Determination of the Amino Acids Involved in the Interactions Between DPY-30 and ASH2L, Key Components of the MLL1 Core Complex

Carey Stuart

Follow this and additional works at: https://surface.syr.edu/honors_capstone

Part of the Biochemistry Commons

Recommended Citation
https://surface.syr.edu/honors_capstone/260

This Honors Capstone Project is brought to you for free and open access by the Syracuse University Honors Program Capstone Projects at SURFACE. It has been accepted for inclusion in Syracuse University Honors Program Capstone Projects by an authorized administrator of SURFACE. For more information, please contact surface@syr.edu.
Determination of the Amino Acids Involved in the Interactions Between DPY-30 and ASH2L, Key Components of the MLL1 Core Complex

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

Carey Stuart

Candidate for B.S. Degree and Renee Crown University Honors

May 2011

Honors Capstone Project in ______Biochemistry_______

Capstone Project Advisor: _____________________________
(Dr. Michael Cosgrove)

Honors Reader: _____________________________
(Dr. Ramesh Raina)

Honors Director: _____________________________
James Spencer, Interim Director

Date: _____________________________
Abstract

The human version of the DPY-30 protein is homologous to the DPY-30 protein in *Caenorhabditis elegans* (nematode), along with other DPY-30 homologous proteins in other organisms. This protein is involved in dosage compensation of X-linked genes, balancing the levels of expression of these genes between the sexes. The mechanism by which the balancing is carried out varies from organism to organism. For example, in *C. elegans* hermaphrodites (XX), transcript levels of the X-linked genes are cut in half. In *Drosophila*, the genes on the male’s (XY) X chromosome are transcribed at twice the rate of the female’s (XX) genes. In human females (XX), one of the X chromosomes in each cell is inactivated at random. If DPY-30 is absent from the organism, this can lead to XX-lethality. In humans, DPY-30 forms a complex with four other proteins: MLL1, WDR5, RbBP5, and ASH2L. This complex is responsible for methylating histones, particularly histone H3 Lysine 4 (H3K4). Methylation of H3K4 promotes transcription of genes. Similar complexes are found in other organisms. The fact that these proteins are conserved across species indicates how important they are. Within the complex itself, DPY-30 binds to ASH2L. The amino acids responsible for this interaction, however, remain unknown. The purpose of this project is to identify which amino acids are responsible for the binding between DPY-30 and ASH2L. A further extension of this project is its potential anticancer applications. When MLL1 is activated improperly and forms the complex, it can lead to the development of Acute Lymphoblastic Leukemia (ALL). It is thought that if the complex were to be disrupted and broken apart or prevented from forming, the cancerous cell would stop proliferating and die. If the amino acids responsible for the DPY-30 and ASH2L binding were identified, a drug or peptide could be designed to bind to DPY-30 or ASH2L, preventing it from completing formation of the MLL1 Core Complex.

The first step in this experiment was to compare DPY-30 homologs from different species, to see if any amino acid residues were either completely conserved or mostly conserved. Once the conserved residues were identified, one was selected to change, Arginine (R) 54. It was changed to an Alanine (A). Three other single amino acid mutations were made: Arginine (R) 76, Leucine (L) 66, and Leucine (L) 65. All three were changed to Alanines. Once primers with the correct mutation were made, the DNA sequence was put through PCR and transformed into *E. coli* cells. The DNA was extracted, sequenced, and transformed into another *E. coli* strain. A large culture was grown, expression of protein was induced, the cells were lysed, and the protein was collected. Once the mutant protein was purified, it was subjected to multiple tests to determine its binding affinity for ASH2L. For the R54A mutant, the binding interactions were weakened, but not completely inhibited. For the L66A mutant, there was no measurable effect on the binding interactions. It was concluded that Arginine 54 was a much more important residue than Leucine 66, as far as binding affinity was concerned.
# Table of Contents

Acknowledgments...........................................................................................................i

Capstone Project Body....................................................................................................1

Sources Cited and Consulted..........................................................................................22

Appendices....................................................................................................................24

Written Summary of Capstone Project.............................................................................26
Acknowledgements

I would like to thank my Research and Honors Advisor, Dr. Michael Cosgrove, as well as Anamika Patel, Valarie Vought, Venkatasubramanian Dharmarajan, Ilona Chepak, Benny Howard, Diana Zhang, Kyle Fahey, Uchenna Mbawuike, and Melody Sanders for their time, energy, advice, support, and help on this project.

