Distribution Agreement

In presenting this thesis as a partial fulfillment of the requirements for a degree from Emory University, I hereby grant to Emory University and its agents the non-exclusive license to archive, make accessible, and display my thesis in whole or in part in all forms of media, now or hereafter now, including display on the World Wide Web. I understand that I may select some access restrictions as part of the online submission of this thesis. I retain all ownership rights to the copyright of the thesis. I also retain the right to use in future works (such as articles or books) all or part of this thesis.

Shahdabul Faraz April 1st, 2014

Purification, Activity and Crystallization of a SET-domain Protein

by

Shahdabul Faraz

Anamika Patel, PhD
Research Adviser

Xiaodong Cheng, PhD
Thesis Adviser

Department of Biology

Patricia Marsteller, PhD
Committee Member

Matthew Weinschenk, PhD
Committee Member

2014

Purification, Activity and Crystallization of a SET-domain Protein

By

Shahdabul Faraz

Anamika Patel, PhD
Research Adviser

Xiaodong Cheng, PhD
Thesis Adviser

An abstract of
a thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2014

Abstract

Purification, Activity and Crystallization of a SET-domain Protein

By Shahdabul Faraz

The protein SET10, found originally in *Schizosaccharomyces pombe,* is a member of SET domain proteins known for its methylation of protein (particularly histone) lysine residues. Although researchers have uncovered the activity and structure of various SET domain proteins, not much is known about SET10. We, therefore, began to work with this protein, focusing three main goals: to successfully purify the protein, to characterize its activity, and to uncover pertinent structural information. We first transformed a plasmid containing the SET10 gene into *Escherichia coli* cells. We then induced the expression of this protein within the cells. After cellular lysis via sonication, we obtained an impure solution containing the protein. We used several methods of fast protein liquid chromatography to obtain purified protein.

To characterize activity, we first performed a peptide-pull down assay that showed that SET10 was able to bind to the histone H3 peptide. We also performed mass-spectrometry experiments, which showed that there was an increase in H3 peptide methylation activity with an increase in reaction time and temperature. The monomethylation peaks, however, were small, suggesting that perhaps SET10 was a slow-acting enzyme and/or that the H3 peptide was not the ideal *in vitro* substrate. Lastly, we performed a radiometric assay that showed that individual free peptides were not ideal substrates for SET10. Rather, there was an increased methylation activity when using individual histones or histone octamers as substrates.

We then performed a trypsin digestion experiment, which showed that the SET10 protein did not have a smaller, more stable domain. We then set up crystallization screens

and were able to obtain crystal “hits” in two distinct conditions. Our crystals diffracted to a resolution of 7Å, which was not high enough to obtain atomic details. Our data, however, allowed us to identify that the cubic crystal space group was P23. Moving forward, our goal is to obtain higher quality crystals that provide higher resolution diffraction data. In order to improve crystal quality, we can perform additive screens, slow down crystal growth, use a different cryopreservant or dehydrate the crystals.

Purification, Activity and Crystallization of a SET-domain Protein

By

Shahdabul Faraz

Anamika Patel, PhD
Research Adviser

Xiaodong Cheng, PhD
Thesis Adviser

A thesis submitted to the Faculty of Emory College of Arts and Sciences
of Emory University in partial fulfillment
of the requirements of the degree of
Bachelor of Sciences with Honors

Department of Biology

2014

Acknowledgements

This research project would not be possible without the unwavering guidance of many people around me. I would like to thank Dr. Xiaodong Cheng, who graciously invited me to work in his laboratory. I have received a warm welcome from the very beginning, and I am immensely grateful for that. I would like to also thank my research advisor Dr. Anamika Patel, who has been a very accommodating and helpful mentor. She trained me diligently and always took time to answer my questions. If I am fortunate enough to mentor other students in the future, I will surely look to her as an exemplary model of how to conduct myself. I would also like to thank the other members of the laboratory, especially Yusuf Olanrewaju and Dr. John Horton, for their advice and support.

Additionally, I owe a large debt of gratitude to Dr. Patricia Marsteller and Dr. Matthew Weinschenk. Over the past four years, I have worked closely with both of them. They are two of the most enthusiastic and genuinely caring faculty at Emory University. I thank them for their mentorship over the years and for agreeing to be a part of my thesis committee.

Table of Contents

INTRODUCTION

The overall purpose of this study can be divided into three main categories: to purify the protein SET10, to characterize its *in vitro* activity and to uncover important structural information. The protein SET10 in *Schizosaccharomyces pombe* is a SET-domain containing protein. Although many other SET domain-containing proteins have been widely investigated, the relevant literature pertaining to purification, structure and activity of the SET10 protein is virtually non-existent. As such, a principle aim of this study is to narrow the knowledge gap that exists within the current literature.

The SET domain is a well-conserved sequence that was initially characterized in three *Drosophila melanogaster* proteins: Su(var)3-9, Enhancer-of-zeste and Trithorax (Dillon et al., 2005). This 130-140 amino acid sequence has been found in all studied eukaryotic organisms as well as in some bacteria and viruses (Dillon et al., 2005). The domain aids in the transfer of a methyl group from S-adenosyl-L-methionine (AdoMet) to the side chains of lysine residues in both histone and non-histone proteins (Dillon et al., 2005). Depending on the particular lysine residue that it methylates, the SET domain can be involved in a diverse range of functions including transcriptional silencing and activation, recruitment of proteins to DNA damage sites and regulation of euchromatic and heterochromatic DNA (Dillon et al., 2005).

Since the discovery of the SET domain, the structures and activities of numerous SET domain proteins have been reported. Studies have showcased the roles of some of these proteins in the pathogeneses of various diseases, such as colorectal and hepatocellular carcinomas (Hamamoto et al., 2004). These studies have helped to identify the SET domain as a key linker between the basic sciences and medicine. Our specific

protein of interest, SET10, is 547 amino acids long and weighs ~62.3 kDa (UniprotKB and Pombase databases). Several other pieces of information including the grand average of hydropathicity (GRAVY), estimated half-life and extinction coefficients have been reported in existing databases (UniprotKB and Pombase databases).

SPECIAL OBJECTIVES AND STEPS

The first objective of this study was to extract the SET10 protein so that subsequent activity assays and structural experiments could be performed. The *E. coli* cells were used to grow the protein. The techniques and procedures that were used as part of this objective were transformation, small and large-scale culture growth, fast protein liquid chromatography (FPLC) and sodium dodecyl sulfate-polyacrylamide gel electrophoresis (SDS-PAGE).

The second objective of this study was to characterize the *in vitro* activity of the SET10 protein. The techniques and the procedures that were used as part of this objective are matrix-assisted laser desorption/ionization (MALDI), peptide pull-down assays and radiometric assays.

The third, and final, objective of this study was to uncover relevant structural data for the SET10 protein. After purification of the protein, crystallization trays were set up. In the event of successful retrieval of crystals, X-ray crystallography and structure prediction software were used. A trypsin digestion experiment was performed to determine the stable domains of the protein.

RESULTS

Purification

Recombinant Gene Expression: We performed the steps of recombinant gene expression and were able to express the SET10 protein in *E. Coli* cells. The plasmid we used during transformation (uptake of exogenous genetic material from the surroundings through the plasma membrane) contained the SET10 gene under the control of a lac operon promoter. After successful transformation and eventual colony growth, we successfully grew a "small scale culture" and a "large scale culture." With respect to cellular growth during the "large scale culture," we were able to reach an optical density of 0.5 at 600 nm. The optical density, or absorbance, of our cell-containing mixture was used as a convenient proxy for the concentration of *E. coli* in that mixture.

Protein Induction: We were able to induce the SET10 gene via the addition of Isopropyl β-D-1-thiogalactopyranoside (IPTG). ITPG's structure mimics that of allolactose, which naturally induces the transcription of the lac operon. As previously mentioned, the SET10 gene in our plasmid was under the control of a lac operon promoter.

Sonication: The sonicator has a probe that was in direct contact with our sample. The probe releases ultrasonic sound waves that cause pressure changes within the liquid (Desjouy et al., 2013). This leads to what is known as cavitation, the formation of small gas bubbles within the liquid (Desjouy et al., 2013). These bubbles are continuously forming and collapsing, creating enough energy to break the cells (Desjouy et al., 2013).

Fast Protein Liquid Chromatography: The supernatant collected after sonication and centrifugation contained the protein SET10, but it also contained a variety of *E. coli* proteins and impurities that needed to be filtered out. We were able to achieve a very high purity of our protein mixture by using FPLC. This method of purification relies on the fact that different components of our mixture solution have their unique affinities to the mobile phase of the column (flowing buffer) and the solid stationary phase (Sheehan & O'Sullivan, 2004). The unique affinities are due to differences in size, charge, shape and hydrophobicity (Sheehan & O'Sullivan, 2004). Although all liquid chromatography employ the above basic principles, there are many types of columns that may be used.

