction

Polymerase chain reaction was used to amplify genes from plasmid DNA. Template plasmid DNA was diluted to 10 ng/µL. 10 µL of this template solution was combined with 10 µL of 10x ThermoPol buffer (New England BioLabs), 10 µL of 2.5mM dNTP, 5 µL of 10 µM forward and reverse primers, 0.5 µL of 2 units/µL of Pfu polymerase, and MilliQ water to a final combined volume of 100 µL. This reaction mixture was inserted into the thermocycler and incubated at 95ºC for 2 minutes to ensure full denaturation. The sample was then subjected to 5 cycles of 2 minutes at 95ºC, 30 seconds at T_m-5ºC (where T_m is the lowest primer annealing temperature), and 75ºC for an appropriate extension time. Typically, the extension time was 1 minute for every 1,000 bp in the expected product's length (e.g. 1 minute 30 seconds for a 1500 bp gene). Following this, the

sample was subjected to 20 cycles of 30 seconds at 95ºC, 30 seconds at 60ºC, and the appropriate extension time at 75ºC. Following these cycles, the sample was held at 75ºC for 3 minutes to ensure DNA renaturation. The sample was then incubated at 4ºC. Following thermal cycling, 10 µL of PCR product was combined with 2 µL of 6x GLB and was analyzed via electrophoresis on an appropriate-concentration agarose gel. The rest of the reaction mixture was extracted with 100 µL phenol/CIA and then 100 µL CIA to remove contaminants such as the polymerase. The sample was then ethanol precipitated and resuspended in 30 µL of TE (10, 0.1).

2.2.6 Subcloning

Much of the preparation of the polycistronic expression plasmid required the use of subcloning, the transfer of genes from one plasmid to another. This involved the preparation of insert and vector DNA by restriction mapping, "sticky-ended" ligation of the insert and vector DNA, transformation of the plasmid into competent cells, PCR screening of colonies, and plasmid preparation. Proper subcloning was then confirmed by restriction mapping and/or sequencing.

2.2.7 Restriction Enzyme Digestion

Insert and vector DNA must have "sticky ends," the cohesive ends achieved via digestion with certain restriction enzymes, for proper ligation. Vector DNA was prepared by combining 5 µL of plasmid DNA at ~0.2 µg/µL, 3 µL of an appropriate 10x New England BioLabs buffer, 3 µL of 1 mg/mL BSA, 1 µL of 100 mM DTT, 1.5 µL of the two restriction enzymes at 10-20 units/µL, and MilliQ water to a final combined volume

of 30 µL. For insert DNA created via PCR amplification, 15 µL of PCR product in TE (10, 0.1) was combined with the buffer, BSA, DTT, restriction enzymes, and water. The mixture was then incubated at 37ºC for 1-2 hours. In the event that both enzymes were not compatible with the same New England BioLabs buffer, a sequential digest was performed, where the restriction enzyme that is more active in lower-salt solutions digested for 1.5 hours prior to the addition of salt and the second enzyme. An example of a sequential digest is double digestion with BamHI and BsrGI. In this case, the DNA was first digested with 2 µL of 10 units/µL BsrGI for 1.5 hours, after which 0.6 µL 5 M NaCl and 1 µL of 20 units/µL BamHI were added. The mixture was then incubated for an additional 1.5 hours. 6 µL of 6x GLB was added to the sample in order to stop the reaction and prepare it for gel electrophoresis. The desired digested sample was then analyzed by electrophoresis on an appropriate agarose gel, and the desired fragment was purified from the gel via centrifugation.

2.2.8 Isolation of DNA from Agarose via Centrifugation

Agarose gel purification is necessary to isolate and recover restriction enzyme-digested DNA fragments. This procedure has been known to recover 50-70% of the DNA in a gel band between 100-8,000 bp. The agarose gel was run as usual. When electrophoresis was complete, the desired DNA fragment was cut out of the gel with a razor blade over a UV transilluminator. The gel fragment was then placed into a filter assembly, consisting of a 0.5 mL Eppendorf tube with a hole in the bottom and stuffed with siliconized glass wool, resting inside a 1.7 mL Eppendorf tube. The gel and filter assembly were spun at 7,000

rpm for 5 minutes in a microcentrifuge. The purified DNA filtered into the 1.7 mL Eppendorf tube and the filter assembly was discarded.

2.2.9 "Sticky-Ended" Ligation

Purified vector and insert fragments were ligated into plasmids by combining 1 μL of 10x T4 DNA polymerase buffer, 0.5 μL of 100 mM DTT, 2 μL of gel-purified vector DNA, 1.5 μL of gel-purified insert DNA, 1 μL of 40 units/μL T4 DNA polymerase (expressed and purified by the Tan Lab), and water to a final combined volume of 10 μL. A negative control reaction was performed without the inclusion of insert DNA. The ligation mixture was incubated at room temperature for 15-60 minutes before being transformed into TG1 competent cells. The negative control mixture was also transformed into TG1 cells. Successful ligation resulted in a 2-4 times more colonies on the ligation plate than the negative control plate. Similar numbers of colonies on both the ligation and vector plates signified unsuccessful subcloning.

2.2.10 Plasmid Transformation

A 100 μL aliquot of frozen TG1 competent cells was thawed on ice. 2 μL of ligation mix (or negative control reaction mix) was added to the cells and incubated on ice for 15-40 minutes. The cell aliquot was then heat shocked by placing it into a 42ºC water bath for 30 seconds followed by immediate incubation on ice to allow the DNA to enter the cells. 0.3 mL of 2x TY media was added to the aliquot and the culture was placed in a shaking incubator at 200 rpm at 37ºC for 15-45 minutes to allow for the expression of antibiotic

resistance. 0.3 mL of the culture was plated onto a TYE plate with appropriate antibiotics to select for the desired plasmid and incubated at 37ºC for 10-18 hours.

2.2.11 PCR Screening of Selected Colonies

Following transformation, colonies from the ligation mix plate were screened using PCR to identify clones whose plasmids contain the correct insert. Primers were selected such that their PCR products would be different for the correct insert and the parent vector plasmid without the insert. 4-12 colonies from the transformation plate were swirled in 100 µL aliquots of water and restreaked onto a new TYE plate with appropriate antibiotics. The plate was then incubated at 37C for 10-16 hours. For a 12-colony PCR screen, 28 µL ThermoPol buffer (New England BioLabs), 28 µL of 2.5 mM dNTP, 14 µL each of 10mM forward and reverse primers, and 1.4 µL of 2 units/µL Pfu polymerase were combined into a PCR reaction mix. 19 µL of this PCR reaction mix was combined with 1 µL of each colony's cell suspension in PCR tubes. The samples were then placed into a thermocycler and incubated at 95C for 2 minutes, followed by 25 cycles of 30 seconds at 95ºC, 30 seconds at T_m-10ºC, and 45 seconds at 75ºC. Following the PCR cycle, 3 µL of 6x GLB was added to each sample prior to analysis on via agarose gel electrophoresis. Two clones that produced the correct PCR product were selected for plasmid preparation.

2.2.12 Plasmid Preparation

Plasmid preparations purify the plasmid DNA from *E. coli*. Two PCR screen-positive colonies were inoculated into 100 mL of 2x TY media with appropriate antibiotics, which

were then placed in a shaking incubator at 200 rpm at 37ºC for 10-18 hours. Following incubation, the cell cultures were spun at 4,000 rpm (Heraeus Megafuge 1.0R, #7570G rotor) at room temperature for 5 minutes. The supernatant was discarded and the cell pellet was resuspended in 5 mL of LYSIS buffer (50 mM glucose, 25 mM Tris-Cl pH 8.0, 10 mM EDTA, Na). 10 mL of NaOH/SDS solution (0.2 M NaOH, 1% SDS) was added to lyse the cells. The sample was shaken vigorously and incubated on ice for 3-5 minutes. 10 mL of cold 5 M potassium acetate/2.5 M acetic acid was added, shaken vigorously, and incubated on ice for 3-5 minutes. The sample was then spun at 4,000 rpm (Heraeus Megafuge 1.0R, #7570G rotor) at room temperature for 3 minutes and filtered through a sinctured glass funnel. The supernatant was combined with 12.5 mL isopropanol and incubated at room temperature for 10 minutes before being spun at 13,000 rpm (Sorvall SS-34 rotor) at room temperature for 7 minutes. The supernatant was discarded and the pellet resuspended in 150 µL of TE (10, 50) (10 mM Tris-Cl pH 8.0, 50 mM EDTA). 2.5 µL of RNase A was added, and the sample was incubated at 37ºC for 15 minutes. Following this digestion, the sample was extracted twice with 150 µL phenol/CIA and once with 500 µL CIA.

A Sephacryl S400 spun column was prepared by placing the top half of a 1.7 mL Eppendorf tube onto a 5 mL polypropylene tube. A Gilson blue micropipettor tip was stuffed with siliconized glass wool and placed onto the half-Eppendorf tube. The blue tip was then filled with Sephacryl S400 high-resolution resin and spun at 2,000 rpm (Heraeus Megafuge 1.0R, #7570G rotor) at room temperature for 5 minutes. The sample was then loaded onto the Sephacryl S400 spun column and the apparatus was spun at

2,000 rpm (Heraeus Megafuge 1.0R, #7570G rotor) at room temperature for 5 minutes. The plasmid, eluted in TE (10, 0.1), was collected from the bottom polypropylene tube.

2.2.13 UV Quantitation of DNA

UV absorbance was used to determine the concentration of purified plasmids. To ensure accurate quantitation, the plasmid solution was diluted such that its A_{260} fell between 0.1 and 1.0 (in most instances, a 1:100 dilution). The Cary 50 Bio UV-Visible Spectrophotometer was blanked with TE (10, 0.1). The sample's absorbance spectrum was recorded from 320-220 nm. The absorption values at 320 nm, 280 nm, and 260 nm were obtained. The adjusted A_{260} (A) was calculated by subtracting A_{320} from A_{260}. The concentration was then calculated using Beer's law ($A = \varepsilon cl$), where the extinction coefficient (ε) for DNA was 20 $(mg/mL)^{-1}cm^{-1}$ and the path length (l) was 1 cm.

2.2.14 Restriction Mapping of Prepared Plasmids

Restriction mapping was used to determine if the desired plasmid was successfully subcloned. Restriction enzymes were selected such that their digestion patterns for parent and desired plasmids could be distinguished via agarose gel electrophoresis. Typically, selected enzymes cut the insert once and the vector once. Reaction mixes containing 2 µL of plasmid DNA at ~0.2 ug/uL,1 µL of appropriate 10x New England BioLabs buffer, 1 µL of 1 mg/mL BSA, 0.5 µL of 100 mM DTT, 0.5-1 µL of the restriction enzyme(s) at 10-20 units/µL, and MilliQ water to a final combined volume of 10 µL. The reaction mix was then incubated at 37ºC for 1-2 hours, combined with 2 µL 6x GLB, and analyzed via electrophoresis on an appropriate-concentration agarose gel. For plasmids created from

PCR-amplified genes and PCR mutagenesis products, restriction map-positive samples were then sequenced.

2.2.15 DNA Sequencing

DNA samples were sequenced courtesy of the Pennsylvania State University Nucleic Acid Facility. 5 µL of each sample, concentrated or diluted to 0.2-0.3 µg/µL, were submitted with 5 µL of 1 µM sequencing primer for each sequencing reaction. The resulting electropherogram and computer interpretation of the sequence were compared to the expected plasmid sequence and analyzed for mutations.

2.2.16 Site-Directed PCR Mutagenesis and DpnI Selection

The Tan Lab procedure for site-directed mutagenesis is largely based on the QuikChange method available from Stratagene. This mutagenesis protocol was employed to make site-specific base pair mutations in plasmids in order to remove internal restriction sites that interfered with subcloning. Two complimentary oligonucleotide primers were used to introduce mutation(s) using linear PCR amplification. 0.5 µL of 10 ng/µL template plasmid DNA, 2.5 µL of 10x Pfu polymerase buffer, 2.5 µL of 2.5 mM dNTP, 0.7 µL of 10 µM forward and reverse mutagenesis primers, 0.4 µL of 2.5 units/µL PfuTurbo polymerase, and water to a final combined volume of 25 µL. This reaction mixture was inserted into the thermocycler and incubated at 95ºC for 2 minutes to ensure full denaturation. The sample was then subjected to 12 cycles of 2 minutes at 95ºC, 30 seconds at 55ºC, and 68ºC for an appropriate extension time. Typically, the extension time was 1 minute for every 1,000 bp in the plasmid's length (e.g. 3 minutes seconds for

a 3,000 bp plasmid). Following these cycles, the sample was held at 75ºC for 3 minutes to ensure DNA renaturation. The extension step was performed at 68ºC to avoid strand displacement by PfuTurbo. Following thermal cycling, 2 µL of reaction mix was removed for later use as a DpnI digestion negative control.

