Spring 5-1-2014

The Purification and Characterization of the Drosophila melanogaster Trithorax Protein and its Implications in the Studies of the SET domain Family of Proteins

Kelsey Monteith

The Purification and Characterization of the *Drosophila melanogaster* Trithorax Protein and its Implications in the Studies of the SET domain Family of Proteins

A Capstone Project Submitted in Partial Fulfillment of the Requirements of the Renée Crown University Honors Program at Syracuse University

Kelsey Monteith
Candidate for B. S. Degree
and Renée Crown University Honors
May 2014

Honors Capstone Project in Biochemistry

Capstone Project Advisor: _____
Dr. Michael Cosgrove

Capstone Project Reader: _____
Dr. Stephen Dorus

Honors Director: _____
Stephen Kuusisto, Director

Date: April 23, 2014

Abstract

Methylation at histone H3 lysine 4 (H3K4) is a post-translational modification often associated with transcriptional regulation through altering the structural state of chromatin. The human mixed lineage leukemia protein-1 protein (MLL1) is a known histone methyltransferase that catalyzes the transfer of methyl groups to H3K4. MLL1 works in a core complex with other essential components, proteins WDR5, RbBP5, Ash2L, DPY-30 (WRAD), which is required for H3K4 dimethylation. Trithorax (TRX) protein is the *Drosophila melanogaster* ortholog to human MLL1, and although structurally similar is unable to perform dimethylation when in complex with the human components. The goal of this study is to understand the structural basis for this difference. We systematically mutated 20 amino acids in TRX the equivalent amino acid in human MLL1 and tested for a gain-of-function H3K4 dimethylation activity. We found 20 amino acid positions in TRX that were highly conserved among intertebrates but were different in vertebrates. Out of the 20 amino acids mutated, 5 showed a gain of dimethylation activity. All of the mutations that showed a gain of dimethylation activity localized to a common SET domain surface. The identified mutations on the common surface identify a location of the dimethyltransferase active site on MLL1.

Table of Contents

Acknowledgements

I would like extend a special thank you and express how grateful I am to those who have helped and guided me through my thesis in its entirety.

Thank you to Dr. Michael Cosgrove, Capstone advisor and Principal Investigator, for allowing me to join his wonderful and dedicated lab group. Dr. Cosgrove's extended patience, guidance and support through my academics has been invaluable, and for that I will be forever grateful. I cannot fully encapsulate the amount that I have learned and gained from the last three years. Thank you to Dr. Anamika Patel, former post-doc, for acting as a patient and extraordinary mentor during my first years in the lab. Thank you to all of the other Cosgrove lab members at SUNY Upstate Medical University, Stephen Shinsky, Dr. Nilda Alicea-Velazquez, Susan Viggiano, Kevin Namitz, and Michael Hu, for guiding and helping me through all facets of my time in the lab.

Thank you to Dr. Stephen Dorus for acting as my Capstone reader and an additional source of support for my thesis.

Thank you to the Department of Biochemistry and Molecular Biology at SUNY Upstate Medical University for providing endless opportunity to participate in undergraduate research, throughout the academic year as well as the summer.

Thank you to the Renée Crown University Honors program for providing initial connections, endless support and encouragement and the opportunity to present my thesis research in early April 2014 in Pittsburgh, PA at the ACC Meeting of the Minds Undergraduate Research Conference. An additional thank you to the Honors program for the Crown Award, which helped fund essential materials for experiments required for my thesis.

Thank you to the Departments of Biology and Chemistry at Syracuse University for providing the essential academic and laboratory tools for research and throughout my undergraduate career.

To those not mentioned here, thank you for your support, guidance and patience.

Introduction

In eukaryotes, gene expression is regulated by structural alterations of chromatin states. Although most cells in a multi-cellular organism contain equivalent genetic information, differential gene expression programs allow for each cell to have its own identity[Error! Reference source not found.]. How genes are regulated and inherited without changes to the DNA sequence is studied through the field of epigenetics. Epigenetics could explain how through development each cell-type maintains a distinct and lasting gene expression profile[Error! Reference source not found.]. The mechanisms of gene regulation alter the structure of chromatin, a molecule composed of DNA and histone and non-histone proteins.

Chromatin is composed of repeating nucleosome subunits[3]. The subunits contain a histone octamer consisting of two copies of each histone (H2A, H2B, H3 and H4) around which ~147 base pairs of DNA are wrapped[4]. Histones are small, conserved globular proteins that contain basic flexible histone "tails" which extend outward from the histone octamer. The mobile position of nucleosomes on the DNA helix allows the chromatin to assume a higher order of organization[5]. Alterations or modifications to the nucleosomal subunits cause the chromatin structure to alternate between two chromatin states, heterochromatin and euchromatin. Heterochromatin, associated with gene repression, is condensed and compact, making the DNA inaccessible to transcription factors and proteins for processing. In contrast, loosely packed chromatin, euchromatin, allows the DNA to be more open and accessible which is generally correlated with gene activation[6,7].

The regulation of chromatin states between active and repressive gene expression is maintained by a variety of post-translational modifications on the histone proteins of the nucleosome[5]. Identified and studied histone modifications include methylation, acetylation, phosphorylation, ubiquitylation and more[8]. Histone methylation involves the transfer of methyl groups from the methyl donor S-adenosylmethionine to positively charged amino acids lysine and arginine[9]. Due to the positively charged amino group on the lysine side chain, lysine residues can accept up to three methyl groups (mono-, di-, and trimethylation). (**Figure 1**).

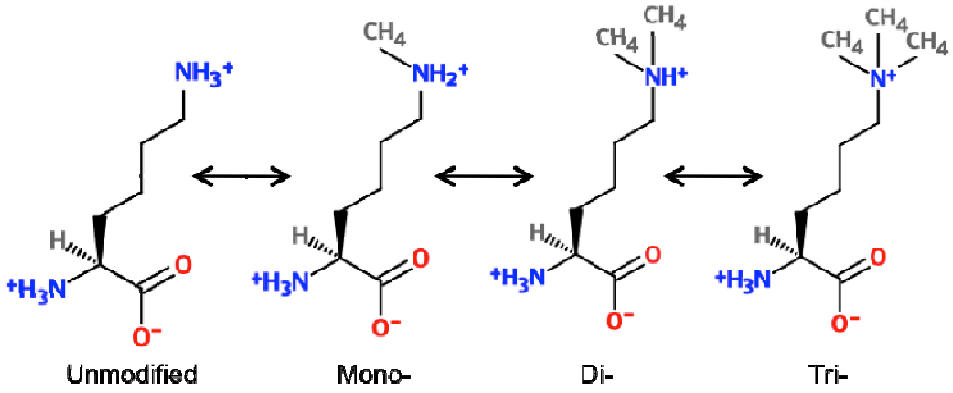


Figure 1. Basic amino acid lysine can have varying degrees of methylation on the positively charged amino group of the 'R' side-chain

Different post-translational modifications to histones at various residues has been suggested to constitute a "histone code"[10]. The "histone code" refers the fact that different histone modifications can be recognized by different effector proteins that signal a specific cellular response. In order to regulate gene expression, eukaryotic cells have evolved a series of enzyme complexes that modify the histones post-translationally, thereby regulating the chromatin structure with transcriptional implications[11].

Maintaining gene expression through histone modification

In multicellular organisms, the regulation of genomic transcription is essential for body segmentation, cell division, cell patterning and tissue homeostasis. Epigenetic mechanisms regulate these processes by modulating chromatin structure by the action of two main groups of proteins, the Trithorax group (TrxG) and the Polycomb group (PcG)[1]. TrxG proteins maintain active transcription while the PcG proteins maintain transcriptional repression[12,13].

The first determined gene member of the TrxG was the Trithorax (*TRX*) gene initially discovered in *Drosophila melanogaster*[14]. The *TRX* gene was initially identified through observations of mutations resulting in homeotic transformations[15, 16]. Homeotic transformations are mutations in which a region of the body changes into the likeness of another, such as extra sets of wings or legs in anatomical positions where they do not naturally occur[17]. Homeotic phenotypes result from the misexpression of a group of genes called Homeotic (HOX) genes. *HOX* genes are essential in directing regional body patterning and segment orientation[18]. The *trx* gene encodes a gene product, the trithorax (TRX) protein, which regulates the expression of homeotic genes[19,20,21]. Although the regulation mechanisms are not well understood, a *trx* gene mutation has given insight into possible mechanisms of regulation.

The trx^{z11} mutation leads to homeotic transformations resulting from a single amino acid change in the SET (SU(VAR)3-9, E(z), and Trx) domain found at its C-terminus[22,23]. The SET domain is a conserved structural motif that is involved in chromatin regulation and protein-protein interaction. The trx^{z11}

mutation has been suggested to strongly reduce the capability of the SET domain to bind to histones, negatively affecting the regulative capabilities of the TRX protein[24]. Indeed the Cosgrove lab has shown that the TRX^{Z11} mutation is catalytically inactive[25].

SET domain and the core complex family of proteins

The majority of the SET domain containing proteins have been shown to catalyze histone lysine methylation, which as discussed previously play significant roles in transcription, the cell cycle and cell differentiation[26]. The gene orthologous to Trithorax in humans is called human Trithorax (HRX) or the Mixed Lineage Leukemia (MLL) gene. Although many of the SET domain proteins are able to catalyze histone methylation alone, many require interaction with other proteins for optimal enzymatic activity[27,28]. MLL family members interact with conserved group of proteins called WRAD (WDR5, RbBP5, Ash2L and DPY-30), which is required for mono- and dimethylation of H3K4. MLL1 when in complex with WRAD is called the MLL1 core complex. The complex is conserved throughout eukaryotic evolution. In budding yeast it is called COMPASS, which catalyzes mono-, di-, and trimethylation.

Mixed lineage leukemia protein-1 and leukemia

Alterations, mutations and misregulations of the *MLL1* gene have been correlated with aggressive leukemias[29]. Like TRX in *D. melanogaster*, the human MLL1 protein is responsible for regulation of HOX genes throughout vertebrate

development. In adults and throughout embryonic development, MLL1 is widely expressed throughout the body, as well as in myeloid and lymphoid cells[29]. Myeloid and lymphoid cells are hematopoietic stem cells that divide to give red and white blood cells. Alterations of the *MLL1* gene lead to an abnormal increase of immature white blood cells, which crowd out the healthy white and red blood cells ultimately leading to death.

As seen in **Figure 2**, the *MLL1* gene contains the SET domain, the domain responsible for its histone methyltransferase activity, which specifically methylates H3K4[31]. As described previously, lysine residues can accept up to three methyl groups. Unmodified and mono-methylated H3K4 is associated with gene repression while di- and tri-methylated H3K4 is associated with gene transcription[30].

