5-8-2020

Investigation of potentially catalytic residues of Uba5 through mutagenesis, purification, and structural characterization

Grant Bradley

Investigation of Potentially Catalytic Residues of Uba5 Through Mutagenesis, Purification, and Structural Characterization

---

An Honors College Project Presented to

the Faculty of the Undergraduate

College of Science and Mathematics

James Madison University

---

by E. Grant Bradley

Accepted by the faculty of the College of Science and Mathematics, James Madison University, in partial fulfillment of the requirements for the Honors College.

FACULTY COMMITTEE: HONORS COLLEGE APPROVAL:

Project Advisor: Chris E. Berndsen, Ph.D. Associate Professor of Biochemistry	Bradley R. Newcomer, Ph.D., Dean, Honors College
Reader: Jonathan Monroe, Ph.D. Professor of Biology	Reader: Ray Enke, Ph.D. Associate Professor of Biology

Table of Contents

Table of Figures

Abstract

Ubiquitin-fold modifier 1 (Ufm1) is a member of the Ubiquitin (Ub) family of proteins whose primary function is degradation of proteins through a sequential mechanism of chemical reactions. Though Ufm1's specific roles are largely unknown, this family of proteins has shown to play a part in a wide variety of processes, including regulation of the cell cycle[1], secretory functions of cells[2,3], and blood clotting[4]. Ufm1's mechanism of action proceeds with the aid of three enzymes: an E1, E2, and E3. Uba5 is the E1 activating enzyme that is specific to Ufm1, and its mechanism of action and active site chemistry is not well understood[5,6]. We do know that the enzyme contains a catalytic cysteine residue that is thought to be activated by another structure within the protein[5–7]. The goal of this study was to investigate potentially catalytic residues within Uba5 through modeling with YASARA[8], generate mutant proteins with point mutations within these residues, and purify these mutants for structural modelling and activity assays, to compare to the wild type. Such residues were identified and selected for purification as part of this study. These proteins were expressed; however, problems with the expression and purification protocol prevented us from successfully purifying the mutations, and SAXS data showed that what we did purify contained significant aggregation. As future directions, we suggest various adjustments to the expression and purification method that we feel will lead to better yields, and therefore continuation of these aims.

Introduction

Post translational modification of proteins

Post translational modification of proteins is an essential process that are a component of nearly all biological pathways and regulate a variety of biological functions. These modifications largely involve the addition or removal of a functional group, such as an inorganic molecule or a peptide, to a protein to modify its activity in some way. Many well-known biological processes such as energy storage and release, DNA transcription, blood clotting, regulation of the cell cycle, and protein degradation involve post translational modifications[4].

Ubiquitination

Protein degradation is an essential type of post translational modification in biological systems that is controlled by the addition of ubiquitin (Ub) or ubiquitin-like proteins (Ubl), also called ubiquitination. Ubiquitin is a small, 76 residue protein[2] that is the most widely studied and the founding member of the Ub-fold family of proteins. Initially thought to be a hormone involved in lymphocyte differentiation and adenylate cyclase activation, ubiquitin's crystal structure was first solved in 1985. Ubiquitin was described as being small and a very tightly hydrogen bonded protein, with 87% of the amino acids being involved in hydrogen bonding[9,10]. Its major structural characteristic is the beta-grasp fold, which distinguishes this family of proteins. This structure consists of five anti-parallel beta sheets that form one large sheet and one alpha helix that appears to be "grasped" by the beta sheet structures[9,10]. Since its discovery, this structure has been found in other proteins, and has been shown to perform a variety of functions by providing an open surface for weak molecular interactions and by forming barrel-like structures[11]. Eukaryotes have evolved a complex and refined system of protein degradation using Ub and a cascade of enzymes involved in its biochemical pathway[2,11,12].

Ub molecules are typically attached to lysine (Lys) residues on the target substrate. In addition to the target substrate, ubiquitin can be attached to lysines within other ubiquitin molecules forming chains. These different attachment modes signal different actions by the cell. Monoubiquitination refers to the attachment of a single Ub to a single Lys within a protein, whereas multiubiquitination refers to the attachment of single Ub molecules to multiple Lys resides within a target substrate. Both been shown to have a role in marking proteins for certain endocytic and secretory processes[3,2]. Polyubiquitination, the attachment of Ub chains to target proteins, serves the more widely known function of ubiquitination, which is marking the protein for degradation. Ub chains act as the tag that marks a target protein for certain death. Once tagged by the polyubiquitin chain, the protein is degraded by the proteasome into its amino acid components, and the Ub is recycled.

Organisms have a variety of reasons to break down proteins using this process. Proteins can become damaged, and therefore no longer able to function correctly. They also may be needed for a particular stage in development, such as in the case of cyclin B[1]. Cyclin B is an abundant intracellular protein during metaphase of the eukaryotic cell cycle, but is degraded after polyubiquitination during anaphase to prevent the initiation of a new cell cycle[1]. In many cases, proteins are broken down so that their amino acid components can be utilized by the cell for energy, for the synthesis of other amino acids, or as a part of routine protein recycling.

The mechanism of ubiquitination, considered to be one of the most conserved mechanistic frameworks across all eukaryotic organisms[12], involves a cascade of three enzymes, E1, E2, and E3. Each Ubl has its own unique E1 enzyme, and several different E2 and E3 enzymes that allow for specificity to each protein target[13]. The E1, or ubiquitin activating enzyme, activates the C-terminus of the Ub or Ubl through adenylation. The E1enzyme

facilitates the hydrolysis of ATP for this step, yielding a molecule of AMP bound to the Ubl and pyrophosphate. An activated cysteine residue of the E1 enzyme then attacks the C-terminus of the Ubl, removing the AMP and forming a thioester bond between the E1 enzyme and the Ubl.

E1 enzymes are divided into two categories: canonical and non-canonical. Two distinct pseudo symmetric adenylation domains, the inactive acetylation domain (IAD) and active acetylation domain (AAD), distinguish the canonical E1 enzymes from the non-canonical enzymes[14,15]. Canonical E1 enzymes also contain a domain called the Ub fold domain (UFD) that aids in E2 enzyme specificity and binding. Mechanistically, canonical E1 enzyme cascades contain an adenylation of an additional Ubl during the transfer to the E2 enzyme[16]. Non-canonical E1 enzymes contain homodimeric adenylation domains with symmetrical orientation and lack the distinct Cys domain. The difference in structure also reportedly changes the mechanism and non-canonical E1 enzymes reportedly lack the extra adenylation step[17].

After the E1~Ubl thioester is formed, the E1 enzyme transfers the Ubl to an active site cysteine residue of a conjugating enzyme, or E2 enzyme, in a step called transthioloation[6,17]. E2 enzymes contain a unique active site domain, called the UBC, that allows them to perform the function of Ub conjugation. Each different E2 enzyme contains a different structural variation of this domain that allows for specificity. This is the site of E1 and E3 interaction, and contains the active site, called the catalytic groove or cleft, which are the amino acids that surround the Cys residue where the Ub thioester is formed[13,18]. E2 enzymes also contain C-terminal and N-terminal domains that allow for specific interaction with its specific E1 enzyme and a range of E3 enzymes[13].

After the E2~Ubl thioester is formed, the E2 enzyme, along with an E3 ligase enzyme, transfers the Ubl to the target protein substrate. E3 enzymes are the most functionally diverse

enzyme family in this cascade, and its primary role is substrate targeting. Their mechanisms of action are varied, with some E3 enzymes forming thioester bonds with the Ubl, while others do not[19]. An overview of the three-step mechanism is shown in Figure 1.

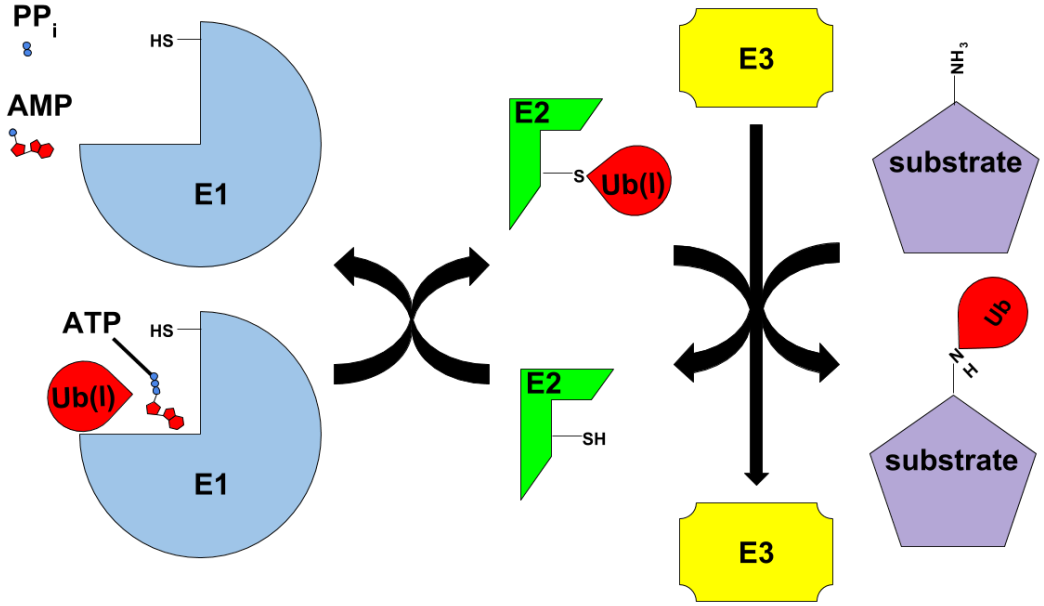

Figure 1. Ubiquitination of a target substrate proceeds through a mechanism involving three enzymes: E1, E2, and E3.

Ubiquitin-fold modifier 1 (Ufm1)

Discovered in 2002, Ubiquitin-fold modifier 1 is a Ubl that is conserved in animals and plants but apparently not in fungi[7]. While its amino acid sequence shows only 16% sequence identity[20] to Ubiquitin, Ufm1 forms a similar beta grasp fold structure that allows it to function as a Ubl in a cascade, called "Ufmylation". Ufm1 contains a C-terminal glycine residue that is activated by Ubl-activating enzyme 5 (Uba5) through an adenylation followed by thioester formation at a cysteine residue (Cys250). Ufm1 is then transferred to Ub fold modifier conjugating enzyme 1 (Ufc1) without the second adenylation reaction required by canonical E1

enzymes. Finally, Ufm1 is conjugated to a target protein by the concerted action of Ufc1 and the Ufm1-specific ligase 1 (Ufl1) substrate, the only known E3 ligase for Ufm1[7,15].