The first one we employed was called the His-Tag column, which is a nickel-based column that can bind to polyhistidine tags. The SET10 protein was tagged with 6 histidine amino acid residues, making the column an appropriate choice for purification (Schmitt et al., 1993). During this run, the FPLC machine mixes our impure solution with a low-imidazole washing buffer. This allows our protein and other histidine-rich proteins to bind to the column, while the flow though (containing several impurities) is collected and eventually disposed as waste (Schmitt et al., 1993). The elution buffer used during this column contains a much higher concentration of imidazole as compared to the washing buffer (Schmitt et al., 1993). Imidazole's structure mimics the side chains of histidine residues (Schmitt et al., 1993). As such, they compete with the binding sites on the column, leading to displacement of the SET10 protein and other histidine-rich proteins (Schmitt et al., 1993). We successfully collected our fractions from this column.

The second column run that was used is known as the Q column. The low salt C1 buffer was first used by the machine to mix and load our impure protein mixture. The

isoelectric point of our protein is pI=5.03, meaning that the protein has no net charge at this pH (ProtParam Database). If the pH of the buffer is below the isoelectric point, then the protein has a net positive charge. If, on the other hand, the pH of the buffer is above the isoelectric point, then the protein has a net negative charge. Since the pH of our buffer was around 7.4, our protein had a net negative charge. Since the Q column is positively charged, SET10 and other negatively charged proteins were able to bind to the column (Kopaciewicz et al., 1983). Other proteins were washed through and eventually disposed as waste (Kopaciewicz et al., 1983). The high salt elution buffer disrupted the electrostatic interactions between the protein and the column, allowing the proper collection of the SET10 protein (Kopaciewicz et al., 1983). We had to be careful in order to ensure that our peak did not represent deoxyribonucleic acid (DNA) impurities (Sweder et al., 1988). DNA, being more negatively charged than the protein, binds more strongly to the Q column (Chandra et al., 1992). As such, a higher concentration of salt is required to disrupt the DNA-column interactions (Sweder et al., 1988). This means that the DNA, if present, will likely appear after the protein peak (Sweder et al., 1988). Even though this mixture was much more pure than our starting mixture, some protein impurities were still be present.

After running the Q column, we ran one last step of purification known as the sizing column run. The purpose of this column is to separate the proteins by their size (Regnier, 1983). Larger proteins elute out faster and earlier than the smaller proteins (Regnier, 1983). This is because these proteins are too large to be able to travel through the bead channels inside the column (Regnier, 1983). By bypassing these channels, they are able to move quickly through the column because they “see” less volume within the

column (Regnier, 1983). The smaller proteins go through the bead channels and, therefore, take longer to flow out (Regnier, 1983). In effect, they "see" more volume within the column. We were able to successfully purify our protein as can be seen by the chromatograms and the gel analysis.

Gel and Chromatogram Results: How do we know that the peaks on our chromatogram actually correspond to the SET10 protein? Also, how can we assess the purity of our samples after doing the column runs? In order to answer the above questions, we took samples corresponding to different positions on the chromatogram and ran them on SDS-PAGE.

As can be seen from the gel picture (Figure 1A), there are 15 wells for samples to run on. The first well on the left represents the standard lane that contains molecules of known sizes. The second and third lanes contain the uninduced sample and the sample induced with IPTG. The fourth lane contains the crude lysate, which is the supernatant that was left over after sonication. The next 11 lanes represent different samples that were collected and correspond to the points across the entire absorbance peak. Based on the standard lane and the fact that our protein is ~62.3 kDa, we saw our protein SET10 in all of the sample lanes. We collected the samples from lanes 10 to lanes 14. Lanes 5 to 8 had significant impurities, so those samples were not used during the Q column run. We suspected that the protein impurities in lanes 5 to 8 may be due to the presence of heat shock proteins, whose expression is increased in the presence of high temperatures or other stress signals.

After the Q column run, samples were again taken corresponding to the clear peak seen in Figure 1A. Looking at the picture of the gel (Figure 1B), the right most lane represents the standard lane. The first lane to its left represents the Q column input, which was the partially purified mixture obtained after the His-Tag column run. The next lane was the "flow through" that represents the impurities that did not bind to the Q column during the loading stage of the run. The rest of the samples (lanes 4 to 15) represent purified samples that were collected from the run. As can be seen, there are still some impurities especially in lanes 10 to 15. Overall, however, the samples obtained are significantly more pure than our pre-FPLC mixture or the one obtained after the His-Tag column run. The samples corresponding to lanes 6 to 10 were collected and used for the sizing column. After the sizing column, a gel was run using the samples corresponding to the observed peak in Figure 1C. Looking at the gel (Figure 1C), the right most lane represents the standard lane. All of the other lanes represent samples from within and around the absorbance peak. Samples corresponding from lanes 4 to 10 were collected because they represented the most pure mixtures. The final SET10 concentration we obtained was 10.6 mg/mL.

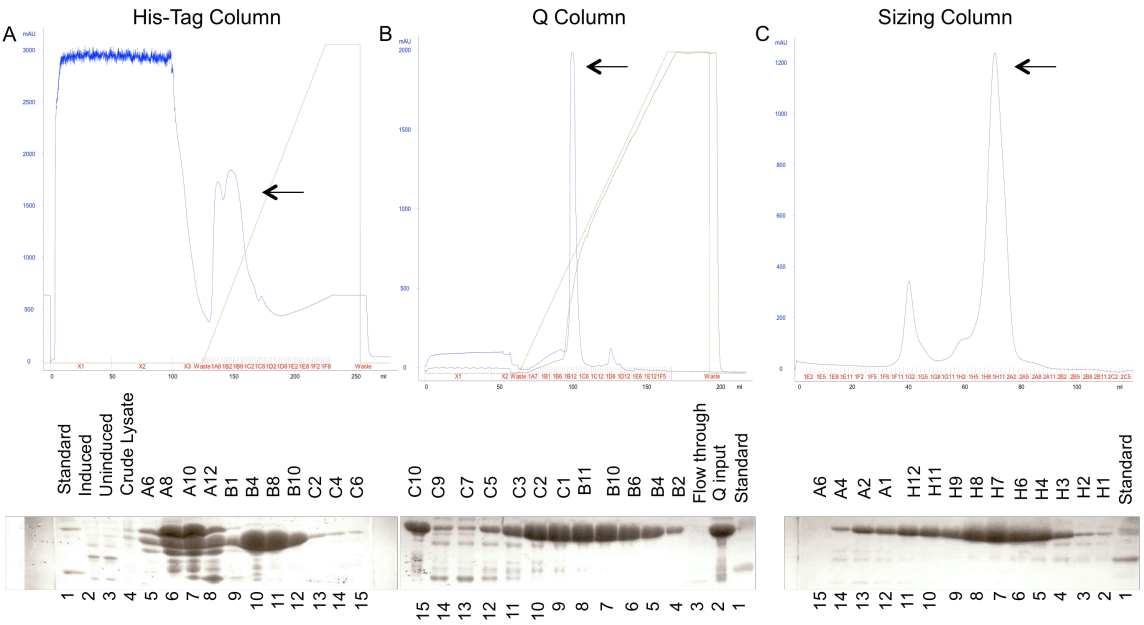

Figure 1. The chromatograms and gels associated with the following: A) His-Tag run B) Q-column run and C) Sizing column run. The gel corresponding to each run is shown below each chromatogram.

Characterizing Protein Activity

Once we were able to successfully purify the protein, our next goal was to characterize its activity. In order to devise experiments, we asked ourselves several important questions. What is this protein's ideal substrate? Is it a slow-acting or a fast-acting protein? Under which temperatures does it work most efficiently? We then devised experiments to answer each of these questions.

Peptide Pull Down Assay: Peptide pull-down assays were performed in order to establish whether or not SET10 was actually able to bind to the histone H3 (1-21)

peptide. These assays are an invaluable tool that researchers use to elucidate protein-protein interactions. Our H3 peptide (bait protein) was tagged with a biotin group that allowed it to bind to the beads used during the experiment. If the SET10 protein was indeed able to bind to the H3 peptide, then a bead-biotin-peptide-protein complex would form. It can, therefore, be seen when gels are run. Light SET10 protein bands were seen in all the trials except in the control tube, which had no peptide (data not shown). This suggests peptide-dependant binding, meaning that the SET10 protein cannot bind to the beads on its own. Rather, it binds to the H3 peptide, which then uses its attached biotin group to bind to the beads. This confirmed to us that our protein was able to bind to the H3 peptide.

Matrix-Assisted Laser Desorption/ Ionization (MALDI) Experiments: Knowing that the protein SET10 actually binds the H3 peptide, our next goal was to uncover whether SET10 was a fast-acting or slow-acting protein and under what conditions (time and temperature) lysine methylation takes place. If monomethylation is present, each peak within this region should have a corresponding peak 14 amu (weight of a methyl group) downstream. If dimethylation and trimethylation are present, we should notice peaks 28 amu and 42 amu downstream, respectively. We did not notice any such discernable methylation peaks when the reaction was allowed to run for 3 hours at room temperature (~25°C). This suggested that perhaps SET10 was a slow-acting protein and more time was required for methylation, that the temperature was too low, that the H3 peptide was not an ideal substrate, or a combination of these factors.