I would also like to thank my Honors Reader, Dr. Ramesh Raina, Dr. Larry Wolf, Dr. William Starmer, Dr. Kari Segraves, and all the people in the Renee Crown Honors Program, especially Hannah Richardson, Steve Wright, and Eric Holzwarth.

I would also like to give a special thanks to my parents, family, and friends for all their constant love and support.

Additionally, I would like to thank the Syracuse University Biology Department, the American Cancer Society, the National Science Foundation, and the Leukemia Research Foundation for providing funding for this research.
Capstone Project Body

Introduction

Human DPY-30 (dumpy-30) is a homolog of DPY-30 found in *Caenorhabditis elegans*. It is found in the nucleus and is expressed in all human tissues. DPY-30 is about 22.4 kiloDaltons (kD) and is a homodimer (Wang et al, 2009). The two identical monomers are each made up of two helical subunits (Wang et al, 2009). This protein earned its name from experiments in which mutations in the protein resulted in a “dumpy” phenotype in the organism, in this case *C. elegans* (Hsu and Meyer, 1994). The human version of the protein is a vital component of dosage compensation machinery and loss of DPY-30 activity results specifically in XX-lethality (Dong et al, 2005). Dosage compensation is a regulatory process that functions at the chromosome level to balance the levels of expression of X-linked genes between the sexes (Hsu et al, 1995). Otherwise, the sex with multiple X chromosomes would receive a two-fold dose of the X-linked genes compared to the other sex, which leads to XX-specific lethality (Dong et al, 2005). This condition, for example, affects *C. elegans* hermaphrodites (XX), *Drosophila* females (XX), and female mammals (XX) (Hsu et al, 1995). Even though the end result is the same, regulation of these X-linked or sex-linked genes, however, is carried out differently from one species to another. In *C. elegans*, the transcript levels produced by each X chromosome in XX hermaphrodites are halved to equal the transcript levels of the X in males (XO) (Dong et al, 2005). In *Drosophila*, the X-linked genes in males (XY) are
transcribed at twice the rate as the two X chromosomes in females (XX) (Dong et al, 2005).

In female mammals (XX), one of the X chromosomes is randomly inactivated so that the X-linked gene expression level matches that of the males (XY) (Dong et al, 2005). This leads to the formation of Barr bodies. A Barr body is the term used to describe the highly condensed, inactivated X chromosome, which is unable to have any of its genes transcribed and/or expressed. In certain animals, DPY-30 is used for other developmental processes: normal body size, mating behavior, correct tail morphology, and coordinated movement (Hsu et al, 1995). The homolog of DPY-30 found in yeast functions as a member of the histone 3 lysine 4 methylation complex, which is the key part of the epigenetic transcriptional control mechanism (Dong et al, 2005). Epigenetics refers to any heritable change in gene expression, meiotic and/or mitotic, that is not actually coded for in the DNA sequence of an organism (Egger et al, 2004). In humans, DPY-30 is able to form a complex with four other proteins, Mixed Lineage Leukemia protein-1 (MLL1), ASH2L, WDR5, and RbBP5 (Crawford and Hess, 2006). ASH2L stands for absent, small, homeotic discs-2-like, WDR5 stands for WD repeat protein-5, and RbBP5 stands for retinoblastoma-binding protein 5 (Patel et al, 2008). The MLL1 complex is necessary for methylation of histones, in particular histone H3 Lysine 4 (H3K4) (Crawford and Hess, 2006). This methylation is important for allowing transcription of target genes to occur within cells (Crawford and Hess, 2006). MLL1 is also vital to the regulation of hox
genes in hematopoiesis and development (Patel et al, 2008). The fact that these core components are conserved across species speaks volumes about their importance in cellular activity (Crawford and Hess, 2006). Within the MLL1 core complex itself, DPY-30 binds specifically to ASH2L (South et al, 2009), yet the identity of the binding site remains unknown (see Figure 1).

