2014-01-01

Biochemical Characterization Of Four Distinct Proteins

Gustavo A. Avila
University of Texas at El Paso, gaavila3@gmail.com

BIOCHEMICAL CHARACTERIZATION OF FOUR DISTINCT PROTEINS

GUSTAVO A. AVILA

Department of Chemistry

APPROVED:

Chuan Xiao, Ph.D., Chair

Mahesh Narayan, Ph.D.

Katja Michael, Ph.D.

Jianjun Sun, Ph.D.

Charles Ambler, Ph.D.
Dean of the Graduate School

DEDICATION

My sincerest gratitude is forever extended to those who have endured this endeavor with me. To my parents, family, friends, and mentor I am eternally grateful for your guidance and support.

BIOCHEMICAL CHARACTERIZATION OF FOUR DISTINCT PROTEINS

by

GUSTAVO A. AVILA, B.S.

DISSERTATION

Presented to the Faculty of the Graduate School of

The University of Texas at El Paso

in Partial Fulfillment

of the Requirements

for the Degree of

DOCTOR OF PHILOSOPHY

Department of Chemistry

THE UNIVERSITY OF TEXAS AT EL PASO

August 2014

ACKNOWLEDGEMENTS

My graduate career has been not a solitary journey – the collaborative spirit in research has been an integral component in my research. I would like to extend my absolute gratitude to my advisor, colleagues, family, and friends for all the guidance and support. Their encouragement has made this endeavor less of a daunting task, which would have otherwise been impossible for me. I would like to take this opportunity to express my sincerest gratitude to all those who have aided me. I would like to thank my advisor, Dr. Chuan Xiao, for always challenging his students to develop not only as researchers, but also as independent thinkers. His dedication to science is everlasting, and philosophy regarding collaboration and dissemination of resources only reinforces his ardent passion for scientific advancement. On a more personal note, I would like to thank Dr. Xiao for his on-going support and patience with me. His dedication to research and motivation will continue to inspire me.

I would also like to extend my appreciation to my committee members, Katja Michael, Jianjun Sun, and Mahesh Narayan, for their contributions to my dissertation. Their distinct backgrounds and unique perspectives have helped propel my research. Moreover, their collaborations, suggestions, and benevolent access to their labs and resources have been essential. In particular, I would like to thank Dr.'s Katja Michael, Jianjun Sun, Susana Chiocca, Erquan Zhang, Dapeng Zhou, and German Rosas-Acosta, with whom we have had an ongoing collaboration. It has been a privilege to have worked with so many different minds and encompass many different projects.

I would also like to thank the Chemistry Department, for providing financial assistance the ability to conduct my research in their department and providing an environment conducive for research. A special thanks to the both the NSF-SMARTS program/coordinators, as well as the UTEP IDR for the financial assistance provided during my academic career. Many thanks to the Biology Department core facility BBRC staff and resources, usage of their instruments were paramount to conduct my research.

I would also like to thank my lab colleagues past and present including: Nancy Rondeau, Duer Bolotaulo, Sayan Chakraborty, Scott Rush, Emmanuel Silva, Ricardo Parra, David Acosta, Christina Alvara, Jasmine Peralta, Lu Xiao, Mason Arbogast, Martin Chacon, Brenda Rodarte, Tiffany Blankenship, Merrit Romeike, Nico Silva, and Jorge Lopez. Also to graduate students from outside lab groups: Dr. Zacariah L. Hildenbrand, Dr. Sudheer Molugu, Pedro Jaquez, Dr. Susmita Bandyophadyay, Ricardo McCreary, Emmanuel Zubia, Sanghamitra Majumdar, and Andrew Pardo. Research has been less daunting to undertake with all of your helpful suggestions, assistance, and discussions.

Finally, I would like to thank my family and friends for their enduring support. I thank my Sister Zayra, for her reassurance and love. I cannot put into words the appreciation I have towards my parents. It did not matter how hopeless something seemed, they could always manage to make a bad day into a very important lesson learned. Their love and dedication to my well being is exemplar of good parenting, and would not have accomplished this feat without them. In particular, I thank my mother Leticia, the voice of reason, optimism, and sanity. She has taught me to work through frustrating situations and that not all problems can be resolved easily. You have my eternal love and adulation. I would also like to thank my father Manuel, for teaching me the value of being prepared, and thinking practically. To my late Grandmother, thank you for your love and nurturing nature. To all my family, it has been a turbulent ride, but with all your love and support, I have had the distinct honor to soar with "Eagles", and for that my profound thanks.

Again, to all the people that have made my graduate career not just successful, but also enjoyable, my sincerest gratitude. Thank you.

ABSTRACT

Proteins are ubiquitous in all living organisms, executing the majority of cellular functions in distinct ways. Understanding a protein's role necessitates investigating its structure and function, which are closely related. My research couples these two aspects by delving into the biochemical and structural characterization of proteins in four distinct systems, all playing central roles in numerous significant disease progressions. These four original research endeavors were all targeted for structural studies with a unifying relationship to establish our new structural biochemistry lab. These four systems are: (1) Gam1, an early adenovirus protein globally inhibiting host SUMOylation; (2) Anthrax toxin complexed with its cellular receptor; (3) hCRY1, a core component of the circadian rhythm; and (4) an antibody against cancer cell-surface glycans. Traditional molecular, biochemical, and biophysical techniques such as molecular cloning, *in vitro* protein-protein assays, and structure determination have been utilized in the above mentioned projects to analyze protein functions. All four projects have obtained results at different stages, two have reached key milestones and two have established knowledge towards the future steps to structural analyses. All of these projects have built a strong foundation for the current and future success of the lab. Each project has contributed in general to understand the biological process at the molecular level. In addition, the four chosen systems involve investigations of protein-protein, protein-carbohydrate and protein-DNA interactions, all are critical for elucidating how biomolecules work collaboratively to perform the function of the cell. The results from these projects will ultimately improve our ability to develop novel biological and biomedical applications in hopes of developing treatments for the diseases.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1:

BIOCHEMICAL CHARACTERIZATION OF AN EARLY GENE PRODUCT OF AN ADENOVIRUS GAM1

1.1 INTRODUCTION TO GAM1

After folding into their proper three-dimensional shapes, many proteins will undergo posttranslational modifications (PTM) in order to fully attain their final function [1, 2]. Moreover, many PTMs are reversible, which have the ability to switch protein functionality among different biological isoforms [3-5]. The protein under investigation is Gam1, an early gene product of adenovirus Chicken Embryo Lethal Orphan (CELO) characterized by a molecular weight of 31 kDa. It is involved in the regulation of viral replication. Gam1 was discovered serendipitously as a novel anti-apoptotic protein upon screening for viral proteins that regulate cellular apoptosis [6]. In 1996, Susana Chiocca and colleagues originated their research endeavor by sequencing and characterizing the genomic organization of avian adenovirus (CELO). Further evidence shows that Gam1, a non-structural protein, plays important roles in controlling host cellular machinery so that the virus may effectively replicate itself [7, 8]. In 2002, Gam1 was also found to disrupt transcriptional regulation by inactivation of Histone Deacetylase 1 (HDAC1), an enzyme that regulates gene expression in cells. Most recently, Gam1 was discovered to have a unique function of modulating SUMOylation (Small ubiquitin-related Modifier) [9-11], a reversible PTM that alters protein function. As a result, Gam1 could manage to accomplish the aforementioned multiple tasks. It is still unclear whether Gam1 performs its function via the direct modulation of gene expression, or through PTMs such as the SUMOylation pathway. Proteins found in subnuclear compartments are heavily SUMO modified, a probable mechanism against viral infections [12-14]. Furthermore, the mechanism was recently elucidated of how Gam1 inhibits host SUMOylation [15, 16] described in detail below. All in all Gam1 is not only the first viral protein discovered to impact host SUMOylation, but also the only viral protein to inhibit SUMOylation

globally. Understanding the mechanism of Gam1's function will build the foundation to utilize it as a tool to obliterate SUMOylation in targeted cells. Aspects of what SUMOylation entail, how viruses intertwine into the modification to this cellular process, and how Gam1's functions can alter the normal process of SUMOylation will be discussed in this chapter. Thereafter, research performed with Gam1 will be demonstrated.