0.5 µL of 20 units/µL DpnI was added to the remainder of the reaction mix and incubated at 37C for 1 hour. DpnI is a Type IIM restriction endonuclease that cleaves methylated DNA. In this instance, it is used to digest wild-type plasmid DNA, rendering it incapable of being transformed. 2 µL of the DpnI-digested reaction mix, as well as 2 µL of the DpnI-undigested negative control were transformed into aliquots of competent TG1 cells. Reaction mix plates containing roughly half the colonies of the DpnI-undigested control indicated successful mutagenesis and DpnI digestion. Colonies from the DpnI-digested plate were PCR screened as outlined above. However, addition of the restriction enzyme whose site was removed will help identify mutated clones. Prior to electrophoresis, 0.5 µL of 10-20 units/µL restriction enzyme was added to each PCR tube and incubated at 37ºC for 1 hour. The samples were then analyzed via agarose gel electrophoresis. Successful mutagenesis was indicated by lack of digestion of the PCR product by the restriction enzyme. Two mutation-containing clones were then used in subsequent plasmid preparations. The sequences of the resultant plasmids were then checked by restriction mapping and sequencing.

2.3 Protein Methods

2.3.1 SDS-PAGE Electrophoresis

Sodium dodecyl sulfate polyacrylamide gel electrophoresis, or SDS-PAGE electrophoresis, was used to separate proteins by molecular weight and analyze samples from expression and purification steps.

18% (w/v) SDS-PAGE gels were prepared by stacking 10-15 sets of glass gel plates into a Mini-Protean II gel-pouring block (BioRad, Richmond, CA). Separating mix was prepared by combining 8 mL of water, 36 mL of 30% acrylamide/0.5% bis-acrylamide, 120 uL of 1% bromophenol blue in ethanol (w/v), 15 mL of 3 M Tris-Cl pH 8.8. The mixture was then deaerated and combined with 600 uL 10% SDS (w/v), 60 uL of tetramethylethyldiamine (TEMED), and 240 uL of 25% AMPS and injected into the gel-pouring block until the glass plates were 70% full. The mixture was covered with water-saturated butanol and allowed to polymerize at room temperature. After the mixture was fully solidified, the water-saturated butanol was removed by washing with ethanol and water. Stacking mix was prepared by combining 5 mL water, 10 mL of 10% acrylamide/0.5% acrylamide, and 4.8 mL of 0.5 M Bis-Tris. The stacking mix was then deaerated and combined with 200 uL 10% SDS (w/v), 15 uL of TEMED, and 80 uL of 25% AMPS and added to the top of the gel-pouring block, followed by the placement of well combs. The gels were allowed to polymerize at room temperature and then were carefully removed from the gel-pouring block. The 10-15 resulting gels were stored at 4C in a damp paper towel for up to 2 weeks.

SDS-PAGE gels were run by assembling a Mini-Protean II electrophoresis apparatus filled with protein gel running buffer, or PGRB (10 mM Tris, 76 mM glycine, 0.02% SDS (w/v)). The gel was loaded with samples that had been suspended in protein gel loading buffer, or PGLB (125 mM Bis-Tris pH 6.8, 20% glycerol, 4% SDS (w/v), 15% 2-mercaptoethanol (v/v), 0.04% bromophenol blue (w/v)), and boiled at 95ºC for 2 minutes. The gel was then electrophoresed at 10 W until the bromophenol blue die had run off of the gel (~30 minutes).

The gel was then removed from the electrophoresis assembly, transferred to a container filled with FIX solution (45% ethanol, 9% acetic acid), and incubated on a rocking platform for 5 minutes. The FIX solution was removed, and the gel was covered with STAIN solution (0.5% Coomassie Blue R (w/v), 45% ethanol, 9% acetic acid) and incubated on a rocking platform for 5 minutes. Tue STAIN solution was removed, and the gel was covered with DESTAIN solution (7% ethanol, 5% acetic acid) and incubated at 60C until the purple background staining of the gel was removed. The DESTAIN solution was replaced with water, and the gel was incubated at room temperature for several hours prior to drying in a cellulose membrane.

2.3.2 100 mL Small Scale Expression

All expression and coexpression vectors used in these experiments were developed by Song Tan and were based on the T7 expression system (Tan et al, 2005). In order to determine optimal parameters for soluble expression, proteins were expressed in 100 mL quantities at 37ºC and 28ºC for times ranging from 0-6 hours and 23ºC and 18ºC for

times ranging from 0-24 hours. Generally, higher-temperature expressions yield more expression, whereas lower-temperature expressions show increased solubility of expressed proteins.

The appropriate T7-based expression or coexpression plasmid was transformed into BL21(DE3)pLysS or BL21-CodonPlus(DE3)-RIL competent cells as described in Section 2.2.10 above. Following plate incubation, 100 mL of 2x TY media containing appropriate antibiotics (usually 50 μg/mL ampicillin and 25 μg/mL chloramphenicol) was inoculated with 3-5 colonies from the transformation plate. For 28ºC, 23ºC, and 18ºC expressions, the culture was then incubated in a shaking incubator at 37ºC and 200 rpm until the culture's OD_{600} was 0.05-0.15, at which point it was transferred to a shaking incubator at 28ºC, 23ºC, or 18ºC, respectively, and 200 rpm. For 37ºC expressions, no change in temperature was necessary. The culture was grown at the appropriate temperature until its OD_{600} measured 0.4-0.7, at which point 500 μL of uninduced culture was removed and the culture was induced by the addition of 100 μL of 0.2 M IPTG. The uninduced sample ("0 hour time point") was placed in an Eppendorf tube, spun at full speed in a microcentrifuge, the supernatant was aspirated and discarded, and the pellet was resuspended in 100 μL of PGLB.

Time points were collected hourly for up to 6 hours for 37ºC and 28ºC expressions and every two hours up to 24 hours for 23ºC and 18ºC expressions. This was accomplished by removing 250 μL of cell culture, placing it in an Eppendorf tube, spinning it at full speed in a microcentrifuge, discarding the supernatant via aspiration, and resuspending

the cell pellet in 100 µL of PGLB. All time points were analyzed via SDS-PAGE electrophoresis (as described in Section 2.3.1 above) to identify protein expression levels at various times.

For 37ºC and 28ºC expressions, after 3 hours of incubation post-induction, 50 mL of culture was removed and spun at 4,000 rpm (Heraeus Megafuge 1.0R, #7570G rotor) at room temperature for 10 minutes. The supernatant was discarded, and the cell pellet was resuspended in 10 mL of P300-EDTA (50 mM NaPO4 pH 7.0, 300 mM NaCl, 1 mM benzamidine, 5 mM 2-mercaptoethanol). For 23ºC and 18ºC expressions, a sample was similarly prepared at 12 hours post-induction. The cell resuspension was then flash frozen with liquid nitrogen and stored at -20ºC. This sample was later used to determine the solubility of the expressed protein.

2.3.3 Determination of Solubility from Small Scale Expression

To determine the solubility of the expressed protein, resuspended cells harvested at the 3 hour time point of the 100 mL small-scale expression were thawed in lukewarm water. The cells were then sonicated for 10 seconds at 50% duty cycle, 30% maximum power, incubated on ice for 30 seconds, and sonicated again. 25 µL of the sonicated whole-cell extract, or "WCE," was removed and mixed with 25 µL of PGLB. 500 µL of the WCE was removed and spun at full speed for 5 minutes in a microcentrifuge. 25 µL of the supernatant was combined with 25 µL of PGLB. The remaining supernatant was aspirated away and the pellet was resuspended in 500 µL of P300-EDTA. 25 µL of the pellet resuspension was removed and combined with 25 µL of PGLB. The WCE,

supernatant, and pellet samples were then analyzed via SDS-PAGE electrophoresis. Solubly expressed protein appeared in the WCE and supernatant, while insolubly expressed protein appeared in the WCE and pellet. Extreme amounts of cellular proteins in the pellet indicated poor cell lysis.

2.3.4 Small Scale Talon® Metal Affinity Batch Purification

Hexahistidine-tagged proteins can be purified from contaminating *Escherichia coli* proteins by exploiting their affinity for cobalt-based metal affinity resins such as Talon® resin (Clontech, Mountain View, CA). For hexahistidine-tagged proteins and complexes, 5 mL of cell extract (as prepared in Section 2.3.3) was used for small scale purification.

To prepare the Talon® resin, 1 mL of resin was washed with 10 mL of water and then equilibrated with 10 mL of P300-EDTA. The cell extract was spun at full speed for 5 minutes in a microcentrifuge. The washed resin and the supernatant were combined and incubated at room temperature for 1 hour on a rotating stand. The mixture was then spun at 4,000 rpm (Heraeus Megafuge 1.0R, #7570G rotor) at room temperature for 5 minutes. The flow-through was separated from the resin and collected, collecting a 25 µL sample to combine with 25 µL of PGLB. The resin was washed twice with 10 mL of P300-EDTA and was transferred to a BioRad BioSpun column. The protein or complex was then eluted using P300-EDTA + 100 mM imidazole, collecting the solution 6 500 µL fractions. A 25 µL sample of each fraction was collected and mixed with 25 µL of PGLB. The WCE, supernatant, pellet, flow-through, and fraction samples were analyzed using

SDS-PAGE electrophoresis. Fractions containing the desired protein(s) were stored in 20% glycerol, flash frozen, and kept at -20ºC.

2.3.5 Nitrocellulose Membrane Immunoblotting

SDS-PAGE electrophoresis provides a means for visualizing proteins, but in order to identify specific proteins, Western blotting was performed with protein-specific antibodies. An SDS-PAGE gel was electrophoresed as described in Section 2.3.1, but was not fixed and stained. Following electrophoresis, it was incubated in 5 mL of transfer buffer (25 mM Tris, 192 mM glycine, pH 8.3). The gel was then blotted to a Hybond ECL nitrocellulose membrane (Amersham Pharmacia, Piscataway, NJ) in a blotting device. During assembly of the blotting apparatus, components were kept wet with transfer buffer to avoid air bubbles. The blotting apparatus was assembled by carefully layering a fiber pad, a piece of filter paper, the buffer-equilibrated gel, the nitrocellulose membrane, another piece of filter paper, and a second fiber pad into the device's blot chamber. The apparatus was positioned in such a way that the gel was closer to the negative electrode and the nitrocellulose membrane was closer to the positive electrode. The blot chamber was loaded into the device with an ice pack and filled with methanol-containing transfer buffer (25 mM Tris, 192 mM glycine, pH 8.3, 20% methanol) and electrophoresed at 100 V for 90 minutes with constant stirring.

After blotting, the membrane was removed from the blotting assembly and incubated in TBS (25 mM Tris, 0.15 M NaCl, pH 8.0) for 5 minutes, and then preincubation buffer (TBS, 2% nonfat dry milk (w/v)) for 30 minutes on a rocking platform. The addition of

nonfat dry milk prevented nonspecific binding of proteins to the nitrocellulose membrane. The membrane was washed twice with TTBS (25 mM Tris, 0.15 M NaCl, 0.05% Tween 20, pH 8.0) and then incubated with the primary antibody at an appropriate dilution for 1 hour on a rocking platform. The membrane was washed three times with TTBS and then incubated with the secondary antibody at an appropriate dilution for 1 hour on a rocking platform. The membrane was washed three times with TTBS and incubated with a 1:1 mixture of ECL detection solutions 1 and 2. Excess ECL detection solution mixture was removed and the membrane was covered in clear plastic wrap for film exposure.