Another aspect of this project is its potential therapeutic applications. The MLL1 gene itself has a tendency to undergo reciprocal translocations (Srinivasan et al, 2004), where the gene breaks at a cleavage point and rejoins with another gene from another chromosome. Another occurrence is that MLL1
can undergo partial tandem duplications (MLL-PTD), where certain parts of the
gene are duplicated. Both situations result in fusion proteins, which have been
directly linked to leukemogenesis, specifically Acute Lymphoblastic Leukemia
(ALL) (Srinivasan et al, 2004). When functioning normally, MLL1 is critical for
hematopoiesis and the development of the organism, as the loss of function
results in death. The involvement of this protein in cancer is the ultimate driving
force behind this project. These MLL-PTD proteins still end up forming the
complex with the other four proteins. The idea is that if the formation of the
complex was disrupted, the cancerous cell would not be as active or as potent,
which would help in treating people with the disease. This is why discovering
the amino acids that make up the binding site between DPY-30 and ASH2L is so
important. A peptide or drug could then be used to block formation of the
complex. This same strategy could be employed with any of the other binding
sites within the complex.

The purpose of my research is to determine which amino acid residues
are involved in the binding process between the DPY-30 protein and the ASH2L
protein. A sequence alignment will be performed using various homologs of
DPY-30 in order to determine which residues have been conserved over the
course of evolution. Such homologs include the DPY-30 homolog in C. elegans
and the Saf19p protein, the DPY-30 homolog in S. cerevisiae (Dong et al, 2005).
The rationale behind this is that if certain residues have been completely or
mostly conserved, they must have some importance for the function of the
protein and/or the overall survival of the cell/organism. Once these conserved residues have been identified, wild type DPY-30 protein will be expressed and purified. Then the conserved residues will be systematically mutated one at a time, via site-directed mutagenesis. The mutant DPY-30 proteins will be purified and an assay will be performed after each individual mutation to determine how it affects the binding of DPY-30 to ASH2L. To yield qualitative results, Native (agarose) gel assays will be performed to observe if binding is affected. Analytical Ultracentrifugation (AUC) will be used to obtain quantitative results regarding to the binding affinity the mutated DPY-30 has for ASH2L. Additional analysis of the mutant might include an enzymatic activity assay to see what effect the mutant protein has on the ability of the MLL1 Core Complex to methylate H3K4. This project is expected to provide some insight into the specific role that DPY-30 plays in the MLL1 core complex.

METHODS

Sequence Alignment and Protein Expression and Purification

DNA sequences of DPY-30 homologs/DPY-30-like proteins in 13 different organisms were collected and compared using the National Center for Biotechnology Information’s database (NCBI). A sequence alignment was performed, using Clustal W (Clustal 2.0.8, 2008), to identify completely conserved or highly conserved amino acid residues (see Figure 15, Appendix A). The organisms used for the alignment were *Caenorhabditis elegans* (nematode)
(Accession number: Q10661), *Saccharomyces cerevisiae* (Baker's yeast) (Accession number: NP_010757), *Mus musculus* (house mouse) (Accession number: Q99LT0), *Homo sapiens* (human) (Accession number: Q9C005), *Bos taurus* (cattle) (Accession number: AAI11635), *Ornithorhynchus anatinus* (platypus) (Accession number: XP_001508793), *Monodelphis domestica* (gray short-tailed opossum) (Accession number: XP_001371479), *Gallus gallus* (red jungle fowl) (Accession number: XP_419530), *Strongylocentrotus purpuratus* (purple urchin/purple sea urchin) (XP_001189793), *Pan troglodytes* (chimpanzee) (Accession number: XP_001164263 and XP_00164291), *Apis mellifera* (honey bee) (Accession number: XP_001120012), *Macaca mulatta* (rhesus monkey) (Accession number: XP_001105547 and XP_001105621), and *Canis lupus familiaris* (dog) (Accession number: XP_532923). The 54th amino acid, Arginine, was chosen as the first target for site-directed mutagenesis. The forward and reverse primers with the appropriate mutation were designed and ordered. The primer sequences were amplified using Polymerase Chain Reaction (PCR). The reaction mixtures were put on ice for 2 minutes to cool them below 37°C and then subjected to Dpn I restriction enzyme digestion to get rid of any parental or nonmutated DNA. Each amplification reaction was incubated at 37°C for 1 hour to allow for proper digestion. The mutated DNA was then transformed into *Escherichia coli* DH5α cells, which were then incubated on Lysogeny Broth (LB) agar plates. Once colonies were formed, the blue ones, which represented successful transformations, were selected to use for growing
an overnight 5mL culture in LB broth/media. Five microliters of 1000x carbenicillin were added to each tube used and the tubes were incubated at 37°C in a shaker, making sure the cap was slightly loose to allow oxygen to reach the cells. Only one colony was used for inoculating each tube. Once the cultures were grown, a Miniprep was performed to lyse the cells and extract the DNA. The DNA was stored at -20°C.