There is little known about the biological roles of Ufm1, with only 5 target proteins having been identified since its discovery[6]. However, some studies have postulated possible roles of the protein based on its expression in certain tissues. Ufm1 was shown to be highly expressed in secretory cells, including pancreatic beta cells[21]. Endoplasmic reticulum (ER) stress, relatively common in this cell type and was shown to increase Ufm1 expression. Cells deprived of Ufm1 or Ufl1, its cognate E3, showed an increased level of apoptosis following prolonged ER stress. Research has been done into whether high levels of transcription of these proteins leads to deficiency in insulin secretion. In a recent study[22], transgenic mice containing a mutation that led to skeletal muscle insulin resistance, pancreatic beta cell compensation, and finally beta cell failure, were studied to see if they showed a difference in protein expression. Their results showed that Ufm1 as well as VCP, an endoplasmic reticulum ATPase that aids in protein degradation, were expressed at higher levels in the transgenic diabetic mice[22]. The research suggests that increased transcription of Ufm1 plays a factor in beta cell response to insulin resistance, with this response leading to decreased insulin secretion. The study, however, found 159 differentially expressed proteins in the transgenic mice[22], so the involvement of Ufm1 in the development of type 2 diabetes still needs to be studied further. Mutation of Ufm1 has also been associated with leukodystrophy. A deletion mutation located in the promoter region of Ufm1 was detected using homozygosity mapping and whole exome sequencing in individuals with hypomyelinating leukodystrophy in an affected family of Roma descent[23]. Affected individuals had symptoms of lack of development and severe epileptic encephalopathy.

Ufmylation has also been shown to influence breast cancer formation. In a recent study, the nuclear receptor coactivator ASC1 was identified as a target protein for Ufm1[24], and ASC1 Ufmylation led to an increase in gene expression of estrogen receptor target genes. Overexpression of these associated genes, as well as knockout of the Ufm1-specific protease 2 (UfSP2), was associated with increased tumor formation. Tumor formation decreased in Ufmylation-deficient cell lines of ASC1, as well as Uba5-knockout cell lines[24]. This demonstrates a link between Ufmylation of ASC1 and development of breast cancer.

Currently in the Berndsen lab at James Madison University, work is being done to study the role of Ufmylation in the development of the leishmaniasis infection, caused by *Leishmania donovani* (Ld), a trypanosomatid parasite. The structure and function of Ld Ufm1 is being studied in comparison with Human Ufm1. Collaborators in the Berndsen lab expressed and purified Ld Ufm1, as well as an N-terminally truncated version of Ld Ufm1. The study[25] has demonstrated the compatibility of Ld and human systems, with human Uba5 binding to full-length Ld Ufm1, but not the N-terminally truncated version, suggesting a role for the N-terminus. Students working on this aim of the project intend on doing more work looking into compatibility between the human and Ld Ufm1 systems. Optimized purification of human Uba5 and further understanding of the human Ufmylation cascade will support this study to better understand the mechanism of leishmaniasis infection.

Ubiquitin-activating enzyme 5 (Uba5)

Uba5 is the E1 enzyme that is specific to Ufmylation, and it has some key differences when compared to other E1 enzymes. It is a non-canonical enzyme with a homodimeric structure, and it is the smallest E1 that is known[6], with a molecular weight of 45 kDa. In contrast to canonical E1 enzymes, Uba5 does not contain a distinct catalytic Cys domain. Rather, it has a

catalytic Cys residue in the active site that forms the thioester bond with Ufm1. How the homodimer structure and lack of Cys domain are related to the function of non-canonical E1 enzyme's like Uba5 is currently not understood[6]. Uba5 also contains a Ufm1-interacting sequence (UIS) which is required for Ufm1 binding. The mechanism of Ufm1 binding to Uba5 is often called a "trans-binding mechanism", because Ufm1 interacts with both subunits of the Uba5 dimer[5], In this mechanism, Ufm1 interacts with the adenylation domain of one Uba5 and the UIS of the other Uba5. This trans binding mechanism is unique to Uba5. Other E1's either lack the UIS entirely (such as those involved in the Ubiquitin, NEDD8, and SUMO mechanisms), or contain an analogous structure that interacts in cis (such as that in the ATG8 mechanism)[5]. Also found in Uba5 and some other E1 enzymes is a structure called a crossover loop (CL) which contains the active site Cys. This loop has been suggested to be the site of conformational changes upon Ufm1 binding and mutant versions of Uba5 with mutations in this region were able to bind Ufm1 but not activate it[6].

Despite these distinctions between Uba5 and other E1's, the general reaction mechanism and intermediates are similar to those of other E1's. Uba5 activates Ufm1, a Ubl, through a catalytic mechanism that contains an adenylation and a thioesterification step. In the adenylation step, the C-terminus of Ufm1 attacks the alpha phosphate of ATP, producing a pyrophosphate molecule, which yields a Ufm1-AMP intermediate[16,17]. Next, a cysteine, (Cys250), is known to form a thioester bond with the C-terminus of Ufm1 as a part of this mechanism[6,7]. The sulfur atom of the cysteine attacks the carbon atom of the C-terminus of Ufm1, and an AMP molecule is then cleaved to yield the final product, Uba5 bound to Ufm1 by a thioester bond at Cys250 of Uba5. An overview of this proposed mechanism is shown below in Figure 2.

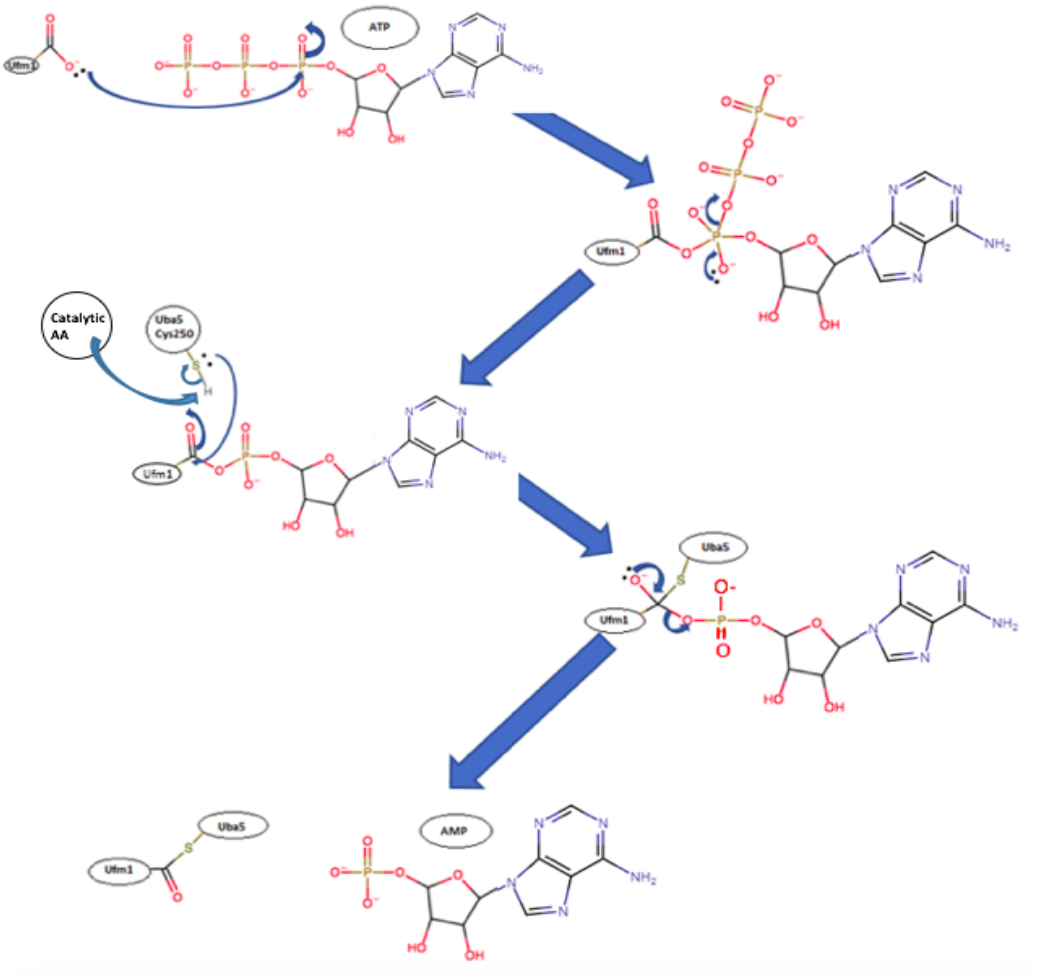

Figure 2. Complete proposed arrow-pushing mechanism of adenylation and thioesterification steps of the UFM1-UBA5 complex formation.

In order for the proposed mechanism to be accurate, the catalytic Cys residue of Uba5 is likely activated by another amino acid in order to attack the C-terminus of Ufm1. Such a residue has yet to be identified, and there are no conserved amino acids in the active site that could serve as a base [14,15,26]. Furthermore, no amino acids that function to stabilize transition states of the mechanism have been identified. The goal of this study is to demostrate how the active site amino acids facilitate the attacking of Ufm1 by the catalytic Cys of Uba5.

Goal of the study

A study to determine how the active site of Uba5 facilitates the mechanism of Ufm1 activation has not been performed. The long-term goal of this study is to identify potential key amino acids in the active site of Uba5 that facilitate its chemical mechanism. This will be achieved through expression and purification of Uba5 mutant contructs, and analyzing the structural and functional properties of these mutant proteins through activity assays and structural studies to determine the effect(s) of the mutations. The resulting data will be part of a larger study into the Ufmylation mechamism with applications to the broader E1 family. Having a better understanding of these processes will provide information about chemistry that is essential for life, with numerous applications in the study human biology and disease.

Methods

Preliminary mechanism study

Before beginning any wet lab techniques, computational techniques with YASARA[8] were used to determine potential catalytic amino acids for further study. For the first model, a structure of Uba5 in complex with Ufm1[6] was aligned with that of Uba5 bound to ATP[27]. After alignment, the structures were joined and the bonds were deleted, yielding a 3D model of Uba5 and Ufm1 in complex, with ATP in the active site. The structure was then energy minimized, to produce the local lowest energy conformation and calculate a ΔG value. The goal of making the ternary complex was to model the active site environment with both substrates bound to identify potential residues that are holding ATP in the active site and stabilize this conformation of starting materials.