1.1.1 SUMOYLATION AND VIRUSES

SUMO (Small ubiquitin-related Modifier) was discovered nearly twenty years ago, when investigation of yeast proteins [17], soon afterward a human protein [18] was altered by SUMO as well. From these studies, it was determined that SUMO would effectively be covalently bound to target proteins impinging on their ability to establish new protein-protein interactions. This attachment of SUMO is a reversible modification. As a result, SUMOylation can switch proteins' function as well as redirect their localization within the cell [19]. In the time since these discoveries, hundreds of experiments conducted have demonstrated numerous, a quantity now in the hundreds, targets of SUMO. Each of these SUMOylated targets are capable of assuming a completely different role in their functionality. Experiments have shown that SUMOylation inhibits protein-protein interactions in a variety of ways including the additional surface of SUMO itself, obstruction of binding sites, or potential structural modifications that have occurred due to the SUMO modification [19]. As a result, SUMO can stabilize, de-localize, or label a protein for degradation. In terms of molecular size, SUMO tags are typically around 10 kilo Daltons (kDa), and their overall three dimensional structures are similar to those of ubiquitins [20-22]. SUMOylation is found throughout eukaryotes, with variations in the number of SUMO genes. For instance, humans have four genes encoding SUMO tags, while *Drosophila melanogaster* only has one [23]. Similar to the ubiquitination reaction, SUMOylation occurs when an isopeptide bond is formed between the C-terminal glycine (Gly) of SUMO and the ε-

amino group of a lysine residue in the target protein. The enzymatic cascade that ensues for both SUMOylation and ubiquitination processes requires the activities of three classes of enzymes: E1, E2, and E3. The cascade begins with activation of the tags catalyzed by the E1 enzymes in ubiquitination and SUMO Activation Enzymes (SAE 1 and SAE2) in SUMOylation. Afterward, the second step is conjugation catalyzed by the E2 enzymes of each system. Finally, substrate modification is catalyzed by the E3 ligases, enzymes that catalyze the transfer of tags from the conjugating enzymes E2 to the target substrates [9, 24]. It is the different E3 ligases utilized by the two systems that define the specificity of SUMOylation and ubiquitination [19].

As abovementioned, the attachment of SUMO to target substrates has been shown to play a variety of roles in protein stability, as well as regulation of transcription by influencing transcription factors or directly activating the process of transcription [9]. Given that viruses require a host to replicate, viruses must produce an environment suitable for their reproduction. One of the important functions that they must possess is to control components responsible for transcription in host cells. Evidence has shown that viruses can passively adapt to host SUMOylation by modifying cellular and/or viral proteins. An example is the vaccinia virus protein A40R, which is SUMOylated and remains in the cytosolic side of the Endoplasmic Reticulum, where viral replication occurs [25]. Viruses also exploit the SUMOylation pathway for viral assembly, one example is the SUMOylation of M1 protein of Influenza A virus [26]. Conversely, some viruses have been found to produce specific proteins, which directly control host SUMOylation in a concerted effort to survive and proliferate [27]. One example is a viral protein from Kaposi's sarcoma-associated herpes virus (KSHV) that acts as a transcriptional regulator, which arrests the cell cycle effectively preventing apoptosis so that viral replication continues [28]. Another example of viral utilization of SUMOylation is for immune evasion. For instance, the Ebola Zaire Virus (EBOV) protein VP35 increases the SUMOylation of a principle cellular component that produces a pro-inflammatory response, thereby weakening the host immune system [29].

1.1.2 GAM1 AND SUMOYLATION

Gam1 was initially believed to facilitate viral replication by affecting host gene expression [9, 11]. Gam1's expression has been shown to increase transcription from various promoters [24]. A Gam1 mutant was generated to test its effect on transcription. The mutant was unable to activate transcription [8]. Moreover, it has been demonstrated that viruses alter subnuclear structures, such as the chromatin-remodeling enzyme histone deacetylase 1 (HDAC1) [11]. HDAC1 prevents transcription by removing acetyl groups from the amino terminal domain of histones, allowing for a favorable interaction between histones and DNA. Proteins found in subnuclear compartments are heavily SUMO modified, and the modifications taking place are known mechanisms against viral infections. This led to the discovery of the true mechanism of how Gam1 facilitate viral replication via interference with the host SUMOylation. Gam1 disrupts these proteins, which led to the discovery that SUMO-1 is diminished in the nucleus [10]. Gam1 expression diminishes global nuclear SUMOylation, by degradation of the SUMO E1 heterodimer (SAE1/SAE2), which ultimately eliminates the SUMOylation pathway [9]. SAE1/2, mediated by Gam1, is incorporated into ubiquitin ligase complexes leading to degradation [16, 24]. It is noteworthy that viruses exploit these cellular degradation pathways to manipulate normal cell processes—such as SUMOylation by Gam1 [24]. Gam1 was the first viral protein discovered to provoke the disappearance of SUMO E1 and E2 proteins, regulating host cell proteins [9].

In vitro research with Gam1 and loss of SUMO occurred in a dose dependent manner, and only when an intact SUMO-E1 complex was not already established [9]. It was revealed that Gam1 does not compete with SUMO-E1 binding; on the contrary, E1 disappeared [9]. These findings were supported when it was shown that the E1 heterodimer was eliminated *in vivo* while CELO infection was taking place [9]. When a proteasome inhibitor associated with ubiquitination is added, SAE1 and SAE2 heterodimer formation is restored, implicating Gam1's function links to the ubiquitin-like pathway [9].

The Ubiquitin Proteasome System (UPS) is an essential moderator of cellular processes, which is carried out via governing protein degradation [30]. As mentioned above, similar to SUMOylation, ubiquitination is carried out by the same enzymatic cascade containing: E1 activating enzyme, the E2 conjugating enzymes, and the E3 ligases. The usage of different E3 ligases is where differences are seen between the SUMOylation and ubiquitination pathways [31, 32]. Given Gam1's association in these PTM pathways, bioinformatical investigations were carried out to define a conserved domain utilized in these modifications [33]. Further analyses of Gam1 demonstrated two motifs: suppressor of cytokine signaling (SOCS) and BC-Box-like-motif (BC box), which are required for establishing a complex containing ubiquitin like proteins, elongin B/C and E3 Ligases, Cullins 2 and 5 [33]. The N-terminal motif termed BC Box interacts with Elongins B, an ubitquitin-like protein, and C, an S-phase kinase-associated protein 1 (Skp1), both of which are involved in protein degradation [34]. Evidence suggests that Elongin BC is central to SOCS complexes and renders Elongin BC as a multifunctional complex capable of regulating cell processes in a variety of ways [34]. The SOCS box, typically found at the C-terminus of a protein, serves as a link between the substrate binding domain and that of the protein(s) associated with aforementioned ubiquitination E3 ligase complexes [24, 33, 35]. The discovery of SOCS in Gam1 is an important breakthrough. Given that SOCS containing proteins play a grand role in mediating protein elimination by targeting for polyubiquitination and their eventual removal via the proteasome, it is rational to conclude that Gam1 destroys SUMOylation via protein degradations. Another major contributor in the complex is the E3 ligase. Given the multiple existences of E3 ligases in both SUMOylation and ubiquitination pathways, members have been classified into different families based on their structure and functionality. One of the classes is the Cullin RING (really interesting new gene) Ligases (CRL's). These families of ligases are the most studied yet are poorly understood. They are not as frequently used in ubiquitin processes, albeit garner ubiquitin-ligase type activity [30]. There are six types of CRLs organized based upon the substrate receptors utilized. The CRLs that are pertinent