The concentration of each DNA sample was measured and a sample of each DNA sequence, along with the T7 primer, was sent to be sequenced at SUNY Upstate Medical University to check for sequence accuracy. Five microliters of sample and one microliter of T7 primer (1 picomole/microliter) were mixed in a PCR tube. Once it was confirmed that the site was successfully mutated from an Arginine into an Alanine, while maintaining the wild-type sequence throughout the rest of the DNA, the mutant DNA was then transformed into the Rosetta 2 PlyS strain of *E. coli*. A small scale culture was grown using Terrific Broth II (TBII, 50mL). Fifty microliters each of carbenicillin (50mg/mL stock solution) and 1x-chloramphenicol (20mg/mL stock solution) were added to the culture. Then 10mL from the small scale culture were used to start a large scale (1L) culture. Expression of the protein was induced through the addition 1 mL of Isopropyl β-D-1-thiogalactopyranoside (IPTG, 1M stock solution). The large flasks were placed on a shaker and incubated at 37°C. Once the proper levels of growth were achieved, the cultures were incubated at 4°C for 1 hour. The cells were then centrifuged and stored as a pellet in the -80°C
freezer, if necessary. The pellet was thawed, if necessary, and was broken open using Lysis buffer and a Microfluidizer.

RESULTS

The mutant protein was collected and purified through a series of steps. The first was a Nickel Affinity Chromatography column (see Figures 2 and 3).

This technique is also called a His-trap run because DPY-30 has a His-tag (six histidines) attached to its N-terminus. This structure has an affinity for Nickel, allowing the protein to bind to the column, trapping it there, while other proteins were washed away. The fractions with the protein were collected and run on an SDS polyacrylamide gel to check the protein’s purity. The gel was stained (coomassie Stain, 30 seconds in microwave, 10 minutes on rocker) and

Figure 2: DPY-30 R54A His-trap Run 1. The R54A protein sample (lanes 1-8) after its first nickel affinity chromatography run. There are a lot of impurities (other proteins) present. The protein marker (PM) or ladder is a set of standards with known molecular weights and is used as a reference point for other proteins run on the gel. The crude extract (CE) is taken from the protein sample prior to purification.
then destained (Destain 1-30 seconds in microwave, discard, Destain 2-30 seconds in microwave, overnight on rocker) to allow the bands to be viewed.

Dialysis was then performed on the collected samples, in order to purify the sample further, remove Imidazole from the sample, but mostly to cleave off the His-tag present on the protein. The cleavage of the His-tag was achieved by the addition of 500 microliters of GST-TEV. The fractions with the most DPY-30 present were combined and then run over another Nickel Affinity Chromatography column. The samples with protein were collected and run on a second SDS polyacrylamide gel to again check purity (see Figures 4 and 5). The gel was stained using the above procedure.

Figure 3: DPY-30 L66A His-trap Run 1. The L66A protein sample (lanes 1-13) after its first nickel affinity chromatography run. Notice the large amount of impurities present.
Figure 4: DPY-30 R54A His-trap Run 2. The R54A protein sample (lanes 1-8) after its second nickel chromatography run. There are considerably fewer impurities present. The crude extract and protein marker were run alongside the samples to use as reference points.

Figure 5: DPY-30 L66A His-trap Run 2. The L66A protein sample (lanes 1-12) after its second nickel chromatography run. There are considerably fewer impurities than after the first run. The crude extract for the second run is a sample of the protein after dialysis is completed.
The fractions were combined and the sample was concentrated down to 5 or less milliliters, if necessary. The concentration of the protein was measured using the Nanodrop spectrophotometer and then the extinction coefficient (0.397) was calculated using the NCBI website, which then led to the calculation of the true concentration of the protein. The sample was then run on a size exclusion chromatography column. The fractions were then analyzed using SDS PAGE (see Figures 6 and 7) and the staining/destaining procedure used above. Some samples were placed in the -80°C freezer, while another was used for performing analysis.