Next, models of the two potential transition states (the two structures on the right in Figure 2) were built starting with the model of starting materials, adding and removing the appropriate bonds. The resulting structure was energy minimized to optimize the conformation, and the ΔG value was measured. The purpose of this step was to identify potential residues in Uba5 that may be acting to stabilize these transition states (or lack thereof), which would provide information on the plausibility of the proposed mechanism.

The last model that was constructed was of the Uba5~Ufm1thioester, with the molecule of AMP in the active site. This was done by adding and removing the appropriate bonds and energy minimizing the resulting structure. The ΔG value was recorded in order to show the change in energy during each step of the proposed mechanism.

Biochemistry lecture course integration

In Dr. Berndsen's Fall 2018 Biochemistry course, students conducted a lecture project on the mechanism of the adenylation and thioesterification steps of the UFM1-activating mechanism. Many of the projects took a similar approach as the previously described computational methods. From the class projects, including my own, a list of potential catalytic amino acids was compiled.

Protein expression

After potentially catalytic amino acids were identified, we ordered Uba5 containing these mutations from GenScript. This DNA was then transformed into competent BL21 *E. coli* cells. After transformation, the cells were grown overnight on selection media plates containing ampicillin. Once colonies formed, cells from the agar plate were transferred to 100 mL of 2XYT media containing 1x ampicillin, and grown overnight. After overnight growth, 4 mL of culture was transferred into 1 L cultures containing 1x ampicillin, and shaken at 37°. Absorbance at 600

nm was measured until 0.8 A, at which time protein expression was induced using 0.01 g IPTG per liter flask. After induction, cultures were incubated overnight at 20° with shaking. Cultures were then centrifuged, and the pellets were collected and stored at -80° C.

Gravity column purification

Cell pellets were suspended in lysis buffer containing 50mM sodium phosphate (NaH_2PO_4), 500 mM NaCl, and 10 mM imidazole (pH 8.0), and vortexed until homogeneous. The resulting mixture was then sonicated at 45% amplitude to lyse cells. The resulting mixture was then centrifuged, and the supernatant was collected. The supernatant was then passed through a Corning 0.2µm bottle-top filter before the purification began and was diluted 1:1 with DI water.

The first attempt at purification was done using a gravity column with Gold Biotechnology nickel resin. Gold biotechnology cobalt resin was also used in an attempt of purification using gravity column in an attempt to obtain cleaning purification. The supernatant was applied to the resin, and the flow through was collected. Lysis buffer from above was washed over the resin and collected. The sample was then washed with buffers identical to the lysis buffer, but containing 50 mM and 100 mM imidazole. These wash buffers were passed through the column and the flow through was collected. Finally, an elution buffer containing 50mM sodium phosphate (NaH_2PO_4), 500 mM NaCl, and 300 mM imidazole (pH 8.0) was passed through the column, and the flow through was collected. Purity was assessed via SDS PAGE gels, of the supernatant flow through, washes, and elution steps.

FPLC purification

Fast protein liquid chromatography (FPLC) with the GE ÄKTA system was also used for purifying Uba5. FPLC uses an automated, high-pressure system as opposed to purifying

protein by hand using a gravity column. The system uses a continuous gradient of increasing mobile phase concentration, with adjustable fractionation volume to allow increased precision. The system also measures the conductivity and UV absorbance of the outflow to be able to pinpoint which fractions contain protein. In the first purification protocol, the His tagged Uba5 was purified using affinity chromatography, utilizing Nickel resin in a GE Healthcare 1 mL Histrap^{TM} HP column. The mobile phase utilized was the lysis buffer from above, with a continuous imidazole gradient of 10 mM-300 mM. After issues with sustaining high pressures in the column, a new 5 mL Histrap^{TM} was purchased and used with the above protocol.

Presence of protein was confirmed using UV absorbance at 280 nm, and the corresponding fractions were analyzed for protein using sodium dodecyl sulfate polyacrylamide gel electrophoresis (SDS-PAGE). Fractions showing bands at 40 kD were concentrated, and protein concentration was determined using UV absorbance values using the calculated extinction coefficient of 21430[28]. The protein was then divided into aliquots and frozen.

After purification, the column was washed with water and 20% ethanol, and stored at 4° C. The column was stripped periodically with a stripping buffer consisting of 20 mM sodium phosphate, 0.5 M NaCl, 50 mM EDTA, pH 7.4[29]. Once stripped, the resin was reapplied and the column was stored in ethanol.

Small angle X-ray scattering (SAXS)

In order to obtain preliminary structural data to visualize Uba5 in solution, samples of wtUba5 in lysis buffer were sent to SiBYLS beamline 12.3.1. Fifty frames were obtained, with frames collected every 0.3 seconds for a total of 15 seconds. Exposures of buffers were also collected to subtract from the sample exposure. Using PRIMUS[30], a Guinier approximation plot was generated from a single exposure due to suspected aggregation. From the Guinier plot, Pair

Distribution function (P[r]) plots were generated automatically by the program and later manually. In manual data fitting, we removed an aggregated fraction of the data by altering the data range that was used to generate the plot.

Results

Preliminary mechanism study

In the model of the initial reactants, Lys127, Arg246, and Asn112 of UBA5 were identified as components of a possible enzyme cave that functions to hold ATP in place, allowing reaction to commence by ATP hydrolysis. The proposed interactions are ionic in the cases of Lys and Arg, which carry a positive charge. Asn appears to hydrogen bond with the terminal oxygen on the phosphate end of ATP. Figure 3 shows the enzyme cave and its amino acid components. The ΔG of this structure determined by energy minimization is -23881.772 kJ/mol. Arg246 was identified as a potential critical amino acid by members of Dr. Berndsen's Biochemistry course as well. For these reasons, we wanted to investigate a mutation of this amino acid to see its effects on Uba5's ability to bind to Ufm1.

The next structure that was modeled was that of the pentavalent transition state, shown as the second structure in Figure 2. In this structure, Gly83 and Val84 are situated directly next to the oxygen that holds the negative charge, and their hydrogens are pointed toward the ATP molecule (Figure 4). Based on this model, these amino acids likely act as an oxyanion hole that stabilizes the unstable pentavalence transition state, which is shown in Figure 2 as the second structure, with the α phosphate of ATP having five bonds. These two amino acids were also identified by Dr. Berndsen's Biochemistry lecture class as being potentially important to the mechanism. The ΔG was determined to be -23448.639 kJ/mol, a slight increase in energy from the starting materials.

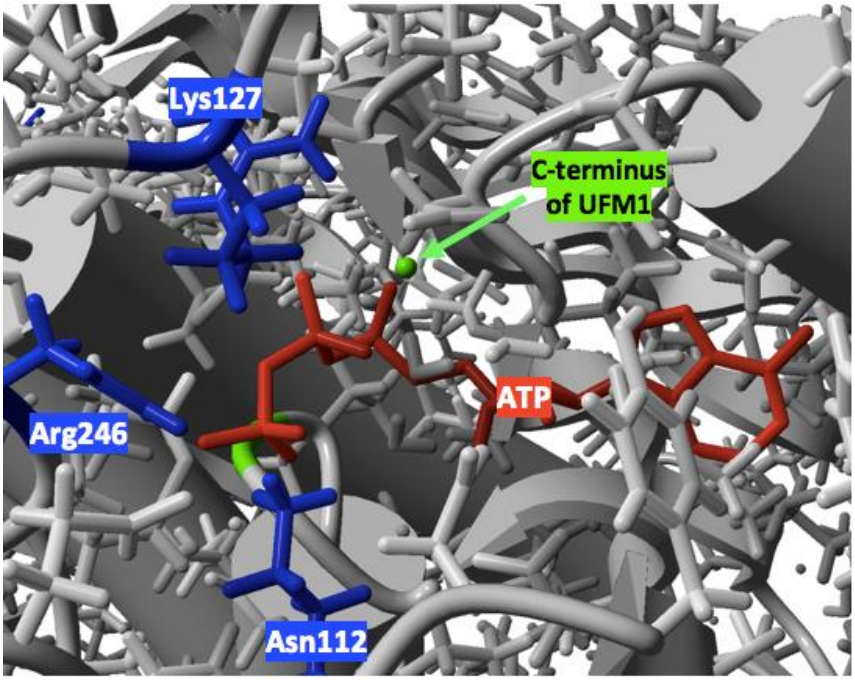

Figure 3. Model of the Uba5~Ufm1 thioester complex with ATP positioned in the active site. The positively charged amino acids highlighted in blue.

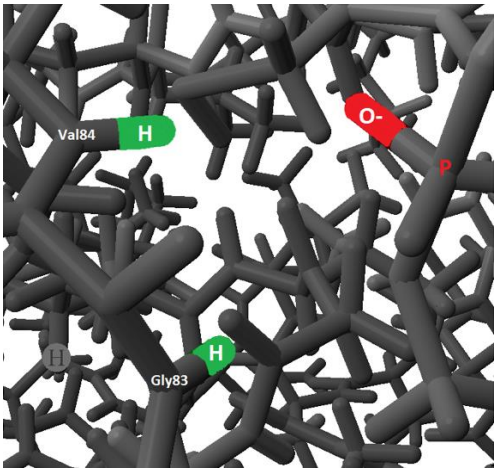

Figure 4. Model of oxyanion hole formed by Val84 and Gly83, which are shown on the left in green. Shown also is the oxygen molecule of ATP, which holds electrons as part of the transition state.

The next structure modeled was that of the first intermediate, Ufm~AMP, after the pyrophosphate has been cleaved. It is at this mechanistic step that the catalytic Cys (Cys250) attacks Ufm1[17]. Upon analysis of the structure for potential residues that could deprotonate the Cys, none were apparent. In this model, Cys250 is 13.4 Å from the C-terminus of Ufm1. The ΔG

value is -25195.985 kJ/mol. This leads us to believe that some structural change must take place in order for the Cys residue to be activated, or it already exists in the thiolate state. Mutants of residues Gly83, Gly82, Arg115, Arg55, Pro255, Lys245, Arg246, Lys242, C250, Glu247, V84, Pro232, Glu241, and Tyr53 were then synthesized by Genscript for purification and biochemical study.

Gravity column purification of Uba5

The first purification attempt used was a gravity column purification, using ion exchange chromatography with nickel resin. As shown in Figure 5, the gel showed no bands at the molecular weight of Uba5, 40 kD, in any of the wash or elution lanes. The only lane with a band where one would expect to see Uba5 is the supernatant. Multiple purifications using gravity column method all produced this same result. This suggests that the protein did not bind to the resin in the gravity column.