2.3.6 Large Scale Expression

Once a protein's expression had been optimized by comparing the results of different small scale expressions (as described in Section 2.3.2), a 6-12 L large scale expression was conducted. The appropriate T7-based expression or coexpression plasmid was transformed into BL21(DE3)pLysS or BL21-CodonPlus(DE3)-RIL competent cells as described in Section 2.2.10 above. Following plate incubation, a 100 mL starter culture of 2x TY media containing appropriate antibiotics (usually 50 µg/mL ampicillin and 25 µg/mL chloramphenicol) was inoculated with 3-5 colonies from the transformation plate. The culture was incubated in a shaking incubator at 200 rpm at 37ºC until it reached and OD_{600} of 0.1-0.2, at which point 12-24 large flasks containing 500 mL of 2x TY media and the same antibiotics were inoculated with 3-5 mL of the starter culture. The large flasks were incubated in a shaking incubator at 200 rpm at 37ºC until their OD_{600} measurements reached 0.1-0.2, at which point the temperature was lowered to the

appropriate temperature of expression or kept the same if the optimum temperature was 37C. The cells were incubated until reaching an OD_{600} of 0.4-0.6. At that point, a 500 uL sample was collected and prepared as a 0 hour time point as described in Section 2.3.2. The flasks were then inoculated with 0.5 mL of 0.2 M IPTG. At the appropriate time, a 250 uL sample from each flask was collected, spun at full speed in a microcentrifuge, and its pellet resuspended in 100 uL of PGLB. The cells were then spun at 7,000 rpm (Sorvall SLA-3000 rotor) at room temperature for 5 minutes. The resulting cell pellet was then resuspended in 150-300 mL of P300-EDTA, and the resuspension was flash frozen in liquid nitrogen and stored at -20ºC. The 0 hour time point and the flask samples were then analyzed by SDS-PAGE electrophoresis.

2.3.7 Large Scale Extract Preparation

Cells harvested from the large scale expression were allowed to thaw in a lukewarm water bath and were sonicated three times for 10 seconds at 70% maximum power, 50% duty cycle with incubation on ice for 30 seconds between sonication steps. A 25 µL sample of the sonicated extract (WCE) was mixed with 25 µL of PGLB. The remaining extract was spun at 18,000 rpm (Sorvall SS-34 rotor) at 4ºC for 20 minutes. The supernatant was transferred to a new flask and the pellet was resuspended in its original volume. 25 µL samples of the pellet and supernatant were mixed with 25 µL of PGLB. The WCE, pellet, and supernatant samples were then analyzed by SDS-PAGE electrophoresis.

2.3.8 Metal Affinity Column Purification

Large scale extracts of the MWRA complex containing a hexahistidine affinity tag were first purified over a 20 mL low-pressure Talon® resin (Clontech) chromatography column at 4ºC. The column was equilibrated in P300-EDTA prior to sample loading. The large scale extract (~200 mL) was loaded onto the column at 3 mL/min, and the flow-through was collected. The column was then washed with 60 mL of P300-EDTA at 3 mL/min, 40 mL of P300-EDTA + 10 mM imidazole at 3 mL/min, and 60 mL of P300-EDTA + 15 mM imidazole at 3 mL/min. The complex were then eluted using P300-EDTA + 100 mM imidazole at 3 mL/min. 5 mL fractions were collected and analyzed for MWRA complex content by SDS-PAGE electrophoresis. Fractions containing the greatest amount of complex and the least amount of contaminants were pooled. The pooled sample was then quantitated via UV absorption.

2.3.9 UV Quantitation of Protein

UV absorbance was used to determine the concentration of purified protein. The Cary 50 Bio UV-Visible Spectrophotometer was blanked with P300-EDTA. The sample's absorbance spectrum was recorded from 320-220 nm. The absorption values at 320 nm, 280 nm, and 260 nm were obtained. The adjusted A_{280} (A) was calculated by subtracting A_{320} from A_{280}. The concentration was then calculated using Beer's law ($A = \varepsilon cl$), where the extinction coefficient (ε) was based on the sample being quantitated (see Table 2.3 below) and the path length (l) was 1 cm.

Table 2.3: Extinction Coefficients of MLL WMRA Subcomplex Versions

WMRA Version	Extinction Coefficient
Contains STRHIS-tagged hWDR5	1.319 $(mg/mL)^{-1}cm^{-1}$
Contains untagged hWDR5	1.307 $(mg/mL)^{-1}cm^{-1}$

2.3.10 Protease Cleavage of Purification Tags

Purification tags such as the hexahistidine tag can interfere with catalytic activity and should be removed after successful purification. Tobacco etch virus Nuclear Inclusion a (TEV NIa) protease recognition sites were engineered into expression plasmids containing purification tags to allow for cleavage. A 1:100 substrate to protease ratio was used to ensure tag cleavage. The reaction was incubated overnight to further ensure complete cleavage.

2.3.11 Source™ S Cation-Exchange HPLC Purification

Following initial purification via Talon® metal affinity resin, additional purification of the Talon®-purified pool was achieved by high pressure liquid chromatography (HPLC) over a cation exchange column, which consisted of 10 mL of Source™ S resin (Amersham Pharmacia) at room temperature. The column was first equilibrated using H20 (10 mM HEPES pH 7.5, 20 mM NaCl). Elution consisted of a linear gradient over 20 column volumes (200 mL), from 20 mM NaCl to 500 mM NaCl. 6 mL fractions were collected. The conductivity and A_{280} were recorded throughout. 25 µL samples of select fractions, which corresponded to chromatogram UV absorbance peaks, were mixed with 25 µL of PGLB and analyzed via SDS-PAGE electrophoresis. Fractions containing the greatest amount of MWRA complex and the least amount of contaminants were pooled. This HPLC procedure was repeated until the entire Talon®-purified pool was loaded. The

selected fractions from the Source™ S HPLC runs were pooled together and quantitated by UV absorbance. The Source™ S fraction pool was dialyzed 3 times against 1 L of H200 (10 mM HEPES pH 7.5, 200 mM NaCl, 10 mM 2-mercaptoethanol).

2.3.12 Protein Concentration

MWRA complex samples were concentrated using Amicon Ultra (Millipore, Billerica, MA) centrifugation units with a 7 kD molecular weight cutoff membrane. The membrane was washed with 5 mL of water and then 5 mL of H200. The complex was concentrated to 11.36 mg/mL. The sample was then tested for protein aggregation by dynamic light scattering.

2.3.13 Dynamic Light Scattering

Dynamic light scattering (DLS) was employed to identify protein aggregation by measuring polydispersity and hydrodynamic size. A Protein Solution Dynapro™ (Wyatt Technology, Santa Barbara, CA) system was used. The system was blanked using MilliQ water and H200 before measurement of the protein sample. All DLS cuvettes were cleaned with water, ethanol, and compressed air between measurements.

2.4 Enzymology

2.4.1 Methyltransferase Activity Assay

A qualitative methyltransferase assay was used to determine the catalytic activity of the purified MWRA complex. It was adapted from the methyltransferase assays employed by the Cosgrove Laboratory at Syracuse University. (Patel et al, 2009) Purified MLL

MWRA complex was incubated for 8 hours at 15ºC under Cosgrove Laboratory methyltransferase conditions (50 mM Tris pH 8.5, 200 mM NaCl, 3 mM DTT, 5 mM $MgCl_2$, 5% glycerol, 500 µM S-adenosyl methionine, 250 µM *Xenopus* histone H3). 20 µL samples from each reaction were combined with 20 µL of PGLB and electrophoresed on an SDS-PAGE gel. The gel was then blotted to a nitrocellulose membrane as described in Section 2.3.5. The membrane was probed with rabbit anti-H3K4me2 (Active Motif, Carlsbad, CA) as a primary antibody and donkey HRP-linked anti-rabbit IgG (Pharmacia) as a secondary antibody.

2.4.2 Single Turnover Kinetics Assay

Single turnover kinetics assays were used to determine the substrate specificity and activity of purified MLL MWRA complex. The assay was performed by Dr. Michael Cosgrove and Dr. Anamika Patel at Syracuse University. 2 µg of MLL MWRA complex was incubated at 15ºC with 250 µM residues 1-20 of histone H3 and ^{3}H-radiolabeled S-adenosyl methionine (AdoMet) under Cosgrove Laboratory methyltransferase conditions (50 mM Tris pH 8.5, 200 mM NaCl, 3 mM DTT, 5 mM $MgCl_2$, 5% glycerol). Samples taken from the reaction were analyzed via MALDI-TOF mass spectrometry. Resultant data was plotted to create reaction progress curves tracing the relative amounts of unmethylated, monomethylated, dimethylated, and trimethylated H3K4.

Chapter 3: Results

3.1 Preparation of Polycistronic Vector

3.1.1 Single Polycistronic Vector

The original cloning strategy was to subclone the hMLL1Δ1, hWDR5, hRbBP5, and hAsh2L genes into a pST69, a polycistronic expression vector, as illustrated in Figure 1.12. Plasmids containing the four genes were obtained from Dr. Michael Cosgrove of the Department of Biology at Syracuse University. Internal restriction sites in the hAsh2L and hWDR5 genes that interfered with the planned subcloning steps were removed via PCR-based mutagenesis. The genes were then amplified via PCR using oligonucleotide primers which incorporated the restriction sites necessary for later cloning steps. Gene inserts were subcloned into transfer plasmids based on the pST44 system developed by Song Tan. A Strept peptide/hexahistidine fusion affinity tag (Tan et al, 2005) was added to the N-terminus of hWDR5 for future use in affinity purification. The genes were then cloned into the pST69 polycistronic expression vector (Song Tan, personal communication). All constructed plasmids were verified by restriction mapping and, where appropriate, by DNA sequencing.

Subcloning proceeded as planned until the insertion of the hAsh2L gene into the pST69 vector using RsrII and SgrAI restriction sites. Multiple attempts at this subcloning step yielded no viable colonies growing on the transformation plate. At this point, a different subcloning scheme was developed as outlined in Section 3.1.2. Eventually, improvements in how DNA fragments were purified by gel electrophoresis resulted in the construction

of the desired single plasmid to coexpress the 4 subunits of the MWRA complex. My fellow Schreyer Honors student Slater Nair assisted in the creation of this vector.

3.1.2 Dual-Plasmid Polycistronic System

Difficulties incorporating the hAsh2L subunit into the pST69 polycistronic plasmid spurred the development of an alternative expression system. A dual-plasmid system was devised using ampicillin- and kanamycin-resistant plasmids. This provided selection criteria to verify the successful transformation of both plasmids. The pST69 construct containing hMLL1Δ1, hWDR5, and hRbBP5 and a pST73 construct (Song Tan, personal communication) containing hAsh2L were prepared as outlined in Figure 1.13. The creation of the pST69 construct follows the initial subcloning scheme, but bypasses the insertion of the hAsh2L gene into the vector using RsrII and SgrAI restriction sites.

In order to express all four subunits of the MWRA complex, both plasmids would need to be sequentially transformed into competent cells. One of the two constructs was transformed into a competent cell aliquot and grown on a TYE + appropriate antibiotics. These transformed colonies were then used to prepare a line of plasmid-containing competent cells, into which the second construct was then transformed. Resultant colonies would only proliferate in the presence of ampicillin and kanamycin if they successfully incorporated both plasmids.

The gene sequences and oligonucleotide primer sequences used during the subcloning of these constructs are provided in Chapter 5.

3.2 Optimization of Small Scale MLL MWRA Complex Expression

After the creation of the pST69 ("single polycistronic") vector and the pST69/pST73 ("dual-plasmid polycistronic") expression plasmids were created, the coexpression of hMLL1Δ1, hWDR5, hRbBP5, and Ash2L subunits was tested using both schemes, at various temperatures, and using different bacteriological strains. A summary of these subunits can be found in Table 3.1 below.

Table 3.1: Subunits composing the MLL MWRA complex

Name	Identity	MW	pI
hMLL1Δ1	hMLL residues 3769-3969	26,214	8.82
hWDR5	untagged hWDR5	36,588	8.40
STRHISNhWDR5	hWDR5 with cleavable N-terminal Strept peptide and hexahistidine tags	40,252	8.27
hRbBP5	hRbBP5	59,152	4.71
hAsh2L	hAsh2L	60,110	7.04

Small-scale (100 mL) coexpression of the MLL MWRA complex was first attempted in BL21(DE3)pLysS cells at 37ºC and 28ºC using both the single polycistronic and dual-plasmid expression schemes. Very low expression levels were observed at both temperatures. However, it was determined that single-plasmid polycistronic expression was more efficient than the dual-plasmid system (Figure 3.1). It was also determined that the complex's solubility increased with lower expression temperature (Figure 3.2).