**Figure 6: DPY-30 R54A Gel Filtration Run.**
The R54A protein sample (lanes 1-6) after its size exclusion chromatography run. The sample is virtually impurity free.
Native Gel Assay, Sedimentation Velocity Analytical Ultracentrifugation, and [³H] Methyltransferase Assay

The mutant protein’s binding ability with ASH2L was determined using Native gel assays and Sedimentation Velocity Analytical Ultracentrifugation (SV AUC). These techniques tested the hypothesis that the mutated Arginine (Arg54 or R54) was involved in DPY-30’s interaction with ASH2L. If it was, then the ASH2L and DPY-30 should either not form a complex at all and will remain two separate proteins or have a weakened interaction, as determined by SV AUC. The Native gel was used to yield qualitative results by visually determining whether or not that the two proteins were interacting as normal. The AUC was used to yield quantitative results by measuring the sedimentation coefficient of the proteins and protein complexes. The R54A mutant was then analyzed using

Figure 7: DPY-30 L66A Gel Filtration Run. The L66A protein sample (lanes 1-8) after its size exclusion chromatography run. Essentially all impurities have been removed by this point.
an enzymatic assay, so see what effect, if any, the mutant had on the methyltransferase activity of the WRAD complex. The mutant DPY-30 was combined with the WDR5, RbBP5, and ASH2L and this complex was compared to the wild-type WRAD complex. A radioactive methyl group donor (radioactive S-adenosyl methionine, SAM) (see Figure 8) was used to measure the amount of methylation that occurred.

![Figure 8: S-adenosylmethionine (SAM). The methyl group (-CH₃) on the Sulfur (S) is the methyl group that is transferred from one molecule to another. The regular Hydrogens (¹H) were replaced with (³H), or Tritiums, which are radioactive. An enhancer solution was then used so that the activity could be visualized.](image)

Similar methods and techniques were used for the other mutant DPY-30 proteins that were generated: R76A, L65A, and L66A. The R76A and L65A mutants were carried through to the phase where they were expressed and then frozen as pellets in the -80°C freezer, while the L66A mutant was fully purified and analyzed, with the exception of using it in the enzymatic activity assay. Future experiments would revolve around inducing other single amino acid or maybe even multiple mutations in the DNA and repeating the methods used for
the R54A site-directed mutagenesis. Other areas of research could include determining the mutant’s effect on the activity of the MLL1 core complex and/or what the structure of DPY-30 is while bound to ASH2L.

*The R54A Mutant Negatively Affects DPY-30’s Interactions With ASH2L While the L66A Mutant Does Not*

The sequence alignment identified 13 completely conserved sites between the DPY-30 homologs/DPY-30-like proteins (see Figure 15, Appendix A). Based on the native gel assay and the AUC data, the R54A DPY-30 mutant was still able to bind with wild-type ASH2L. The Native gel showed slight differences between the lane with the wild-type DPY-30 and ASH2L (lane 3) and the lane with the mutant DPY-30 and ASH2L (lane 5). There was a greater amount of free ASH2L left in the lane with the mutant DPY-30 than in the lane with wild-type DPY-30 (see Figure 9). This implies that the binding was weakened as a result of the mutation.
The AUC results showed that the mutant protein had a similar sedimentation coefficient to the wild-type protein for its main peak, but there were also additional peaks present for the mutant protein that were absent from the wild-type protein (see Figure 10a). The peak that normally corresponds to wild-type ASH2L binding with wild-type DPY-30 shifted when the R54A DPY-30 mutant was run on the AUC with wild-type ASH2L (see Figure 10b).

Figure 9: DPY-30 R54A Native gel assay. Lanes, from left to right – wild-type DPY-30 (lane 1), wild-type ASH2L (lane 2), wild-type DPY-30 interacting with wild-type ASH2L (lane 3), mutant DPY-30 (lane 4), and mutant DPY-30 interacting with wild-type ASH2L (lane 5). In the last lane, there is more unbound ASH2L leftover with the mutant DPY-30 than in the lane with the wild-type DPY-30. Concentrations of ASH2L, wild-type DPY-30, and mutant DPY-30 were 0.5 milligrams/milliliter. These results suggest that the R54A mutant is still able to bind with wild-type ASH2L but that the interaction is weaker.
Figure 10a (top): DPY-30 R54A SV AUC. The AUC data for the DPY-30 R54A mutant at three different concentrations: 1.0 milligrams/milliliter, 0.6 milligrams/milliliter, and 0.2 milligrams/milliliter. The main peak for all three runs has a sedimentation coefficient of 1.3. Wild-type DPY-30 has about the same sedimentation coefficient. The wild-type protein does not have any additional peaks, however, as opposed to the mutant protein. These additional peaks might be caused by other conformations of the protein, which were most likely brought on by the mutation that was made. These results suggest that mutation affects DPY-30’s ability to fold properly, which could result in multiple conformations.