We replicated the above experiment, but we increased the reaction time to 5 hours and performed trials at both room temperature and 30°C. As seen from Figure 2, we saw very small monomethylation peaks around ~2729 amu when the reaction was allowed to continue for 5 hours at both 30°C and room temperature. The peaks, however, did seem to be more defined in the 30°C trial. Since an increase in temperature and time showed promising results, we performed another experiment with three trials. Each of the three reactions was allowed to go overnight (24 hours) and was conducted at three different temperatures (room temperature, 30°C and 37°C). As seen from Figure 3, the trials occurring at 30°C and 37°C showed higher monomethylation peaks than the trial at room temperature. An increase in temperature, in effect, seemed to increase the rate of the reaction, an effect that is consistent with the principles of enzyme kinetics. Dimethylation or trimethylation peaks were not observed.

Lastly, we wanted to note the effect of an increase in time on the methylation status of the H3 peptide. As mentioned before, we suspected SET10 to be a slow-acting protein. If this is true, an increase in time should allow more peptides to be methylated, thereby increasing the size of the methylation peaks. We allowed the reaction to run for the following time periods: 15 minutes, 30 minutes, 1 hour, 5 hours and overnight. For the sake of completeness, each of these time trials was performed at three temperatures (room temperature, 30°C and 37°C). As seen from Figure 4, no methylation peaks were observed after 15 or 30 minutes. The mass-spectrometry identified very small peaks in the 1 hour and 5 hour trials. More visible peaks, however, were seen in the overnight trial. This seemed to confirm our suspicion that SET10 is a s1ow-acting protein that requires a long time to methylate the H3 peptide. Because none of the methylation peaks

were large relative to the parent unmodified peptide peak, we suspected that perhaps the H3 peptide was not the ideal substrate for SET10.

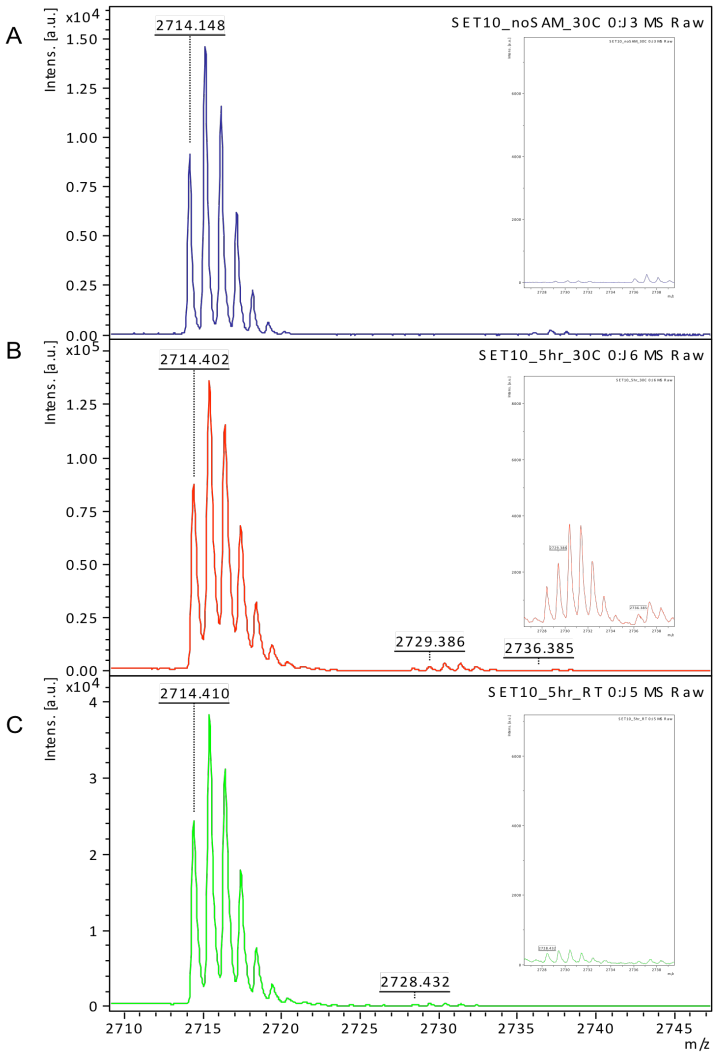

Figure 2. The methylation activity of SET10 during the following conditions: A) No AdoMet added in 30°C temperature, B) 5 hour reaction time at 30°C temperature and C) 5 hour reaction time at room temperature. The smaller panels show a magnified version of the monomethylation peaks.

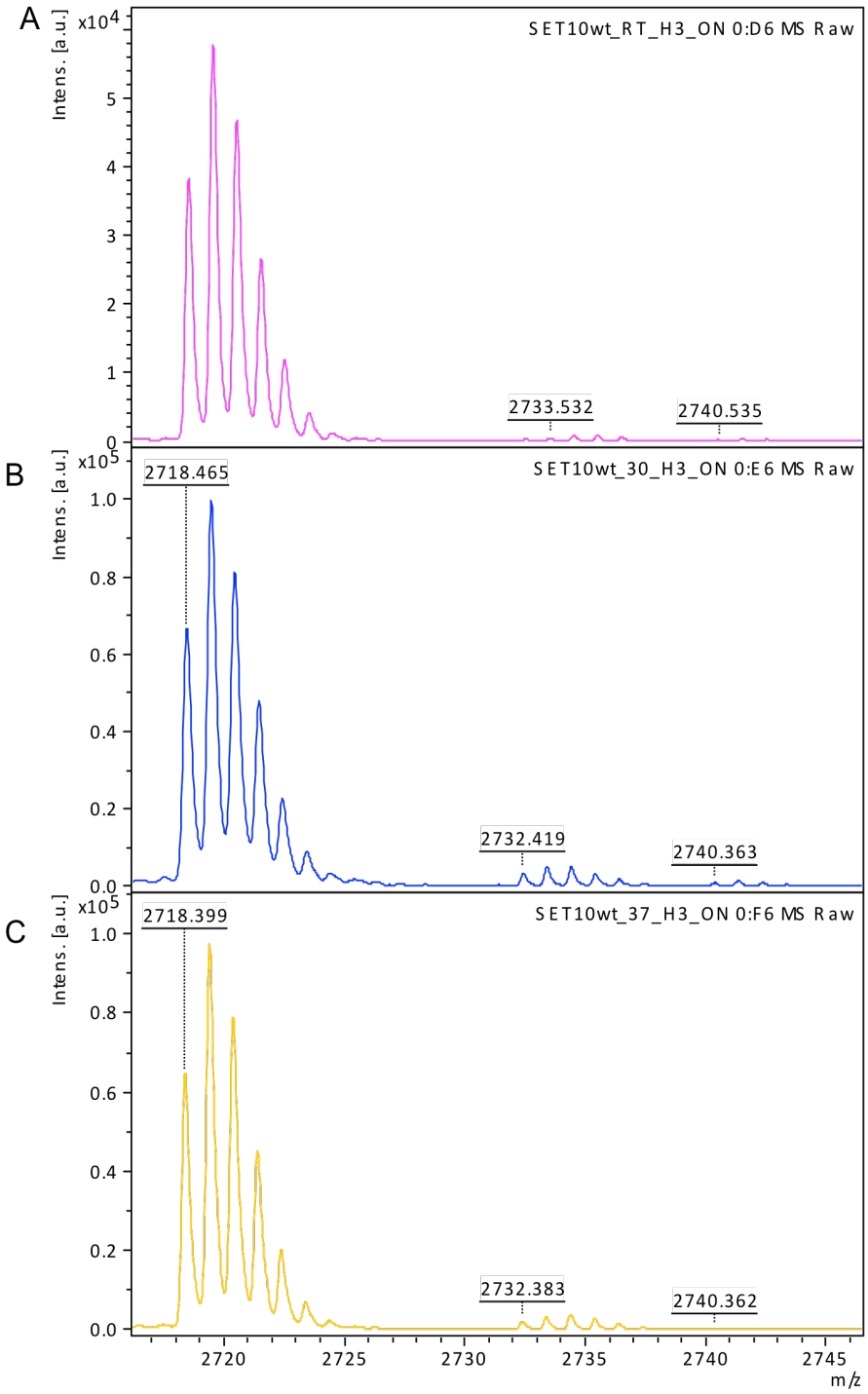

Figure 3. The methylation activity of SET10 during the following conditions: A) overnight reaction at room temperature, B) overnight reaction at 30°C temperature and C) overnight reaction at 37°C temperature.