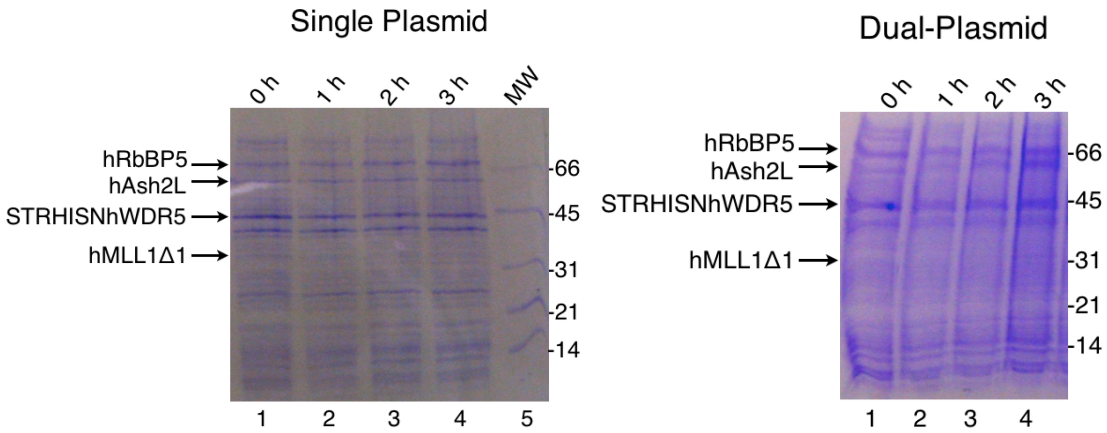

Figure 3.1: SDS-PAGE analysis of small-scale coexpression of the MLL MWRA complex (single plasmid expression and dual-plasmid expression) in BL21(DE3)pLysS cells at 37ºC

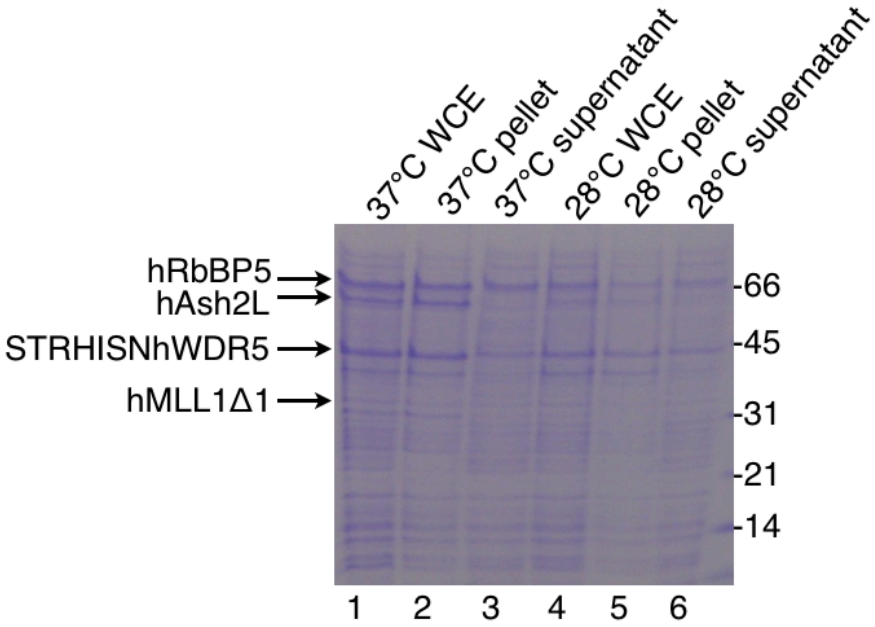

Figure 3.2: SDS-PAGE analysis of small-scale solubility test of the MLL MWRA complex (single plasmid expression) in BL21(DE3)pLysS cells at 37ºC and 28ºC

In order to select for cells BL21(DE3)pLysS cells containing both expression constructs, it was necessary to use media containing kanamycin, ampicillin, and chloramphenicol. These cells grew substantially slower than those containing the 4-subunit pST69 construct did in the presence of ampicillin and chloramphenicol. Dual-plasmid cells only

expressed approximately half of the protein that single-plasmid polycistronic cells did at 37ºC.

The low expression level of the proteins led to use of BL21-CodonPlus(DE3)-RIL cell, which contain extra copies of the *argU*, *ileY*, and *lewW* tRNA genes for coexpression of the complex. It is well documented that recombinant proteins tend to express in larger quantities at higher temperatures, but tend to express more solubly at lower temperatures. (Schein and Noteborn, 1988; Tolia and Joshua-Tor, 2006) The coexpression of the MLL MWRA complex was tested at 37ºC, 28ºC, 23ºC, and 18ºC. At all four temperatures, the complex was expressed in moderate quantities, with each of the subunits being expressed at an approximately 1:1:1:1 stoichiometric ratio (Figure 3.3).

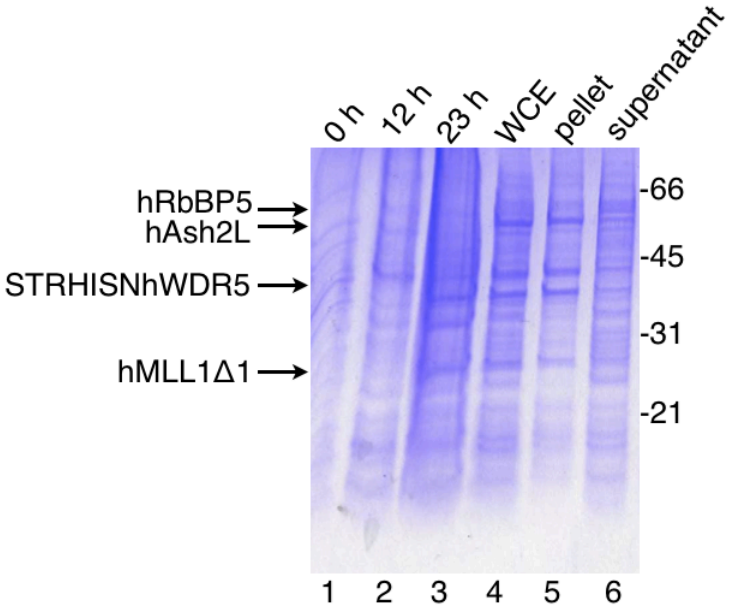

Figure 3.3: SDS-PAGE analysis of small-scale coexpression and solubility test of the MLL MWRA complex in BL21-CodonPlus(DE3)-RIL cells at 18ºC

A crude protein extract was prepared from these cells, and the complex's solubility was analyzed. Increased solubility was observed at lower expression temperatures, with

approximately 60% solubility when expressed at 18ºC and harvested after 12 hours. The complex was then purified via small-scale Talon® metal-affinity batch purification (Figure 3.4).

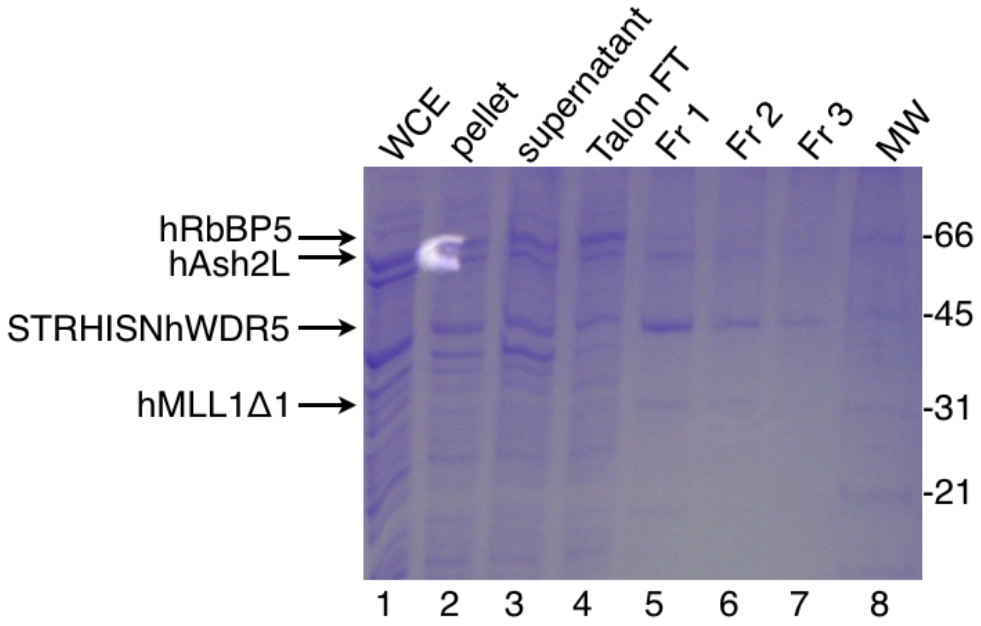

Figure 3.4: SDS-PAGE analysis of small-scale Talon® metal-affinity batch purification of the MLL MWRA complex

In order to further confirm the identity of the band corresponding to STRHISNhWDR5, a Western blot was performed using anti-hexahistidine primary antibodies (Figure 3.5). The Western blot confirmed that hWDR5 was being expressed, was migrating as expected, and was minimally lost (5-10%) in the flow-through during Talon® metal-affinity batch purification.

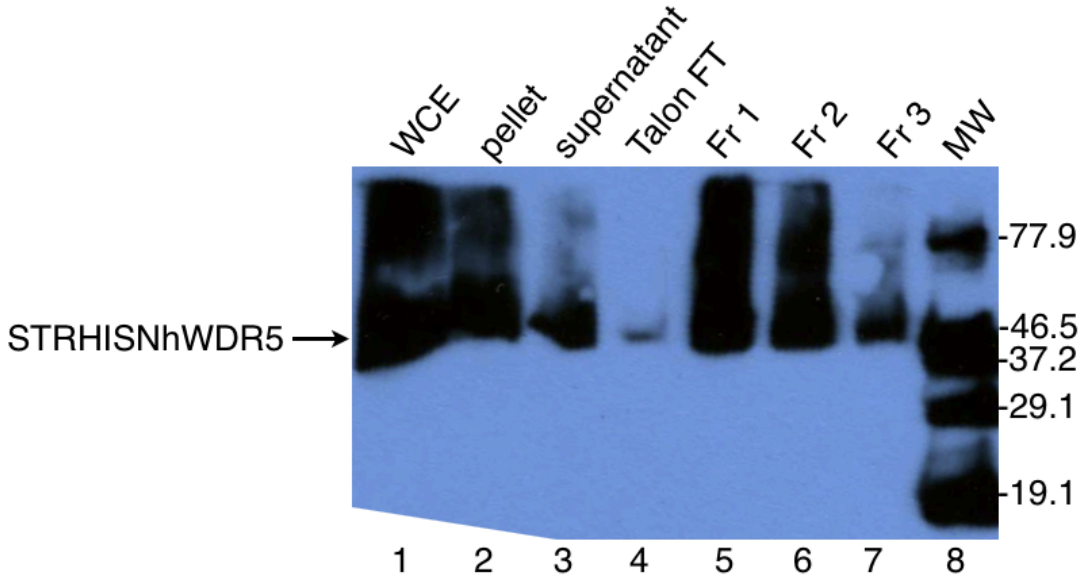

Figure 3.5: Western blot of small-scale Talon® metal-affinity batch purification of the MLL MWRA complex using anti-hexahistidine antibodies

It was determined that the MLL MWRA complex optimally expresses in BL21-CodonPlus(DE3)-RIL cells at 18ºC harvested 12 hours after induction.

3.3 Large Scale Expression and Purification of the MLL MWRA Complex

After optimal expression conditions were identified for the MLL MWRA complex, the cells were expressed on a larger scale. 6 L of cells were grown at 18C, induced when the OD_{600} measured 0.3-0.4, and were harvested 12 hours after induction. Cells were resuspended in 150 mL of P300-EDTA buffer and flash frozen with liquid nitrogen.

I assisted Allen Minns, a research technician in the Tan Laboratory, in the purification of the MLL MWRA complex. A crude extract was prepared by allowing the flash-frozen cell resuspension to thaw, sonicating, and centrifuging the sonicated extract as outlined in Section 2.3.7. Approximately 150 mL of crude extract was obtained. The 150 mL of

extract were loaded onto a 20 mL Talon® metal affinity column equilibrated in P300-EDTA. Selected fractions were analyzed for MLL MWRA complex content by SDS-PAGE electrophoresis (Figure 3.6). Fractions 27-32 were pooled (Talon® pool).