Figure 10b (bottom): The AUC data for wild-type DPY-30 by itself, wild-type ASH2L by itself, wt DPY-30 + wt ASH2L, and wt ASH2L + DPY-30 R54A. The concentration for all the proteins analyzed was 0.5 milligrams/milliliter. The peak for the mutant complex has a different sedimentation coefficient than the wild-type complex. This result backs up the result from the native gel assay (above) suggesting that the mutant DPY-30 is still able to bind to wt ASH2L but that the interaction is not quite the same as compared to the wild-type complex.
The enzymatic activity assay that was subsequently performed revealed that the WRAD complex with the mutant DPY30 was not active with the H3 peptide (see Figure 11).

![Coomassie and ³H Fluorography](image)

**Figure 11:** DPY-30 R54A Enzymatic activity assay. Left lane has WRAD complex with wild-type DPY-30. Right lane has WRAD complex with mutant DPY-30. Wild-type complex shows activity with histone H3 peptide while mutant complex does not. Results suggest that DPY-30 does play a role in the complex’s enzymatic functionality.

The results for the L66A mutant, on the other hand, were not nearly as noteworthy. The Native gel did not appear to show any significant differences between the mutant DPY-30-ASH2L complex (lane 5) and the wild-type DPY-30-ASH2L complex (lane 3). This suggested that the binding was not affected by this particular mutation (see Figure 12).
This conclusion was confirmed by the AUC results, in which the mutant DPY-30-wt ASH2L complex had the same sedimentation coefficient as the wild-type complex (see Figure 13). Additionally, the mutant DPY-30 had about the same sedimentation coefficient as the wild-type protein (see Figures 14a-c).
DISCUSSION

The identification of the conserved residues implied that they were of some importance. If not, they would have been less likely to be conserved over the course of evolution across different species. This led to the belief that mutating one of them would have some negative effect on DPY-30’s ability to dimerize and/or bind with ASH2L. The R54A mutation appears to be somewhat

Figure 14a (top): DPY-30 L66A SV AUC. AUC data for DPY-30 L66A, at 0.125mg/mL.
Figure 14b (middle): AUC data for DPY-30 L66A, at 0.25mg/mL.
Figure 14c (bottom): AUC data for DPY-30 L66A, at 0.5mg/mL. All three samples have peaks at 1.31, which is about the sedimentation coefficient for wild-type DPY-30. There are also no additional peaks of any significance present. This suggests that the mutation has no effect on the conformation of the protein.
important regarding the binding sites between DPY-30 and ASH2L. This mutation additionally seems to affect DPY-30’s ability to fold properly, as the mutant by itself had an altered sedimentation coefficient than its wild-type counterpart. One possible explanation is that the mutation caused different conformations of the protein to form, such as an alternate dimer or perhaps even a tetramer. Sedimentation coefficients are dependent upon both the mass and shape/conformation of the molecule being analyzed. In this case, both variables were changed, but the conformation probably had a greater effect than the mass. The mutant DPY-30 also appears to affect the enzymatic activity of the WRAD complex. The inactivity of the WRAD complex with the H3 peptide suggests that DPY-30 might play an important role in the enzymatic activity of this complex.

The L66A mutant, it was concluded, did not seem to play a significant role, or any role, in either DPY-30’s ability to fold properly or DPY-30’s ability to bind to ASH2L. There were no additional peaks present in the mutant AUC data, suggesting that no alternate conformations of the protein were formed. The difference in mass between the mutant and wild-type was also not enough on its own to illicit a change in the sedimentation coefficient.