to Gam1 are Cullins 2 and 5, which utilize a BC box to recruit target proteins [30]. Gam1, however, does lack two consensus aspects found in SOCS-box containing proteins: it does not contain a P/L rich C-terminus, nor does it contain a highly conserved Cysteine residue at position 6 [24]. Bioinformatical analysis conducted on a purported alpha helical region of Gam1 was overlaid an Elongin BC complexed von Hippel-Lindau supporting the possibility that Gam1's BC Box is indeed functional [36]. Further analysis shows that there exists the possibility of Gam1's C-terminal hydrophobic patch of residues to interact with EloC in its hydrophobic pocket. The interaction between these proteins has been supported by both *in vivo* and *in vitro* assays [24]. Currently multiple experiments have clearly demonstrated that Gam1 functions as a mediator to eliminate the SUMOylation E1 via protein degradation. It has been revealed that Gam1 also causes the loss of E2 within SUMOylation, by a yet unexplained mechanism. No linkage has been established between Gam1 and E3 in SUMOylation.

1.1.3 SUMMARY

In essence, the focal point of the above introduction pertains to the initial discovery of Gam1, whose function renders an environment more conducive for viral replication. Gam1 is the first viral protein discovered to play a role in the global disappearance of SUMOylation. Viruses employ various mechanisms, including manipulation of posttranslational modifications (PTM), in order to quell host elimination. SUMOylation is one of such PTM pathways exploited by viruses, which has a known effect on activating or de-activating gene production so that the virus can reproduce. The protein under investigation Gam1 was the first viral protein discovered to globally inhibit SUMOylation, via the degradation of E1 and E2 proteins in the pathway. SUMO-modified proteins have multifunctional effects on protein stability, function, and also to affect either positively or negatively regarding activation of genes. By impinging on host cell SUMOylation, Gam1 plays an essential role in viral survival. In order for Gam1 to perform its function, it must interact with particular proteins and disrupt

the formation of protein complexes pertinent to SUMOylation. Collectively, these findings of Gam1 were the earliest ground breaking evidence showing one viral protein is capable of manipulating important biochemical pathways, such as SUMOylation, in order to reproduce. While the disappearance of E1 proteins in SUMOylation caused by Gam1 has been biochemically characterized, the exact molecular details by which this occurs is still undetermined. The ultimate goal of this research is to obtain a high-resolution crystal structure of Gam1. Determining the three-dimensional structure of Gam1 is crucial in understanding its functions at a molecular level, particularly the mechanism by which this gene regulates cellular SUMOylation enzymatic activity. In order to obtain a crystal structure, a large amount of highly homogenous and functional Gam1 is critical, which is the main focus in this chapter.

1.2 LONG TERM GOALS AND SPECIFIC AIMS

As mentioned in the introduction, Gam1 was the first viral protein discovered to globally inhibit host SUMOylation via the elimination of SAE1/SAE2 and E2. However, the exact molecular mechanism by which Gam1 alters SUMOylation is still unknown. The **long-term goal** of this research is to determine the structure of Gam1 and understand its function within viral infection. Structural elucidation of Gam1 will provide a better understanding of its function at the molecular level. Within this dissertation, recombinant Gam1 has been cloned, expressed, and purified from bacterial expression system. Furthermore, functional analysis has been carried out to verify the interaction between Gam1 and its cellular targets. These results will establish a foundation for the future structural determination of Gam1 with or without the binding components. Two specific aims have been proposed in this research to test the *main hypothesis that Gam1's function serves as a bridge from SUMOylation to ubiquitination*.

Specific Aim One: To perform *in vitro* analysis of the interaction between recombinant Gam1 and SUMOylation pathway proteins. SUMOylation assays were utilized to analyze the association of purified recombinant Gam1 and SAE1 and determine if SUMOylation is inhibited. Results obtained from these experiments will show that recombinant Gam1 is not only properly folded but also has the ability to bind SUMOylation proteins. This will clearly demonstrate one aspect of Gam1's functionality, which is the capability to interact with SUMOylation SAE1 directly. Compared to previous *in vivo* experiments indicate SAE1 becomes targeted into Ubiquitin E3 ligase complexes through Gam1, here we aim to show the in vitro direct interaction of SAE1 with Gam1, which will be coupled with future structural analysis of the key region of the interface.

Specific Aim Two: To conduct *in vitro* studies of the interaction between Ubiquitination system protein interactions and recombinant Gam1 to test its functionality. Similar biochemical techniques utilized in Specific Aim One will aid in the analysis of interactions between recombinant Gam1 and EloC, a Ubiquitin E3 ligase. BC-Box, a sequence required for the interaction with EloC binding was found in Gam1 at its hydrophobic C-terminus. Results from this aim will provide direct evidence whether Gam1's BC-Box is functional. Gam1's direct ability to interact with EloC will demonstrate the key idea that Gam1 can navigate through both SUMOylation and ubiquitination. Future experiments will include co-incubation with both SAE1 and EloC.

The Results obtained from both Aims will pave the way for future structural studies that will provide essential information regarding the details of Gam1's function as a bridge between these abovementioned PTMs interactions. Information from the structure together with the biochemical characterization will shed light on Gam1's function in the viral replication cycle.

1.3 MATERIALS AND METHODS

1.3.1 MOLECULAR CLONING OF GAM1 INTO AN EXPRESSION VECTOR(S)

Bacterial expression remains as one of the most popular methods to produce recombinant protein; given its low cost, high yield, and relative ease of manipulation. One of the caveats however, is the proper selection of a host and selection of a suitable promoter [37, 38]. All of the constructs selected were for bacterial expression with different bacterial hosts and multiple vectors. The full-length gene sequence was: amplified via PCR, restriction endonuclease digested, and ligated into an expression vector(s) and transformed into competent bacterial cells. During the PCR primer design, expression tags were added to either the N-terminal or C-terminus of Gam1 gene. These tags are for affinity purification, enhancing solubility, assisted folding, and functional studies. Antibiotics were utilized to select transformed cells. Ligated construct within the transformed cell was purified and verified by restriction enzyme digestions and DNA sequencing.