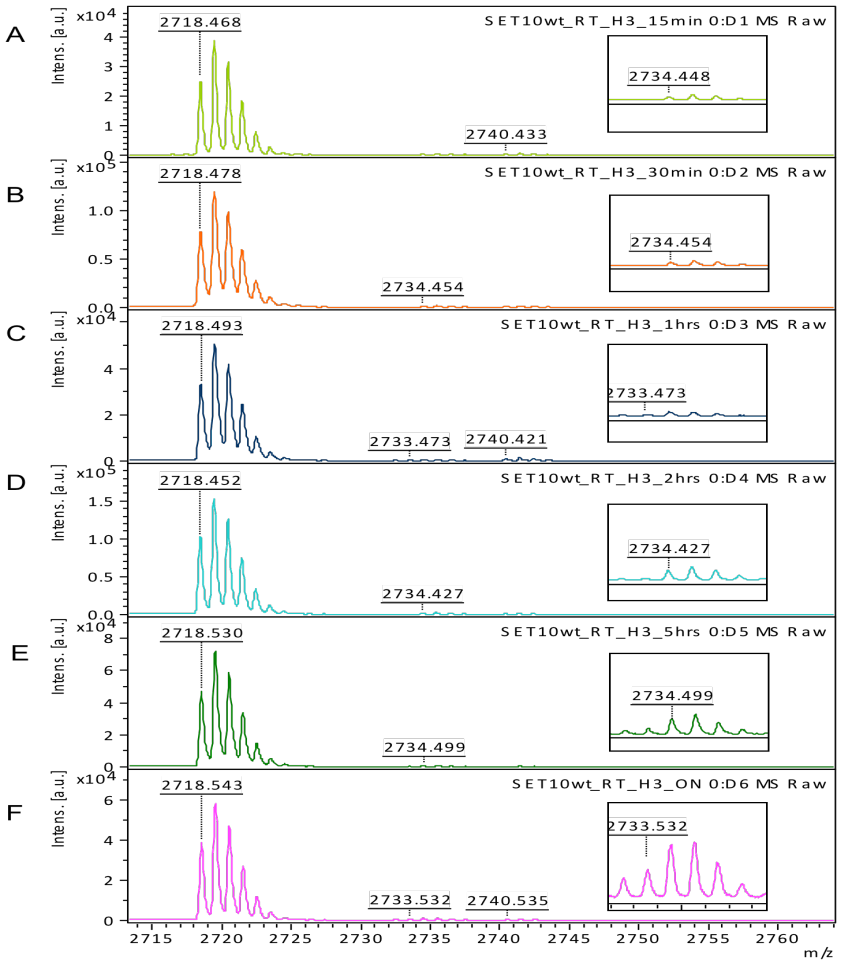

Figure 4. The methylation activity of SET10 at room temperature with increased reaction time: A) 15 minute, B) 30 minute, C) 1 hour, D) 2 hours, E) 5 hours and F) overnight (~24 hours). The smaller panels show a magnified version of the monomethylation peaks.

Radiometric Assay Experiment: A radiometric assay experiment was performed in order to characterize the differential binding abilities of SET10 to different substrates.[1] Different free peptides as well as whole histones were used as substrates during the experiment. In the first experiment, we allowed the reaction to continue for only one hour. As can be seen from Figure 5A, bands were only present for the whole histones,

[1] This radioactive assay was performed by Dr. Anamika Patel.

core histones and octamer. The free peptides, regardless of their modified status, were not methylated by SET10 after one hour, as seen by the absence of any bands. When the reaction was allowed to run for 24 hours, the free peptides did become methylated.

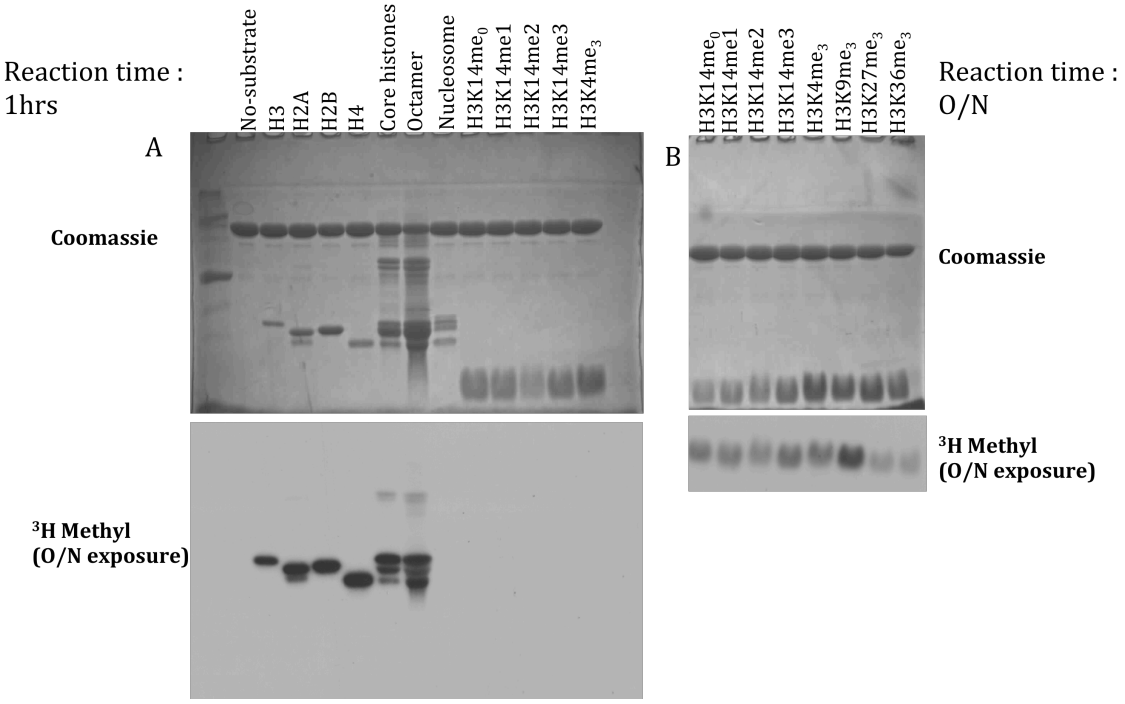

Figure 5. The use of various histone substrates by SET10. The reactions was allowed to run for two reaction times: A) 1 hour and B) overnight (~24 hours).

Crystallization

Trypsin Digestion Experiment: The first aspect of the structure that we wished to solve was to uncover the smallest and most stable domain of the protein. Oftentimes, a small, more stable domain can be isolated that can retain the activity of the larger protein. We set out doing this by using the protease trypsin, which is a serine protease (Hedstrom, 2002). This means that the amino acid serine serves as the nucleophilic amino acid in the enzyme's active site (Hedstrom, 2002). This protease contains the catalytic triad of

histidine-57, serine-195 and aspartic Acid-102 (Hedstrom, 2002). These three amino acids function together at the active site and is the crucial mechanistic driver of trypsin's function. As mentioned, serine's –OH groups acts as the nucleophile by attacking the carbonyl group of the amino acids lysine and arginine (Hedstrom, 2002). The electrons on one of histidine's imidazole nitrogen (N1) can accept the hydrogen from serine's –OH, leaving the resulting oxygen sufficiently negatively charged (Hedstrom, 2002). This facilitates the attack of the carbonyl by the negatively charged oxygen. The carboxyl group of the aspartic acid forms a hydrogen bond with the N-H (N2) in histidine's imidazole functional group (Hedstrom, 2002). This makes the N1 nitrogen sufficiently electronegative, such that it is able to accept the hydrogen from serine's –OH (Hedstrom, 2002). We predicted that the less stable portions of SET10, if such portions actually existed, would be cut by the trypsin, leaving smaller, most stable domain intact. As seen from Figure 6, an additional band below the parent band was not seen, indicating that the SET10 was not cut by trypsin. This suggested that perhaps the entire protein was stable and that a smaller domain did not exist.

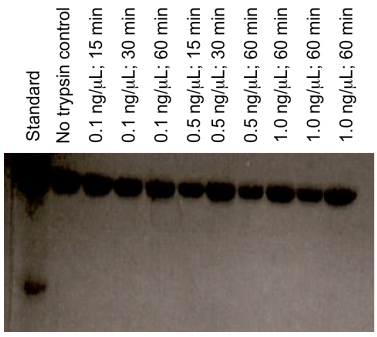

Figure 6. The gel was obtained from a trypsin digestion experiment. The absence of a band below the parent band indicates the absence of a smaller domain of the protein.

Protein Crystallization: To uncover the molecular structure of SET10, we crystallized the protein and then collected X-ray crystallography data. The atoms within the crystal cause the X-ray beam to diffract in different directions (Smyth & Martin, 2000). Based on the angles and intensities of these diffracted beams, researchers are able to get valuable information regarding the three-dimensional electron density within the protein (Jones et al., 1991). In order to obtain crystals, we tested a variety of commercial crystallization screens. These screens contained hundreds of unique conditions. Proteins are unique with regards to their temperature sensitivity, pH preference and size. As such, it is quite difficult to predict *a priori* what conditions will lead to the successful crystallization of a particular protein. Rather, several conditions are usually tested in the hopes that a particular condition will facilitate crystallization.