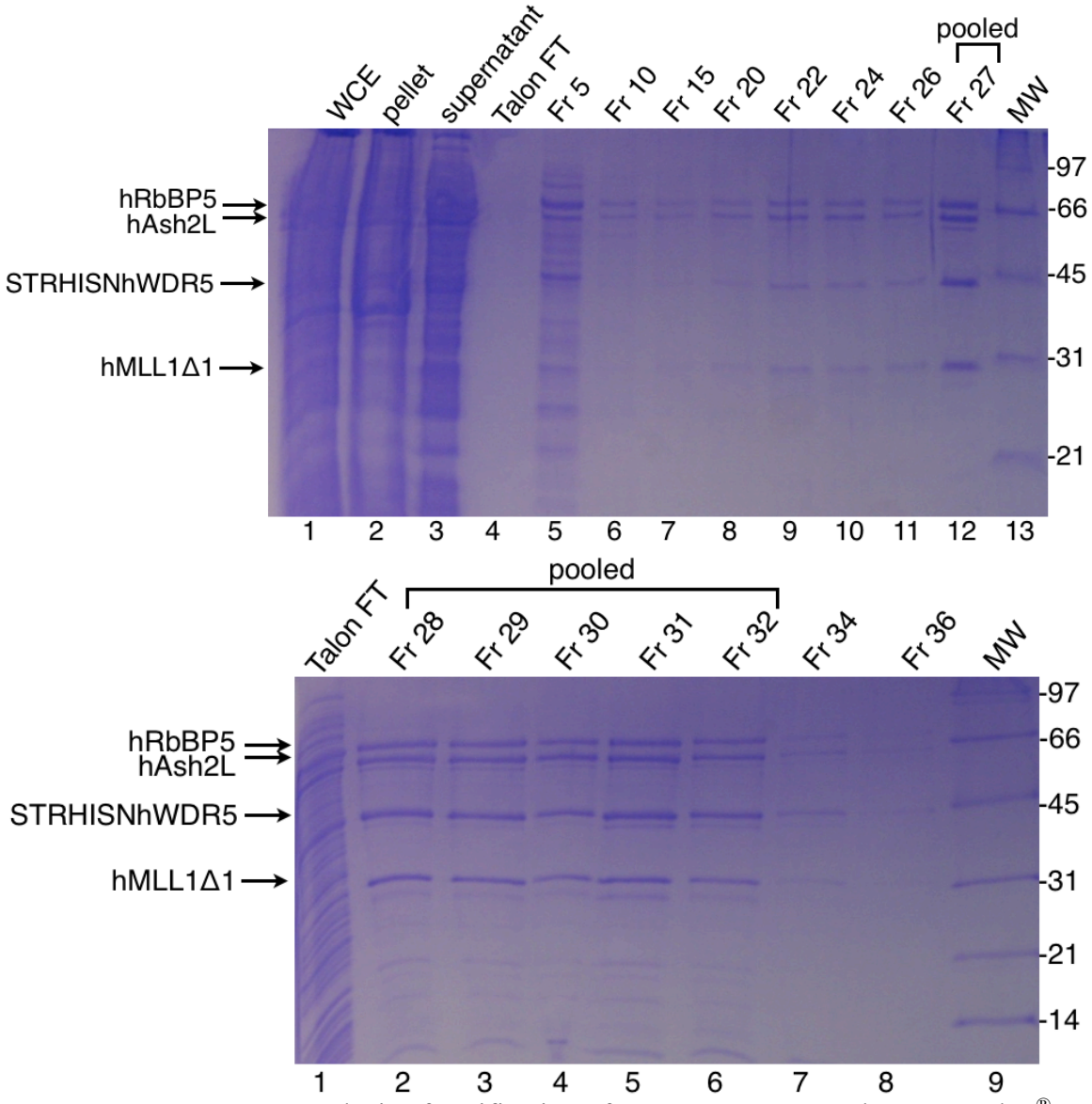

Figure 3.6: SDS-PAGE analysis of purification of MLL MWRA complex over Talon® metal affinity column. The chromatography system's chart recorder was malfunctioning, preventing the generation of a chromatogram.

The Talon® pool was then quantitated by UV absorption. Assuming that all protein in the sample was MLL MWRA containing a Strept peptide/hexahistidine-tagged WDR5 subunit, the Talon® pool contained 0.569 mg/mL (20.5 mg total). This assumption is incorrect; these fractions clearly contain contaminating proteins. Analysis of the Talon® flow-through indicates that approximately 50% of the MLL MWRA complex did not bind the resin and was therefore not purified.

Prior to further purification, the Strept peptide/hexahistidine affinity tag was cleaved from the WDR5 subunit using Tobacco Etch Virus (TEV) NIa protease. 28 µmol of TEV NIa protease was added to the Talon® pool estimated to contain 280 nmol of MWRA complex. The protease digest mixture was incubated at room temperature overnight while dialyzing against 2L of T100-EDTA (20 mM Tris-Cl pH 8.0, 100 mM NaCl, 1 mM benzamidine, 10 mM 2-mercaptoethanol). SDS-PAGE analysis confirmed that the affinity tag was successfully cleaved without affecting the other subunits of the complex (Figure 3.7).

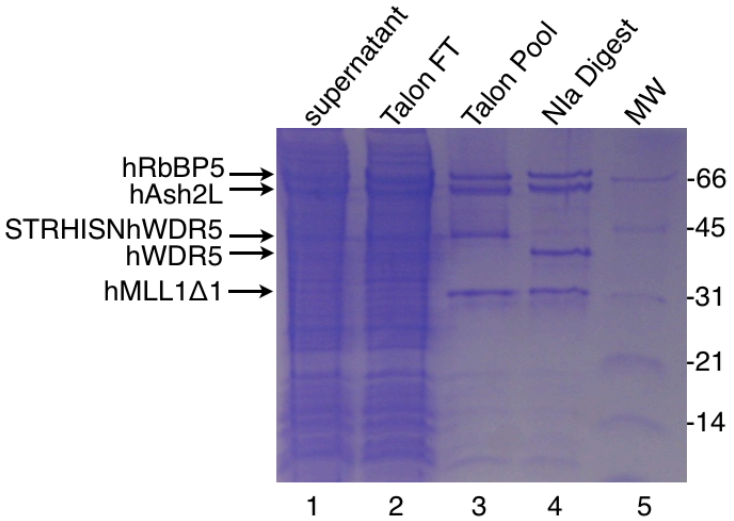

Figure 3.7: SDS-PAGE analysis of TEV NIa cleavage of affinity tag from WDR5 subunit of MLL MWRA complex

Additional purification of the MLL MWRA complex was achieved by cation exchange Source™ S10 chromatography. Two separate preparative runs were performed with an increasing salt gradient from 20 mM to 500 mM NaCl in 10 mM HEPES pH 7.5 (H20 to H500). The chromatograms from both runs were nearly identical (Figure 3.8). As such, SDS-PAGE analysis of fractions was conducted only on fractions from the first run. (Figure 3.9).

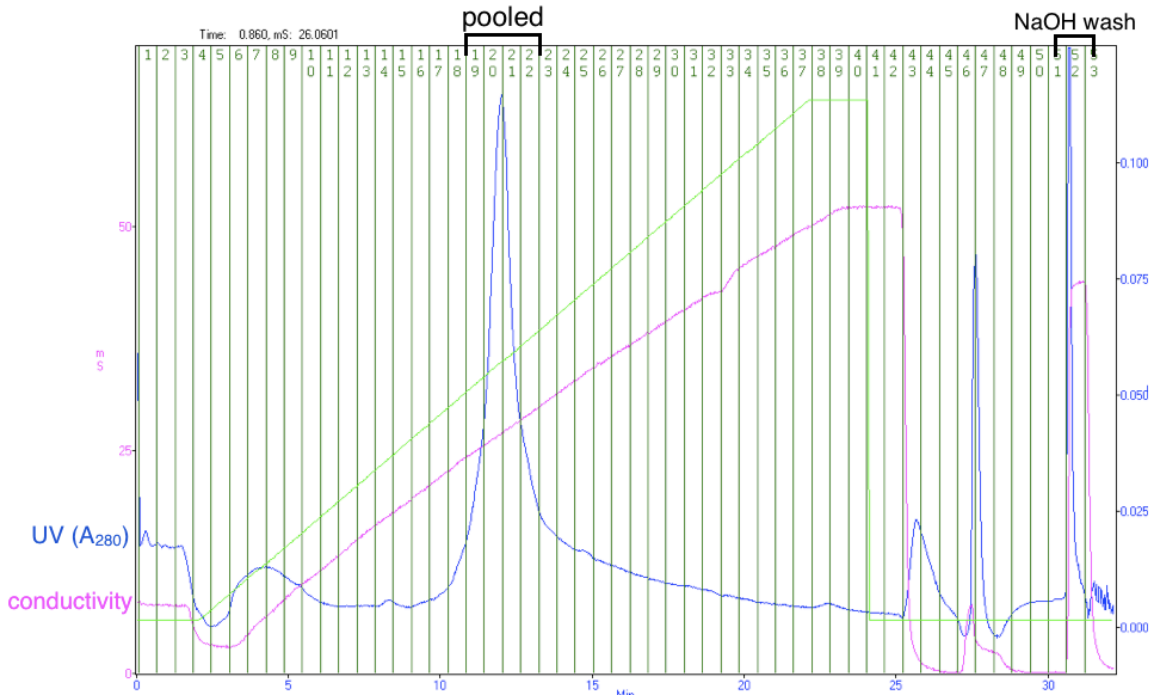

Figure 3.8: Purification of MLL MWRA complex over Source™ S10 column

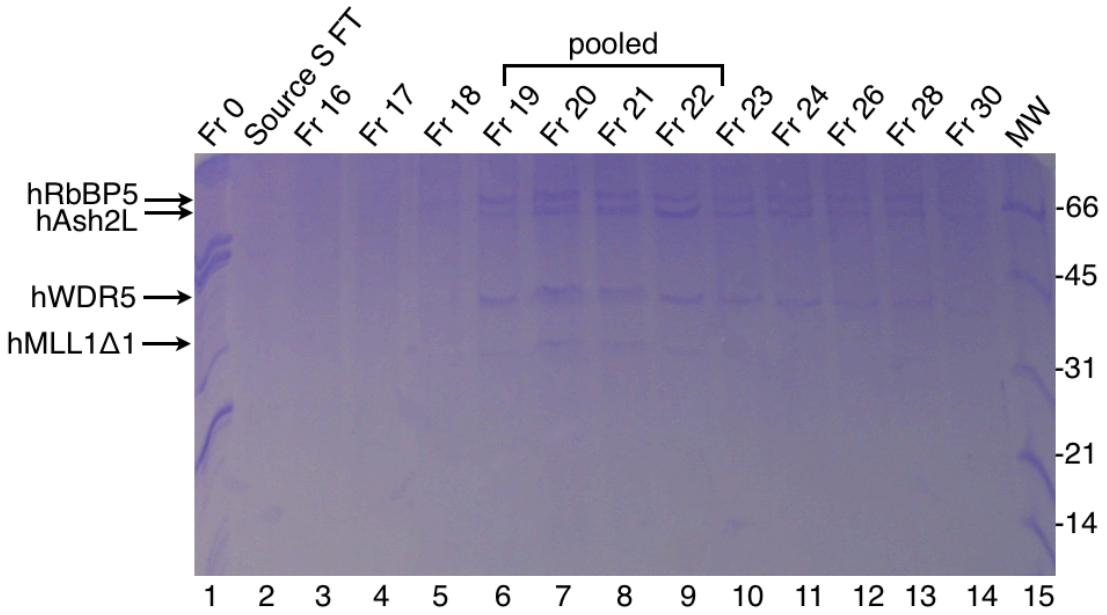

Figure 3.9: SDS-PAGE analysis of purification of MLL MWRA complex over Source™ S10 column

Fractions 19-22 from both runs were pooled as purified MLL MWRA complex, and dialyzed against 1 L of H100-EDTA three times. The pool was then quantitated by UV absorption. Assuming that all protein in the sample was MLL MWRA containing an untagged WDR5 subunit, the pool contained 0.103 mg/mL protein (4.1 mg total). The sample was then concentrated using a 30 kD cutoff centrifugation concentration device and the NaCl concentration of the concentrated pool was adjusted to 200 mM. UV quantitation determined that the pool contained 14.2 mg/mL protein (3.13 mg total). Dynamic light scattering analysis of the sample showed a single monodispersed peak (11.9% polydispersity, 95.2% intensity), indicating that the complex had not aggregated. Glycerol was added to 20%, which brought the final concentration to 11.36 mg/mL. Dilutions of the purified, concentrated MLL MWRA sample were analyzed by SDS-PAGE (Figure 3.10) A dilution of the concentrated complex was also compared to purified hMLL1Δ1, hWDR5, hRbBP5, and Ash2L provided by Dr. Michael Cosgrove at

Syracuse University via SDS-PAGE (Figure 3.11). While the MLL MWRA complex was not completely pure, it was believed to be pure enough to be used in enzymatic reactions.