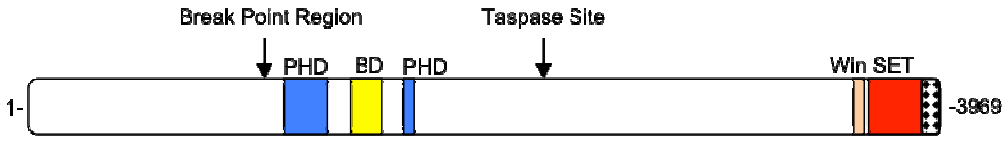


Figure 2. Gene structure of the *MLL1* gene. The schematic includes a few of the other important domains as well as the *Win* and SET domains. The figure is adapted from *Dharmarajan et al. (2012).*

Previous studies by the Cosgrove lab have shown MLL1 forms a core complex with WRAD through the *Win* (WDR5 Interaction) motif[32]. The *Win* motif is defined as a core motif in the interaction between MLL1 and WDR5 containing a conserved arginine residue that is essential for allowing the complex to form (**Figure 3**)[27]. In the absence of WRAD, MLL1 is only able to transfer one methyl group, while the MLL1-WRAD core complex can transfer two methyl groups to H3K4[27]. Previous studies as well as a crystal structure of MLL1 with a

histone H3 peptide and methyl donor S-adenosylmethionine (SAM) have identified the location of the active site for the first methyl group transfer[33]. The active site location and mechanism for the second methyl group transfer remains relatively unknown. While there are varied theories, our current experimental results suggest a portion of MLL1 is required for formation of the second active site, but its location remains unknown. To develop a greater understanding of the biological mechanism of MLL1 histone methyltransferase activity in the core complex, we turned to the TRX ortholog of *D. melanogaster*.

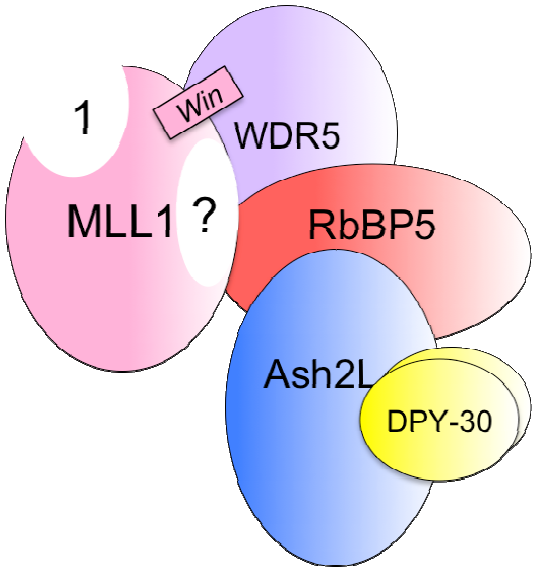


Figure 3. Proposed structure (not to scale) of the MLL1 core complex with the *Win* motif of MLL1 required for complex formation. Previous studies have identified a first active site labeled 1, where monomethylation occurs. Preliminary studies done by the Cosgrove lab have suggested there is a second active site on MLL1 where dimethylation occurs. The exact location of the second active site is unknown (signified by "?") Adapted from Patel *et al.(2009).*

Trithorax can form a hybrid core complex

Although TRX and MLL1 are orthologous, by primary structure examination it is clear that there have been evolutionary changes over time. To visualize these

differences, we created a primary sequence alignment of twenty-six different MLL1 orthologs, highlighting conserved residues. A section of the alignment is seen in **Figure 4** (***See Appendix for full alignment***).

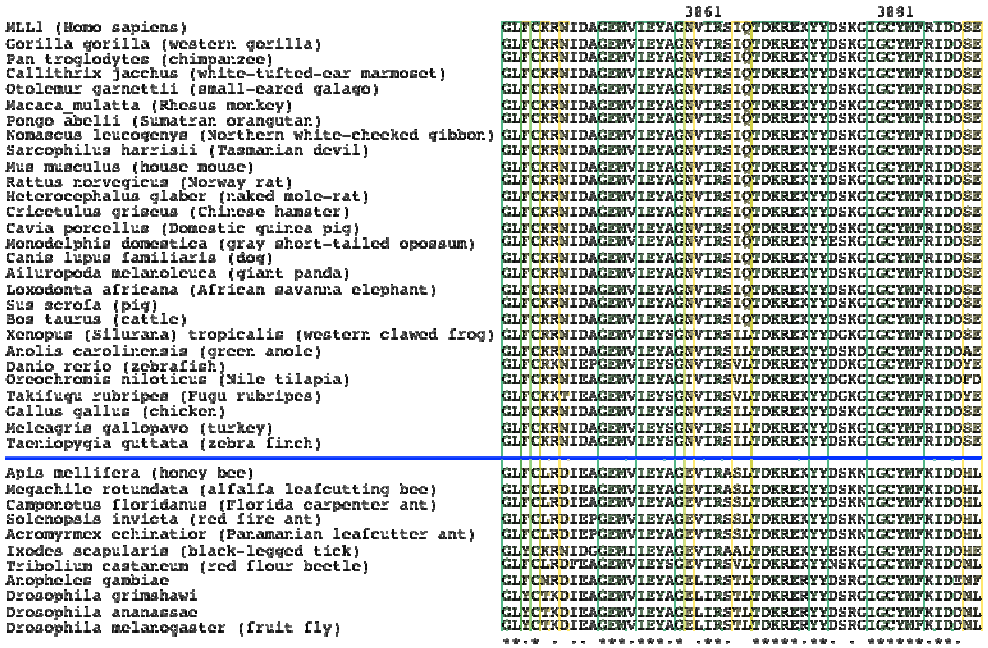


Figure 4. Sample of a ClustalW multiple sequence alignment showing the conserved residues (green) and residues (yellow) proposed to be essential for dimethylation activity of MLL1, among vertebrates and invertebrates. The residues proposed as essential for dimethylation activity were chosen based on their differences among the vertebrate and invertebrate categories, shown by the blue dotted line. (*Full alignment in Appendix*).

Upon observing the conserved residues in the alignment, we recognized a high degree of conservation, including complete conservation of the arginine residue in the *Win* motif region (***Appendix***). We hypothesized that due to the conserved arginine residue, we would be able to form a hybrid core complex with TRX and human WRAD (hWRAD). We performed a qualitative GST-pull-down assay and observed that the complex formed. This means that the mechanisms for complex formation have been conserved through ~780 million years of evolution.

With successful formation of the hybrid TRX core complex, we then hypothesized it would display similar enzymatic activity as the MLL1 core complex. However, using our quantitative mass spectrometry time course assay, we were surprised to find differing activity. Like the MLL1 core complex, the hybrid complex performed H3K4 monomethylation. However, the results showed that the hybrid TRX core complex was unable to perform H3K4 dimethylation. With these new results to consider, we returned to the sequence alignment to determine the factors responsible for the differences in activity. We hypothesized that the difference in activity must be due to amino acids that are not conserved between TRX and MLL1.

Upon further examination of alignment, we identified a unique conservation pattern at twenty positions in the alignment when vertebrates and invertebrates are divided into separate groups (**Figure 4**). At these locations, there is a noticeable evolutionary switch where the residue in vertebrates is relatively conserved amongst all vertebrates, but is a different conserved residue in invertebrates. We theorized that over the course of evolution, at one of these locations, a residue was altered that changed the activity of MLL1 and TRX. We hypothesized that if we mutate the residue at one of these twenty locations in the primary structure of TRX to the residue found at the equivalent location in MLL1, we would be able to observe a gain of dimethylation activity. In this study, we have mutated each of the twenty identified residues and identified several that confer a gain of dimethylation activity at H3K4.

Our results provide great insight into the structural arrangement of MLL1 core complex subunits. In addition, these results suggest a putative location for the missing active site that confers H3K4 dimethylation. As the interactions among the proteins in the MLL1 core complex are not well understood, a greater understanding of the regulation of MLL1 enzymatic activity could lead to innovative strategies to manipulate gene expression patterns in leukemic cells.

Methods and Materials

Site-directed mutagenesis and protein expression

DNA for *Drosophila Melanogaster* constructs of TRX 2976-3726 ($TRX^{2976-3726}$) and TRX 3475-3726 (TRX^{C251}) pGEX vectors as GST fusion were obtained as gifts from Dr. Peter Harte at Case Western Reserve University. To prepare the mutants, the shorter GST-TRX^{C251} construct was subjected to site-directed mutagenesis (QuickChange II) using manually designed primers ordered from Integrated DNA Technologies.

Each GST-TRX^{C251} mutant DNA was expressed individually in competent *Escherichia coli* DH5alpha cells on Lysogeny Broth (LB) agar plates containing 50μg/mL carbenicillin antibiotic at 37°C overnight. Single bacterial colonies were selected and grown in a 5mL solution of LB at 50μg/mL for 18h. Crude bacterial solution was pelleted and subjected to the Wizard® Plus SV Minipreps DNA Purification System (Promega). The purified plasmids were then expressed in *Escherichia coli* (Rosetta II, Novagen), grown on 50μg/mL carbenicillin and 20μg/mL chloramphenicol containing LB agar plates at 37°C overnight. Single colonies of cells were then grown at 30°C in a 5mL solution of Terrific Broth II media (TBII) with 50μg/mL of carbenicillin and 20μg/mL chloramphenicol for 18h. 250μL of these cells were then grown in TBII in the presence of 50μg/mL of carbenicillin at 37°C until the optical density reached approximately 0.80. The cells were then chilled at 4°C for 1h. The cells were then induced with 500μM isopropyl-1-thio-β-D-galactopyranoside (IPTG) for 18h (later optimized for

induction with 1mM IPTG for 24h). The cultures are then centrifuged at 5000xRPM for 30 minutes at 4°C. Cells were obtained and resuspended in a lysis buffer (50 mM Tris, pH=7.3, 150 mM NaCl, 10% Glycerol, 3 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride, Deoxyribonuclease I (10mg/mL), and 1x Bug Buster) lysed for 4h at 4°C on a rocker, and further define by centrifugation. The supernatants of the mutants were collected and frozen at -80°C.

Initial expression of the protein by use of IPTG was observed with a longer wild-type GST-TRX$^{2976\text{-}3726}$ construct (amino acids 2976-3726 at the C-terminal end) at approximately 120kDa in molecular weight (**Figure 5**).

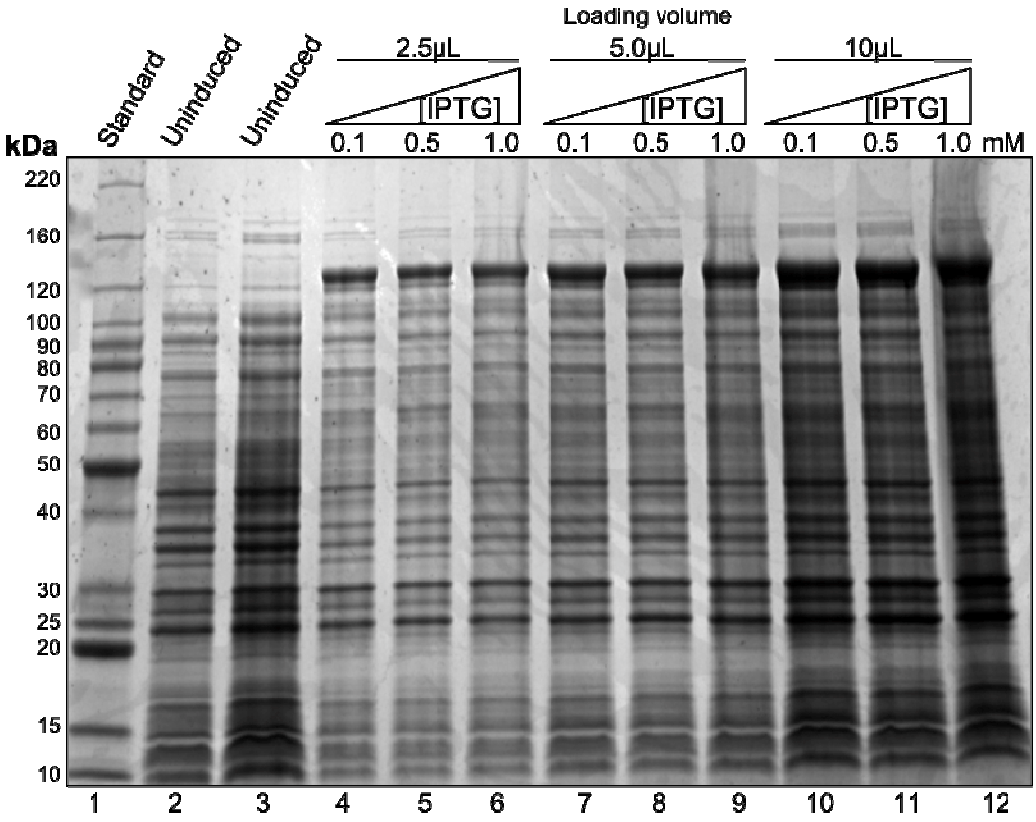


Figure 5. Optimization protein expression with isopropyl-1-thio-β-D-galactopyranoside (IPTG). Varying levels of IPTG were added to induce expression of the GST-Trx$^{2976\text{-}3726}$ construct to observe optimal levels of expression.