There are two main methods that we employed for the purposes of protein crystallization: sitting drop vapor diffusion and hanging drop vapor diffusion. The chemistry involved in both of these methods is the same, and the main difference is the location of the protein solution (Smyth & Martin, 2000). The protein is mixed with a unique condition containing precipitants that are meant to encourage crystallization (Smyth & Martin, 2000). In our case, the protein and precipitants were mixed in a 1:1 ratio, and the volume of this drop was 2 μL. This solution is placed on the cover slip in the hanging drop method or on a pedestal in the sitting drop method. There is a larger reservoir solution that contains merely the unique precipitants. The volume of our larger reservoir was 60 μL. The precipitants in the smaller drop are in lower concentration since it is mixed with protein than the precipitants in the pure reservoir solution. Although there is diffusion of water vapor in and out of both the small solution and the reservoir

solution, there is more vapor diffusion out of the smaller volume (Smyth & Martin, 2000). The net movement of vapor is from the small drop to the reservoir solution. Vapor diffusion continues, leading to a decrease in the volume of the smaller drop (Smyth & Martin, 2000). This occurs until equilibrium is reached with regards to the concentrations of the precipitants in the two solutions.

As the concentration of the precipitants and the protein in the small drop increases, the protein becomes supersaturated (Chayen & Saridakis, 2008). Given the correct conditions, the protein molecules may appear in the form of crystals (Chayen & Saridakis, 2008). This occurs when the protein molecules pack in a repeating pattern and are held together by non-covalent interactions. When the protein is initially dissolved, it interacts with the water molecules in the small drop via hydrogen bonds (Arakawa & Timasheff, 1984). As the salt concentration in the precipitant increases within the small drop (due to vapor diffusion), it interferes with the protein-solvent interactions (Arakawa & Timasheff, 1984). The protein-protein interactions overpower the protein-solvent interactions, leading to "salting out" of the protein (Arakawa & Timasheff, 1984). Different precipitants have unique abilities to interfere with the protein-solvent interactions, which is why we get crystal "hits" with some precipitants and not others.

Looking at a protein crystallization phase diagram with changeable parameters (perhaps precipitate concentration vs. protein concentration), four different areas can be identified (Chayen & Saridakis, 2008). There is an area of high supersaturation, where protein precipitation occurs (Chayen & Saridakis, 2008). There is an area of medium supersaturation, where a process known as nucleation takes place (Chayen & Saridakis, 2008). Nucleation occurs at the beginning stages of crystallization and is characterized as

the initial formation of a crystalline solid (Chayen & Saridakis, 2008). This solid serves as the surface for additional crystal growth. Thirdly, there is an area of low supersaturation (metastable zone) that is most conducive to the growth of large and high quality crystals (Chayen & Saridakis, 2008). Lastly, there is an area of undersaturation where the protein is dissolved and no crystallization or precipitation occurs (Chayen & Saridakis, 2008). With our vapor diffusion method, the protein solution is initially undersaturated. With net vapor diffusion out, the small drop becomes supersaturated with protein. Ideally, nucleation should first occur allowing the solution to "move" from the medium supersaturated nucleation zone to the low supersaturated metastable zone. This should ideally lead to high-quality crystal growth.

After setting up crystallization trays with a variety of different screens, we were able to receive crystal "hits" in two unique conditions (Figure 7). The first condition that facilitated crystallization was 2.0 M ammonium sulfate, 0.1 M Bis-Tris at pH 5.5. The second successful condition was 35% tert-butanol, 0.1 M sodium citrate at pH 5.6. The crystals were then successfully cryopreserved in 20% glycerol. This allows for two distinct advantages. Crystals can get damaged from radiation damage, such as from X-ray (Henderson, 1990). Cryoprotection lowers such damage and likely allows the collection of more effective data (Henderson, 1990). Secondly, it helps to prevent the formation of ice that can damage the protein crystals (Boutron, 1986). Since proteins have irregular shapes with a lot of channels, the crystals typically contain a high amount of solvent within them (Boutron, 1986). When the temperature decreases, ice may form and damage the crystals (Boutron, 1986). This decreases the quality of the crystals, resulting in poor diffraction. The cryopreservant lowers the freezing temperature, preventing ice

formation. Instead of freezing when the temperature is lowered, the cryopreserved solution will turn into a supercooled glass, a process known generally as vitrification (Boutron, 1986).

These crystals were subject to X-ray diffraction. The crystal is first placed on a goniometer and rotated while it is being bombarded by X-rays (Smyth & Martin, 2000). The crystal is rotated such that we are able to collect all of the information that is needed to reconstruct the protein structure (Smyth & Martin, 2000). The incident X-rays are being diffracted by the atoms within the crystal into various directions (Smyth & Martin, 2000). Since the protein crystals have a regular array of proteins arranged in a particular way, it scatters the incident X rays in a certain way (Smyth & Martin, 2000). This creates a pattern of diffracted pattern (looks like spots) known as the reflections (Smyth & Martin, 2000). These reflections are observed on a CCD detector behind the crystal (Figure 8). Using these two-dimensional diffracted patterns at different rotations, one can produce a three-dimensional view of the crystal's electron density. Using this electron density data, the position of the atoms and bonds can be decoded (Smyth & Martin, 2000). So why specifically are X-rays used? The wavelength of X-rays is similar to the spacing between the electron scatterers in a crystal. To optimize diffraction, the spacing of the scatterers should be similar to the wavelength of the incident rays.

The diffraction pattern shown in Figure 8 was obtained from the crystals that were grown on the 35% tert-butanol condition. After brief data analysis, we were able to determine the space group, which describes the packing of molecules in a particular way. Researchers have shown that there are merely 230 space groups in which repeating units can be arranged in a three-dimensional network (Smyth & Martin, 2000). The space

group of our crystals was P23, which represents a simple cubic shape. This is evident from the unit cell parameters a, b and c, which represent the lengths of the unit cell's axes. All of the unit cell lengths are equal to 184.5 Å. Furthermore, the angles between the axes are 90°. The information that all cell lengths are equal and that each angle is 90° to each other tells us that the unit cell is simple cubic. We were also able to determine that there were four molecules per asymmetric unit. The asymmetric unit refers to the smallest occupied location within a unit cell.

Although we were able to uncover the identity of the space group, the unit cell parameters and the number of protein molecules per asymmetric unit, we are unable to solve the atomic structure at this point. This is because the resolution of our data is not robust enough to allow for the atomic level elucidation of the structure. To obtain a meaningful and relatively accurate electron density map, a resolution of 3Å or below is required (Smyth & Martin, 2000). As this value decreases for our dataset, our resolution improves enabling us to see a more detailed and accurate final picture (Smyth & Martin, 2000). Our best dataset had a resolution of 7Å.

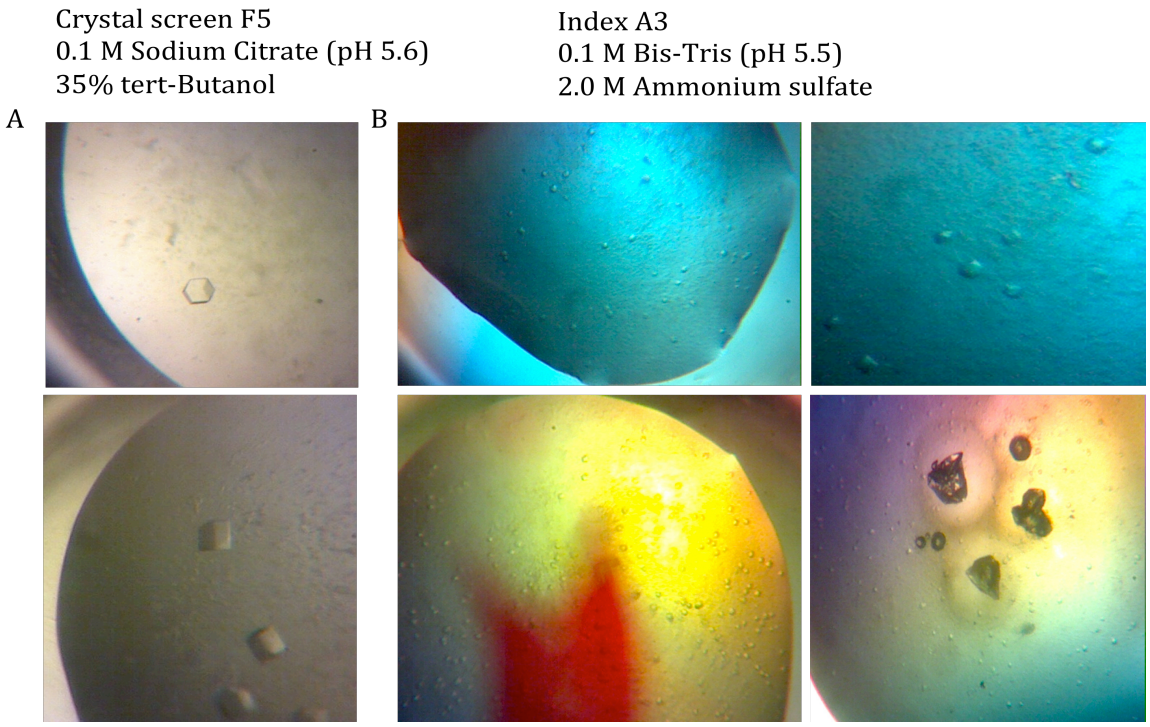

Figure 7. The pictures of crystals obtained from the A) 35% tert-butanol condition and B) 2.0 M Ammonium Sulfate condition.