1.3.1.1 POLYMERASE CHAIN REACTION AND CLONING

Polymerase chain reaction (PCR) is a widely used tool for molecular biology to amplify specific DNA fragments using two unique primers specific for the targeted DNA (template) [39, 40]. Pairs of primers were designed to target the 5' and 3' regions of the DNA sequence to be cloned. PCR primers (table 1.1) with additional restriction enzyme sequences at the end were designed to amplify the gene of Gam1 from a plasmid construct pSG9m (courtesy of Dr. Susana Chiocca) [41]. The final optimized PCR reaction conditions utilized for each construct are listed in Table 1.1

Final PCR products were dual digested with respective enzymes (see table 1.1). The digested products were gel harvested and ligated into the vector(s) following Invitrogen's T4 DNA Ligase protocol [42, 43]. Competent cells of *E. coli* strains of BL21(DE3) or Rosetta 2(DE3)pLysS were

produced chemically following a lab protocol from Cold Spring Harbor [44]. The ligated products were then transformed into either BL21(DE3) or Rosetta 2(DE3)pLysS competent cells following the traditional 42°C heat shock method [44]. Ligated constructs were selected on ampicillin LB plates. Individual bacterial colonies were picked, a glycerol stock was made, and subsequently stored in the -80°C. Purified plasmids from the glycerol stocks were verified by either PCR or restriction enzyme digestion by the cloning enzymes. The constructs were also further verified by DNA sequencing (UTEP Biology Core Facility). In these constructs affinity tags such as N terminal 6X, 10X His tag or C-terminal 6X His tags were incorporated during PCR, or acquired from the vector plasmids. Furthermore, some of the vectors contain fusion tags, which facilitate protein folding, hence, increase solubility and enhance further purification (described in detail later). Also, some cloning of Gam1 was performed utilizing Invitrogen's Gateway ® Directional Cloning system, which is based on DNA recombination to directionally clone blunt end PCR products without the usage of restriction enzymes [45, 46].

1.3.1.2 Molecular Cloning in Multiple Vectors

Five expression vector systems were selected with different tags for affinity chromatography purification, as well as increasing the successful folding of the recombinant protein (See table 1.1). The first one was a novel Gateway® cloning system from Invitrogen. In this system, the PCR product of Gam1 is directionally recombined into an *entry* vector (pENTR/SD/TOPO), utilizing topoisomerase I from vaccinia virus [47] without restriction enzyme digestion and ligation (refer to manufacturers protocol). Correct entry colonies are screened via PCR and purified plasmid was used in the next procedure. In the Gateway® system, Gam1 gene can be transferred into various *destination* vector systems taking advantage of the site-specific recombination properties from bacteriophage lambda [48].

The bacterial expression *destination* vector (pET-DEST42) was chosen because of its C-terminal 6X His tag that can be used for purification purposes. The recombinant vector was subsequently transformed into competent cells of *E. Coli* BL21(DE3) strain.

Concurrently, Gam1 was also cloned into another vector system pET-49b(+) (Novagen) containing a Glutathione S-Transferase (GST) tag at the N-terminus, for purification and enhanced solubility purposes (see table 1.1). GST is an enzyme found in both eukayotes and prokaryotes that catalyzes the transfer of reduced glutathione (GSH) to a substrate [49]. Both the vector and the PCR product of Gam1 were digested with restriction enzymes SpeI and XhoI and ligated by T4 DNA ligase [© Life technologies] to produce an expression construct. The ligated construct was transformed into *E. Coli* BL21(DE3) strain competent cells, individual colonies were verified by restriction enzyme digestions and DNA sequencing. In Parallel, a C-terminal His tagged construct was generated, using the pSG9m-Gam1 plasmid as the template to conduct PCR with the primers in table 1.1. The vector selected was pET-30b(+) which has a 6X His tag at the C-terminus.

In order to optimize protein production using common codons in bacterial systems, the sequence of Gam1 was codon optimized and synthesized de novo by Genscript ®. In addition, another vector named pCold-Trigger Factor (© Takara, courtesy of Dr. Jianjun Sun, Department of Biological Sciences, UTEP) was utilized to express the codon optimized Gam1. (TF) is a prokaryotic ribosome associated [50, 51] with a cold shock chaperone of 48kDa that helps in polypeptide folding at low temperatures to reduce the chances of forming mis-folded and insoluble proteins [52, 53]. As a native product from prokaryotes, TF is highly expressed in *E.coli* using the strong *cspA* promoter. In addition, the pCold-TF vector contains: translation enhancing elements, an N terminal His-Tag before TF, and multiple cleavage sites to separate from the target protein. Both pCold-TF vector and Gam1 were digested with restriction enzymes NdeI and HindIII and ligated by T4 DNA ligase (© Life technologies) to produce a construct (Table 1.1). The clones were screened and confirmed by restriction enzyme

digestions and DNA sequencing. Gam1 was also cloned into pCold1 (© Takara, courtesy of Dr. Jianjun Sun, Department of Biological Sciences, UTEP), which maintains the same elements as N-pCold-TF, however lacking the TF at the N-Terminus of the recombinant protein. Digestions and ligation reactions were performed following the methods listed above.

In addition, Gam1 was also cloned into a modified pCold-TF vector (referred to as C-pCold-TF). In this modified vector, one of the three cleavage sites as well as the N-terminal 6X His tag were deleted (performed by Dr. Jianjun Sun, Department of Biological Sciences, UTEP). Primers were designed with an additional C-terminal 6X His tag after the recombinant Gam1 gene, for purification purposes. This construct with a C-terminal 6X His tag was designed to ensure full-length expression of recombinant Gam1. C-pCold-TF and PCR product of Gam1 were digested with NdeI and EcoRI and ligated by T4 DNA ligase (© Life technologies) to produce a construct (Table 1.1).

Table 1.1 Primers and PCR Settings For Gam1 Amplification

Clone	Initial Denaturation Time (s) @ Temp	Denaturation Time (s) @ Temp	Annealing Time (s) @ Temp	Extension Time (s) @ Temp	Final Extension Time (s) @ Temp
pET30b+-Gam1-His	Primers: Sense: 5’ CGCCCATATGGCCCGCAACCCATTCGCATGTTCCCT -3’ NdeI Antisense 1: 5’- CGCGCTCGAGATGATGATGATGCAGAGAATGGTAGGGGTG -3’ XhoI				
	94 C @ 60 s	94 C @ 10 s	62 C @ 20 s	65 C @ 120 s	65 C @ 600 s

<table>
<tr><td>pDEST42-

Gam1-His</td><td colspan="5">Primers:
Sense:
5’- CACCATGGCCCGCAACCCATTCCGCATGTTCCCTG -3’
CACC overhang for Invitrogen Gateway ® cloning
Antisense:
5’- AGCTGCGGAGCGACCTTCAATCAGAGAATGGTAGGGGTGGTGC -3’</td></tr>
<tr><td></td><td>94 C @ 60 s</td><td>94 C @ 10 s</td><td>68 C for 20 s</td><td>65 C @ 120 s</td><td>65 C @ 600 s</td></tr>
<tr><td>pDEST42-His-Gam1

(the PCR was done in three steps)</td><td colspan="5">Primers:
Sense #1
5’- GCGGCTATTGAAGGTCGCGCCCGCAACCCATTCCGC -3’
Sense #2
5’- CACCACCACCACCACCACCACCACTCCGCGGCTATTGAAGGTCGC- 3’
Sense #3
5’- CACCATGGCAGCAGCACACCACCACCACCACCACCAC -3’
CACC overhang for Invitrogen Gateway ® cloning
Antisense:
5’- TTACAGAGAATGGTAGGGGTGGTGCCGAGACCCGAGTCC -3’
Stop codon</td></tr>
<tr><td></td><td>94 C @ 30 s</td><td>94 C @ 10 s</td><td>68 C @ 20 s</td><td>65 C @ 51 s</td><td>65 C @ 600 s</td></tr>
<tr><td>pET49b+-GST-Gam1