The initial optimal concentration of IPTG for induction was decided at 500μM due to minimal variability in expression level above 500μM. Later during large scale expression the induction concentration was increased to 1mM to maximize protein expression.

Assays to maximize the solubility of the protein were conducted with the identified induction concentration. Although the expression tests were performed with GST-TRX$^{2976\text{-}3726}$, we had difficulties solubilizing the protein. We continued the remaining experimental assays with the shorter ~54kDa GST-TRX^{C251} construct. Although no expression tests were performed for optimization, we assumed the conditions used for GST-TRX$^{2976\text{-}3726}$. Wild-type GST-TRX^{C251} expressed at 500μM was used for these assays, but was expressed at 1mM in the large scale purification. Varying concentrations of NaCl salt and Sarkosyl detergent in lysis buffer (50 mM Tris, pH=7.3, 10% Glycerol, 3 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride) were tested (**Figure 6**).

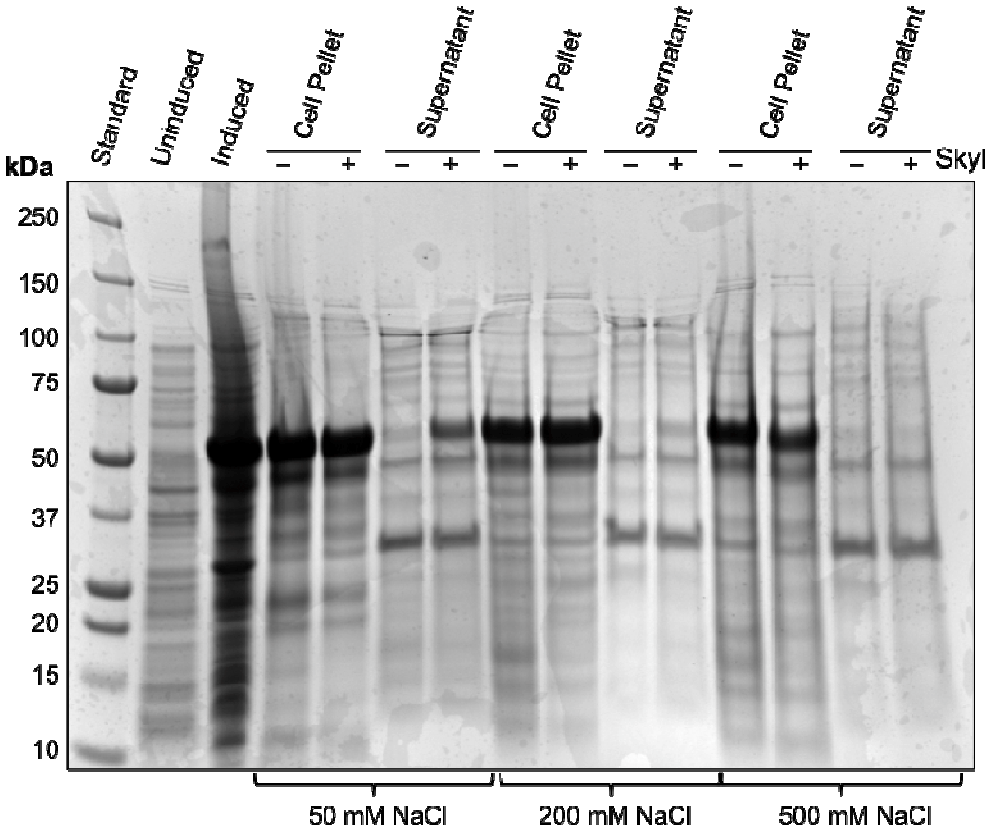


Figure 6. Optimizing the solubility conditions of the GST-TRX^{C251} wild-type protein. Varying conditions included sodium chloride (NaCl) concentrations and the addition of Sarkosyl (Skyl) detergent. GST-TRX^{C251} is expected to run at approximately ~54kDa. The solubilized protein would be expected in the supernatant sample and insoluble fraction would remain in the cell pellet.

The results from this assay indicate 50mM NaCl and 1.5% Sarkosyl detergent optimal for maximum solubility. However, later large-scale purifications suggested that the 50mM NaCl was possibly interfering with the GST tag efficiency and the 1.5% Sarkosyl was disturbing the integrity of the GST-TRX^{C251} protein, thus the Sarkosyl was removed and the salt concentration was raised to 150mM to meet the minimum recommended condition for GST activity.

Small-scale purification of GST-Trx^{C251} mutants by Glutathione S-transferase (GST) pull-down assay

To clarify positive results found in the preliminary radioactive assays with crude protein, small-scale purification procedures by use of glutathione agarose beads (Pierce Glutathione Agarose). The glutathione agarose beads are beads of agarose that have been cross-linked to glutathione and stored in a water and sodium chloride slurry solution. The GST-tag on the protein of interest is able to bind to the glutathione attached to the beads and thus removed from the crude *E. coli* mixture. After a series of washes, the GST-tagged protein of interest can be eluted from the beads by incubation with an eluting buffer containing a competing concentration of glutathione. Small-scale purifications were conducted to obtain a sufficient amount of protein to perform confirmation activity assays. Once confirmed, positive mutants for dimethylation activity would be purified on a large scale.

100μL of glutathione bead slurry was decanted into fresh and chilled 1.5mL centrifuge tubes, centrifuged (4000xG, 4°C, 3 minutes) and then washed three times. The beads were washed with column buffer (50 mM Tris, pH=7.3, 150 mM NaCl, 10% Glycerol, 3 mM dithiothreitol, 1μM $ZnCl_2$), with a centrifugation step (4000xG, 4°C, 3 minutes) following each wash. 200μL of crude *E. coli* lysate of GST-TRX^{C251} fusion proteins were incubated at 4°C, rocking for 4h with 100μL of glutathione bead slurry that had been pre-washed and had the supernatant removed as previously described. After incubation, the agarose bead solution was centrifuged and washed three times with buffer A as

described above. Upon the removal of the supernatant of the final wash, 200µL of eluting buffer B (50 mM Tris, pH=8.0, 150 mM NaCl, 10% Glycerol, 3 mM dithiothreitol, 1µM $ZnCl_2$, 10mM glutathione) was added to the beads and incubated for 1h with rocking, at 4°C. After incubation, the beads were then centrifuged and washed again with the eluting buffer B three times. Each supernatant after incubation and the end beads were run on a Tris-glycine SDS (sodium dodecyl sulfate) gel for analysis. We expect to see purified protein in the supernatant right after the eluting step (**Figure 7**).

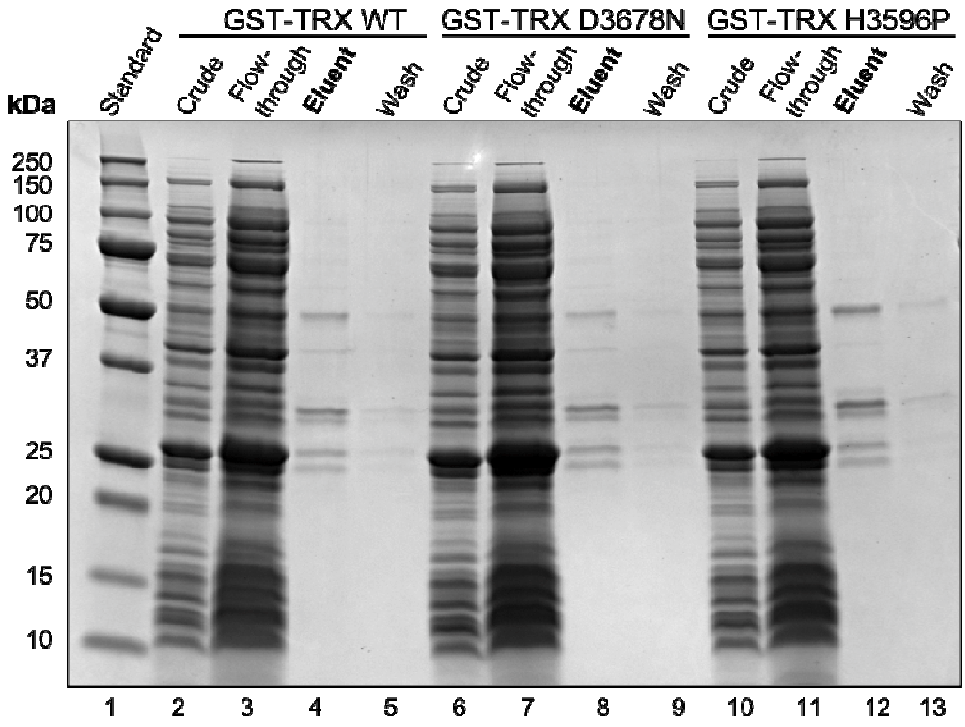


Figure 7. Small-scale purification by use of Glutathione Agarose pull-down. The SDS-PAGE gel shows successful purification with the wild-type and mutant proteins.

Although there was successful purification of the wild-type and mutant protein seen in **Figure 7**, quantitative examination of concentration proved to be inconclusive. There was not sufficient protein to read concentration to use for

experimental assays, thus large-scale purification by affinity column was used for wild-type and mutant GST-TRX^{C251} proteins.

Large-scale protein purification

GST-TRX^{C251} proteins were grown large-scale and were expressed in *Escherichia coli* (Rosetta II, Novagen) by growing cells with the plasmids at 37°C in a one-liter solution of Terrific Broth media with 50μg/mL of carbenicillin. The cells were then induced at 15°C with 1mM isopropyl-1-thio-β-D-galactopyranoside for 24h. Cells were obtained and resuspended in a lysis buffer (50 mM Tris, pH=7.3, 150 mM NaCl, 10% Glycerol, 3 mM dithiothreitol, 0.1 mM phenylmethylsulfonyl fluoride, and EDTA- free protease inhibitor mixture (Roche Applied Science)), lysed by microfluidizer machine, and further clarified by centrifugation. The supernatants of the GST-TRX and GST-MLL1 proteins were collected and passed over a glutathione-sepharose column (GSTrapTM FF column, GE-Healthcare), and eluted with reduced glutathione (**Figure 8**). The fractions containing GST fusion protein were combined and dialyzed with three changes of lysis buffer (without the protease inhibitors)[27,32].