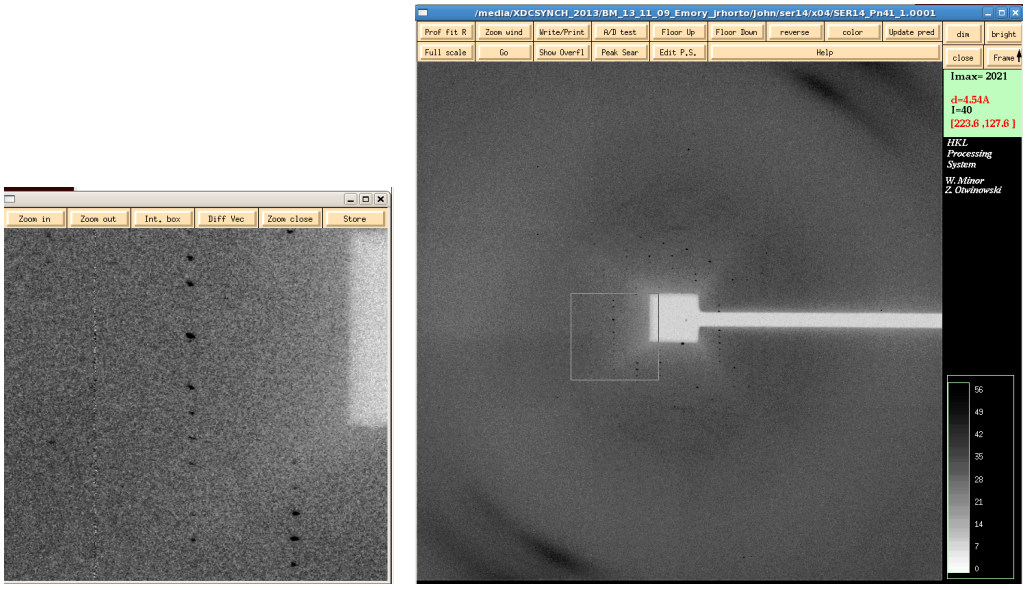

Figure 8. The diffraction pattern obtained from the tert-butanol crystals during X-ray diffraction. The left figure shows the magnified pattern seen in the right figure.

DISCUSSION

The results from the activity assays that we performed allowed us to make several conclusions. First of all, the peptide pulldown assays confirmed that SET10 is, in fact, able to bind to the H3 peptide (data not shown). The MALDI experiments provided more specific details by showing that the H3 peptide is likely not an ideal substrate for the SET10 protein. We also showed that with an increase in reaction temperature or reaction time, there was increase in monomethylation of the H3 peptide. As mentioned, dimethylation and trimethylation peaks were not observed. Perhaps our most enlightening experiment was the radiometric assays (Figure 5), which allowed us to make two unique conclusions. Firstly, we suggest that *in vitro* SET10 is able to methylate the whole H3 histone and other whole histones such as H2A, H2B, H4, core histones and the entire octamer.

Secondly, these results also suggest that free peptides are not ideal substrates for SET10. In order to investigate these findings further, we conducted a follow-up experiment with many more free peptides, allowing the reaction to continue overnight (24 hours). For this experiment, the following free peptides were used: unmodified H3 peptide, H3 peptide monomethylated, dimethylated, trimethylated at the K14 position, H3 peptide trimethylated at the K4 position, H3 peptide trimethylated at the K9 position, H3 peptide trimethylated at the K27 position, and H3 peptide trimethylated at the K36 position. All of the radioactive peptide bands when the reaction was allowed to continue for 24 hours. This does confirm that the free peptides are not ideal substrates for the SET10 protein, as the rate of reaction is slower with them when compared to whole histones. These experiments provide context to better interpret the results from our

MALDI experiments. The reason that we observed small methylation peaks could be due to the fact that the rate of reaction was very slow with the H3 peptide. This overnight experiment also corroborates our previous conclusion that the SET10 protein is nonspecific.

Moving forward, we must focus the bulk of our efforts on improving the quality of our crystals. Higher quality crystals will have proteins aligned identically, leading to the protein molecules diffracting the X-rays in the same way. This produces a preferred diffraction pattern with increased resolution. So the main question moving forward is this: how can we obtain higher quality crystals that provide diffraction data with a higher resolution?

First of all, crystals may not diffract well due to internal structural heterogeneity. The proteins in the crystal may be flexible, preventing the formation of high quality crystals (Wolfová et al., 2007). One potential way to increase rigidity is to add substrates or co-factors that may allow the protein to assume a certain functionally active (and presumably more rigid) orientation (Wolfová et al., 2007). During some of our screens, we added peptide or peptide and Adenosyl-L-homocysteine (AdoHcy) to the protein in order to see if this would help crystallization. We, however, did not specifically test the diffraction from these crystals, nor did we compare it to the data obtained from crystals without such added molecules.

Another source of structural heterogeneity may be due to differential post-translational modifications that may have occurred on the protein while it was expressed within the bacteria (Kim et al., 2001). Additionally, such heterogeneity may be due to the incomplete cleavage of the polyhistidine tag that was initially attached to the protein

(Birtley & Curry, 2005). A variety of steps during our entire expression and purification stages may have promoted structural heterogeneity (Birtley & Curry, 2005). It is, therefore, important to carefully revisit such steps and alter these steps, if needed. A general rule to follow would be to reduce the number of steps and the time from expression to crystallization (Birtley & Curry, 2005).

Another path that usually leads to the formation of low quality crystals is rapid crystal growth (Ketrane et al., 2009). Several steps can be tried in order to slow down crystal growth, which usually results in higher quality and larger crystals that diffract to a higher resolution (Ketrane et al., 2009). We can potentially dilute the drop in which our protein and precipitants are mixed. For example, instead of simply mixing protein and precipitants in a 1:1, some water may be added as well. This would take a longer time (more vapor diffusion needed) to get to the point where the protein would become supersaturated. This would also likely decrease the likelihood of the protein entering the "high supersaturation" zone, where precipitation occurs (Chayen & Saridakis, 2008). It is also possible to cover the reservoir solution with silicon oil or paraffin as this would also reduce the vapor diffusion rate between the reservoir and the drop (Chayen, 1999). Using larger drops of the mixed protein-precipitant solution may also help. These larger drops have a smaller surface area to volume ratio compared to smaller drops (Chayen, 1999). As such, equilibration will take longer.

The cryopreservants that are used may be contributing to the poor diffraction patterns. As such, it is important to experiment with different substances. It may be helpful to scan the literature and identify the substances used as cryopreservants when dealing with similar SET-domain containing proteins. Another method that may be tried

is the use of additional commercial screens with more unique conditions (Birtley & Curry, 2005). Some conditions will facilitate the production of higher quality crystals. Experimental trial and error and creativity may also be necessary. For example, mixing two unique conditions in different ratios may produce some beneficial results (Birtley & Curry, 2005). During our experiment, we mixed our two initial "hit" conditions and set up crystallization. Using this approach, we were able to obtain more crystals.

An additional method that has often led to a vast improvement in the resolution of the diffraction pattern is the dehydration of the crystals (Heras et al., 2003). As mentioned earlier, the protein crystals contain many solvents within them due to the irregular shape of proteins. This may lead to spatial heterogeneity, preventing the perfect stacking of protein molecules in a regular pattern. As a result, the quality of the diffraction will be lower. By reducing this aqueous environment within the crystal (dehydrating), one can significantly improve the resolution of the diffracting pattern (Heras et al., 2003). One method of dehydration involves the transfer of crystals from its drop into a sitting drop containing a dehydration solution (Heras et al., 2003). This drop may be equilibrated overnight with a larger reservoir solution containing the same dehydration solution (Heras et al., 2003).

Using the above methods, our first goal will be to obtain higher quality crystals whose diffraction patterns will provide a higher resolution. After obtaining these crystals, we can begin to use the necessary software and techniques so that we can solve SET10's atomic structure.