(the PCR was done in two steps)</td><td colspan="5">Primers:
Sense #1
5’- GGTAGTATTGAAGGTCGCGCCCGCAACCCATTCCGC -3’
Sense #2
5’- GCGCACTAGTGGTGGTGGTAGTATTGAAGGTCGCGCC -3’
SpeI</td></tr>
</table>

	Antisense: 5’- GCGCCTCGAGTTACAGAGAATGGTAGGGGTGGTG -3’ XhoI				
	94 C @ 30 s	94 C @ 10 s	68 C @ 20 s	65 C @ 51 s	65 C @ 600 s
pCold-His-TF-Gam1	*De novo* synthesized construct (Genscript) in pUC57 digested with 5’-NdeI and 3’ HindIII				
pColdI-Gam1	*De novo* synthesized construct (Genscript) in pUC57 digested with 5’-NdeI and 3’ HindIII				
pCold-TF-Gam1-His	Primers: Sense 5'- GCGCTACATATGGCTCGTAATCCGTTTCGTATGTTTCCGGGCGACCTGCCG TACTACATGG -3' NdeI Antisense 5' GGTCGAATTCTTAATGATGATGATGATGATGGCGACCTTCAATCAGCGAAT GATACGGATGAT 3' EcoRI				
	98 C @ 30 s	98 C @ 10 s	66.1 C @ 15 s	72 C @ 5 s	72 C @ 600 s

1.3.2 EXPRESS TARGET PROTEIN:

E.coli expression systems were utilized to produce large quantities of Gam1 due to its low cost, fast growth, and easy manipulation. BL21(DE3) and Rosetta 2(DE3)pLysS cells (both from Novagen) are modified *E.coli* bacterial strains that have T7 promoter region, which can produce more proteins than normal bacteria. Rosetta 2(DE3)pLysS strain is derived from BL21(DE3). Both can effectively express protein due to the absence of two genes, *lon* and *ompT,* that code for two key proteases, which degrades misfolded protein and removes extracellular proteins, respectively [54]. Reduced protease degradation and tight regulation of expression make these strains optimal for *E. Coli* production. Additionally, Rosetta 2(DE3)pLysS contain tRNA's for seven rare bacterial codons, which will facilitate the expression of eukaryotic proteins [55, 56]. Moreover, the addition of T7 lysozyme in this strain helps suppress basal expression levels of the T7 RNA Polymerase before the induction by IPTG, preventing leaky expression of target protein.

For expression trials, a few of the clones were grown in media maintaining the same conditions to determine the most suitable clone (demonstrating the best expression) for large-scale purification. Once the best clone was selected, further small-scale expression trials have been carried out to optimize conditions relating to protein expression: such as bacterial growth, temperature for induction, and amount of time for post-induction expression. After the optimal conditions have been established, large-scale expression were conducted to purify the target protein to high homogeneity utilizing various chromatographic techniques.

1.3.2.1 PROTEIN EXPRESSION FOLLOWING TRADITIONAL BACTERIAL GROWTH

The constructed plasmids control the expression of nascent protein under the induction with IPTG. Multiple optimization trials were performed including differences in media composition, temperature, optical density (OD) before induction, IPTG concentration, and time post induction to establish the optimal conditions for protein expression. An auto-induction system (MagicMedia™ Medium, see protocol © Life technologies) was also tried. Constructs including GST-Gam1, pET-30b(+), and pET-DEST42 were expressed under the optimized condition described below. Transformants of GST-Gam1 in pET-49(b)+ (constructed in house) were grown in 100 μ g/mL ampicillin LB liquid media at 37°C agitating at 220 rpm until an OD_{600} of 0.8 is reached. Then the culture was immediately induced with 1 mM IPTG and incubated four hours at 30°C with 200 rpm agitation. The cells are harvested by centrifugation at a speed of 5000xg and the pellets are re-suspended in lysis buffer (50mM HEPES pH 7.5, 10mM $MgCl_2$, 150mM NaCl, 0.02% NaN_3 and 1mM phenylmethylsulphonyl fluoride PMSF). The re-suspended pellets were then subjected to freeze-thaw three times to lyse the cell. Lysozyme was added to final concentration of 0.5 mg/ml to assist in cell lysis. Lysozyme is an enzyme that digests the peptidoglycan of the rigid cell wall of E. coli, leaking the cellular contents including the expressed target protein. The cell lyses are then centrifuged at a high speed of 35,000xg at 4°C for 40 minutes to remove cell debris, large cellular organisms and un-lysed cells. To verify protein production, SDS polyacrylamide gel electrophoresis (SDS-PAGE) was used.

1.3.2.2 COLD INDUCTION PROTEIN EXPRESSION

Conditions for protein expression were optimized with the C-pCold-TF, which were subsequently utilized for all future expression protocols. In contrast to expression protocols mentioned above, the pCold-TF cultures were grown to an OD_{600} of 0.6, at which point agitation was stopped and temperature was dropped to 16°C for 1 hour. Culture was then induced with 0.1 mM IPTG and

maintained at 16°C for 14 hours. Cells are harvested by centrifugation at 5250xg for 40 minutes at 4°C and the pellets are re-suspended in lysis buffer (50mM Tris-Cl pH 7.5, 300mM NaCl, 0.02% NaN_3, 1mM PMSF, and 10mM imidazole) and is subsequently treated with freeze-thaw process with lysozyme (0.2 mg/mL) and harvested as described above. Centrifugation at a high speed of 35,000xg at 4°C for 40 minutes (repeated twice) removes cell debris, large cellular organisms, and un-lysed cells. Cleared lysate is then used for protein purification.

1.3.3 PROTEIN PURIFICATION TRIALS

Purification of recombinant protein is essential for usage in biochemical characterization. Traditional methods of purification are largely based on the downstream usage of the target protein, which include: affinity chromatography respective to the expression clones affinity tag, followed by refinement chromatographic techniques such as ion exchange, and size-exclusion to reach high homogeneity. In affinity chromatography, the properties of the target recombinant protein are well defined (such as an additional 6X-His tag) that can be exploited in the purification process. This will result in separation from other non-specific proteins yielding a relatively pure protein. Immobilized metal ion affinity chromatography (IMAC) is based on the affinity of the amino acid Histidine for bivalent metal ions, such as cobalt, copper and nickel[57]. GST tagged proteins follow the same premise of affinity purification by both increasing the solubility of the target protein as well as being captured via immobilized glutathione while impurities remain unbound [58]. Refinement techniques such as ion exchange chromatographies take advantage of the difference in protein charge at different pH. This is based on the electrostatic attraction and repulsion between the immobilized charged molecules and the charge on the target protein [59]. There are two types of exchange columns (cation and anion) that can be used based on the charge on the protein. Negatively charged proteins interact

with positively charged ions in anion exchange chromatography (AEC). The reverse is true for positively charged proteins.

Gel filtration, also referred to as size exclusion chromatography is based on the molecular weight of the protein. The matrix of the column contains tiny pores, which freely interact with smaller weight molecules, whereas larger molecules do not enter the pores and are thus excluded in decreasing molecular size [60]. Therefore, samples applied to the SEC will elute larger proteins first and retain smaller proteins within the matrix longer.