Full-length human WDR5, RbBP5, Ash2L and DPY-30 proteins were each expressed in *E. coli* (Rosetta II, Novagen) and then purified by passing over a nickel affinity column, GST- TEV cleavage during dialysis, a second pass over the nickel affinity column followed by gel filtration chromatography[27,32].

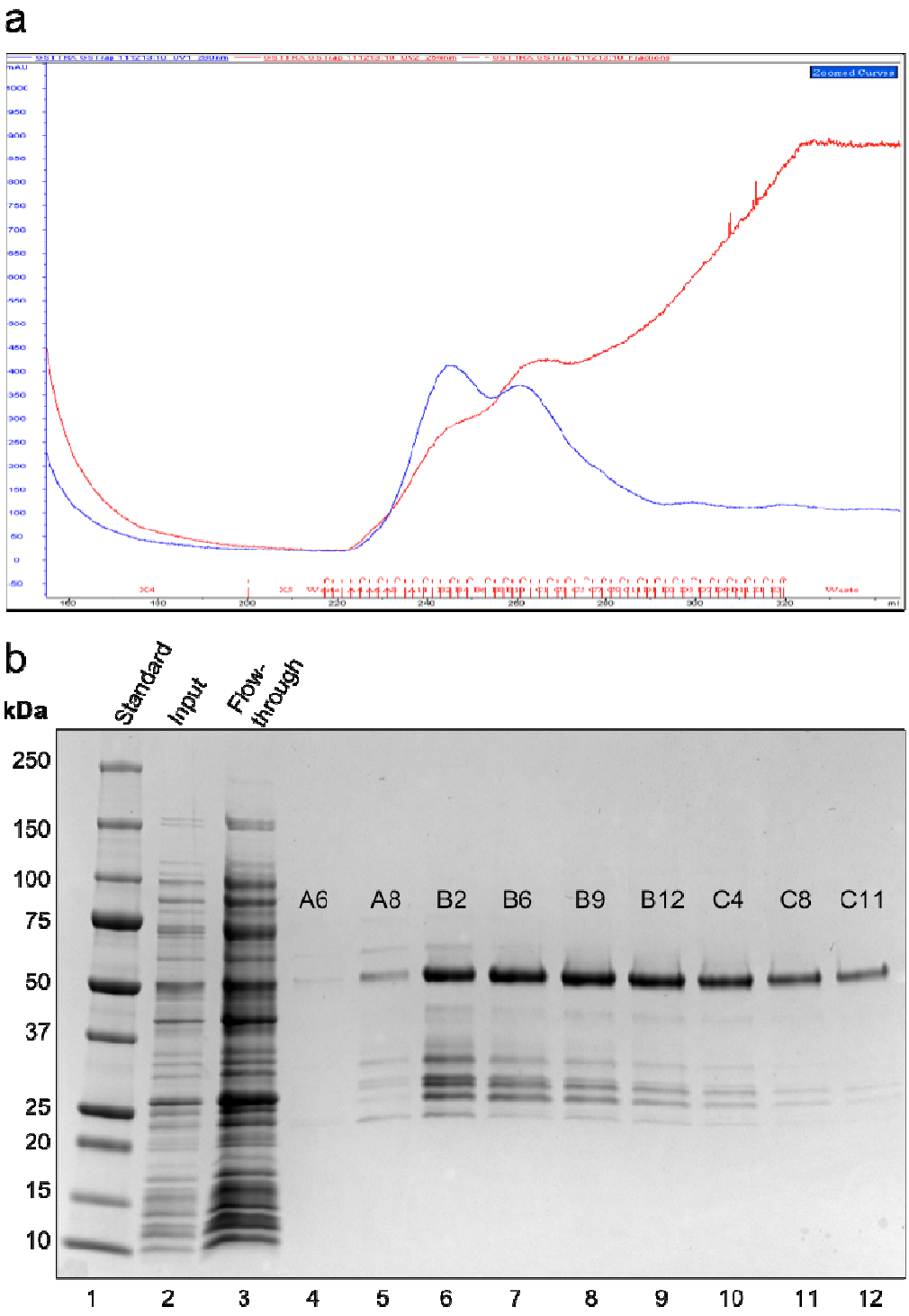


Figure 8. GST-TRX^{C251} wild-type purification by GST affinity column. a) Chromatogram of the first run of the crude protein lysate over the GSTrap 5mL column b) Corresponding fractions from peak seen on chromatogram on gel. Fractions A8-C11 were combined for dialysis.

Glutathione S-Transferase (GST) Pull-Down Assay

GST-pull-down assays were conducted by incubating 100μL GST fusion protein to 100μL of pre-washed (three times with buffer containing 50 mM Tris, pH=7.3, 150 mM NaCl, 10% Glycerol, 3 mM dithiothreitol, 1μM $ZnCl_2$, centrifuged at 4000xG for 3 minutes) glutathione-agarose beads for 3h at 4°C. Then the beads were washed again with the same wash buffer three times, then 200μL of eluting buffer (50 mM Tris, pH=8.0, 150 mM NaCl, 10% Glycerol, 3 mM dithiothreitol, 1μM $ZnCl_2$, 10mM glutathione) was added to the beads and incubated for 1h at 4°C. The beads were then centrifuged and washed again with the same wash buffer three times. Each supernatant after incubation and the end beads were run on a Tris-glycine SDS gel for analysis.

MALDI-TOF Mass Spectrometry Methyltransferase Assays

MALDI-TOF (Matrix-assisted Laser Desorption/Ionization-Time of Flight) assays were conducted with incubating 7.56μM GST-TRX^{C251} with 250μM S-adenosylmethionine and 10μM histone H3 1-20 residue peptide at 16°C in 1x assay buffer (50 mM Tris, pH=8.5, 200 mM NaCl, 3 mM dithiothreitol, 5 mM $MgCl_2$, 5% glycerol). The reactions were quenched at nine different time points with 0.5% trifluoroacetic acid. The quenched samples were then mixed with α-cyano-4-hydroxycinnamic acid. The mass/charge intensity was measured by MALDI-TOF mass spectrometry for each time point at State University of New York-Environmental Science and Forestry, Syracuse, NY. mMass was used to integrate peak areas and to define relative amounts of unmodified, mono-, di- and trimethylated peptides. The total percent methylation represented by mass/charge

over time was plotted using Microsoft Excel. Samples at the 24 hour time point were plotted for comparison using SigmaPlot.

Radioactive screen for dimethylation activity

TRX mutants and controls were incubated with ^{3}H S-adenosylmethionine and 250μM unmethylated or monomethylated histone H3 1- 20 residue peptide, 4μM human WRAD at 15°C in 1x assay buffer (50 mM Tris, pH=8.5, 200 mM NaCl, 3 mM dithiothreitol, 5 mM MgCl2, 5% glycerol) and an EDTA-free protease inhibitor mixture for 8h. Reactions were quenched with 1x SDS buffer. Reaction mixtures were then loaded into a pre-cast Nu- PAGE BisTris Gel and run at 200 Volts for 30 minutes. The gel was then stained, destained and a picture was taken. The gel was then soaked in enhancer fluid, dried and then exposed to Kodak Biomax film.

Results

MLL1 is highly conserved among Vertebrates and Invertebrates

A ClustalW multiple sequence alignment was performed to identify the conserved residues among the orthologs of MLL1 in a variety of vertebrates and invertebrates (**Figure 4 and Appendix**). As shown in the alignment, the conserved residues are highlighted to represent the high level of conservation amongst the orthologous organisms of MLL1. The majority of the conservation was seen in the SET/post-SET domain, as well as in the *Win* (WDR5 Interaction) motif, previously identified as a core motif in the interaction between MLL1 and WDR5[32]. A conserved critical arginine residue found in the *Win* motif that is essential for the complex to form is completely conserved throughout the organisms in the alignment[31].

Upon further observation of the sequence alignment (**Figure 4 and Appendix**) 20 residues were identified to have high conservation among vertebrates and differential high conservation among invertebrates at the same position exemplifying an interesting evolutionary switch. These identified amino acids, shown in the sequence alignment, have been mapped onto the TRX homology model based on the human MLL1 crystal structure **(shown in yellow and marine blue in Figure 9**).

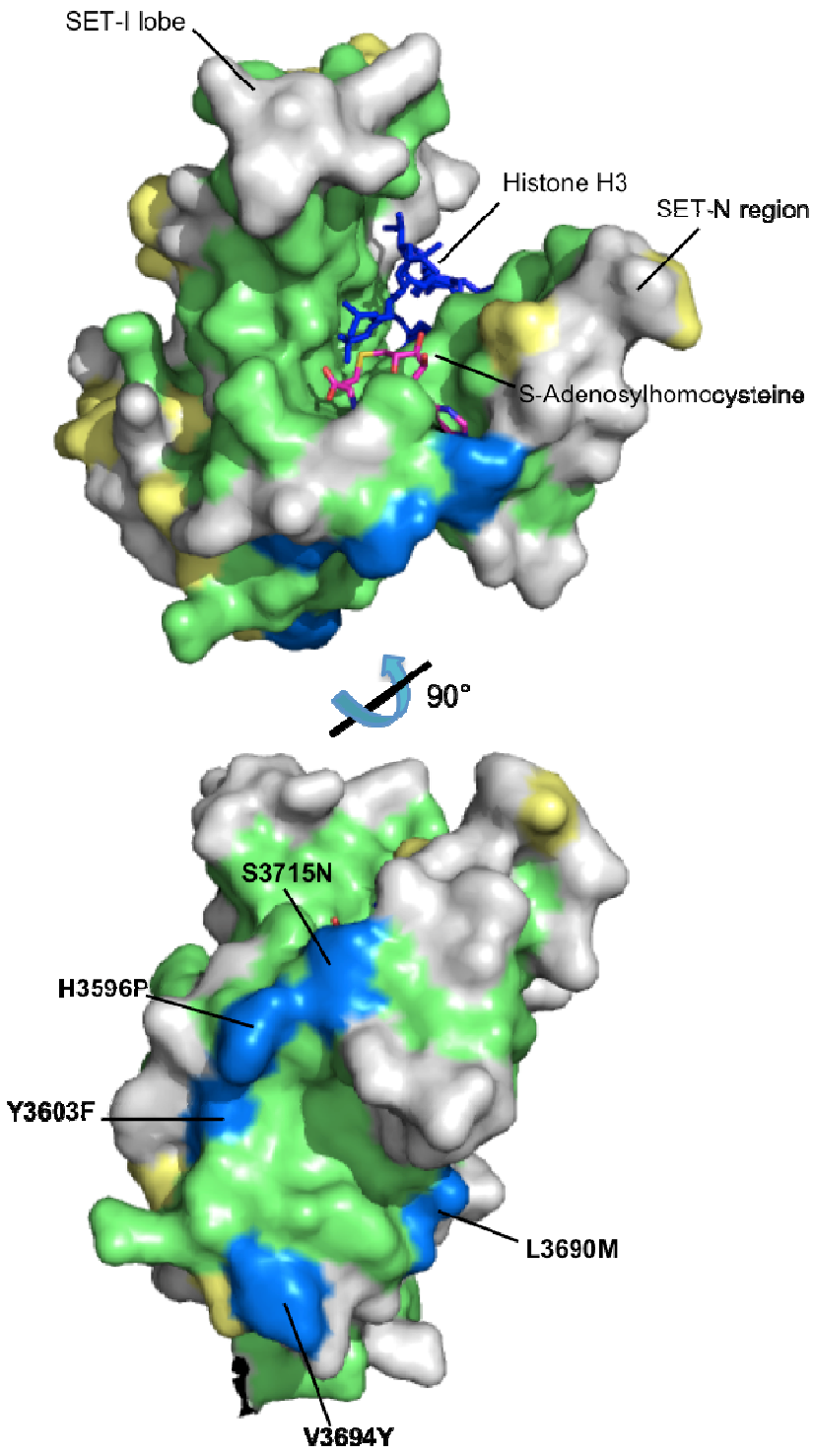


Figure 9. Crystal structure homology model of TRX based on the known crystal structure of MLL1. Methyl donor end-product S-adenosylhomocysteine (magenta) and histone H3 peptide (blue) are labeled in the known first active site. The surface of MLL1 (grey) has conserved residues (green), evolutionary switch mutants (yellow) and identified gain-of-function mutants (marine blue) mapped on the surface.