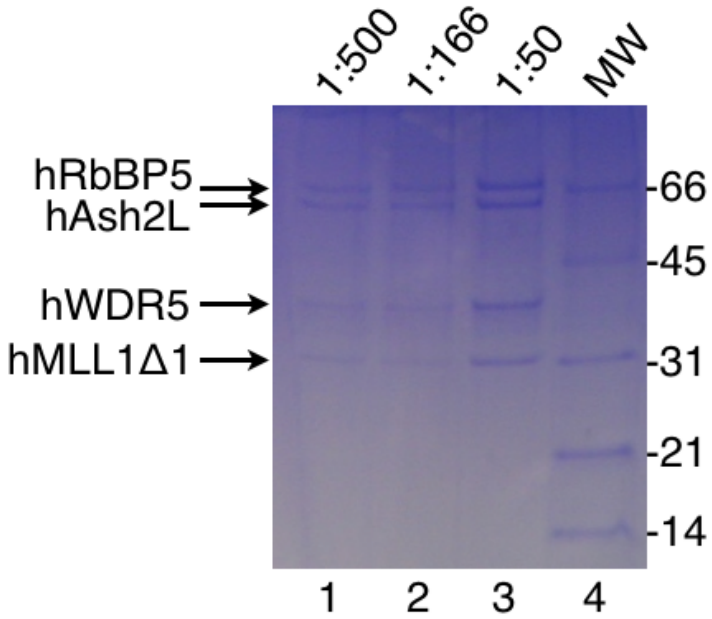

Figure 3.10: SDS-PAGE analysis of dilutions of concentrated MLL MWRA complex

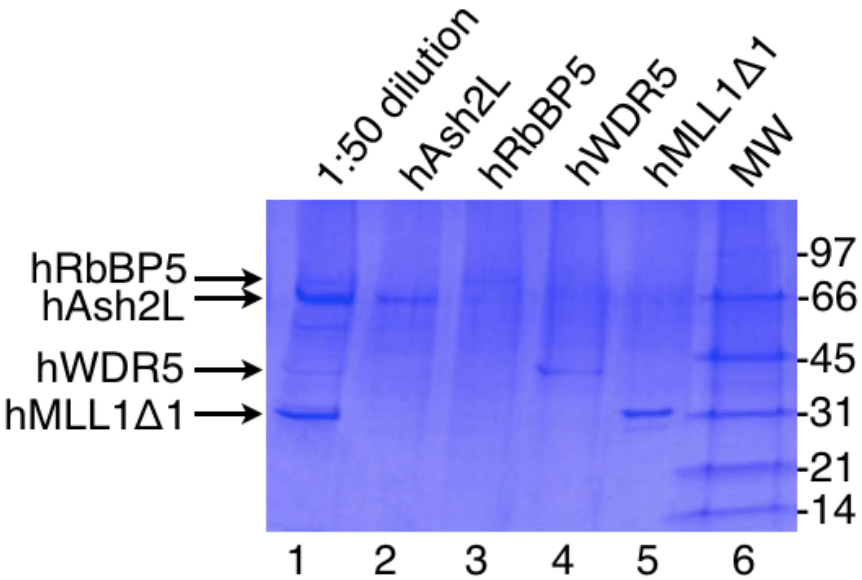

Figure 3.11: SDS-PAGE comparison of Talon® pool to purified subunits provided by Michael Cosgrove

3.4 Methyltransferase Activity of the MLL MWRA Complex

The purified MLL MWRA complex was then analyzed for its methyltransferase activity. It was expected that, if properly prepared, the complex would be capable of dimethylating lysine 4 of histone H3. Dr. Michael Cosgrove and Dr. Anamika Patel at Syracuse University compared the enzymatic activity of my bacterially coexpressed complex with their complex reconstituted *in vitro*. The complexes were incubated at 15ºC with a peptide corresponding to residues 1-20 of histone H3 and S-adenosyl methionine. MALDI-TOF mass spectrometry analysis of the reaction over time (Patel et al, 2009) indicated that the polycistronically expressed, purified MLL MWRA complex exhibits the same dimethylase activity and specificity as MWRA that was reconstituted *in vitro* by the Cosgrove Laboratory at Syracuse University. (Figure 3.12, Michael Cosgrove, personal communication, 14 Mar 2011)

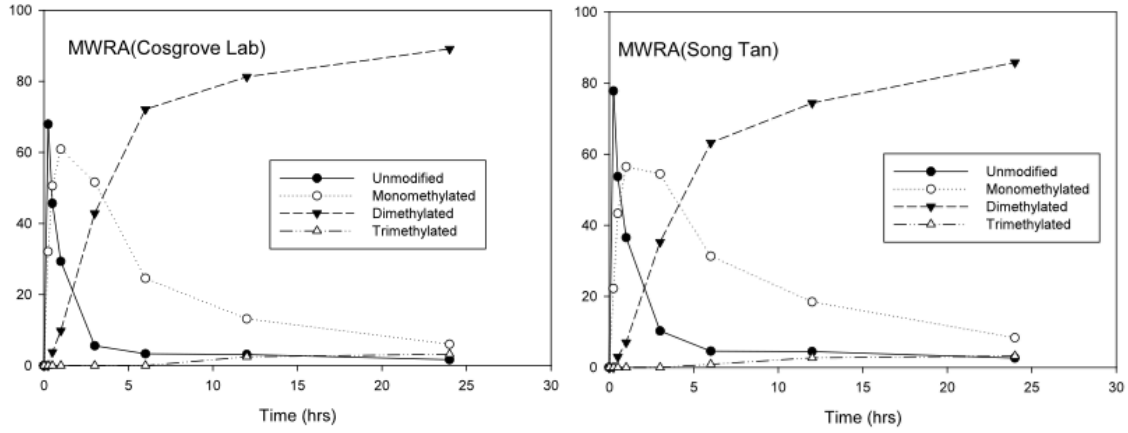

Figure 3.12: Single-turnover progress curves for dimethylation of H3K4 by *in vitro*-reconstituted MLL MWRA versus polycistronically coexpressed MLL MWRA complex as determined by Dr. Michael Cosgrove and Dr. Anamika Patel (Michael Cosgrove, personal communication, 14 Mar 2011)

Chapter 4: Discussion

4.1 Preparation of Polycistronic Expression Vector

I was able to successfully subclone the four MLL MWRA subunit genes into pST50Tr and pST66Tr transfer plasmids. With these plasmids, I was able to create the quadricistronic expression plasmid pST69-hAsh2Lx1-STRHISNhWDR5x1-hMLL1Δ1-hRbBP5. However, later subcloning steps (in particular the incorporation of hAsh2Lx1) proved to be difficult. The dual-plasmid system was devised to circumvent this step.

At first, I hypothesized that the size of the plasmid (8,008 bp) impeded transformation efficiency due to the lack of viable colonies produced from transformation. In the past, transformation of large plasmids, particularly those near 10,000 bp, yield a small fraction (2%) of the transformants obtained from the transformation of 3,000 bp plasmids (Siguret et al, 1994). However, anecdotal evidence from previous experiments in the Tan Laboratory indicated that our transformation protocol is robust for plasmids up to at least 10,000 bp. As a result, the techniques I used in my subcloning work were analyzed for potential problems. One possible issue was the way I gel purified DNA fragments. Factors that appear to improve the construction of the desired plasmids include using fresh electrophoresis buffer and using clean razor blades to excise the desired band from the agarose gel.

Although it ultimately proved unnecessary, I was also able to create a dual-plasmid expression system to coexpress the MWRA complex. In particular, I created the tricistronic plasmid pST69-STRHISNhWDR5x1-hMLL1Δ1-hRbBP5 and the

monocistronic plasmid pST73x1-hAsh2Lx1 were created. I determined a protocol to transform both plasmids into the same cell. I transformed in the first plasmid, prepared competent cells from that transformation plate, and then transformed the second plasmid into those competent cells.

While pST69-hAsh2Lx1-STRHISNhWDR5x1-hMLL1Δ1-hRbBP5 allows for the successful coexpression of the MLL MWRA complex, it can be optimized. Internal restriction sites should be removed from the four cassettes. This will facilitate the easy replacement of subunits with mutant and deletion variants in future studies. These resultant versions of the MWRA complex could be used in crystallization trials and nucleosomal binding assays to further analyze the structure of the complex and characterize its interaction with the nucleosome.

In addition, the sequence of hMLL1Δ1 should be optimized to increase expression of the protein. About 25% of the protein contained WDR5-RbBP5-Ash2L heterotrimer without MLL1Δ1, as evidenced by Source™ S fractions 22-30 (Figure 3.7). This may be due to decreased expression of MLL1Δ1. Rare codons could be silently mutated via PCR mutagenesis to improve bacterial translation of the gene. (Hannig and Makrides, 1998) The creation of pST69-hAsh2Lx1-STRHISNhWDR5x1-hMLL1Δ1-hRbBP5 was an important first step in order to express and study the MLL MWRA complex, but additional work on the plasmid may yield improved coexpression.

4.2 Small Scale Expression and Purification of the MLL MWRA Complex

The four subunits of the MLL MWRA complex coexpress optimally in BL21(DE3)-CodonPlus-RIL cells at 18°C. Solubility tests revealed that the subunits are roughly 60% soluble under these conditions. The four subunits were also successfully copurified using Talon® batch purification.

That these conditions yielded the most soluble complex does not mean that the conditions cannot be improved. The complex's unimpressive solubility may be a result of poor lysis. Additional or more forceful sonication of cells, along with the addition of lysozyme, may improve lysis and therefore complex solubility. Other expression strains should also be explored as potential candidates for MLL MWRA coexpression. For example, Stratagene ArcticExpress competent cells contain Cpn10 and Cpn60 chaperonins from *O. antarctica*, which allow for substantially better protein processing at low temperatures and demonstrate improved recombinant protein expression and solubility. Overall, though, BL21-CodonPlus(DE3)-RIL cells produced an adequate amount of soluble protein. Batch purification over Talon® resin also provided good preliminary purification and allowed for easy identification of the complex subunits in SDS-PAGE experiments.

4.3 Large Scale Expression and Purification of the MLL MWRA Complex

The MLL MWRA complex was expressed under these conditions in 6 L of cells. This crude extract was initially purified over a Talon® resin column by virtue of the Strept peptide/hexahistidine affinity tag on the WDR5 subunit. The affinity tag was then

cleaved before the complex was further purified via HPLC over Source™ S resin. This method allowed for sufficient purity, although some unknown contaminants did remain after purification and concentration.

However, these experiments were not without their problems. Approximately 50% of the expressed MLL MWRA complex eluted in the flow-through during Talon® column purification, despite the hexahistidine tag on WDR5. This suggests that the complex is not binding the resin as tightly as one would hope. In order to rectify this problem and increase yield, the 6x histidine tag should be replaced with a 10x histidine tag. Preliminary experiments performed in the Tan Laboratory by my fellow Schreyer Honors student Viktor Tollemar demonstrated that use of the 10x histidine tag improves purification of recombinant proteins over Talon® resin. Other affinity tags may also be employed to purify the complex with different resins and techniques.

4.4 Methyltransferase Activity of the MLL MWRA Complex

The most significant result from these experiments is that the polycistronically expressed MLL MWRA complex displays the same activity as that which was reconstituted *in vitro* by other members of the gene regulation community. The polycistronic expression of the complex eliminates the need for individual subunit purification and reconstitution, all of which are time-, labor-, and resource-intensive steps. We were able to demonstrate that this technique can provide complex of similar quality in considerably larger quantities than originally believed possible. This complex can be used in a multitude of experiments from biochemical assays to structural analyses inside and outside the Tan Laboratory.