Since the constructed clone(s) contain multiple affinity or solubility factors, such as 6X Histidine and GST, Nickel-NTA affinity chromatography as well as purification via the GST fusion Tag was used to purify recombinant protein. Purifications were performed using either a HisTrap HP or GSTrap FF (both from GE) to identify the binding of the recombinant protein and purify from non-specific contaminating proteins. Refinement chromatographic techniques included: (HiPrep Sephacryl s200, HiLoad 26/600 Superdex 200) as size exclusion, HiTrap Heparin HP as a cation exchanger, and HiTrap Q HP anion exchanger.

GST-Gam1 constructs were initially purified using following manufacturer specifications. Lysate was prepared as mentioned in the previous section, and the supernatant was passed through a 0.45 μ m filter (Millipore) and then loaded onto a 5 mL GSTrap FF (GE Life Sciences), equilibrated with buffer A (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4) adjusted to pH 7.4. The column was washed with 50 mL buffer A to remove nonspecifically bound proteins. Target protein was then eluted with buffer B (140 mM NaCl, 2.7 mM KCl, 10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 10 mM Glutathione, and 1 mM EDTA) pH 8.0. Given that the binding efficiency to the column is flow rate sensitive, slow flow rate trials at a sample application of 0.1 mL/min were attempted Figure 1.3.2.

His-Hitrap HP column was purchased from GE healthcare Life Science. HisTrap HP columns contain cross-linked sephorase beads with a chelating group attached to it. Histidine has an affinity towards Ni^{2+} and 6X His will have increased affinity for Ni^{2+}. It has been found that Histidine tagged proteins bind strongly with Ni^{2+} thus allowing for affinity chromatography.

Purification of constructs containing a 6 or 10X His tag were conducted on Nickel columns. Small scale Gravity flow affinity chromatography was used to optimize the purification conditions. 5 mL Ni-NTA agarose beads (Qiagen) suspended in a gravity column was equilibrated with 25 mL of binding buffer (50 mM Tris pH 7.5, 300 mM NaCl, 10 mM Imidazole, 0.02% NaN_3, and 1mM DTT or 50mM HEPES pH 7.5, 100mM NaCl, 10 mM $MgCl_2$, 20 mM Imidazole, 1 mM PMSF, and 0.02% NaN_3). 50 mL cell lysates were filtered by 0.45µm filter and loaded onto the column via gravity flow and any unbound sample was washed out with the same buffer, while allowing only His tagged protein to bind to the column. The flow through was collected in a separate sample tube. Elution's were performed by utilizing step gradient from 20-250 mM imidazole obtained by mixing the binding buffer above with different percentage from a stock elution buffer (50 mM Tris pH 7.5, 150 mM NaCl, 1M imidazole, 0.02% NaN_3 and 1mM DTT, or 50mM HEPES pH 7.5, 100mM NaCl, 10 mM $MgCl_2$, 1 M Imidazole, 1 mM PMSF, and 0.02% NaN_3). Optimal washing (125mM imidazole) and eluting (210 mM imidazole) conditions were established and used in chromatographic purification using AKTA FPLC (Amersham) and 5mL His-Hitrap column (GE). 90 to 100 mL cell lysate was filtered and loaded to the column, followed by washing with 30 mL 125 mM imidazole buffer and eluting at 15 mL 210 mM imidazole elution buffer. Further chromatographic techniques such as Ion exchange, and Size exclusion were utilized for all constructs. Salt precipitation with ammonium sulfate was also tried to remove contaminants from the target protein. The reasoning was that the ionic strength of the target protein would be different than that of the contaminants, which could be later used for other chromatographic

techniques Figure 1.3.4. Affinity purified Gam1 was subjected to both ion exchange and size-exclusion chromatography for further purification.

The eluent from affinity purifications were loaded onto a SEC (HiPrep Sephacryl S200) equilibrated in (50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, and 0.02% NaN_3) (Figure 1.3.5). All other SEC experiments were on (HiLoad Superdex 26/60 200) equilibrated with (25mM Tris pH 8.0, 150mM NaCl, 0.02% NaN_3). Gam1-GST constructs as well as Gam1-pET vectors were run on cation exchange with HiTrap SP FF was equilibrated with Buffer A (50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 150 mM NaCl, and 0.02% NaN_3). Protein was eluted with a linear gradient with Buffer B (50 mM HEPES pH 7.5, 10 mM $MgCl_2$, 1 M NaCl, and 0.02% NaN_3) HiTrap Heparin HP equilibrated in Binding buffer (10 mM Na_2HPO_4, and 0.02% NaN_3). Elutions were performed with a linear gradient using elution buffer (10 mM Na_2HPO_4, and 0.02% NaN_3, 1M NaCl).

The eluent from HiLoad Superdex 26/60 200 was loaded onto IEX (HiTrap Q HP, GE) pre-equilibrated in Buffer A (50 mM Tris pH 8.0, 100 mM NaCl, 0.02% NaN_3), and eluted with Buffer B (50 mM Tris pH 8.0, 1 M NaCl, 0.02% NaN_3, and 1mM DTT) using a linear gradient.

All samples were analyzed for purity on an SDS-PAGE. Samples were prepared with 4X SDS PAGE dye then heated at 100°C for 5 minutes and run on a 10% SDS PAGE stained by Coomasie Blue. Silver staining technique was also utilized to determine the purity of MCP. To determine Protein concentration a standard protein assay kit using BSA (©Thermo scientific) as a reference was utilized.

1.3.4 BIOCHEMICAL CHARACTERIZATION

1.3.4.1 DLS IN GAM1 CHARACTERIZATION

Dynamic light scattering (DLS) is a widely used technique to determine the size, shape and distribution of particles. DLS utilizes a monochromatic laser beam whose wavelength changes when it

passes through a solution with particles in random motion (Brownian motion). This change of wavelength is correlated to the largest dimension size of the particle. When a particle is rotating very fast in a solution, the shape of the particle can be assumed to be spherical with the longest dimension of the molecule considered the diameter. Therefore, DLS can provide an estimate of the protein size. For instance, DLS can be used to characterize protein interactions with a visible shift in the size distribution. This is very useful when investigating multimeric proteins as well as other macromolecular interactions. If the protein is temperature sensitive, or has an increased propensity towards aggregation, DLS may be utilized to analyze the size difference under various conditions. Therefore, DLS can be also used to monitor particle size distribution and species, such as aggregation formation, in real time. In addition, the newer models of DLS instruments can also measure surface charge potential of the protein. This surface charge measurement is very useful for refinement purification procedures, providing insight as to what type of ion exchange chromatography techniques to further purify your target protein. DLS is a really useful technique that does not waste sample, measures in real-time, and does not require large concentrations of sample.