The mutants that appeared positive for dimethylation in the radioactive screen have been mapped onto the homology model using PyMOL, shown in **Figure 9**. It is important to note that the suspect mutants appear to cluster on the backside of the SET-N region. Other research in the Cosgrove lab concerns mutations in MLL2 that are related to Kabuki Syndrome, a rare disorder characterized by a spectra of physical abnormalities and intellectual disabilities[34]. However, the Kabuki mutations cluster on the back of the SET-I helix, opposite to the area of the identified dimethylation mutants[35]. While unexpected, there is a visible crevice in the region of the dimethylation mutations, possibly indicating the active site of dimethylation. In order to confirm the positive mutations, the mutant proteins were grown, purified and will be tested for dimethylation both qualitatively and quantitatively. Identification of the sites for dimethylation activation in MLL1 could lead for possible targets for treatment as well as a greater understanding of enzymatic mechanisms of MLL1. The residues shown to be conserved in the *Drosophila melanogaster* ortholog of human MLL1 also may indicate the necessary residues needed to interact with the components of the human complex, WDR5, RbB5, Ash2L and DPY-30.

GST-TRX^{C251} Interacts with human WDR5 and the human WRAD Complex

As observed in the sequence alignment, the TRX protein contains the essential residues present in MLL1 responsible for binding to the WRAD complex. As previously discussed, the *Win* motif and the conserved arginine residue of MLL1 are significant for the binding interaction. MLL1 binds to the WRAD complex

primarily through the arginine residue and WDR5, which is bound to the other three components[32]. Due to the conservation of the arginine residue amongst the orthologs and TRX proteins, we hypothesized the TRX protein would be capable of binding to the human WRAD (hWRAD) complex.

A GST-pull-down assay was conducted to compare binding of TRX^{C251} and MLL1 to hWRAD with the hWRAD pull-down control lane (lane 7) (**Figure 10**).

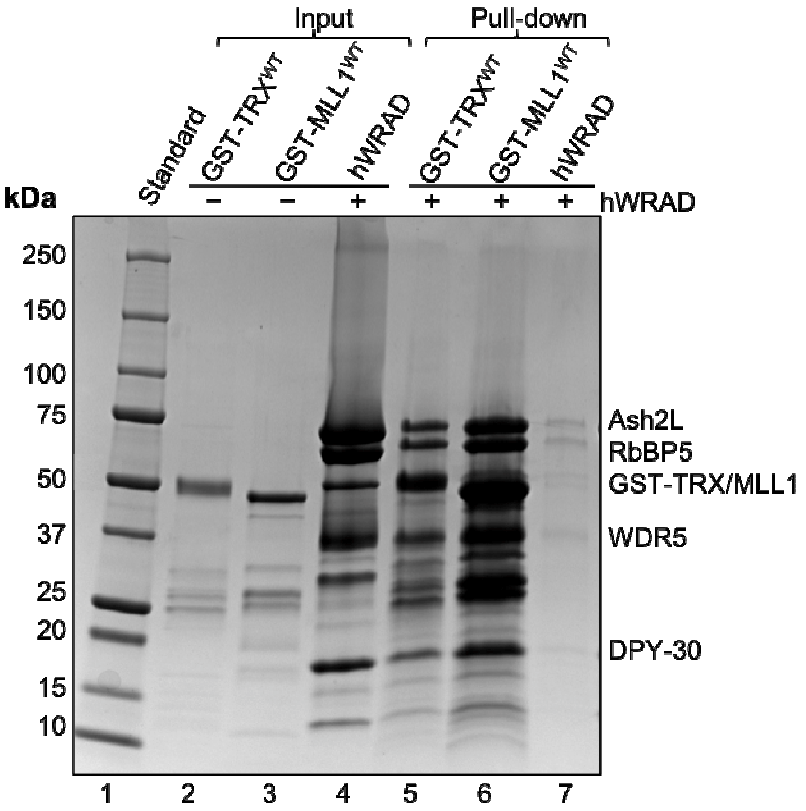


Figure 10. GST-tagged TRX^{WT} with WDR5, RbBP5, Ash2L and DPY-30 (WRAD) showing positive complex formation.

In comparison to the lane containing the input of GST-TRX^{C251}, the lane containing the end beads after having been incubated with GST-TRX^{C251} and hWRAD and washed contains bands at 68kDa, 60kDa, 37kDa, and 20kDa,

representing Ash2L, RbBP5, WDR5, and DPY-30, respectively, as well as a band at 54kDa representative of GST-TRX^{C251}. In contrast, very little hWRAD is pulled down in the absence of GST-TRX^{C251} or GST-MLL1 (lane 7). The presence of these bands leads to the conclusion that TRX^{C251} has the necessary conserved residues in order to bind to the hWRAD complex.

TRX^{C251} has different enzymatic activity than MLL1 in presence of hWRAD

The human MLL1 protein catalyzes H3K4 monomethylation in the presence of methyl donor SAM, optimal buffer conditions and unmodified histone H3 1-20 peptide. In the core complex with hWRAD, the complex performs dimethylation[32]. Due to the highly conserved primary structure between MLL1 and TRX and the observed binding of TRX^{C251} to hWRAD, it was hypothesized TRX^{C251} would have similar dimethylation activity as MLL1 and hWRAD.

MALDI-TOF mass spectrometry time course assays were performed with GST-TRX^{C251} in the presence and absence of hWRAD in order to observe the methylation activity of TRX^{C251} (**Figure 11**). There is little observable difference between the results of the assays with MLL1 and GST-TRX^{C251} without hWRAD (**Figure 11b,f**). The monomethylation activity of MLL1 is detected stronger using the more sensitive radioactive assays, which are shown in the following results section. It is important to note that previous studies showed there was no apparent difference in enzymatic activity observed between GST-MLL1 and untagged MLL1, suggesting there would be no alteration in activity of GST-TRX^{C251} and a cleaved GST-TRX^{C251}.

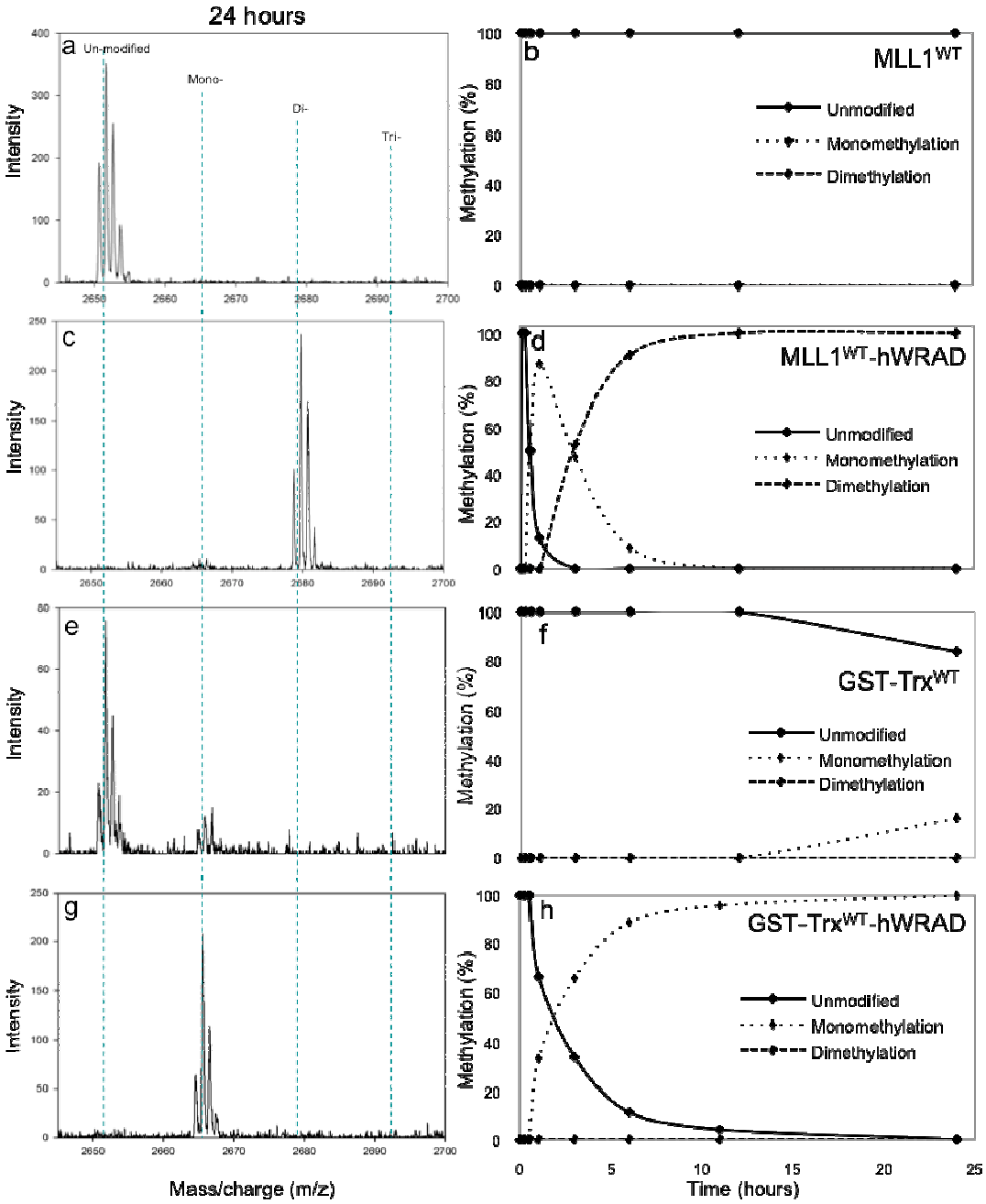


Figure 11. MALDI-TOF spectrometry time course assays. Spectra (left) show methylation present at 24 hours and graphs (right) show unmodified, monomethylation and dimethylation activity as a function of time. a,b) MLL1 wild-type showing no methylation activity c,d) MLL1 wild-type with human WRAD (hWRAD) showing some monomethylation and stimulated dimethylation e,f) GST-TRX wild-type showing some monomethylation at 24 hours g,h) GST-TRX with human WRAD showing stimulated monomethylation, but no dimethylation.