4.5 Future Experiments

Our understanding of the MLL complex is, at best, incomplete. While we understand that it plays a role in the transcriptional regulation of *Hox* genes, there is still much to discover. While certain residues have been identified as critical to the integrity and catalytic activity of the complex, we still know little about its ability to recognize and bind the nucleosome. Future binding and crosslinking experiments will be able to shed light on which residues are critical for substrate recognition.

Previous structural studies have determined the crystal structures of the SET domain of MLL1, as well as of WDR5 bound to peptides from MLL1 and RbBP5. However, crystal structures of Ash2L, RbBP5, and the complex as a whole still elude researchers. Hopefully, the expression and purification outlined in these experiments will assist in the determination of these structures. It may also be possible to determine the structure of the MWRA complex bound to the nucleosome. The determination of the structure of RCC1 bound to the nucleosome by members of the Tan Laboratory has demonstrated that nucleosome-enzyme complexes can be crystallized (Makde et al, 2010).

A similar structure featuring the MLL MWRA complex would allow us to expand our understanding of its recognition of the nucleosome, as well as that of other SET-family histone methyltransferase complexes. Given that fusion proteins containing the MLL1 gene are linked to forms of leukemia, these findings may contribute to increased understanding of MLL1 fusion-mediated leukogenesis. It is also possible that additional insight into the structure and function of the MLL MWRA complex, especially regarding

its role in the larger MLL complex, may someday contribute to a potential treatment to combat these diseases.

4.6 Summary

I have determined ways to express multi-subunit complexes using polycistronic vectors like pST69. The genes that code for Ash2L, RbBP5, WDR5, and the C-terminal domain of MLL1 were successfully incorporated into a polycistronic expression plasmid. This plasmid can be used to express soluble and enzymatically active MLL MWRA complex in large quantities. Our ambition is that this method of producing the complex will contribute to advances in our understanding of the MLL complex, histone methyltransferases, leukogenesis, and the gene regulatory mechanisms of eukaryotic organisms.

Chapter 5: Appendix

5.1 Sequences of MLL MWRA Subunit Genes

hMLL1Δ1

```
   1 GGATCCTTCC GTTTCCACAA GCCAGAGGAG GCCAATGAAC CCCCCTTGAA
  51 CCCTCACGGC TCAGCCAGGG CTGAAGTCCA CCTCAGGAAG TCAGCATTTG
 101 ACATGTTTAA CTTCCTGGCT TCTAAACATC GTCAGCCTCC TGAATACAAC
 151 CCCAATGATG AAGAAGAGGA GGAGGTACAG CTGAAGTCAG CTCGGAGGGC
 201 AACTAGCATG GATCTGCCAA TGCCCATGCG CTTCCGGCAC TTAAAAAAGA
 251 CTTCTAAGGA GGCAGTTGGT GTCTACAGGT CTCCCATCCA TGGCCGGGGT
 301 CTTTTCTGTA AGAGAAACAT TGATGCAGGT GAGATGGTGA TTGAGTATGC
 351 CGGCAACGTC ATCCGCTCCA TCCAGACTGA CAAGCGGGAA AAGTATTACG
 401 ACAGCAAGGG CATTGGTTGC TATATGTTCC GAATTGATGA CTCAGAGGTA
 451 GTGGATGCCA CCATGCATGG AAATGCTGCA CGCTTCATCA ATCACTCGTG
 501 TGAGCCTAAC TGCTATTCTC GGGTCATCAA TATTGATGGG CAGAAGCACA
 551 TTGTCATCTT TGCCATGCGT AAGATCTACC GAGGAGAGGA ACTCACTTAC
 601 GACTATAAGT TCCCCATTGA GGATGCCAGC AACAAGCTGC CCTGCAACTG
 651 TGGCGCCAAG AAATGCCGGA AGTTCCTAAA CTAA
```

hWDR5

```
   1 TCTGCGACGG AGGAGAAGAA GCCCGAGACC GAGGCCGCCA GAGCACAGCC
  51 AACCCCTTCG TCATCCGCCA CTCAGAGCAA GCCTACACCT GTGAAGCCAA
 101 ACTATGCTCT AAAGTTCACC CTTGCTGGCC ACACCAAAGC AGTGTCCTCC
 151 GTGAAATTCA GCCCGAATGG AGAGTGGCTG GCAAGTTCAT CTGCTGATAA
 201 ACTTATTAAA ATTTGGGGCG CGTATGATGG GAAATTTGAG AAAACCATAT
 251 CTGGTCACAA GCTGGGAATA TCCGATGTAG CCTGGTCGTC AGATTCTAAC
```

```
 301 CTTCTTGTTT CTGCCTCAGA TGACAAAACC TTGAAGATAT GGGACGTGAG
 351 CTCGGGCAAG TGTCTGAAAA CCCTGAAGGG ACACAGTAAT TATGTCTTTT
 401 GCTGCAACTT CAATCCCCAG TCCAACCTTA TTGTCTCAGG ATCCTTTGAC
 451 GAAAGCGTGA GGATATGGGA TGTGAAAACA GGGAAGTGCC TCAAGACTTT
 501 GCCAGCTCAC TCGGATCCAG TCTCGGCCGT TCATTTTAAT CGTGATGGAT
 551 CCTTGATAGT TTCAAGTAGC TATGATGGTC TCTGTCGCAT CTGGGACACC
 601 GCCTCGGGCC AGTGCCTGAA GACGCTCATC GATGACGACA ACCCCCCCGT
 651 GTCTTTTGTG AAGTTCTCCC CGAACGGCAA ATACATCCTG GCCGCCACGC
 701 TGGACAACAC TCTGAAGCTC TGGGACTACA GCAAGGGGAA GTGCCTGAAG
 751 ACGTACACTG GCCACAAGAA TGAGAAATAC TGCATATTTG CCAATTTCTC
 801 TGTTACTGGT GGGAAGTGGA TTGTGTCTGG CTCAGAGGAT AACCTTGTTT
 851 ACATCTGGAA CCTTCAGACG AAAGAGATCG TACAGAAACT ACAAGGCCAC
 901 ACAGATGTCG TGATCTCAAC AGCTTGTCAC CCAACAGAAA ACATCATCGC
 951 CTCTGCTGCG CTAGAAAATG ACAAAACAAT TAAACTGTGG AAGAGTGACT
1001 GCTAA
```

hRbBP5

```
   1 ATGAACCTCG AGTTGCTGGA GTCCTTTGGG CAGAACTATC CAGAGGAAGC
  51 TGATGGAACT TTGGATTGTA TCAGCATGGC TTTGACTTGC ACCTTTAACA
 101 GGTGGGGCAC ACTGCTTGCA GTTGGCTGTA ATGATGGCCG AATTGTCATC
 151 TGGGATTTCT TGACAAGAGG CATTGCTAAA ATAATTAGTG CACACATCCA
 201 TCCAGTGTGT TCTTTATGCT GGAGCCGAGA TGGTCATAAA CTCGTGAGTG
 251 CTTCCACTGA TAACATAGTG TCACAGTGGG ATGTTCTTTC AGGCGACTGT
 301 GACCAGAGGT TTCGATTCCC TTCACCCATC TTAAAAGTCC AATATCATCC
 351 ACGAGATCAG AACAAGGTTC TCGTGTGTCC CATGAAATCT GCTCCTGTCA
 401 TGTTGACCCT TTCAGATTCC AAACATGTTG TTCTGCCAGT GGACGATGAC
 451 TCCGATTTGA ACGTTGTGGC ATCTTTTGAT AGGCGAGGGG AATATATTTA
```