MATERIALS AND METHODS

Recombinant Gene Expression: The first step that was completed was transformation of the SET10 containing plasmid into *E. coli* cells. To achieve this, 1 μL of the plasmid DNA was mixed with 50 μL of *E. coli* cells. The mixture was left in ice for 30 minutes. The cells were then heat-shocked at 42°C for 60 seconds. This step allowed the plasmids to enter the cells, achieving transformation. After this step, the cells were put in ice for 2 minutes. 200 μL of lysogeny broth (LB) media was added to the cells. The mixture was placed at 37°C for 1 hour in an incubator in order to promote cell growth. 100 μL of these cells were taken out and put into an LB agar plate containing the antibiotic kanamycin. Since our transformed plasmid contained a gene that conferred kanamycin resistance, only cells that had been successfully transformed with the plasmid were able to grow.

Afterwards, we grew the bacteria in what is known as a "small scale culture." 30 mL of LB media was mixed with 30 μL of kanamycin. Using the tip of the pipette, one colony of the overnight-grown cells were taken and placed into the flask. The tube was placed in a spinning 37°C incubator overnight. This ensured that air was circulated throughout the flask such that the bacteria was able to grow. We used 1000X stock kanamycin, meaning that the ratio of the kanamycin volume to mixture volume should be 1:1000. After the completion of this step, the "large scale culture" step was next. Six large vials were taken and 100 mL of 10X LB media was mixed with 900 mL of sterile water and 1 mL of kanamycin. Each of the large vials was inoculated with 5 mL of the small-scale culture mixture containing the grown *E. coli* cells. The cells were grown until an optical density of 0.5 was reached.

Protein Induction: Once the desired optical density (OD_{600} ~ 0.5) was achieved, the temperature of the incubator was changed to 16°C. Each of the six vials was induced with 2 mL of IPTG. After induction with IPTG, the vials were kept spinning in the 16°C incubator overnight. The next day, each flask's contents were placed in 1L centrifuge bottles and were centrifuged at 4000 rotations per minute (rpm) for 30 minutes at 4°C. This low temperature was used in order to ensure that more cells do not grow and to promote protein stability. After centrifugation, the supernatant was thrown out as waste. The cells at the bottom were resuspended using a lysis buffer. In order to make the lysis buffer, we first made the wash buffer. The wash buffer contained the following: 300 mM sodium chloride (NaCl), 20 mM imidazole, 2 mM dithiothreitol (DTT) and 50 mM tris(hydroxymethyl)aminomethane (Tris). We mixed 50 mL of this wash buffer with a tablet of protease inhibitor and 50 μL of phenylmethylsulfonyl fluoride (PMSF) to convert it into the lysis buffer. The lysis buffer was added to the centrifuge bottles with the cell pellets. The bottles were vortexed in order to allow for resuspension. The contents were transferred to a 150 mL beaker to perform the next step of sonication.

Sonication: For sonication, the 150 mL beaker was placed in a larger plastic plate containing a mixture of ice and water. After tuning and activating the relevant settings, a total sonication time of 10 minutes was set on the sonicator. This achieved cellular lysis. After this process was completed, the mixture was centrifuged at 19000 rpm for 45 minutes at 4°C in order to separate the cell debris and the supernatant. The supernatant was collected since it contained the proteins that came out into solution during sonication. The other cellular debris was discarded.

Fast Protein Liquid Chromatography: During the His-tag run, the FPLC machine mixes the impure solution with the washing buffer (300 mM NaCl, 20 mM imidazole, 2 mM DTT and 50 mM Tris at pH 7.5). As the run continues, the proportion of the elution buffer (300 mM NaCl, 500 mM imidazole, 2 mM DTT and 50 mM Tris at pH 7.5) is linearly increased from 0% to 100%. This elution buffer allowed displacement of our protein. The obtained mixture is more pure than our starting mixture, but there are still some protein impurities such as other histidine-rich proteins. This mixture was taken and treated with thrombin in order to cleave off the polyhistidine tag.

For the Q column, the relevant buffers were made. Buffers A1 (50 mM Tris at pH 7.5, 2 mM DTT, 5% glycerol) and B1 (50 mM Tris at pH 7.5, 2 mM DTT, 5% glycerol, 1M NaCl) were first made. Buffer C1 (50 mM Tris at pH 7.5, 2 mM DTT, 5% glycerol, 100 mM NaCl) was made by using a 9:1 volume ratio of buffers A1:B1. First, the low salt C1 buffer was used. The elution buffer B1, whose proportion was linearly increased from 0% to 100%, was then used to collect SET10 in the effluent.

After running the Q column, we ran one last step of purification known as the sizing column run. Before the run, our protein mixture was concentrated to 2 mL by using a 30 kilodaltons (kDa) centrifugal concentrator. We placed our protein mixture into the concentrator and centrifuged it. Any proteins or molecules that weighed less than 30 kDa passed through a filter and was collected as waste. The molecular weight of our protein was about 62.3 kDa, so it was unable to pass through the filter. This method increased the concentration of our protein and also represented a method of purification since some smaller-sized unwanted proteins were being filtered out. The 2 mL of our concentrated protein solution was manually injected into the FPLC system for the sizing

column run. The buffer used here contained 20 mM Tris, 300 mM NaCl and 1 mM tris(2-carboxyethyl)phosphine (TCEP). The column was allowed to run, allowing the collection of our protein.

Gel Analysis: The protein samples were mixed with 3 X loading dye and loaded onto SDS-PAGE (15 % acrylamide-bisacrylamide). The gel was run for 40 minutes at 200 V. After the run, the gel was incubated with fixing solution (10 % acetic acid, 50 % ethanol) for 5 minutes. It was then stained with coomassie solution for 10 minutes and put on a rocker, followed by destaining with the destaining solution (5 % ethanol, 10 % acetic acid) until the protein band became visible on the gel.

Characterizing Protein Activity

Peptide Pull Down Assay: These assays are an invaluable tool that researchers use to elucidate protein-protein interactions. Our H3 peptide (bait protein) was tagged with a biotin group that allowed it to bind to the beads used during the experiment. We first incubated our 2 μL H3 peptide with 8 μL of SET10 protein (prey protein) and left them on ice. We used a variety of H3 peptides, both modified and unmodified. We also employed a control tube that only contained SET10 and water, but not peptide. After setting up these five tubes, we washed our beads three times with buffer. This process of washing involved adding buffer to the beads, spinning it down in a centrifuge and throwing away the supernatant. We then mixed the beads and the protein-peptide mixture and put the mixture on a rotator at 4°C for 30 minutes. This mixture was then washed with the buffer three more times to wash away any unbound proteins. If the SET10

protein was indeed able to bind to the H3 peptide, then a bead-biotin-peptide-protein complex would form. In other words, since we expect SET10 to bind to the peptide, the protein should not be present in the supernatant and should not be washed out. 20 μL of 1X loading dye was added to each of the 5 tubes containing the beads. The samples were run on a gel.

Matrix-Assisted Laser Desorption/ Ionization (MALDI) Experiments: *We* set up the three following tubes: no AdoMet control (contained protein and peptide), no enzyme control (contained SAM and peptide) and our main trial (contained peptide, AdoMet and enzyme). 2μL of the protein, 1μL of a 10X buffer, 1μL of SAM and 1μL of H3 peptide and 5μL of water was added. Whenever AdoMet or protein was not added, the same amount of extra water was added. The reactions in these three tubes were allowed to run, for example, for 3 hours at room temperature. In the meantime, a matrix containing acetonitrile, water, trifluoroacetic acid (TFA) and alpha-cyano-4-hydroxycinnamic acid (CCA) was prepared. After 3 hours, we stopped the reaction by adding 5μL of the above matrix to 2μL of the reaction solution. The contents of the matrix help stop the protein activity by promoting protein unfolding, ensuring that the reaction does not continue for longer than the desired amount of time (3 hours, in our case). TFA, for example, promotes an acidic environment, which causes our SET10 protein to unfold (Zhong et al., 2005). This protein and other methyltransferases typically require a basic environment to function. The matrix also contains a saturated solution of CCA, which promotes the formation of peptide crystals during drying (Zhong et al., 2005). The acetonitrile and water are added to dissolve the CCA (Zhong et al., 2005). From the above 7μL solution,

2μL was spotted on the mass spectrometry metal plate. 10-15 minutes was allowed for the spots to dry. The dried spots were then analyzed using MALDI.

Radiometric Assay Experiment: A radiometric assay experiment was performed in order to characterize the differential binding abilities of SET10 to different substrates. In effect, we wanted to identify the optimal substrate for the protein. Different free peptides as well as whole histones were used as substrates during the experiment. We used the following substrates: H3 histone, H2A histone, H2B histone, H4 histone, core histones, octamer, nucleosome, unmodified H3 peptide, H3 peptide monomethylated, dimethylated and trimethylated at K14, and H3 peptide trimethylated at the K4 position. We also had a control reaction that had no substrate. In the reaction tubes, we also added our SET10 protein and the radioactive ^{3}H-AdoMet. If SET10 was able to bind to the potential substrate and transfer the radioactive methyl group, then the recipient substrate would be seen as a dark band on the X-ray film. To test if this transfer occurs, the reaction mixture was run on an SDS-PAGE. After staining and destaining the gel, it was placed in an enhancer solution (cocktail of various organic molecules) that increased the fluorescence signal. The gel was then dried, and an X-ray film was placed on it. If SET10 transferred a radioactive methyl group to the substrate, then the substrate would fluoresce onto the film. The film was developed by placing it into a developer solution, allowing the radioactive signal to be seen.