In the assay of GST-TRX^{C251} in the presence of hWRAD, a peak level of monomethylation was observed at 10 hours (**Figure 11h**). In comparison, wild-type MLL1 with hWRAD shows monomethylation peaking at three and a half hours followed by dimethylation, which plateaus at twelve and a half hours (**Figure 11d**). While evidence of dimethylation is weak and inconclusive, it could be that dimethylation activity has a slow rate of activity or that TRX^{251} does not have the necessary residues to be enzymatically active with the hWRAD complex.

Five mutations confer gain-of-function dimethylation activity in TRX^{C251}

As discussed previously in the context of the MALDI-TOF assay, wild-type TRX protein by itself is unable to catalyze dimethylation, catalyzes robust monomethylation activity in the presence of hWRAD. The human MLL1 protein catalyzes H3K4 monomethylation on its own while in the context of hWRAD, catalyzes dimethylation.

To understand their difference, we conducted an assay with monomethylated H3K4 peptide, hWRAD and ^{3}H-SAM, to identify TRX mutants with a gain of dimethylation activity. Gain of function mutations would be detected by an additional radioactive methyl group transferred to the peptide. After screening 20 mutations with the H3K4 monomethylated peptide, we found five TRX mutants that were positive for an increase in dimethylation compared to the wild-type TRX protein (**Figure 9**).

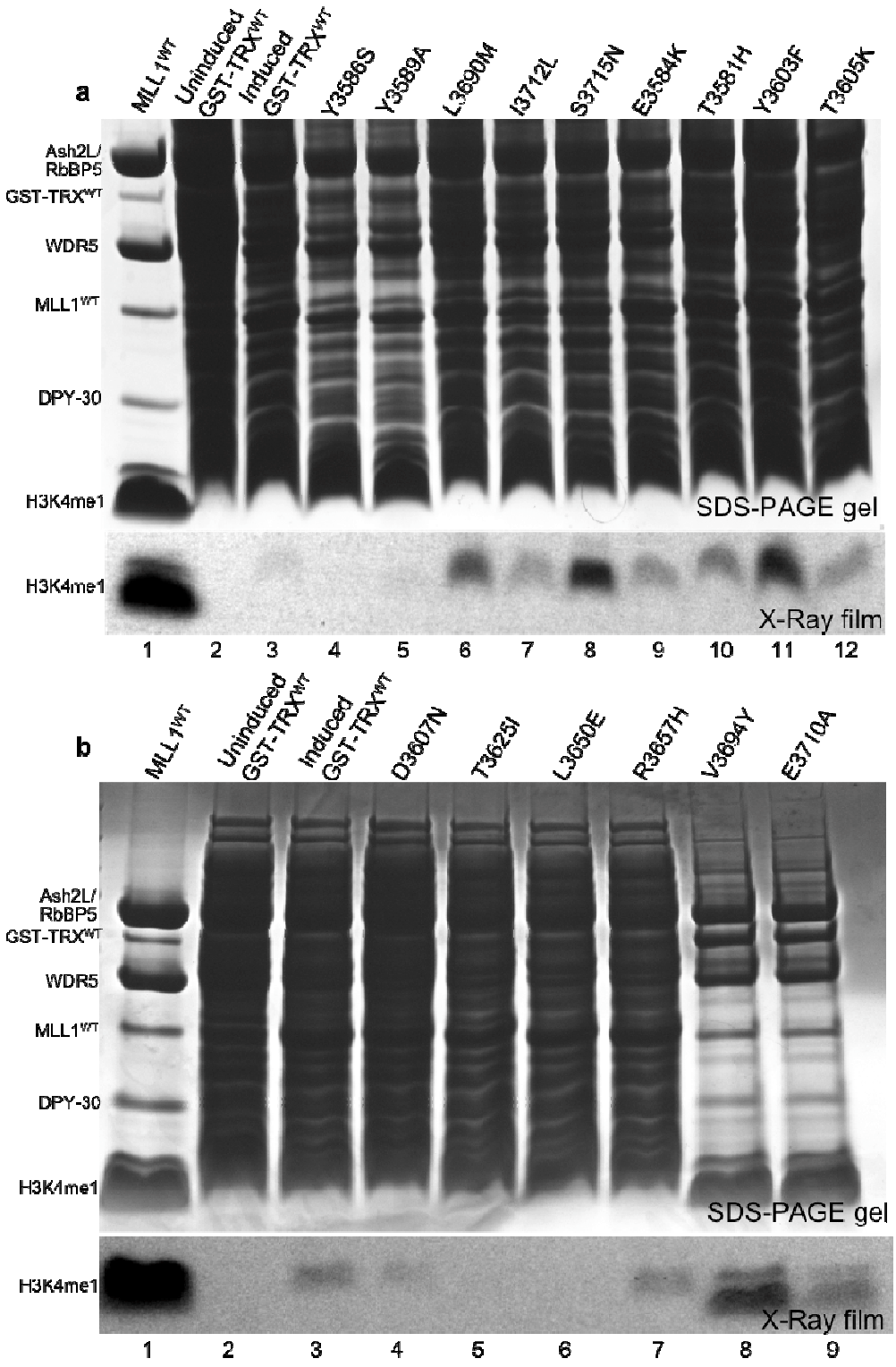


Figure 12 legend on following page

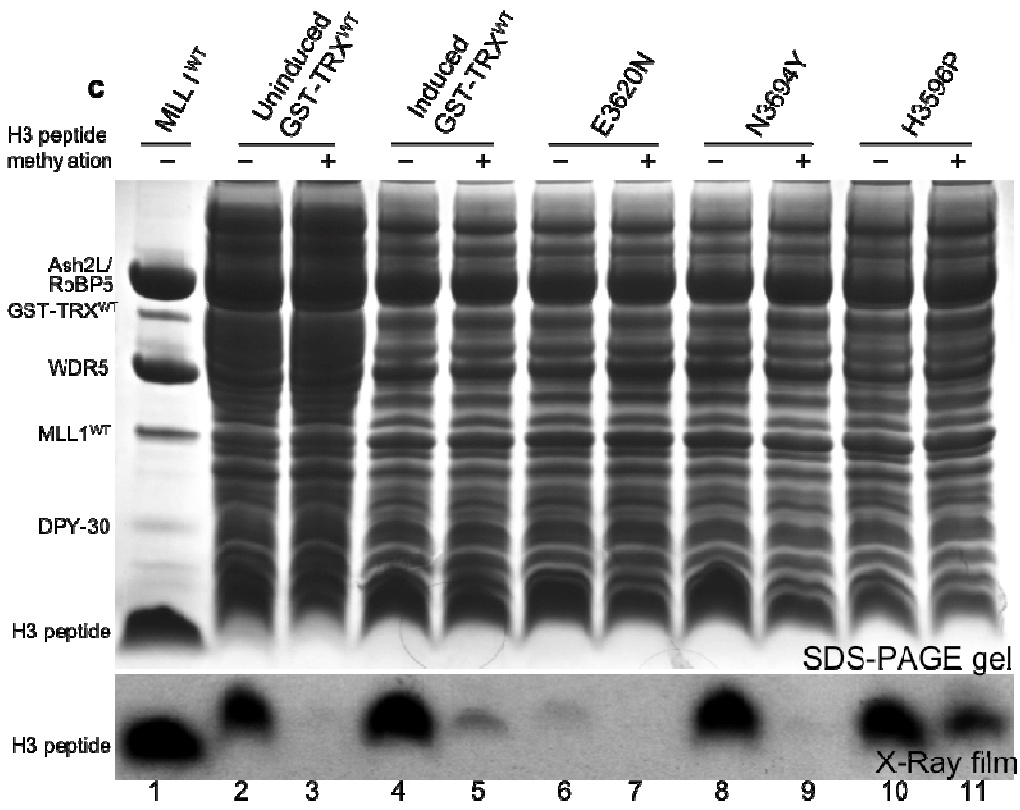


Figure 12. Radioactive gels and X-ray film for screen for dimethylation activity. a,b) Gels and films showing uninduced and induced TRX controls and mutants with human WRAD and monomethylated peptide. c) Gel and X-Ray film showing MLL1^{WT}, uninduced / induced GST-TRX^{C251} and three mutants. The (-) indicates a reaction including an unmodified histone peptide while the (+) indicates a reaction where a monomethylated peptide was added. All reactions were in the presence of human WRAD.

TRX mutants L3690M, S3715N, and Y3603F (**lanes 6, 8 and 11 respectively, Figure 12a**), TRX mutant V3694Y (**lane 8, Figure 12b**), and H3596P (**lane 11, Figure 12c**), show an increase in dimethylation compared to the induced wild-type TRX control with a monomethylated peptide.

However, it is possible that the variability of radioactive signal strength seen in the reactions could be due to expression variability. In future studies to

control for expression level a western blot will be used to normalize the amount of each mutant in the assay.

The positive results from the radioactive assay with the gain-of-function mutants will be optimized with the addition of more controls, the use of affinity column purified protein. We will also probe for gain-of-function activity quantitatively by use of MALDI-TOF assays.

Discussion

Throughout this study, the TRX^{C251} protein was successfully purified and partially characterized in a core complex with human WRAD as a GST fusion protein. Although we were able to perform a variety of assays and experiments to characterize TRX, there are many optimizations to be addressed for purification. The expression assays were performed using the longer GST-$TRX^{2976-3726}$ construct, thus to be sure of the correct conditions for GST-TRX^{C251} we must conduct another expression test. Another purification step to address is cleavage of the GST tag from the fusion GST-TRX^{C251} protein. Although we are able to perform many experiments with the uncleaved tagged protein, the cleaved protein is more similar to what would be found in nature. Preliminary data shows that although we are able to clave the TRX^{C251} protein from the GST tag, the TRX^{C251} precipitates out of solution and becomes undetectable. One possible method to be used is to incubate the GST-TRX^{C251} with hWRAD to form a core complex and then cleave, assuming TRX^{C251} is more stable in the context of hWRAD.

In the future we plan to return to the longer construct to characterize a construct of TRX with a greater number of residues. However the next steps are to optimize the purification and cleavage conditions to increase protein yield and to obtain cleaved TRX protein.

We have shown that MLL1 and that mechanisms for interaction are conserved through billions of years of evolution. The *D. melanogaster* TRX protein is able to form a hybrid core complex with hWRAD. Although TRX is

able to form a hybrid core complex with hWRAD, only stimulation of monomethylation is conserved. This result could be because TRX requires the *D. melanogaster* WRAD (dWRAD) for stimulation of dimethylation activity. The loss of dimethylation stimulation in the hybrid complex may reflect TRX activity *in vivo*, but we do not know its activity without dWRAD. It is possible that the WRAD components in *D. melanogaster* have compensating mutations to retain dimethylation activity seen with the human proteins, which we plan to examine in future investigations. Another possibility is TRX has lost dimethylation activity through the course of evolution. It is also possible MLL1 has evolved to gain dimethylation activity. These questions will be answered through the characterization of the mutants described before as well as by expression and purification of the dWRAD.