The results of these experiments illustrated that Arginine 54 was located along the DPY-30 dimer interface as well as the DPY-30-ASH2L interface, whereas Leucine 66 was not. This might explain the two differing results between the mutants. Another explanation might be the nature of the
substitution that was made in each mutant. Arginine is a large, positively charged molecule while Alanine is a small, neutral, and hydrophobic molecule. Leucine is a large molecule also, but is neutral and hydrophobic. If time had permitted, an enzymatic assay would have been performed with the L66A mutant, even though it most likely would not have shown any significant difference in activity from the wild-type MLL complex. It would have also been beneficial to also test the other two mutants that were made, L65A and R76A, to see what kind of results these proteins yielded.
Sources Cited and Consulted


Appendix A

Figure 15: Sequence Alignment of DPY-30 homologs. The 13 completely conserved amino acids have a star beneath their column and are boxed in.
Appendix B

Buffer Preparation

50 mg/ml stock solution of Carbenicillin: 10mL doubly deionized water (ddH₂O) + 500mg Carbenicillin

1x-Chloramphenicol (20mg/mL): 10mL Ethanol + 200mg 1x-Chloramphenicol

1M stock solution of IPTG: 10mL ddH₂O + 2.3831g IPTG

Column Buffer (4L, pH 7.4): 50mM Tris (24.228g), 300mM NaCl (70.128g), 3mM DTT (1.851036g), 30mM Imidazole (8.1696g) + ddH₂O

Elution Buffer (500mL, pH 7.4): 50mM Tris (3.0285g), 300mM NaCl (8.766g), 3mM DTT (0.2313795g), 500mM Imidazole (17.02g) + ddH₂O

Lysis Buffer: 50mL of Column Buffer + 0.1mM Phenylmethanesulfonylfluoride (PMSF) + 1 Ethylenediaminetetraacetic acid (EDTA)-free protease inhibitor cocktail tablet

Destain 1 (2L): 50% 100 proof Ethanol (1L) + 10% Acetic Acid (200mL) + ddH₂O

Destain 2 (2L): 5% 100 proof Ethanol (100mL) + 10% Acetic Acid (200mL) + ddH₂O

Buffer A (2L, pH 8.5): 25mM Tris-HCl (3.0285g) + 19.2mM Glycine (1.42633g)

Agarose gel (0.8%): 50mL Buffer A + 0.4g Agarose

LB (100mL): 3 pellets + ddH₂O

TBII (50mL): 2.5g TBII + ddH₂O

Coomassie Stain (2.5L): 7g Brilliant Blue R + 250mL Acetic Acid + 1125mL 100 proof Ethanol + 1125mL ddH₂O

t-CEP Buffer (600mL, pH 7.5): 20mM Tris-Cl (1.45368g Tris) + 300mM NaCl (10.5192g) + 1μM ZnCl₂ (6μL) + 1mM t-CEP (0.1770g) + ddH₂O
Captions Summary

The human version of the DPY-30 protein is homologous, or very closely related, to the DPY-30 protein in *Caenorhabditis elegans*, along with other DPY-30 homologous proteins in other organisms. This protein is involved in dosage compensation of sex-linked or X-linked genes, balancing the levels of expression between the sexes. The mechanism by which the balancing is carried out varies from organism to organism. For example, in *C. elegans*, hermaphrodites are XX and the transcript levels of the X-linked genes are cut in half. In *Drosophila* (fly), the genes on the male’s (XY) X chromosome are transcribed at twice the rate of the female’s (XX) genes. Transcription is the process of converting a DNA sequence into the corresponding mRNA. In human females (XX), one of the X chromosomes in each cell is inactivated at random. The inactivated X chromosome becomes highly condensed, forming what is called a Barr body. If DPY-30 is not functioning properly or the protein is absent from the organism, this can lead to XX-lethality (death). Death is caused by the overdose of the proteins encoded by the X chromosome genes. This protein is also involved in developmental processes in animals that are XO. Some of these processes include mating behavior, normal body size, correct tail morphology, and coordinated movement.