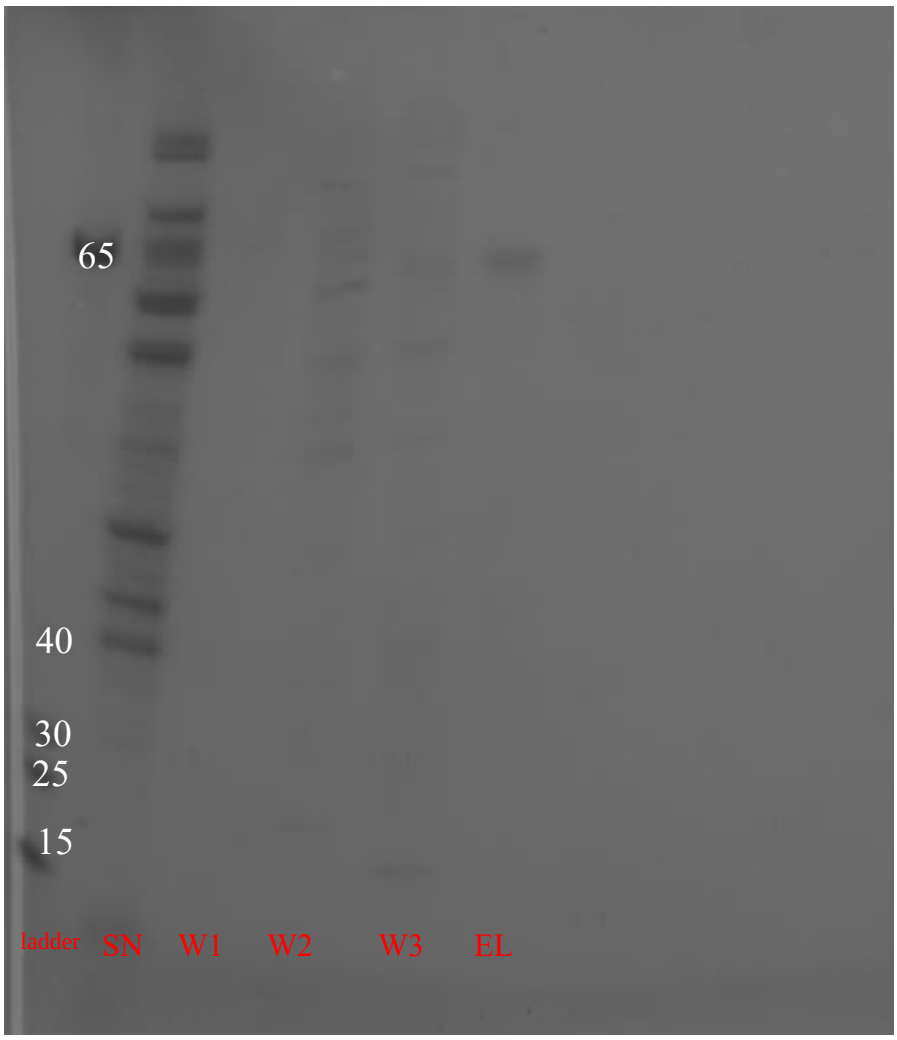

Figure 5. SDS PAGE gel showing attempted purification of wild-type Uba5 using gravity column with nickel resin. Numbers on the left correspond to molecular weights, in kD, as shown by the ladder. Lanes from left to right are as follows: supernatant flow through, wash one (10 mM imidazole), wash two (50 mM imidazole), wash three (100 mM imidazole), and elution buffer (300 mM imidazole).

FPLC purification

The next purification method for wild type Uba5 was FPLC using 1 mL Histrap^{TM} HP nickel column. Shown below in Figure 6 is a representation of changing UV absorbance of the flow through the column as concentration of imidazole increases. Fractions T1-T9 represent the

sample application, with very high UV absorbance due to cellular contents and proteins not binding to the column. Beginning at T11 is the elution at fractionation, beginning at 10 mM imidazole, gradually increasing to 300 mM imidazole. For these fractions, there is a "spike" in the A280 absorption at T11, which is the beginning of the buffer application, and largely steady absorbance values after elution began. The SDS PAGE gel showed that Uba5 eluted at fractions 14, 15, and 16, corresponding to an imidazole concentration of about 65-100 mM imidazole. This was the first purification that was completed with the 1 mL column, and based on the gel, the protein was purified successfully with minimal contamination.

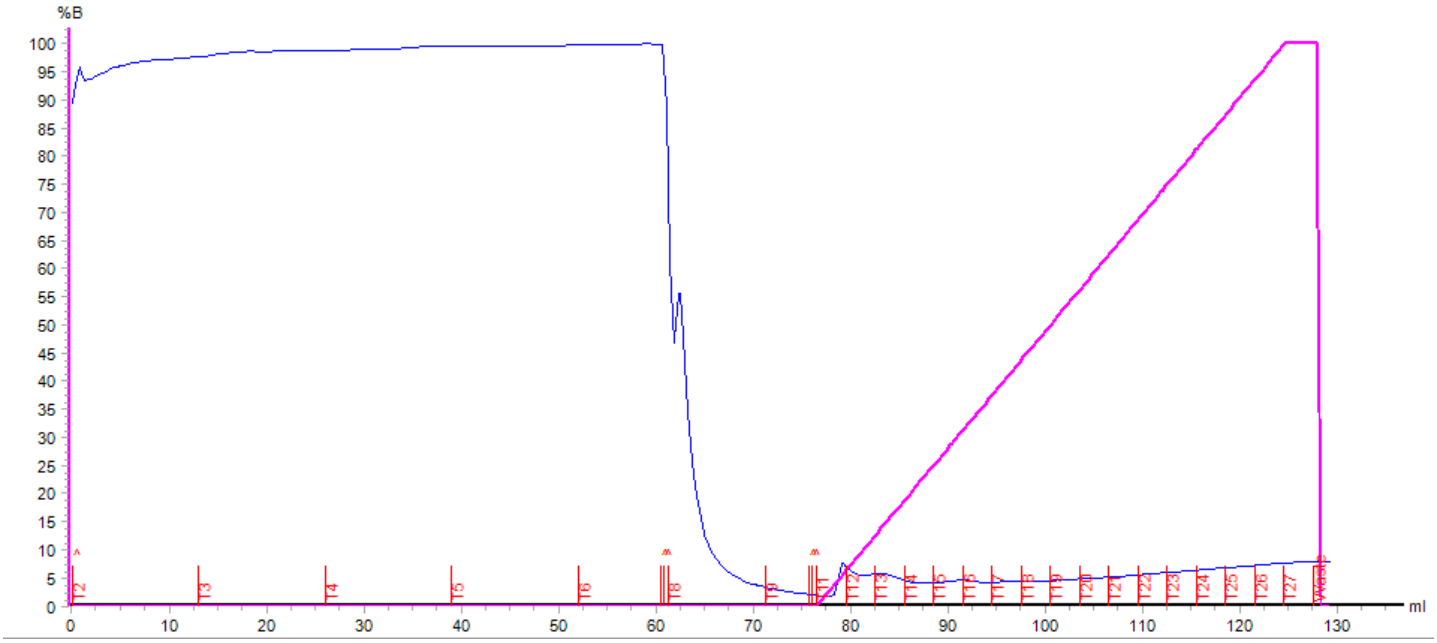

Figure 6. Graph of changing UV absorbance of column flow through with increasing mobile phase concentration, as a function of total flow through volume. Shown in blue is UV absorbance, in mAU. Shown in pink is buffer concentration of imidazole, in %B, where buffer B is 300 mM and buffer A is 10 mM imidazole. Red markers indicate fraction divisions.

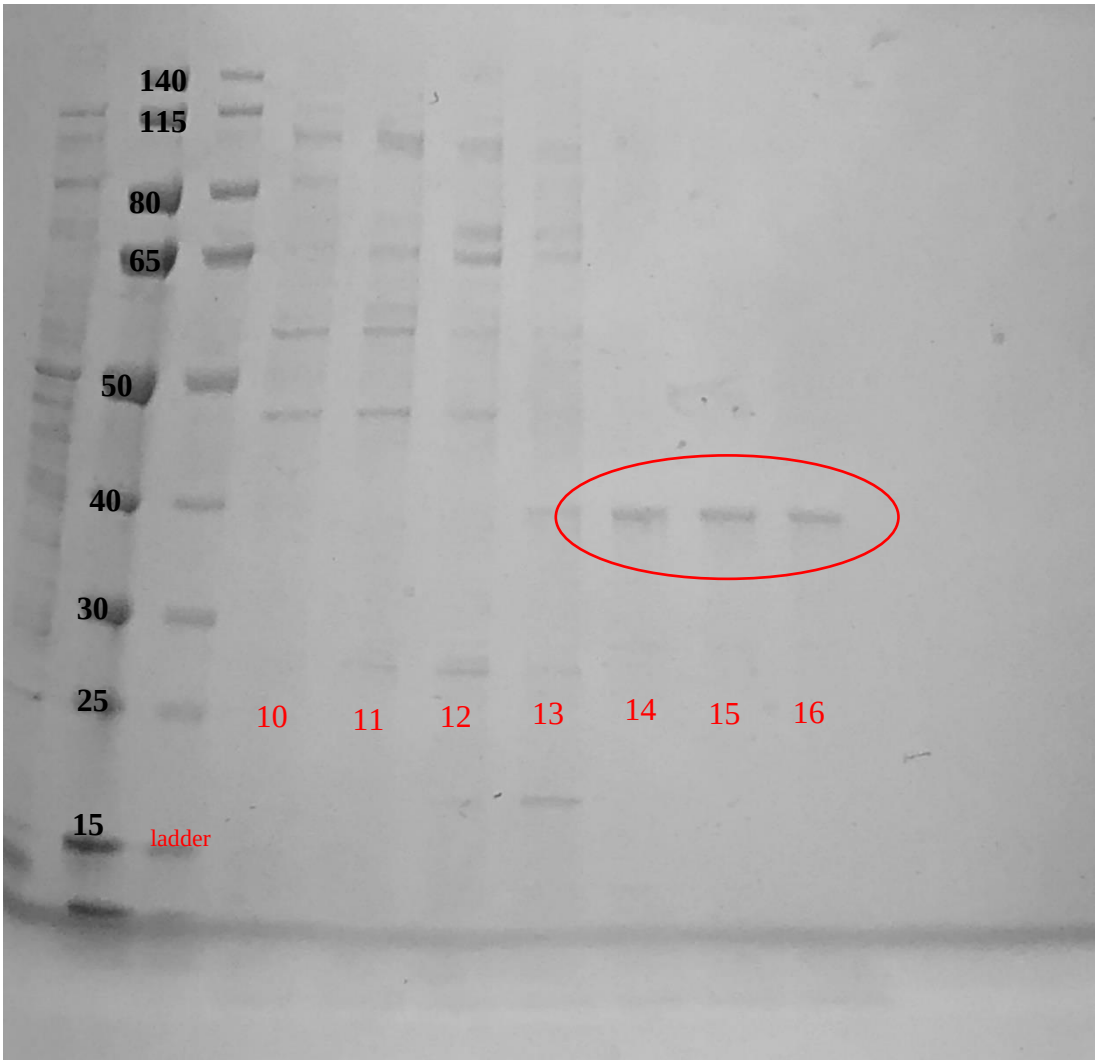

Figure 7. SDS PAGE gel from FPLC purification of wild type Uba5 using 1 mL Histrap™ HP nickel column. Shown on the left is PageRuler Prestained Protein Ladder with markers for each molecular weight in units of kD. Marked in red are fractions from the purification described in Figure 6. Bands corresponding to Uba5 are circled in fractions 14, 15, and 16.