The screen for dimethylation activity has produced five possible sites that are required for H3K4 dimethylation activity: H3596P, Y3603F, L3690M, V3694Y and S3715N. We have successfully purified the positive mutants and plan to conduct confirmation assays both qualitatively and quantitatively as described in the results section. Further assays and studies will be conducted to continue to test the enzymatic capabilities of TRX^{C251} and to gain a further understanding of which structural components of MLL1 are required for activity. A greater understanding of the essential structural components of MLL1 enzymatic activity would give us invaluable insight in gene expression mechanisms, leading to the development of innovative therapeutic strategies to manipulate leukemic cells.

References

1. Schuettengruber, B. *et al*. (2011) Trithorax group proteins: switching genes on and keeping them active. *Nature Reviews* **12,** 799-814

2. Goldberg, A. D., Allis, C. D., and Bernstein, E. (2007) Epigenetics: A landscape takes shape. *Cell* **128**, 635-638

3. Kornberg, R. D., and Thomas, J.O. (1974) Chromatin Structure; oligomers of the histones. *Science* **184,** 865-868

4. Luger, K., Mader, A. W. *et al*. (1997) Crystal structure of the nucleosome core particle at 2.8 A resolution. *Nature* **389,** 251-260

5. Cosgrove, M. S. and Wolberger, C. (2005) How does the histone code work?. *Biochem. Cell. Biol*. **83,** 468-476

6. Grewal, S. I. S., and Moazed, D. (2003) Heterochromatin and epigenetic control of gene expression. *Science* **301,** 798-802

7. Heitz, E. (1929) Heterochromatin, Chromocentren, Chromomeren. *Ber. Dtsch. Bot. Ges.* **47,** 274-284

8. Cosgrove, M. S. (2007) Histone proteomics and the epigenetic regulation of nucleosome mobility. *Expert Rev. Proteomics.* **4,** 465-478

9. Peterson, C. L., and Laniel, M. A. (2004) Histones and histone modifications. *Current Biology* **14,** R546-R551

10. Jenuwein, T., and Allis, C. D. (2001) Translating the Histone Code. *Science* **293,** 1074-80

11. Kusch, T. (2012) Histone H3 lysine 4 methylation revisited. *Landes Bioscience* **3(6),** 1-5

12. Pirrotta, V. (1998) Polycombing the Genome: PcG, trxG, and Chromatin Silencing. *Cell* **93,** 333-336

13. Lyko, F. and Paro, R. (1999) Chromosomal elements conferring epigenetic inheritance. *Bioessays* **21,** 824- 832

14. Ingham, P. W. and Whittle, R. (1980) Trithorax: a new homeotic mutation of *Drosophila melanogaster* causing transformations of abdominal and thoracic imaginal segments. *Mol. Gen. Genet.* **179,** 607-614

15. Ingham, P. W. (1983) Differential expression of *bithorax* complex genes in absence of the *extra sex combs* and *trithorax* genes. *Nature* **306,** 591-593

16. Ingham, P. W. (1998) Trithorax and the regulation of homeotic gene expression in *Drosophila*: a historical perspective. *Int. J. Dev. Biol*. **42,** 423-429

17. Lewis, E. B. (1978) A gene complex controlling segmentation in *Drosophila*. *Nature* **276,** 565-570

18. McGinnis, W. and Krumlauf, R. (1992) Homeobox genes and axial patterning. *Cell* **68,** 283-302

19. Ingham, P. W. (1985) Genetic control of the spatial pattern of selector gene expression in Drosophila. *Cold Spring Harbor Symp. Quant. Biol*. **50,** 201–208

20. Mazo, A. M., Huang, D. H., Mozer, B. A., and Dawid, I. B. (1990) The *trithorax* gene, a trans-acting regulator of the bithorax complex in Drosophila, encodes a protein with zinc-binding domains. *Proc. Natl. Acad. Sci. USA*. **87,** 2112–16

21. Breen, T. R., and Harte, P. J. (1991) Molecular characterization of the *trithorax* gene, a positive regulator of homeotic gene expression in Drosophila. *Mech. Dev.* **35,** 113–127

22. Stassen, M. J., Bailey, D., Nelson, S., Chinwalla, V., and Harte, P. J. (1995) The *Drosophila* trithorax proteins contain a novel variant of the nuclear receptor type DNA binding domain and an ancient conserved motif found in other chromosomal proteins. *Mech. Dev.* **52,** 209-223

23. Breen, T. R. (1999) Mutant alleles of the *Drosophila* trithorax gene produce common and unusual homeotic and other developmental phenotypes. *Genetics* **152,** 319-344

24. Katsani, K. R., Arredondo, J. J., Kal, A. J., and Verrijzer, C. P. (2001) A homeotic mutation in the trithorax SET domain impedes histone binding. *Genes Dev*. **15:** 2197-2202

25. Tie, F., Banerjee, R., Saiakhova, A. R., Howard, B., Monteith, K. E., Scacheri, P. C., Cosgrove, M. S., and Harte, P. J. (2014) Trithorax monomethylates histone H3K4 and interacts directly with CBP to promote H3K27 acetylation and antagonize Polycomb silencing. *Development* **141(5),** 1129-39

26. Qian, C., and Zhou, M. M. (2006) SET domain protein lysine methyltransferases: Structure, specificity and catalysis. *Cell. Mol. Life Sci.* **63(23),** 2755-63

27. Patel, A., Dharmarajan, V., Vought, V. E., and Cosgrove, M. S. (2009) On the mechanism of multiple lysine methylation by the human Mixed Lineage Leukemia Protein-1 (MLL1) core complex. *J. Biol. Chem.* **284(36),** 24242-56

28. Cosgrove, M. S., and Patel, A. (2010) Mixed lineage leukemia: a structure-function perspective of the MLL1 protein. *FEBS J.* **277(8),** 1832-1842

29. Hess, J. L. (2004) MLL: a histone methyltransferase disrupted in leukemia. *TRENDS in Mol. Med.* **10(10),** 500-507

30. Lachner, M., and Jenuwein, T. (2002) The many faces of histone lysine methylation. *Curr. Opin. Cell Biol.* **14,** 286-298

31. Dharmarajan, V., Lee, J., Patel, A., Skalnik, D. G., and Cosgrove, M.S. (2012) Structural basis for *Win* (WDR5 interaction) motif recognition in human SET1 family histone methyltransferases. *J. Biol. Chem.* **287(33),** 27275-89

32. Patel, A., Vought, V. E., Dharmarajan, V., and Cosgrove, M. S. (2008) A conserved arginine-containing motif crucial for the assembly and enzymatic activity of the mixed lineage leukemia protein-1 core complex. *J. Biol. Chem.* **283(47),** 32162-75

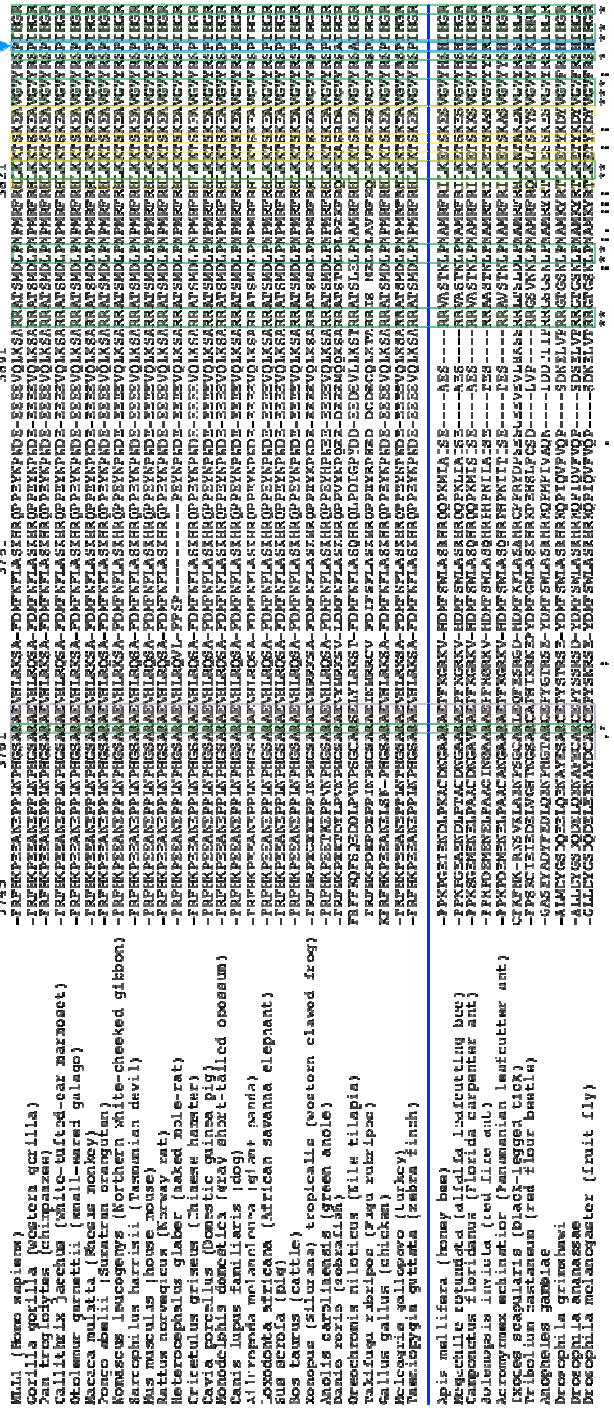
33. Southall, S. M., Wong, P. S., Odho, Z., Roe, S. M., and Wilson, J. R. (2009) Structural basis for the requirement of additional factors for MLL1 SET domain activity and recognition of epigenetic marks. *Mol. Cell* **33,** 181-191