501 TACGGGAAAC GCAAAAGGCA AGATTTTGGT CCTAAAAACA GATTCTCAGG

551 ATCTTGTTGC TTCCTTCAGA GTGACAACTG GAACAAGCAA TACCACAGCC

601 ATTAAGTCAA TTGAGTTTGC CCGGAAGGGG AGTTGCTTTT TAATTAACAC

651 GGCAGATCGA ATAATCAGAG TTTATGATGG CAGAGAAATC TTAACATGTG

701 GAAGAGATGG AGAGCCTGAA CCTATGCAGA AATTGCAGGA TTTGGTGAAT

751 AGGACCCCAT GGAAGAAATG TTGTTTCTCT GGGGATGGGG AATACATCGT

801 GGCAGGTTCT GCCCGGCAGC ATGCCCTGTA CATCTGGGAG AAGAGCATTG

851 GCAACCTGGT GAAGATTCTC CATGGGACGA GAGGAGAACT CCTCTTGGAT

901 GTAGCTTGGC ATCCTGTTCG ACCCATCATA GCATCCATTT CCAGTGGAGT

951 GGTATCTATC TGGGCACAGA ATCAAGTAGA AAACTGGAGT GCATTTGCAC

1001 CAGACTTCAA AGAATTGGAT GAAAATGTAG AATACGAAGA AAGGGAATCA

1051 GAGTTTGATA TTGAAGATGA AGATAAGAGT GAGCCTGAGC AGACAGGGGC

1101 TGATGCTGCA GAAGATGAGG AAGTGGATGT CACCAGCGTG GACCCTATTG

1151 CTGCCTTCTG TAGCAGTGAT GAAGAGCTGG AAGATTCAAA GGCTCTATTG

1201 TATTTACCCA TTGCCCCTGA GGTAGAAGAC CCAGAAGAAA ATCCTTACGG

1251 CCCCCCACCG GATGCAGTCC AAACCTCCTT GATGGATGAA GGGGCTAGTT

1301 CAGAGAAGAA GAGGCAGTCC TCAGCAGATG GGTCCCAGCC ACCTAAGAAG

1351 AAACCCAAAA CAACCAATAT AGAACTTCAA GGAGTACCAA ATGATGAAGT

1401 CCATCCACTA CTGGGTGTGA AGGGGGATGG CAAATCCAAG AAGAAGCAAG

1451 CAGGCCGGCC TAAAGGATCA AAAGGTAAAG AGAAAGATTC TCCATTTAAA

1501 CCGAAACTCT ACAAAGGGGA CAGAGGTTTA CCTCTGGAAG GATCAGCGAA

1551 GGGTAAAGTG CAGGCGGAAC TCAGCCAGCC CTTGACAGCA GGAGGAGCAA

1601 TCTCAGAACT GTTATAA

hAsh2L

1 ATGGATACTC AGGCGGGCTC CGTGGATGAA GAGAATGGCC GACAGTTGGG

51 TGAGGTAGAG CTGCAATGTG GGATTTGTAC AAAATGGTTC ACGGCTGACA

```
 101 CATTTGGCAT AGATACCTCA TCCTGTCTAC CTTTCATGAC CAACTACAGT
 151 TTTCATTGCA ACGTCTGCCA TCACAGTGGG AATACCTATT TCCTCCGGAA
 201 GCAAGCAAAC TTGAAGGAAA TGTGCCTTAG TGCTTTGGCC AACCTGACAT
 251 GGCAGTCCCG AACACAGGAT GAACATCCGA AGACAATGTT CTCCAAAGAT
 301 AAGGATATTA TACCATTTAT TGATAAATAC TGGGAGTGCA TGACAACCAG
 351 ACAGAGACCT GGGAAAATGA CTTGGCCAAA TAACATTGTT AAAACAATGA
 401 GTAAAGAAAG AGATGTATTC TTGGTAAAGG AACACCCAGA TCCAGGCAGT
 451 AAAGATCCAG AAGAAGATTA CCCCAAATTT GGACTTTTGG ATCAGGACCT
 501 TAGTAACATT GGTCCTGCTT ATGACAACCA AAAACAGAGC AGTGCTGTGT
 551 CTACTAGTGG GAATTTAAAT GGGGGAATTG CAGCAGGAAG CAGCGGAAAA
 601 GGACGAGGAG CCAAGCGCAA ACAGCAGGAT GGAGGGACCA CAGGGACCAC
 651 CAAGAAGGCC CGGAGTGACC CTTTGTTTTC TGCTCAGCGC CTTCCCCCTC
 701 ATGGCTACCC ATTGGAACAC CCGTTTAACA AAGATGGCTA TCGGTATATT
 751 CTAGCTGAGC CTGATCCGCA CGCCCCTGAC CCCGAGAAGC TGGAACTTGA
 801 CTGCTGGGCA GGAAAACCTA TTCCTGGAGA CCTCTACAGA GCCTGCTTGT
 851 ATGAACGGGT TTTGTTAGCC CTACATGATC GAGCTCCCCA GTTAAAAATC
 901 TCAGATGACC GGCTGACTGT GGTTGGAGAG AAGGGCTACT CTATGGTGAG
 951 GGCCTCTCAT GGAGTACGGA AGGGTGCCTG GTATTTTGAA ATCACTGTGG
1001 ATGAGATGCC ACCAGATACC GCTGCCAGAC TGGGTTGGTC CCAGCCCCTA
1051 GGAAACCTTC AAGCTCCTTT AGGTTATGAT AAATTTAGCT ATTCTTGGCG
1101 GAGCAAAAAG GGAACCAAGT TCCACCAGTC CATTGGCAAA CACTACTCTT
1151 CTGGCTATGG ACAGGGAGAC GTCCTGGGAT TTTATATTAA TCTTCCTGAA
1201 GACACAGAGA CAGCCAAGTC ATTGCCAGAC ACATACAAAG ATAAGGCTTT
1251 GATAAAATTC AAGAGTTATT TGTATTTTGA GGAAAAAGAC TTTGTGGATA
1301 AAGCAGAGAA GAGCCTGAAG CAGACTCCCC ATAGTGAGAT AATATTTTAT
1351 AAAAATGGTG TCAATCAAGG TGTGGCTTAC AAAGATATTT TTGAGGGGGT
1401 TTACTTCCCA GCCATCTCAC TGTACAAGAG CTGCACGGTT TCCATTAACT
```

```
1451 TTGGACCATG CTTCAAGTAT CCTCCGAAGG ATCTCACTTA CCGCCCTATG
1501 AGTGACATGG GCTGGGGCGC CGTGGTAGAG CACACCCTGG CTGACGTCTT
1551 GTATCACGTG GAGACAGAAG TGGATGGGAG GCGCAGTCCC CCATGGGAAC
1601 CCTAA
```

5.2 Sequences of Primers Used to Prepare MLL MWRA Complex

Amplification of hMLL1Δ1 from pGSTparallel-hMLL1Δ1

STO2924: CGGAATTCGG ATCCTTCCGT TTCCACAAGC C
STO2925: GCCTCTAGAT GTACATTAGT TTAGGAACTT CCGGCATT

Amplification of hWDR5x1 from pET3aTr-hWDR5x1

STO2926: CGGAATTCAG ATCTGCGACG GAGGAGAAGA AGC
STO2927: GCCTCTAGAT GTACATTAGC AGTCACTCTT CCACAGTT

Amplification of hRbBP5 from pET3aTr-hRbBP5

STO1099: TCCCGCGAAA TTAATACGAC
STO2928: GCCTCTAGAT GTACATTATA ACAGTTCTGA GATTGCTCCT

Amplification of hAsh2L from pET3aTr-hAsh2L

STO1099: TCCCGCGAAA TTAATACGAC
STO2929: GCCTCTAGAG GGCCCTTAGG GTTCCCATGG GGG

Mutagenesis to remove internal BsrGI site from hWDR5 to create pET3aTr-hWDR5x1

STO2988: TTCAGACGAA AGAGATCGTA CAGAAACTAC A
STO2989: TGTAGTTTCT GTACGATCTC TTTCGTCTGA A

Mutagenesis to remove internal BglII site from hAsh2L to create pST66Tr-hAsh2Lx1

STO3033: TGTAGTTTCT GTACGATCTC TTTCGTCTGA A
STO3034: GGTCATCTGA GATTTTTAAC TGGGGAGCT

Mutagenesis to remove internal RsrII site from pST73 to create pST73x1

STO3029: CGGCTGGGTG TGGCTGACCG CTATCAG
STO3030: CTGATAGCGG TCAGCCACAC CCAGCCG

References

Alberts, B., Johnson, A., *et al.* Molecular Biology of the Cell, 4th Edition (2002). Garland Science: New York.

Avdic, V., Zhang, P., *et al*. "Structural and Biochemical Insights into MLL1 Core Complex Assembly." Cell Structure 19 (2011): 101-108.

Barski, A., Cuddapah, S., *et al.* "High-Resolution Profiling of Histone Methylations in the Human Genome." Cell 129 (2007): 823-837.

Cao, F., Chen, Y., *et al.* "An Ash2L/RbBP5 Heterodimer Stimulates the MLL1 Methyltransferase Activity Through Coordinated Substrate Interactions with the MLL1 SET Domain." PLoS One 5 (2010): e14102.

Carroll, S. "Homeotic Genes and the Evolution of Athropods and Chordates." Nature 376 (1995): 479-485.

Davey, C., Sargent, D., *et al.* "Solvent Mediated Interactions in the Structure of the Nucleosome Core Particle at 1.9 Å Resolution." Journal of Molecular Biology 319 (2002): 1097-1113.

Dillon, S., Zhang, X., *et al.* "The SET-Domain Protein Superfamily: Protein Lysine Methyltransferases." Genome Biology 6 (2005): 227.

Dou, Y., Milne, T., *et al.* "Regulation of MLL1 H3K4 Methyltransferase Activity by Its Core Components." Nature Structural & Molecular Biology 13 (2006): 713-719.

Grewal, S., and Rice, J. "Regulation of Heterochromatin by Histone Methylation and Small RNAs." Current Opinion in Cell Biology 16 (2004): 230-238.

Guo, H., and Guo, H. "Mechanism of Histone Methylation Catalyzed by Protein Lysine Methyltransferase SET7/9 and Origin of Product Specificity." Proceedings of the National Academy of Science 104 (2007): 8797-8802.

Hannig, G., and Makrides, S. "Strategies for Optimizing Heterologous Protein Expression in *Escherichia coli*." Trends in Biotechnology 16 (1998): 54-60.

Harp, J. Hanson, B., *et al.* "Asymmetries in the Nucleosome Core Particle at 2.5 Å Resolution." Acta Crystallographica Section D: Biological Crystallography 56 (2000): 1513-1534.

Ioshikhes, I., Bolshoy, A., *et al.* "Nucleosome DNA Sequence Pattern Revealed by Multiple Alignment of Experimentally Mapped Sequences." Journal of Molecular Biology 262 (1996): 129-139.

Jenuwein, T. “Translating the Histone Code.” Science 293 (2001): 1074-1080.

Krivtsov, A., and Armstrong, S. “MLL Translocations, Histone Modifications, and Leukaemia Stem-Cell Development.” Nature Reviews Cancer 7 (2007): 823-833.

Liu, H., Cheng, E., and Hsieh, J. “MLL Fusions: Pathways to Leukemia.” Cancer Biology & Therapy 8 (2009): 1204-1211.

Luger, K., Mäder, A., *et al.* “Crystal Structure of the Nucleosome Core Particle at 2.8 Å Resolution.” Nature 389 (1997): 251-260.

Makde, R., England, J., *et al*. “Structure of RCC1 Chromatin Factor Bound to the Nucleosome Core Particle.” Nature 467 (2010): 562-566.

Noma, K., Allis, C., and Grewal, S. “Transitions in Distinct Histone H3 Methylation Patterns at the Heterochromatin Domain Boundaries.” Science 293 (2001): 1150-1155.

Odho, Z., Southall, S., and Wilson, J. “Characterization of a Novel WDR5-Binding Site That Recruits RbBP5 Through a Conserved Motif to Enhance Methylation of Histone H3 Lysine 4 by Mixed Lineage Leukemia Protein-1.” Journal of Biological Chemistry 285 (2010): 32967-32976.

Patel, A., Dharmarajan, V., and Cosgrove, M. “Structure of WDR5 Bound to Mixed Lineage Leukemia Protein-1 Peptide.” Journal of Biological Chemistry 283 (2008a): 32158-32161.

Patel, A., Dharmarajan, V., *et al*. “On the Mechanism of Multiple Lysine Methylation by the Human Mixed Lineage Leukemia Protein-1 (MLL1) Core Complex.” Journal of Biological Chemistry 284 (2009): 24242-24256.

Patel, A., Vought, V., *et al*. “A Conserved Arginine-Containing Motif Crucial for the Assembly and Enzymatic Activity of the Mixed Lineage Leukemia Protein-1 Core Complex.” Journal of Biological Chemistry 283 (2008b): 32162-32175.

Qian, C. and Zhou, M. “SET Domain Protein Lysine Methyltransferases: Structure, Specificity and Catalysis.” Cellular and Molecular Life Sciences 63 (2006): 2755-2763.

Ruthenburg, A., Allis, C., and Wysocka, J. “Methylation of Lysine 4 on Histone H3: Intricacy of Writing and Reading a Single Epigenetic Mark.” Molecular Cell 25 (2007): 15-30.

Schein, C., and Noteborn, M. “Formation of Soluble Recombinant Protein in *Escherichia coli* is Favored by Lower Growth Temperature.” Nature Biotechnology 6 (1988): 291-294.

Shi, Y., Lan, F., *et al.* "Histone Demethylation Mediated by the Nuclear Amine Oxidase Homolog LSD1." Cell 119 (2004): 941-953.

Siguret, V., Ribba, A., *et al*. "Effect of Plasmid Size on Transformation Efficiency by Electroporation of *Escherichia coli* DH5 Alpha." BioTechniques 16 (1994): 422-426.

Slany, R. "Chromatin Control of Gene Expression: Mixed-Lineage of Leukemia Methyltransferase SETs the Stage for Transcription." Proceedings of the National Academy of Science 102 (2005): 14481-14482.

Southall, S., Wong, P., *et al.* "Structural Basis for the Requirement of Additional Factors for MLL1 SET Domain Activity and Recognition of Epigenetic Marks." Molecular Cell 33 (2009): 181-191.

Steward, M., Lee, J., *et al.* "Molecular Regulation of H3K4 Trimethylation by Ash2L, a Shared Subunit of MLL Complexes." Nature Structural & Molecular Biology 13 (2006): 852-854.

Tan, S., Kern, R., and Selleck, W. "The pST44 Polycistronic Expression System fro Producing Protein Complexes in *Escherichia coli*." Protein Expression and Purification 40 (2005): 385-395.

Turner, B. "Cellular Memory and the Histone Code." Cell 111 (2002): 285-291.

Veselkov, A., Demidov, V., Nielsen, P., and Frank-Kamenetskii, M. "A New Class of Genome Rare Cutters." Nucleic Acids Research 24 (1996): 2483-2487.

Whitehouse, I., Flaus, A., *et al.* "Nucleosome Mobilization Catalysed by the Yeast SWI/SNF Copmlex." Nature 400 (1999): 784-786.

Wysocka, J., Myers, M., *et al.* "Human Sin3 Deacetylase and Trithorax-Related Set1/Ash2 Histone H3-K4 Methyltransferase are Tethered Together Selectively by the Cell-Proliferation Factor HCF-1." Genes & Development 17 (2003): 896-911.

Yokoyama, A., Wang, Z., *et al.* "Leukemia Proto-Oncoprotein MLL Forms a SET1-Like Histone Methyltransferase Complex with Menin to Regulate *Hox* Gene Expression." Molecular and Cell Biology 24 (2004): 5639-5649.

Academic Vita

Wil Leonard Santivasi

E-Mail ID	wls5030
Major	Premedicine
Area of Honors	Biochemistry and Molecular Biology
Thesis Title	Polycistronic Coexpression and Copurification of the Human MLL MWRA Histone Methyltransferase Subcomplex
Thesis Supervisor	Dr. Song Tan
Work Experience	***Independent Researcher*** Penn State Center for Eukaryotic Gene Regulation Dr. Song Tan, Associate Professor January 2008 – Present ***Laboratory Teaching Assistant*** Penn State Department of Biology Dianne Burpee, Teaching Laboratory Coordinator August 2009 – Present ***Research and Patient Care Assistant*** Nathan Speare Regional Burn Treatment Center Dr. Mary Lou Patton, Medical Director May 2007 – Present
Honors	*Dean's List* (August 2007 – Present) *Schreyer Honors College Academic Excellence Scholarship* *Andrews Family Scholarship for the Life Sciences* *The President Sparks Award* *The President's Freshman Award*
Professional Memberships	*Alpha Epsilon Delta Pre-Medical Honorary Fraternity* *Phi Beta Kappa Honor Society*
Language Proficiency	*English* *American Sign Language*