After purification of wtUba5, G83P, G82P, R246A, V84A, and V84L mutations were expressed and the purification protocol from above was used in an attempt to purify them. However, purifying the mutants proved more difficult than the wild type. Using run logs from FPLC, purifications initially appeared to have yielded protein. Shown in Figure 8 is a run log of purification of V84A Uba5, in which a significant amount of UV absorption can be seen in fractions 19-21. This elution of protein happened at lower imidazole concentrations than previous successful purification of wtUba5, at between 24.5 mM and 53.5 mM imidazole.

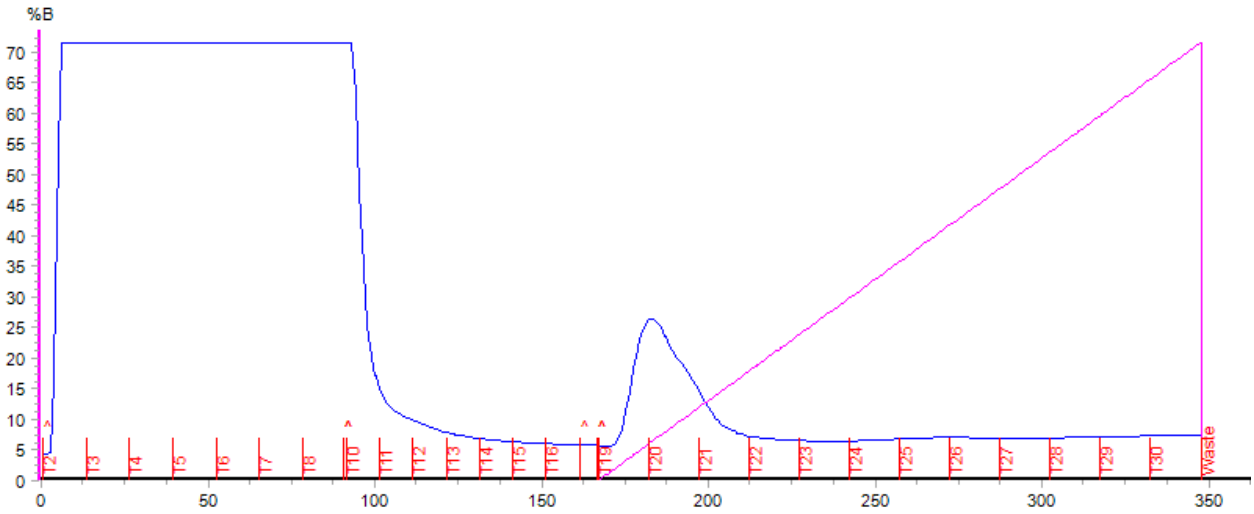

Figure 8. Run log for V84A purification. Fraction markers are shown in red. Shown in pink is increasing imidazole concentration of elution buffer in units of %B, where [buffer B]= 300 mM, and [buffer A]= 10 mM. Shown in blue is UV absorbance, with a spike between fractions 19 and 21.

Despite this initial sign of high protein yield, the band around 40 kD that had been seen in previous gels (Figure 7) was not present, as shown in Figure 9. There were also significantly more bands in the gel than previously, which indicated that our level of purity had decreased. This result began to be seen for every purification that was performed. Shown in Figure 10 is an SDS PAGE gel showing concentrated fractions from mutant purifications, as well as a later purification of wtUba5 completed after mutant purification attempts. Mutations not included in Figure 10 were not concentrated due to lack of any bands around the 40 kDa mark in any fractions. The later wtUba5 and G82P showed faint bands in the target molecular weight, with significant contamination. V84L lacked a band at the target weight entirely, with significant contamination. The G83P sample showed no bands on the gel (Figure 10).

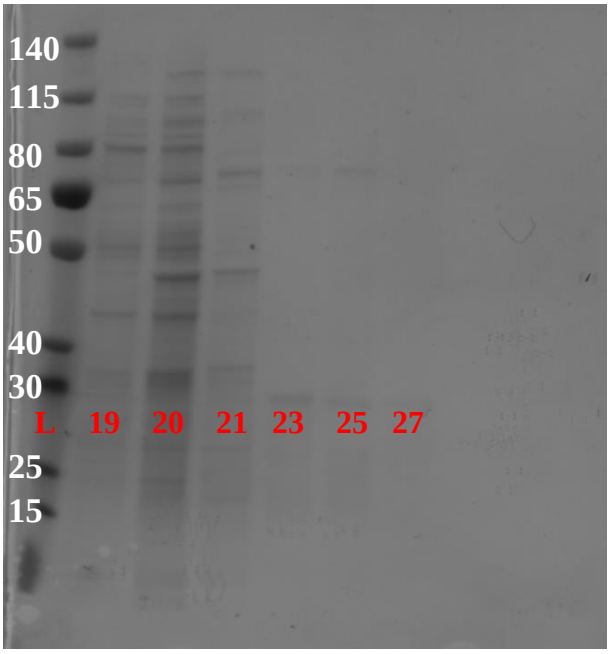

Figure 9. SDS PAGE gel showing fractions from attempted purification of V84A mutation. Fractions marked in red correspond to fractions from the run log shown in Figure 8.

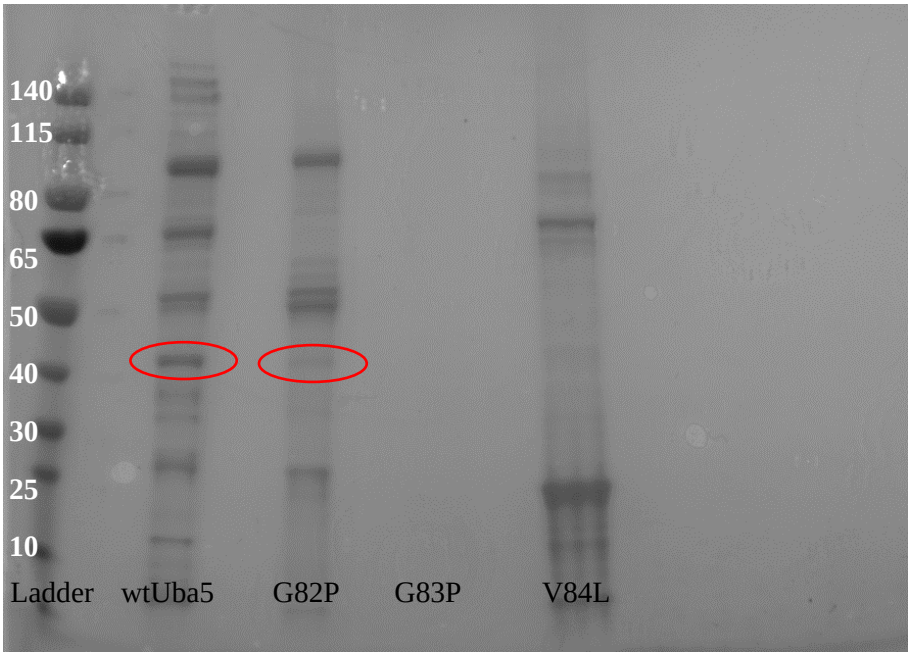

Figure 10. SDS PAGE gel showing concentrated samples of wtUba5 and mutant constructs. Ladder markings in white are in units of kDa.

Troubleshooting

As we began to use the 1 mL column repeatedly, the column became discolored, as shown in Figure 12, and decreasingly able to sustain high pressures. Shown in Figure 11 is one of many run logs in which pressure was unable to be sustained, and the system stopped because pressure became too high. In response to this, the flow rate was decreased until the pressure became sustainable by the column. This was around the time that purifications began to fail, so we thought that lack of pressure, or the lack of ability for the column to handle pressure, was to blame. A new, 5 mL Histrap was obtained and used for future purifications; however, purification results did not improve.

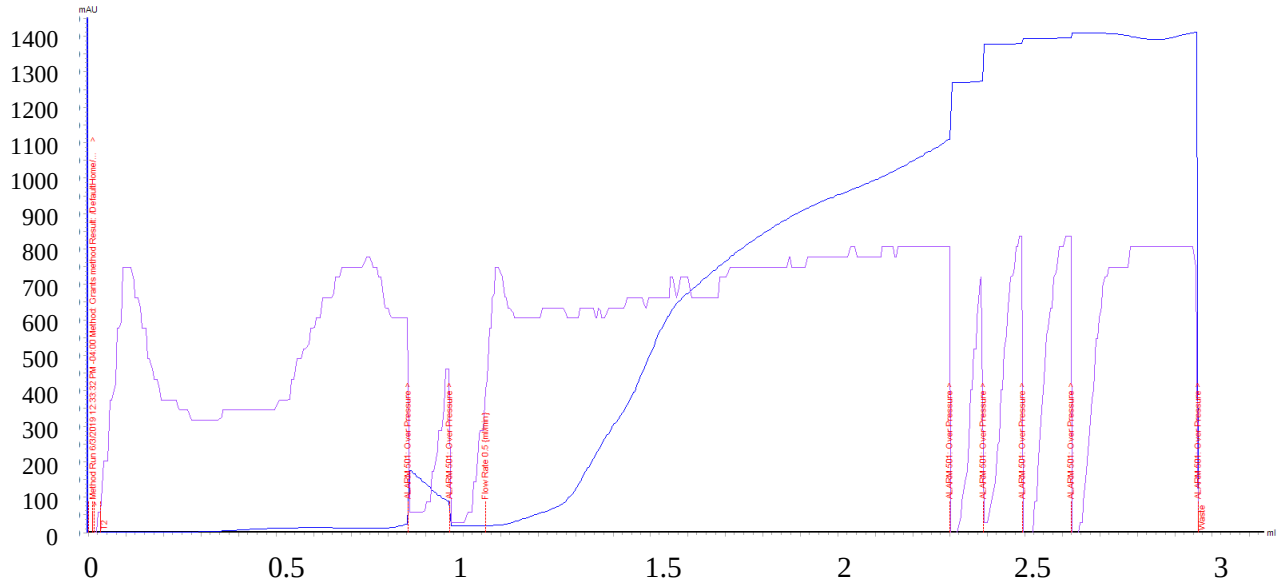

Figure 11. Run log of FPLC purification of wtUba5 using 1 mL column. Shown in blue is change in UV absorbance, in units of mAU, with increasing flow through volume, shown on the x-axis. Shown in purple is pressure of the sample, and Shown in red are error messages indicating that the system was over pressure, and the purification was stopped.