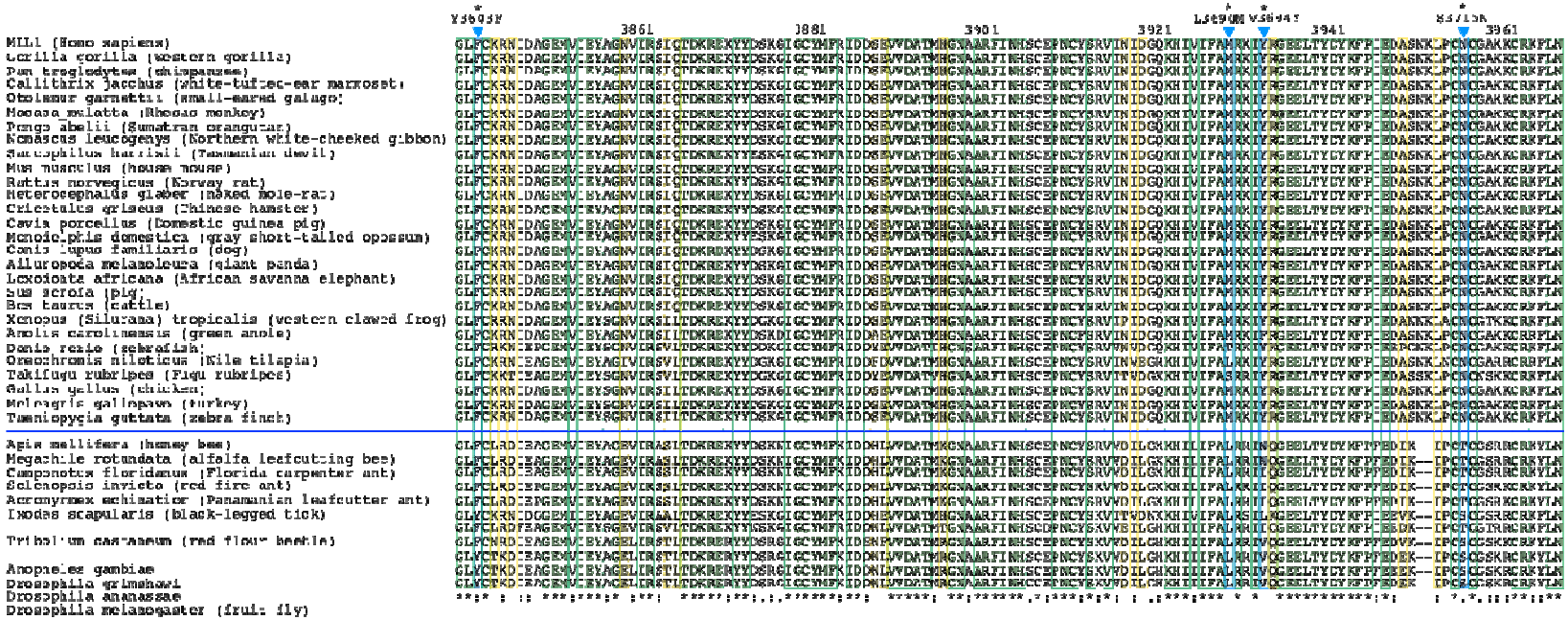
34. Kasdon, B. D., and Fox, J. E. (2012) Kabuki syndrome: diagnostic and treatment considerations. *Ment. Health Fam. Med.* **9(3),** 171-179

35. Shinsky, S. A., Hu, M., Vought, V. E., Ng, S. B., Bamshad, M. J., Shendure, J., and Cosgrove, M. S. (2014) A non-active-site SET domain surface crucial for the interaction of MLL1 and the RbBP5-Ash2L heterodimer within MLL family core complexes. *J. Mol. Biol.* doi: 10.1016/j.jmb.2014.03.011

Appendix

As shown in Figure 4. ClustalW multiple sequence alignment showing the conserved residues (green), proposed evolutionary switch mutants (yellow), and positive gain-of-function mutants (marine blue). The proposed mutants were chosen based on their differences among the vertebrate and invertebrate categories, shown by the blue dotted line. The binding *Win* motif is highlighted (purple). Sequence numbering above the alignment is follows human MLL1 numbering, while each mutant is labeled in TRX numbering (*), showing the amino acid change from TRX to MLL1.




3861
3881
3901
3921
3941
3961
MLL1 (Homo sapiens)
Gorilla gorilla (western gorilla)
Pan troglodytes (chimpanzee)
Callithrix jacchus (white-tufted-ear marmoset)
Otolemur garnettii (small-eared galago)
Macaca mulatta (Rhesus monkey)
Pongo abelii (Sumatran orangutan)
Nomascus leucogenys (Northern white-cheeked gibbon)
Sarcophilus harrisii (Tasmanian devil)
Mus musculus (house mouse)
Rattus norvegicus (Norway rat)
Heterocephalus glaber (naked mole-rat)
Cricetulus griseus (Chinese hamster)
Cavia porcellus (domestic guinea pig)
Monodelphis domestica (gray short-tailed opossum)
Canis lupus familiaris (dog)
Ailuropoda melanoleuca (giant panda)
Loxodonta africana (African savanna elephant)
Sus scrofa (pig)
Bos taurus (cattle)
Xenopus (Silurana) tropicalis (western clawed frog)
Anolis carolinensis (green anole)
Danio rerio (zebrafish)
Oreochromis niloticus (Nile tilapia)
Takifugu rubripes (Fugu rubripes)
Gallus gallus (chicken)
Meleagris gallopavo (turkey)
Taeniopygia guttata (zebra finch)
Apis mellifera (honey bee)
Megachile rotundata (alfalfa leafcutting bee)
Camponotus floridanus (Florida carpenter ant)
Solenopsis invicta (red fire ant)
Acromyrmex echinatior (Panamanian leafcutter ant)
Ixodes scapularis (black-legged tick)
Tribolium castaneum (red flour beetle)
Anopheles gambiae
Drosophila grimshawi
Drosophila ananassae
Drosophila melanogaster (fruit fly)

Summary of Capstone Project

Deoxyribonucleic acid (DNA), and the genes contained therein, contains the entirety of the information the cells in the body require. Although most cells in a multi-cellular organism contain equivalent genetic information, different cells express distinct patterns of these genes allowing each to have their own identity. The differential expression of genetic information is regulated by mechanisms that alter the shape and position of the DNA as opposed to directly altering the genetic information. How the genetic expression profiles are controlled throughout cell division and development is studied in the field of epigenetics, the study of heritable gene expression. Epigenetics could explain how through development each cell-type maintains a distinct and lasting expression profile.

DNA is wrapped around a group of eight small, positively charged proteins called histones. The histone-DNA subunit is called a nucleosome, which regulates the access to DNA. Throughout evolution, eukaryotic organisms have developed mechanisms to regulate the expression of the genetic code by the use of large protein complexes. These large protein complexes are able to add different chemical groups to the histone proteins or onto the DNA itself. The addition of these chemical groups to nucleosomes alters the degree of packaging of the DNA. When tightly packaged, access by other biological molecules responsible for replication or transcription of the genetic material is restricted, and the DNA is therefore considered repressed. However, when loosely packaged, replication and transcription machinery are readily able to access the DNA and

duplicate genetic material for cell division and produce functional genetic products, respectively.

Organisms use a specific group of proteins to transfer of chemical groups to nucleosomes, such as methyl groups. These proteins are referred to as histone methyltransferase and transfer a methyl group from a donor to the receiving histone protein. For example, human Mixed Lineage Leukemia-1 (MLL1) protein is a known histone methyltransferase that transfers methyl groups to histone H3 lysine 4 (H3K4). Lysine is a basic amino acid, a protein building block that can accept up to three methyl groups thus has four varying stages of methylation: unmodified, mono-, di- and tri-methylation. Importantly, different methylation states have different roles; monomethylation of H3K4 is associated with gene silencing, while di-/trimethylation are associated with transcription.

MLL1 is part of an evolutionarily conserved family of proteins and is involved in the regulation of homeotic genes during development through histone methylation. Homeotic genes, or *HOX* genes, are primarily concerned with cell type determination and directing the patterning of body formation during development. The first homeotic genes were discovered in *Drosophila melanogaster* (fruit fly), by the study of the effects of mutations, over expression and under expression of certain genes. These genetic mutations led to a variety of interesting physical mutations of misplaced body segments, thus giving these altered genes the title of homeotic genes. The first gene identified in *Drosophila* as a regulator of homeotic genes was the Trithorax (*TRX*) gene, the *Drosophila* equivalent of the human *MLL1* gene.

Both the MLL1 and TRX proteins are essential during embryonic development. Previous work in the Cosgrove lab has shown that human MLL1 is most active in a complex with at least four other proteins: WDR5, RbBP5, Ash2L and DPY-30 (hWRAD), forming the core complex. Work in other labs on the TRX protein have also shown that there are *Drosophila* homologs of the four WRAD proteins, and is speculated that TRX is likely most active when in complex with these four proteins. MLL1 and TRX share a significant number of conserved residues, many of which make up highly conserved motifs involved in protein binding, suggesting that they behave similar to one another in the biological context. With the high degree of primary conservation of MLL1 and TRX in mind, we hypothesized that TRX would be able to interact with the hWRAD and possibly perform similar reactive abilities as MLL1 in complex with hWRAD. However, before we could examine the capabilities of TRX in the context of hWRAD, we needed to optimize the growth and purification of the TRX protein to have pure material to work with.

The first aim of my project was to optimize the growing and solubility conditions of TRX to ensure we were able to collect as much protein as possible to work with on a reasonable scale. After enough protein was collected, we optimized the purification procedure to establish that we were only testing the protein of interest as opposed to extraneous cell proteins from the growth stage. Once the purified TRX protein had been obtained, I examined the interactive binding capabilities of TRX with hWRAD by the use of pull-down experiments. Pull-downs are qualitative experiments that are used to observe what is binding to

the protein of interest. Our TRX protein has GST (glutathione S-transferase) tag attached to the protein, that can be used to selectively bind the TRX protein to a GST-bead and wash away unwanted cell debris and contaminants. After binding TRX protein, TRX-bead complex is exposed to hWRAD and later visualize by gel electrophoresis what portion of the hWRAD is able to bind to the TRX-bead complex in comparison to MLL1 and the amount of hWRAD that is able to bind to beads. As anticipated, it was observed TRX was able to bind to hWRAD, the formation of the hybrid core complex of human and *Drosophila* components suggest that the molecular mechanisms that account for complex formation have been conserved over 700 million years.

The next stage of the project was to observe the reactivity of TRX in the context of hWRAD. Alone, MLL1 is able to transfer one methyl group to H3K4, while the core complex MLL1-hWRAD is able to transfer two methyl groups. We hypothesized due to the high conservation of TRX and MLL1 subunits, that a complex containing hWRAD and TRX would have similar activity as MLL1. Using a quantitative mass spectrometry experiment measuring the addition of methyl groups to the histone overtime, I found that MLL1 and TRX in the absence of hWRAD were able to transfer one methyl group to H3K4. However, the hybrid TRX-hWRAD complex was unable to transfer two methyl groups.

After observing the unanticipated results of TRX with hWRAD, we wanted to further understand the different in activity between complexes assembled with MLL1 and TRX. I created an amino acid sequence alignment of all the verified MLL1 homologous in other species. The sequence alignment

allowed us to observe the similarities and differences on a primary sequence level. We noticed a clear evolutionary switch between subunits of vertebrates versus invertebrates. For example, at certain positions in the sequence alignment, we noticed that in the majority of vertebrates it was one type of amino acid, whereas in invertebrates it was a different amino acid. I located approximately twenty instances of this evolutionary phenomenon in the alignment. We began to develop a hypothesis suggesting that one of these positions maybe the source of the difference in activity between the human MLL1 and *Drosophila* TRX proteins with hWRAD.

I used an innovative high-throughput screen to mutate each of the twenty amino acids in TRX to the corresponding amino acid in MLL1 to determine if we were able to restore the dimethylation activity of the TRX-WRAD complex. I tested the mutations using a highly sensitive radioactive assay. To our surprise, I identified five out of the 20 mutants that have "gained" the activity of MLL1, to varying degrees. Interestingly, I found that all five mutations cluster on a common surface of the MLL1 protein. We hypothesize that this surface defines the location of the second active that is required for H3K4 dimethylation.

This uncovered knowledge has important implications in our understanding of how the MLL1 protein works. Because the MLL1 protein is involved in development and blood cell differentiation, different mutations and misregulation of the *mll1* gene have been seen in many cases of leukemias, such as acute myelogenic and lymphocytic leukemias, which are common amongst children. With knowledge the MLL1 protein structural features that are required

for dimethylation, we could possibly develop new strategies for treatment of leukemias.