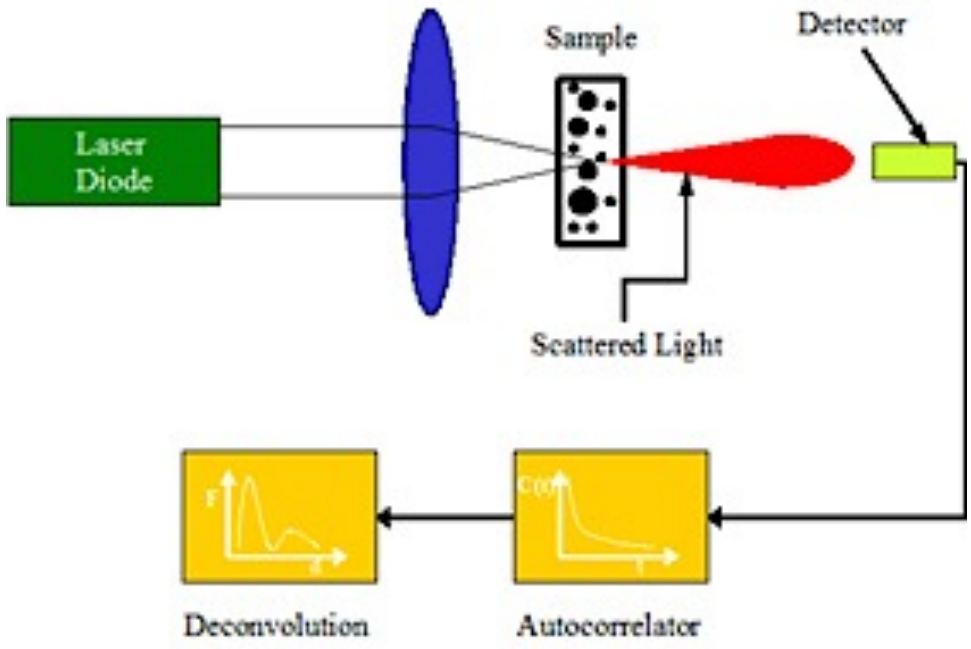

Fig 1. 1 Schematic representations of how scattered waves are correlated to particle distribution using dynamic light scattering detection. Adapted from © Copyright Particle Sizing Systems

DLS measurements of Gam1 in N-pCold-TF and C-pCold-TF were collected. The size distribution helped to distinguish between aggregates and potentially native Gam1. The zeta potential as determined by DLS using a Zetasizer nano ZS (©Malvern) instrument revealed that Gam1 has a slightly negative surface charge.

3 mg/ml Gam1 concentration was used for DLS analysis. 10µl of sample was diluted with 990µl of buffer composed of 25mM Tris pH 8.0, 150mM NaCl and 0.02% NaN_3. Manual settings were set up to collect from 12-20 measurements every 10 seconds. The laser light source utilizes a wavelength at 633.

The same instrument was used to measure the surface charge of Gam1 in the aforementioned constructs. A cuvette containing electrodes at both ends was used to collect data. The premise is based on the charge of the sample, which become attracted to the electrodes with the opposite charge. The zeta potential is a measure of the velocity at which the particle moves (electrophoretic mobility).

1.3.4.2 *IN VITRO* SUMOYLATION ASSAYS IN GAM1 CHARACTERIZATION

Gam1's ability to globally inhibit the SUMOylation pathway has been established using in vitro assays, therefore highly pure Gam1 ~90% at a concentration of 0.5 mg/mL was sent to Dr. German Rosas-Acosta for SUMOylation inhibition assays to verify the previously described disappearance of E1 and E2 from the SUMOylation pathway. *In vitro* SUMOylation reactions were carried out following an established protocol [61], with minor modifications. Briefly, 2 µL of ^{35}S-methionine labeled influenza virus NS1, a SUMOylation target [62-64], produced by a coupled transcription/translation system (Promega Corp., Madison,WI) was incubated with various amounts of recombinant pCold-TF-Gam1, 200 ng GST-SAE2/SAE1, 200 ng GST-UBC9, and 3.4 µg of His-SUMO1 in a final volume of 26 µL solution containing 50 mM Tris pH 7.4, 2.5 mM DTT, 2 mM ATP. All incubations were performed at 37°C for 90 min. The positive control contains all the components listed above without any pCold-Gam1. As negative control, a mock protein expressed in the same pCold-TF vector system was used to show the TF tag does not affect the *in vitro* SUMOylation. Another control of 100% SUMOylation abolishment was obtained by adding an affinity-purified catalytic domain of the yeast SUMO protease Ulp1 (Ulp1403–621) [21] 60 min after the initial reaction began. All reactions were quenched by the addition of 4× SDS-PAGE sample buffer (50 mM Tris pH 6.8, 10% glycerol, 4% SDS, 0.01% bromophenol blue, 2% B-mercaptoethanol). Samples were then boiled for 4 minutes, run on SDS–PAGE, and transferred to Immobilon-FL (0.45 m) membranes (Millipore Corp., Bedford, MA). Membranes were developed by autoradiography using a Personal Molecular Imager™ and Quantity One™ software (both from Bio-Rad Laboratories) or by immunoblotting using IRDye-conjugated secondary antibodies and infrared fluorescence imaging using an Odyssey CLx imaging system (both from Li-Cor Biosciences, Lincoln, NE). Statistical analyses were performed using GraphPad Prism, utilizing unpaired single tailed t-tests with Welch correction.

1.4 RESULTS AND DISCUSSIONS

1.4.1 CLONING OF GAM1

Initial Gam1-GST construct was obtained from Dr. Susana Chiocca, our international collaborator. All pET vector constructs were generated from the plasmid provided by Dr. Chiocca. Amplification of the Gam1 gene for cloning into the aforementioned constructs was performed with a Long Amp PCR kit (© New England Biolabs) (Fig 1.2). Cloning was by either conventional restriction endonuclease reactions or by using the Invitrogen Gateway (©Life technologies) cloning system. The selected Gateway cloning system utilized directional cloning as a method to perform recombination reactions known as LR recombination. This entry vector (pENTR/SD-TOPO) carries the gene, which is then swapped into an expression vector suitable for bacterial expression. Re-cloning into pET-49b(+) again with the aforementioned kit was unsuccessful after multiple trials. Furthermore, DNA sequencing revealed amino acid mutations that would potentially alter Gam1's function. Conditions used for pET-30b(+) cloning resulted in mutations within the gene of interest; therefore, further expression experiments were not pursued. No suitable constructs were produced (Fig 1.3), therefore multiple efforts to troubleshoot the cloning was pursued. After many trials, a handful of clones were produced after following a ligation reaction, which was immediately tested for expression (Fig 1.4). Due to these issues, it was decided to de novo synthesize a bacterial expression codon optimized Gam1 from Genscript. One optimized for cloning into N-pCold-TF and another with an additional 6X His tag at the C-terminus of the Gene for cloning into C-pCold-TF. These clones were sent in pUC57, cloned with NdeI and HindIII for N-pCold-TF; and NdeI and EcoRI for C-pCold-TF. Restriction enzyme reactions were performed from the Genscript clone and ligated into the above-mentioned N-pCold-TF vector. Primers listed in table 1.1 were utilized to amplify Gam1 gene and an additional 6X His tag to generate C-pCold-TF. The expression vectors and Gam1 were successfully cloned using the conditions mentioned above. Reactions performed in the cloning process the digested insertion and vectors

required purification. This was carried out via agarose gel electrophoresis. All products were ligated by T4 DNA ligase. Constructs were then transformed into variants of BL21 (DE3) (BL21 codon+, or Rosetta 2(DE3) pLysS). The plasmids of individual bacterial colonies were purified (Qiagen miniprep kit), and subsequently digested with the restriction enzymes utilized for each respective construct (pET49 and C-pCold-TF). C-pCold-TF with bands at 5.7kb (vector) and ~900 bp (Gam1) demonstrate the successful cloning of Gam1 (Fig 1.5). DNA sequencing performed on pET-DEST42 demonstrated an amino acid mutation at the C-terminus of Gam1 (Data not shown). pCold-TF constructs were found to have no mutations, thus cloning was successful.