Figure 12. Shown in the image on the left is the 1 mL Histrap™ nickel column after multiple Uba5 purifications (left) compared to a new column. Shown on right is the 5 mL Histrap™ column.

Small angle X-ray scattering (SAXS)

Shown in Figure 13 is the Guinier plot for a single frame exposure of wtUba5 in buffer solution. Shown in black are the data points on ln(Intensity) as a funcition of s^2 (which is equivalent to $q^2[Å^{-2}]$), alongsite a linear average shown in red. A deviation from linear, which is mainly seen at low s^2 in the Figure, is indicative that there is aggregation present[31]. This plot produced a high R_G value 152.20 Å, which is also indicative of aggregation as this value is much larger than the predicted 27 Å from crystal structures.

From the curve of scattering data, shown on the left in Figure 14, the program generated an automated P(r) plot, shown on the right in Figure 14, using all the data points. This plot produced a D_{max} value of 563 Å, which suggests there is significant aggregation. The resulting P(r) plot has significant error bars toward higher values of s^2, suggesting that the aggregated segment is found in these data points. Using a truncated data range that minimized the aggregated fraction, we generated the scatter and P(r) plots shown in Figure 15. With the

aggregated fraction eliminated, the Gunier plot produced an R_G value of 41.10, and the P(r) produced a value of 41.13. These values are closer to the expected values suggesting that there is folded protein present. Despite the removal of aggregation from higher r values, there is aggregation that persists, as shown in Figure 15 by the deviation from the blue line at very small angles ($s < 0.025$). Because the presence of any aggregation largely influences the entire data set[31], no further analysis was performed.

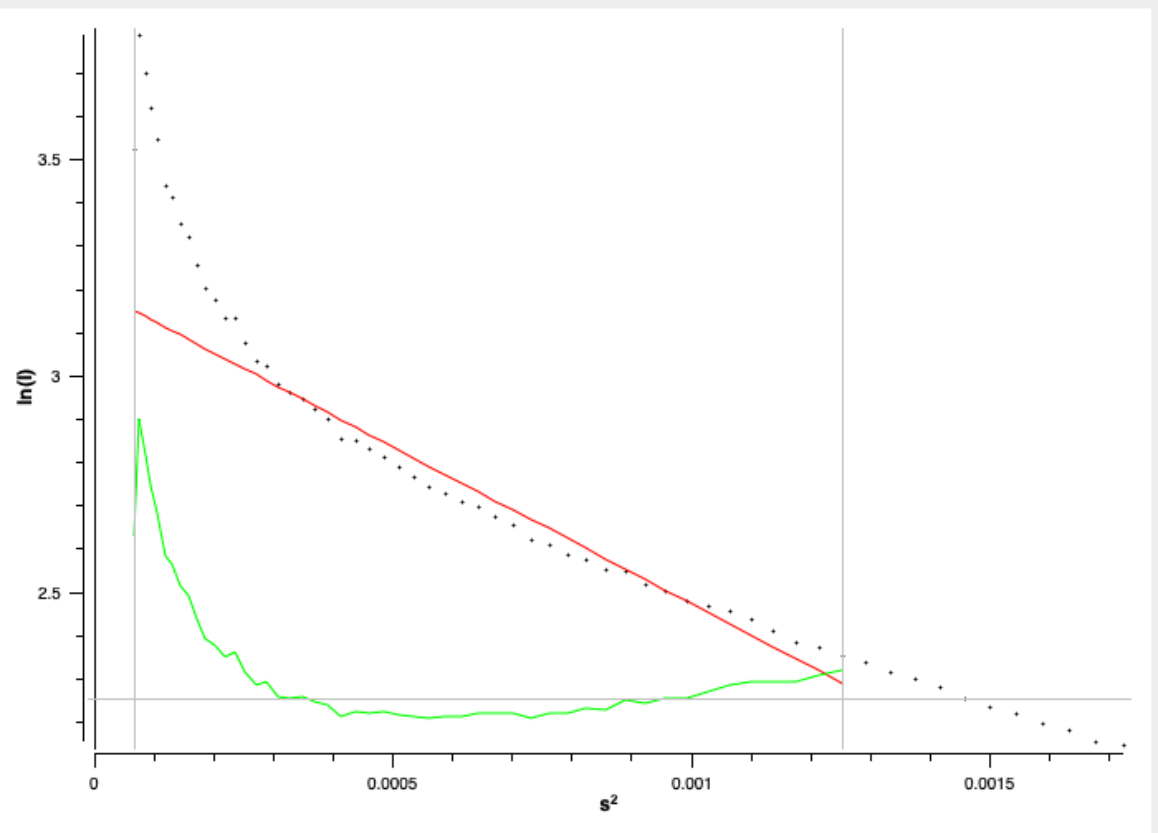

Figure 13. Gunier plot of a single exposure. The variable on the x-axis is a product of the scattering angle (q^2[$Å^{-2}$]) and on the y-axis is Intensity on a logarithmic scale.

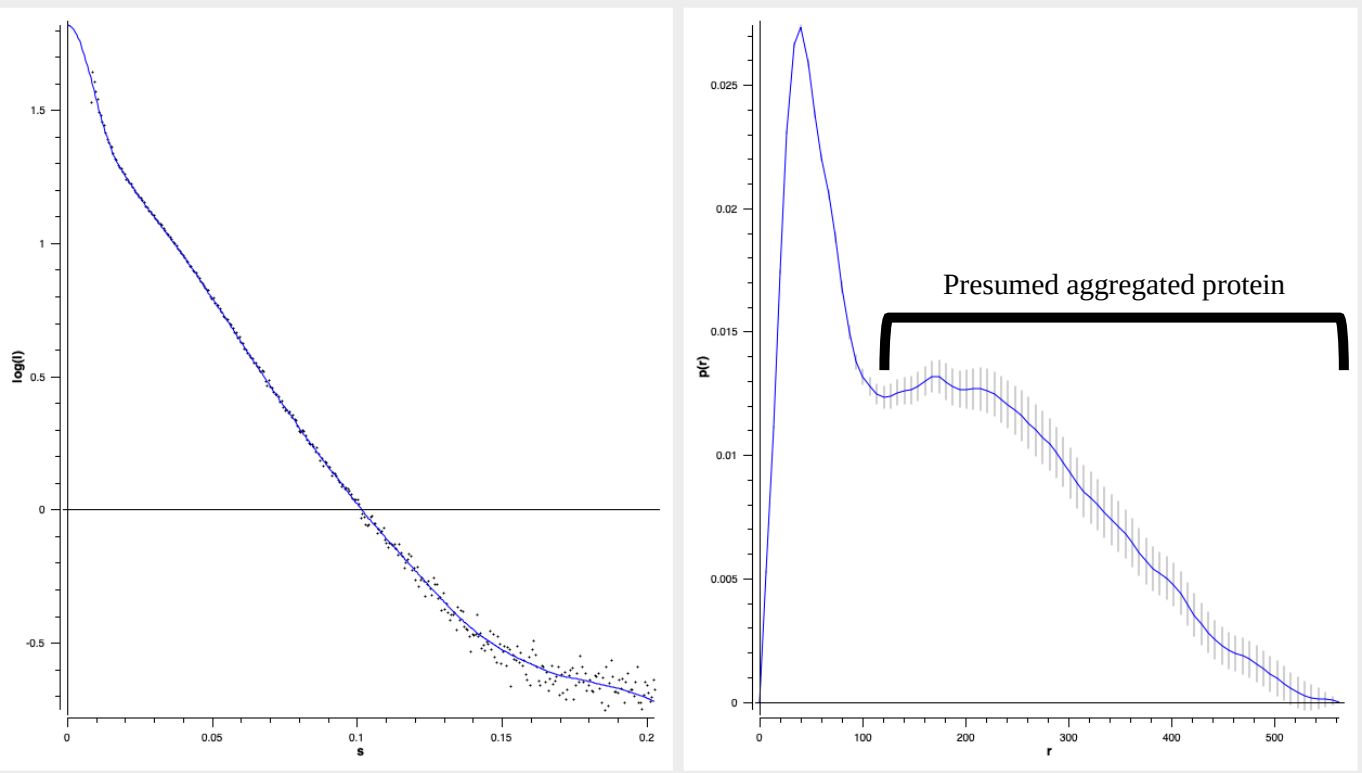

Figure 14. Scattering plot (left) which was automically fitted by PRIMUS software, and the resulting Pair Distribution function plot (right).