In humans, DPY-30 forms a complex with four other proteins: MLL1, WDR5, RbBP5, and ASH2L. This complex is responsible for methylating histones, particularly histone H3 Lysine 4 (H3K4). Lysine is one of the twenty amino acids
used as the building blocks for proteins. A histone is a protein that interacts with DNA. There are 5 types of histones: H1, H2A, H2B, H3, and H4. The last four histones form an octamer (2 copies of each individual histone) that DNA strands wrap around to help condense them. This is called a nucleosome. H1 serves as a linker molecule between nucleosomes. Methylation refers to adding a methyl group (-CH₃) to a molecule. Methylation of H3K4 promotes transcription of genes, some of which are necessary for the organism’s survival. Methylation of DNA is an example of epigenetic regulation, meaning it can bring about changes in the expression of certain genes without altering the DNA sequence of the organism. Epigenetics refers to any heritable change in gene expression, meiotic and/or mitotic, that is not actually coded for in the DNA sequence of an organism. These changes are heritable because they can be passed on to the offspring of an organism. Similar complexes are found in other organisms. The fact that these proteins are conserved across species indicates how important they are. Within the complex itself, DPY-30 binds to ASH2L. The amino acids responsible for this interaction, however, remain unknown. The purpose of this project is to identify which amino acids are responsible for the binding between DPY-30 and ASH2L.

A further extension of this project is its potential anticancer applications. The MLL1 gene had been found to break apart and rejoin with either other MLL1 genes or other genes. This leads to the production of abnormal fusion proteins. These fusion proteins lead to improper activity of MLL1. When MLL1 is activated
improperly and forms the complex, it can lead to the development of Acute Lymphoblastic Leukemia (ALL). It is thought that if the complex were to be disrupted and broken apart or prevented from forming, the cancerous cell would stop proliferating and die. If the amino acids responsible for the DPY-30 and ASH2L binding were identified, a drug or peptide, which is a chain of amino acids, could be designed to bind to DPY-30 or ASH2L, preventing it from completing formation of the MLL1 Core Complex.

The first step in this experiment was to compare DPY-30 homologs from different species, to see if any amino acid residues were either completely conserved or mostly conserved. These conserved residues, or amino acids, would most likely be very important to DPY-30’s ability to function. A conserved residue is one that is the same in all or most of the different species when looking at a similar protein in those different organisms. Once the conserved residues were identified, one was selected to change, Arginine (R) 54. It was changed to an Alanine (A). Three other single amino acid mutations were made: Arginine (R) 76, Leucine (L) 66, and Leucine (L) 65. All three were changed to Alanines. Arginine, Leucine, and Alanine are three more of the twenty amino acids. Once primers, short segments of DNA, with the correct mutation were made, the DNA sequence was put through Polymerase Chain Reaction (PCR) and transformed into \textit{E. coli} cells. PCR amplifies a specific segment of DNA, while a transformation is the term used when foreign DNA is being introduced into an organism. The DNA was extracted, sequenced, and transformed into another \textit{E.}
coli strain. The DNA was extracted by breaking open the cells. Sequencing refers
to determining the exact sequence of a DNA segment to make sure it is correct.
In this case, the DNA was being checked for the presence of the correct
mutation, while the rest of the sequence was wild-type, or not mutated. A large
culture of bacteria was grown, expression of protein was induced, the cells were
lysed or broken open, and the protein was collected. Once the mutant protein
was purified, it was subjected to multiple tests to determine its binding affinity
for ASH2L. The purification served to isolate the mutant DPY-30 protein from all
the other proteins present in the cells, so that experiments could be run on a
fairly pure sample. Otherwise, these other proteins might interfere with the
results, making them unreliable and difficult to interpret. For the R54A mutant,
the binding interactions were weakened, but not completely inhibited. For the
L66A mutant, there was no measurable effect on the binding interactions and
the proteins interacted with each other as they normally do. It was concluded
that Arginine 54 was a much more important residue than Leucine 66, as far as
binding affinity was concerned.

It would have been interesting to examine the results from the other two
mutants that were generated, if there had been time to test them. It certainly
seems that R54 is one of the amino acids involved in the binding of ASH2L by
DPY-30, while L66 is not. If time had permitted testing of the other two
mutations, as well as making additional mutations, more of the amino acids
responsible for the binding might have been identified. This could have
potentially led to complete identification of all the necessary amino acids, which in turn could have led to total inhibition of binding between the two proteins. This might have been achieved by changing all the important amino acids into Alanines or just deleting the important amino acids. If complete inhibition of binding was achieved, a peptide or drug with the appropriate sequence could have been designed and tested to see how well it interacted with DPY-30 and, subsequently, blocked ASH2L from binding with DPY-30.