Crystallization

Trypsin Digestion Experiment: We ran several trials by changing the trypsin concentration and the reaction time. In total, there were a total of ten trials that were all

run at 37°C. One trial had no trypsin and contained only our SET10 protein and buffer. The nine other trials involved three different trypsin concentrations (0.1 ng/μL, 0.5 ng/μL, 1 ng/μL), each of which was tested for three different reaction times (15 minutes. 30 minutes, 1 hour). Each reaction tube contained a total of 30μL (2.4 μL of SET10, 3 μL of trypsin, 24.6 μL buffer). In the trial where no trypsin was used, more buffer was added. The buffer that was used here contained 20 mM Tris, 300 mM NaCl and 1 mM TCEP. In order to stop the reaction, 10 μL of the reaction mixture was added to 5 μL of PMSF. The standard marker and these ten samples were run on a gel.

Protein Crystallization: Initially, we set up the sitting drop vapor diffusion method using unique conditions from several commercials screens (such as Index, Peg/Ion, Hampton Research). The screens were set up using the machine Phoenix. After receiving crystal hits in two unique conditions, we set up several more crystallization trays by altering the above two main conditions. For example, one of our handmade screens contained six different concentrations of Ammonium Sulfate (1 M, 1.2 M, 1.5 M, 2 M, 2.5 M, 3 M), each of which was replicated in four different pH conditions (5.2, 5.5, 5.8, 6.2). The concentration of the Bis-Tris was kept constant at 0.1 M. Another screen was made varying the concentration of ammonium sulfate from 1 M to 3 M, but in smaller increments. Additionally, a handmade screen contained six concentrations of tert-butanol (20%, 25%, 30%, 40%, 45%), each of which was replicated in four different pH conditions (5.2, 5.6, 5.8, 6.2). Another screen was made varying the tert-butanol concentration from 15% to 40% in small increments. The concentration of the sodium

citrate was kept constant at 0.1 M. Several more crystal “hits” were achieved using this method of testing around the original two conditions.

Some crystallization trays were also set up by adding the H3 peptide and AdoHcy. Obtaining crystals in these conditions would allow us to potentially uncover the structure and orientation of the protein when it is in its active state (bound to its peptide substrate and AdoHcy). Lastly, we used a handmade screen that contained both ammonium sulfate and tert-butanol in varying concentrations. The concentration of ammonium sulfate was increased from 0 M to 2.2 M in increments of 0.2 M. The concentration of tert-butanol, however, was decreased from 35% to 0% in variable increments. This resulted in 12 different conditions containing a unique ammonium sulfate: tert-butanol ratio. Each of these 12 conditions were tested in 8 different pHs (5.0 to 5.7 in increments of 0.1). The Bis-Tris buffer was used, and its concentration was kept constant at 0.1 M. More successful “hits” were obtained using these conditions. These crystals were then cryopreserved, meaning they were mixed with certain cryopreservants and then stored at a low temperature. To achieve this protection, our crystals were transferred from their original small drop to another solution that was a combination of the precipitants and cryopreservants (in our case, 20% glycerol). These crystals were then taken out and placed in liquid nitrogen at temperature of 100 K.

References Cited

Arakawa, T., & Timasheff, S. N. (1984). Mechanism of protein salting in and salting out by divalent cation salts: balance between hydration and salt binding.*Biochemistry*, *23*(25), 5912-5923.

Birtley, J. R., & Curry, S. (2005). Crystallization of foot-and-mouth disease virus 3C protease: surface mutagenesis and a novel crystal-optimization strategy. *Acta Crystallographica Section D: Biological Crystallography*, *61*(5), 646-650.

Boutron, P. (1986). Comparison with the theory of the kinetics and extent of ice crystallization and of the glass-forming tendency in aqueous cryoprotective solutions. *Cryobiology*, *23*(1), 88-102.

Chandra, G., Patel, P., Kost, T. A., & Gray, J. G. (1992). Large-scale purification of plasmid DNA by fast protein liquid chromatography using a Hi-Load Q Sepharose column. *Analytical biochemistry*, *203*(1), 169-172.

Chayen, N. E. (1999). Crystallization with oils: a new dimension in macromolecular crystal growth. *Journal of crystal growth*, *196*(2), 434-441.

Chayen, N. E., & Saridakis, E. (2008). Protein crystallization: from purified protein to diffraction-quality crystal. *Nature Methods*, *5*(2), 147-153.

Desjouy, C., Poizat, A., Gilles, B., Inserra, C., & Bera, J. C. (2013). Control of inertial acoustic cavitation in pulsed sonication using a real-time feedback loop system. *The Journal of the Acoustical Society of America*, *134*(2), 1640-1646.

Dillon, S. C., Zhang, X., Trievel, R. C., & Cheng, X. (2005). The SET-domain protein superfamily: protein lysine methyltransferases. *Genome biology*, *6*(8), 227.

HaHamamoto, R., Furukawa, Y., Morita, M., Iimura, Y., Silva, F. P., Li, M., et al., & Nakamura, Y. (2004). SMYD3 encodes a histone methyltransferase involved in the proliferation of cancer cells. *Nature cell biology*, *6*(8), 731-740.

Hedstrom, L. (2002). Serine protease mechanism and specificity. *Chemical reviews*, *102*(12), 4501-4524.

Henderson, R. (1990). Cryo-protection of protein crystals against radiation damage in electron and X-ray diffraction. *Proceedings of the Royal Society of London. Series B: Biological Sciences*, *241*(1300), 6-8.

Heras, B., Edeling, M. A., Byriel, K. A., Jones, A., Raina, S., & Martin, J. L. (2003). Dehydration converts DsbG crystal diffraction from low to high resolution. *Structure*, *11*(2), 139-145.

Jones, T. A., Zou, J. Y., Cowan, S. T., & Kjeldgaard, M. (1991). Improved methods for building protein models in electron density maps and the location of errors in these models. *Acta Crystallographica Section A: Foundations of Crystallography*, *47*(2), 110-119.

Ketrane, R., Saidani, B., Gil, O., Leleyter, L., & Baraud, F. (2009). Efficiency of five scale inhibitors on calcium carbonate precipitation from hard water: Effect of temperature and concentration. *Desalination*, *249*(3), 1397-1404.

Kim, K. M., Yi, E. C., Baker, D., & Zhang, K. Y. (2001). Post-translational modification of the N-terminal His tag interferes with the crystallization of the wild-type and mutant SH3 domains from chicken src tyrosine kinase. *Acta Crystallographica Section D: Biological Crystallography*, *57*(5), 759-762.

Kopaciewicz, W., Rounds, M. A., Fausnaugh, J., & Regnier, F. E. (1983). Retention model for high-performance ion-exchange chromatography. *Journal of Chromatography A*, *266*, 3-21.

Regnier, F. E. (1983). High-performance liquid chromatography of biopolymers.*Science*, *222*(4621), 245-252.

Rolando, M., Sanulli, S., Rusniok, C., Gomez-Valero, L., Bertholet, C., Sahr, T., ... & Buchrieser, C. (2013). *Legionella pneumophila* Effector RomA Uniquely Modifies Host Chromatin to Repress Gene Expression and Promote Intracellular Bacterial Replication. *Cell host & microbe*, *13*(4), 395-405.

Schmitt, J., Hess, H., & Stunnenberg, H. G. (1993). Affinity purification of histidine-tagged proteins. *Molecular biology reports*, *18*(3), 223-230.

Sheehan, D., & O'Sullivan, S. (2004). Fast protein liquid chromatography. In*Protein purification protocols* (pp. 253-258). Humana Press.

Smyth, M. S., & Martin, J. H. J. (2000). x Ray crystallography. *Molecular Pathology*, *53*(1), 8-14.

Sweder, K. S., Rhode, P. R., & Campbell, J. L. (1988). Purification and characterization of proteins that bind to yeast ARSs. *Journal of Biological Chemistry*, *263*(33), 17270-17277.

Wolfová, J., Mesters, J. R., Brynda, J., Grandori, R., Natalello, A., Carey, J., & Kutá Smatanová, I. (2007). Crystallization and preliminary diffraction analysis of Escherichia coli WrbA in complex with its cofactor flavin mononucleotide.*Acta Crystallographica Section F: Structural Biology and Crystallization Communications*, *63*(7), 571-575.

Zhong, H., Marcus, S. L., & Li, L. (2005). Microwave-assisted acid hydrolysis of proteins combined with liquid chromatography MALDI MS/MS for protein identification. *Journal of the American Society for Mass Spectrometry*, *16*(4), 471-481.

List of Figures