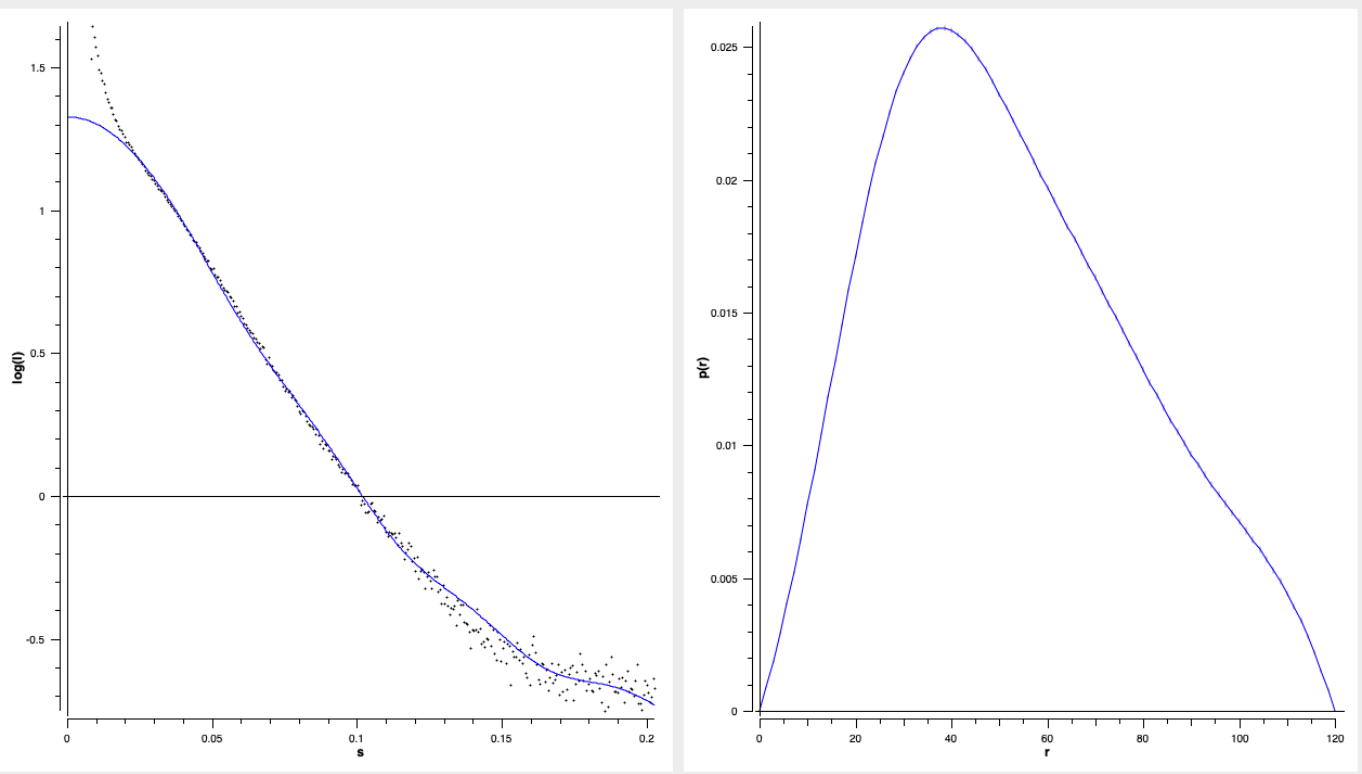

Figure 15. Scattering plot (left) which was manually fitted by removing the aggregated segment of data points, and the resulting Pair Distribution function plot (right).

Discussion

Prior to this project, the Berndsen lab group did not have a reproducible method for purifying wild type Uba5 that consistently yielded protein. The gravity column method was being used, but results were inconsistent. After the acquisition of the automated FPLC system, a reliable method of purifying Uba5 using this new equipment was needed. It appeared initially that we had accomplished this, as evidenced by Figure 7, and other SDS PAGE gels showing similar results. However, purifications became less and less successful over time, eventually looking more like gels in Figures 9 and 10. The lab as a whole, which shares the same FPLC equipment, experienced decreased protein yield and decreased level of purity during this time period. Efforts to iluciate the problem and devise a plan to correct it came to a pause due to the COVID-19 pandemic.

At first, we thought that the issues were caused by pressures being too high in the column and dirty columns. Figure 11 is an example of a run log in which the system shut off multiple times, with an error message that the column was over pressure. Our solution to this was to acquire a larger a 5 mL Histrap^{TM} as opposed to the 1 mL column that was being used. Larger columns have been indicated as a possible cause for contaminant protein elution, due to too many binding sites compared to amount of desired protein, leading to increased binding of other (contaminant) proteins[32]. Contamination of other proteins became a larger problem after we began using the larger column, so this could have contributed to that problem.

Dirty columns could have also contributed to problems with purification. Despite cleaning and storing the column in ethanol, and periodic stripping and recharging, the column frequently became visually discolored, as shown in Figure 12 to the left. We do not really know why this occurred, there are several possible explanations. Firstly, there was likely an insufficient

cleaning and stripping of the column. Stripping and reappying the resin onto the column likely should have been done more often, and other cleaning methods should have been implored to combat contamination in the column. A root cause of this dirtying of the column could have been an insufficent centrifugation procedure prior to loading the column, causing cellular debris or even DNA to contaminate the sample[32].

The most recent purification that took place was the wild-type purification that is shown in Figure 10. The method yielded protein, indicated by the band of 40 kD in the second lane (circled in red). This method took place right after a stripping, thorough cleaning, and recharging of the 5 mL column, in an attempt to reproduce our earlier successful wild-type purifications. We were able to recreate the thick band at 40 kDa, but failed at producing the purity of the previous purifications, as there were several co-eluted proteins. In earlier purifications with the 1 mL column, the problem was that we were unable to produce protein at all. This suggests that the problem of lack of protein yeild is caused by a dirty column, and the problem of contamination is caused by a column that is too large. Combining a lack of adequate cleaning of the column with possible presence of cellular debris, as described above, could have caused the worsening of the resin's ability to bind protein. This would also explain the pressure issues. By switching to a larger column and eventually improving routine cleaning of the column, I postulate that the problem of inability of the column to bind Uba5 was exchanged for a problem of too many binding sites, causing contamination to emerge as our new problem.

Our SAXS data from Figures 14 and 15 contained aggregated segments despite attempts to remove them. Due to the unstable nature of Uba5 that we now know, further precautions should be taken to ensure maximum stabilty of protein, such as keeping protein frozen as much as possible and handling Uba5 in a cold room. However, some proteins tend to aggregate over

time or during concentration[33]. To combat the small amount of aggregation that may be inevitable in these experiements, size exclusion chromatography (SEC) should be used to remove aggregated protein before samples are sent to SIBYLS. Since we appear to have folded protein mixed with aggregates, this approach will allow us to isolate the folded protein and improve our data. SEC-SAXS is a technology used at SIBYLS designed to address this problem. They combine the two techniques, and samples are analysed with SAXS right as they come off of the SEC column. With aggregation of Uba5 being a recurring problem in our lab, this technique should be used with all samples of Uba5.

Future Directions

This project has been a learning experience for everyone involved. We knew at the beginning that Uba5 is a difficult protein to work with, and we acknowledge that there may be other factors unknown to us, or perhaps out of our control that have lead to difficulty working with the protein. I hope that the students who continue this project will learn from our work and be able to accomplish the goals that we aimed for at the start of this project. In this section we will propose methods to optimize the purification protocol, some of which address issues outlined in the discussion section.

First, the lab should verify that Uba5 is still being expressed, and that protein is not denaturing before purification even begins. With little to no yields of Uba5 in any fractions some purifications, issues with expression may be a contributing factor. Using gel electrophoresis, the 2XYT media should be analyzed to see if Uba5 is indeed being expressed. This could be done pre- and post-induction with IPTG to show if production of Uba5 is being induced properly in the BL21 cells. A sample of cells could also be run on a gel just before it is purified, to ensure

that the lysis and sonication processes were done properly, and that there is in fact Uba5 to bind to the column. This can also be done by lysing the cells, adding them to nickel beads, and running the beads on a gel. Protein purifications of late were done using pellets that were kept for months frozen at -80°, so this would show whether issues with Uba5 yeild come from keeping pellets for too long in frozen storage. If post induction media or pellets are shown to lack Uba5, then researchers should attempt look into the IPTG induction protocol, lysis and sonication steps, or shelf life of cell pellets.

Researchers should next test if the His tag is present and functional when purification occurs. Degradation of the His tag would cause Uba5 to not bind to the column, and is listed by Gold technology as a common cause of lack of protein binding[32]. If Uba5 is not binding to the resin at all due to His tag degradation, this could explain binding of a variety of proteins to the column, as Uba5 would not be occupying a significant amount of binding sites. Researchers could confirm the presence of the His tag using Western blotting with an anti His antibody or a Rapid Competitive Assay kit. If these tests suggest that the His tag is damaged, researchers should conduct the lysis and purification procedure at lower temperatures to prevent degradation.

Researcher should take extra steps to ensure that the column is clean. As stated previously, lack of binding or resin specificity to Uba5 could have been caused by an unclean column. Moving forward with the Uba5 project, the lab should build upon and use more cleaning methods for the column. Firstly, stripping and recharging of the column should happen about every 5 purifications and not just when the column fails or becomes discolored. Further cleaning methods should be added to the stripping/recharging process. Before recharging, the column could be washed with 1.5 M NaCl, 1 M NaOH, and 30% isopropanol to clean the column of any contaminant proteins, whether they be ionically or hydrophically bound[29]. To ensure that

needless contamination of the column does not occur, a few adjustments can be made. Firstly, researchers should add a second sonication and centrifugation step in the protocol. Through a second round of centrifugation, we could be more certain that there is no contamination of cellular debris in the purification, which this step is designed to prevent. Seconly, the sample should then be passed through the filter multiple times, or the lab could even aquire a smaller size filter to filter out more potential contaminants that may remain after sonication/centrifugation. Finally, samples should also be diluted even more to prevent high system pressures, and potential aggregation of unwanted proteins or other molecules in the column.

To address the issue of contamination, we suggest that researchers go back to using the 1 mL column. If the pressure issues that were previously encountered were caused by a dirty column, then executing the measures discussed in the above paragraph should lower pressure values to a level that is sustainable by the 1 mL column.

The work presented here is only the start of a continuing project investigating the mechanism of Uba5. Once purification methods are refined, researchers should apply these methods to wild type and mutant Uba5 to obtain clean protein to be used for structural and functional studies. Though these analyses comparing mutant to wild type could not be performed due to obstacles and time limitations, we hope that continuation of the project will lead to our original aspirations coming to fruition, and data advancing our understanding of this important biological process.

References

(1) Berg, J. M.; Tymoczko, J. L.; Stryer, L. Protein Turnover and Amino Acid Catabolism. **2002**.

(2) Grau-Bové, X.; Sebé-Pedrós, A.; Ruiz-Trillo, I. The Eukaryotic Ancestor Had a Complex Ubiquitin Signaling System of Archaeal Origin. *Mol. Biol. Evol.* **2015**, *32* (3), 726–739. https://doi.org/10.1093/molbev/msu334.

(3) Mukhopadhyay, D.; Riezman, H. Proteasome-Independent Functions of Ubiquitin in Endocytosis and Signaling. *Science*. January 12, 2007, pp 201–205. https://doi.org/10.1126/science.1127085.

(4) Berg, J.; Tymoczko, J.; Stryer L. Covalent Modification Is a Means of Regulating Enzyme Activity - Biochemistry - NCBI Bookshelf; 2002.

(5) Kelly, R. T.; Oweis, W.; Wiener, R.; Berndsen, C. E. Conformational Locking of Ufm1 upon Binding to the Ufm1-Interacting Sequence of Uba5. *bioRxiv* **2017**, 161802. https://doi.org/10.1101/161802.

(6) Oweis, W.; Padala, P.; Hassouna, F.; Cohen-Kfir, E.; Gibbs, D. R.; Todd, E. A.; Berndsen, C. E.; Wiener, R. Trans-Binding Mechanism of Ubiquitin-like Protein Activation Revealed by a UBA5-UFM1 Complex. *Cell Rep* **2016**, *16*, 3113–3120. https://doi.org/10.2210/PDB5IAA/PDB.

(7) Padala, P.; Oweis, W.; Mashahreh, B.; Soudah, N.; Cohen-Kfir, E.; Todd, E. A.; Berndsen, C. E.; Wiener, R. Novel Insights into the Interaction of UBA5 with UFM1 via a UFM1-Interacting Sequence. *Sci. Rep.* **2017**. https://doi.org/10.1038/s41598-017-00610-0.

(8) Krieger, E.; Vriend, G. YASARA View - Molecular Graphics for All Devices - from Smartphones to Workstations. *Bioinformatics* **2014**.

https://doi.org/10.1093/bioinformatics/btu426.

(9) Vuay-Kumar, S.; Bugg, C. E.; Wilkinson, K. D.; Cook, W. J. *Three-Dimensional Structure of Ubiquitin at 2.8 A Resolution (Crystal Structure/Protein Conformation)*; 1985; Vol. 82.

(10) Vijay-kumar, S.; Bugg, C. E.; Cook, W. J. Structure of Ubiquitin Refined at 1.8 Å Resolution. *J. Mol. Biol.* **1987**, *194* (3), 531–544. https://doi.org/10.1016/0022-2836(87)90679-6.

(11) Burroughs, A. M.; Balaji, S.; Iyer, L. M.; Aravind, L. Small but Versatile: The Extraordinary Functional and Structural Diversity of the β-Grasp Fold. *Biol. Direct* **2007**, *2*. https://doi.org/10.1186/1745-6150-2-18.

(12) Hochstrasser, M. Evolution and Function of Ubiquitin-like Protein-Conjugation Systems. *Nature Cell Biology*. August 2000. https://doi.org/10.1038/35019643.

(13) Arrigoni, A.; Grillo, B.; Vitriolo, A.; De Gioia, L.; Papaleo, E. C-Terminal Acidic Domain of Ubiquitin-Conjugating Enzymes: A Multi-Functional Conserved Intrinsically Disordered Domain in Family 3 of E2 Enzymes. *J. Struct. Biol.* **2012**, *178* (3), 245–259. https://doi.org/10.1016/j.jsb.2012.04.003.

(14) Schulman, B. A.; Wade Harper, J. Ubiquitin-like Protein Activation by E1 Enzymes: The Apex for Downstream Signalling Pathways. *Nature Reviews Molecular Cell Biology*. May 2009, pp 319–331. https://doi.org/10.1038/nrm2673.

(15) Cappadocia, L.; Lima, C. D. Ubiquitin-like Protein Conjugation: Structures, Chemistry, and Mechanism. *Chemical Reviews*. American Chemical Society February 14, 2018, pp 889–918. https://doi.org/10.1021/acs.chemrev.6b00737.

(16) Haas, A. L.; Warms, J. V. B.; Rose, I. A. Ubiquitin Adenylate: Structure and Role in

Ubiquitin Activation. *Biochemistry* **1983**, *22* (19), 4388–4394. https://doi.org/10.1021/bi00288a007.

(17) Gavin, J. M.; Hoar, K.; Xu, Q.; Ma, J.; Lin, Y.; Chen, J.; Chen, W.; Bruzzese, F. J.; Harrison, S.; Mallender, W. D.; et al. Mechanistic Study of Uba5 Enzyme and the Ufm1 Conjugation Pathway. *J. Biol. Chem.* **2014**, *289* (33), 22648–22658. https://doi.org/10.1074/jbc.M114.573972.

(18) Wenzel, D. M.; Stoll, K. E.; Klevit, R. E. E2s: Structurally Economical and Functionally Replete. *Biochemical Journal*. January 1, 2011, pp 31–42. https://doi.org/10.1042/BJ20100985.

(19) Huang, L.; Kinnucan, E.; Wang, G.; Beaudenon, S.; Howley, P. M.; Huibregtse, J. M.; Pavletich, N. P. Structure of an E6AP-UbcH7 Complex: Insights into Ubiquitination by the E2-E3 Enzyme Cascade. *Science (80-.).* **1999**, *286* (5443), 1321–1326. https://doi.org/10.1126/science.286.5443.1321.

(20) Sasakawa, H.; Sakata, E.; Yamaguchi, Y.; Komatsu, M.; Tatsumi, K.; Kominami, E.; Tanaka, K.; Kato, K. Solution Structure and Dynamics of Ufm1, a Ubiquitin-Fold Modifier 1. *Biochem. Biophys. Res. Commun.* **2006**, *343* (1), 21–26. https://doi.org/10.1016/j.bbrc.2006.02.107.

(21) Lemaire, K.; Moura, R. F.; Granvik, M.; Igoillo-Esteve, M.; Hohmeier, H. E.; Hendrickx, N.; Newgard, C. B.; Waelkens, E.; Cnop, M.; Schuit, F. Ubiquitin Fold Modifier 1 (UFM1) and Its Target UFBP1 Protect Pancreatic Beta Cells from ER Stress-Induced Apoptosis. *PLoS One* **2011**, *6* (4). https://doi.org/10.1371/journal.pone.0018517.

(22) Lu, H.; Yang, Y.; Allister, E. M.; Wijesekara, N.; Wheeler, M. B. The Identification of Potential Factors Associated with the Development of Type 2 Diabetes: A Quantitative

Proteomics Approach. *Mol. Cell. Proteomics* **2008**, *7* (8), 1434–1451. https://doi.org/10.1074/mcp.M700478-MCP200.

(23) Hamilton, E. M. C.; Bertini, E.; Kalaydjieva, L.; Morar, B.; Dojčáková, D.; Liu, J.; Vanderver, A.; Curiel, J.; Persoon, C. M.; Diodato, D.; et al. UFM1 Founder Mutation in the Roma Population Causes Recessive Variant of H-ABC. *Neurology* **2017**, *89* (17), 1821–1828. https://doi.org/10.1212/WNL.0000000000004578.

(24) Yoo, H. M.; Kang, S. H.; Kim, J. Y.; Lee, J. E.; Seong, M. W.; Lee, S. W.; Ka, S. H.; Sou, Y. S.; Komatsu, M.; Tanaka, K.; et al. Modification of Asc1 by Ufm1 Is Crucial for Erα Transactivation and Breast Cancer Development. *Mol. Cell* **2014**, *56* (2), 261–274. https://doi.org/10.1016/j.molcel.2014.08.007.

(25) Lee, A. D.; Berndsen, C. E. Structural and Functional Characterization of the Leishmania donovani Ufm-ylation Pathway | The FASEB Journal https://www.fasebj.org/doi/abs/10.1096/fasebj.2019.33.1_supplement.465.3 (accessed Jan 21, 2020).

(26) Olsen, S. K.; Capili, A. D.; Lu, X.; Tan, D. S.; Lima, C. D. Active Site Remodelling Accompanies Thioester Bond Formation in the SUMO E1. *Nature* **2010**, *463* (7283), 906–912. https://doi.org/10.1038/nature08765.

(27) Bacik, J. P.; Walker, J. R.; Ali, M.; Schimmer, A. D.; Dhe-Paganon, S. Crystal Structure of the Human Ubiquitin-Activating Enzyme 5 (UBA5) Bound to ATP: Mechanistic Insights into a Minimalistic E1 Enzyme. *J. Biol. Chem.* **2010**. https://doi.org/10.1074/jbc.M110.102921.

(28) Gasteiger, E.; Hoogland, C.; Gattiker, A.; Duvaud, S.; Wilkins, M. R.; Appel, R. D.; Bairoch, A. *Protein Analysis Tools on the ExPASy Server 571 571 From: The Proteomics*

Protocols Handbook Protein Identification and Analysis Tools on the ExPASy Server.

(29) Healthcare, G. E.; Hp, H. *Instructions 71-5027-68 AF HisTrap Affinity Columns*.

(30) Konarev, P. V.; Volkov, V. V.; Sokolova, A. V.; Koch, M. H. J.; Svergun, D. I. *PRIMUS* : A Windows PC-Based System for Small-Angle Scattering Data Analysis. *J. Appl. Crystallogr.* **2003**, *36* (5), 1277–1282. https://doi.org/10.1107/S0021889803012779.

(31) Putnam, C. D.; Hammel, M.; Hura, G. L.; Tainer, J. A. X-Ray Solution Scattering (SAXS) Combined with Crystallography and Computation: Defining Accurate Macromolecular Structures, Conformations and Assemblies in Solution. *Quarterly Reviews of Biophysics*. 2007. https://doi.org/10.1017/S0033583507004635.

(32) Biotechnology, G. Affinity His-Tag Purification Troubleshooting Guide. **2014**.

(33) SEC-SAXS Introduction | Small Angle X-ray Scattering/Diffraction https://www-ssrl.slac.stanford.edu/smb-saxs/content/documentation/sec-saxs/introduction (accessed Apr 19, 2020).

Acknowledgements

-JMU REU: NSF Grant No. CHE-1757874

-SAXS data was collected at SIBYLS which is supported by the DOE-BER IDAT DE-AC02-05CH11231 and NIGMS ALS-ENABLE (P30 GM124169 and S10OD018483).

-The Department of Chemistry and Biochemistry, for hosting me, a Biology student, as an undergraduate researcher during REU 2019, and its faculty for helping me discover a love for research.

-Dr. Ray Enke and Dr. Jon Monroe, for serving as committee members and providing support thorughout the project, and to your lab members for collaborating for us.

-The Berndsen Lab Group, current students and alumni, for all the hard work, collaboration, companionship, and memes that made this project possible.

-Dr. Chris E. Berndsen, for going above and beyond as a lecture professor, PI, and mentor to not only myself, but every member of our lab. You have challenged me to do things that I did not think were possible. Your support has played a huge role in making my academic success, in this Capstone and beyond, possible.