Amplified PCR products or digested from pUC57 from above explained PCR and digestion experiments Table 1.1 and both pColdTF vectors were cut with the abovementioned pair of restriction enzymes and was cloned into pColdTF by T4 DNA ligase. Transformed product from the reaction was plated on selectivity plates (Fig 1.6) and the resulting plasmid was purified from BL21 (DE3) cells. To verify the successful insertion of recombinant Gam1 into the vectors, restriction enzyme digestions were performed. Results obtained demonstrated two distinct DNA fragments at the expected sizes of 5.7Kbp and ~900 bp Fig 1.5 when run in an agarose gel. Furthermore, the recombinant DNA was verified by DNA sequencing.

1.4.2 PROTEIN EXPRESSION

The construct provided by Dr. Chiocca (pSG9m) was used for cloning purposes into three pET vector systems mentioned in 1.3.1.2. An uninduced control was utilized to compare with the induced protein, and both were run on the gel (Fig 1.4). A protein band above the 50 kDa marker was apparent in the induced sample, yet not visible in the uninduced sample. Small-scale expression trials for pET 30b(+) as mentioned in section 1.2.1.2 was performed and low detection and non-specific protein was

produced (Fig 1.7-.8). Large-scale expression attempts included the usage of auto-induction low temperatures, differences in expression time and the target protein was determined to be insoluble and was being removed in the centrifugation step as an inclusion body. In an attempt to improve protein production, these constructs were tested in Rosetta 2 pLysS(DE3) [65] (EMD Millipore) that contains plasmids encoding tRNAs for rarely used bacterial codons [56]. However, no improvement of Gam1 expression was observed. Furthermore, cloning into pET DEST 42 produced mutations within the gene leading to the abandonment of these two constructs. No further experiments were attempted (results not shown).

Difference in pCold the N-pCold-TF constructs were grown at 37°C 220 rpm to an OD_{600} of 0.8 in LB medium containing 100 μ g/mL ampicilin, culture was then induced with 1 mM IPTG and allowed to cool for 30 minutes at room temperature. After 30 minutes the culture was refrigerated at 16°C for an additional 18 hours 200 rpm. After 18 hours of incubation at 16°C. Individual sequence verified colonies were grown and induced with IPTG (1mM) to verify the expression of target N-pCold-TF (Fig 1.9). As shown in Fig 1.9, in comparison to the control E. Coli BL21 (DE3) cells, N-pCold-TF+Gam1 generated a protein band above the 75 kDa molecular weight marker. This size (79kDa) corresponds to the TF-Gam1 fusion construct. In order to optimize protein production, different induction conditions were tested (Fig 1.10). Protein in excess was produced in both conditions; therefore, the lower IPTG condition was selected for further large-scale expression. The successful expression of recombinant Gam1 can be attributed to the codon optimization, coupled with the usage of TF in assisted protein folding. This compared to previous expression constructs mentioned above yielded viable protein that was clearly not seen in control *E. Coli*. After successful small-scale protein expression trials, large-scale protein expression with optimized induction conditions was performed. Although pCold-His-TF-Gam1 construct yielded abundant protein (Fig. 1.9-10), significant amounts of lower molecular weight TF-Gam1 protein were observed after several steps of chromatographic purification. These lower molecular

weight TF-Gam1 can be either incomplete translation products of TF-Gam1 or degraded TF-Gam1 from the C-terminus (Fig 1.11). Therefore, in order to purify full length Gam1 from affinity chromatography, the 6xHis tag at the N-terminus was removed and added to the C-terminus of the construct. The same conditions mentioned above were carried out for the expression of C-pCold-TF and pCold1-Gam1. You can see in (Fig 1.12) that recombinant Gam1 is being expressed, albeit is seen in the inclusion fraction. In the case of C-pCold-TF this was due to the incomplete extraction from the whole cell lysate, which was corrected with the addition of Lysozyme. In the Case of pCold1-Gam1, this was not improved suggesting an insoluble recombinant Gam1 protein (Fig 1.13).

1.4.3 PURIFICATION

Initial purification trials were carried out using GST-Gam1 as detailed in 1.3.3, Gam1 inducible expression was achieved (Fig 1.4) albeit protein would always retain at least two other protein contaminants (Fig 1.14-17). Multiple efforts utilizing various chromatographic techniques were carried-out. Three major contaminants were repeatedly produced in protein expression and could not be removed from Target GST-Gam1 construct. This could be a result of endogenous proteins expressed in bacteria, which form an interaction with the fusion protein. During expression trials, one of the major contaminants was GST alone, which was only removed by size exclusion chromatography. The purest protein obtained was via size exclusion (Fig 1.17), albeit protein amount was insignificant and still had a contaminant immediately below the target. Therefore re-cloning into a different expression vector was pursued.

The presence of 6X-His tag at the N or C-terminus of the recombinant Gam1 facilitates the possibility of obtaining relatively pure protein via Nickel NTA column. Given that Nickel is one of the most widely utilized bivalent metals for purifying His tagged proteins [57, 66, 67]. Nickel has been demonstrated to have a high affinity towards histidine residues, however, it has a strong tendency to

bind non-specific proteins [68] if the target protein is low in expression or has the affinity tag hidden. Earlier pET constructs exhibited non-specific protein interactions when purified solely using Ni affinity, and attempts to segregate the non-specific proteins prior to affinity using ammonium sulfate precipitation (Fig 1.18) also proved useless. In order to remove this possibility of nonspecific interactions, 10mM Imidazole was added to the binding buffer. Trial purifications with gravity flow affinity chromatography (Ni-NTA beads) utilizing step gradient elutions to determine the optimal wash and elution of target protein was carried-out. SDS-PAGE analysis demonstrated that recombinant TF-Gam1 represented approximately 40% of soluble proteins in the crude sample before purification (Fig. 1.19). After washing with 40mM, 60mM, 80mM, and 120mM, imidazole concentrations and eluting with 200mM Imidazole, greater than 70% purity Gam1 was obtained (Fig 1.20). Approximately 4 mg/L was generated of target protein for downstream chromatography. From the gravity flow experiments conditions were utilized to perform large-scale purification trials using HisTrap HP columns as a two-step gradient elution. Optimization dependent on the amount of imidazole to utilize for washing out contaminating bands was varied from 100-200 mM and elutions from 200-400 mM. The ideal conditions were determined to be with washes 125 mM and elution at 210-220 mM imidazole respectively. Lysate loading capacity was determined to be 0.6L for pCold-TF (With N-terminal His) N-pCold-TF and 2.5L for pCold-TF (with C-terminal His) C-pCold-TF. Later experiments with pCold-TF utilized Buffer A (Binding Buffer) was 50 mM Tris pH 7.5, 300 mM NaCl, 10 mM Imidazole, and 0.02% NaN_3. Buffer B (Elution Buffer) was. C-pCold-TF was optimized on the HisTrap HP column with ideal washing conditions at 190mM, and elution at 400 mM (Fig 1.21). This was later determined to be too high a concentration; therefore, washing conditions were reduced to 120 mM and elutions at 200 mM Imidazole, respectively.

Most of the contaminants were removed with 125 mM, with protein purity after one step estimated at about 70%. Further purification using size exclusion chromatography (SEC) yields a

protein with a purity exceeding 80% as determined by Coomassie blue stained SDS-PAGE (Fig 1.22-23), effectively removing the smaller contaminants and the majority of the 52 kDa band corresponding to trigger factor alone. BSA standards were utilized to calibrate the elution volumes correlated to the size of the protein. SEC also provides a profile of different Gam1 forms in the sample: aggregated, dimeric and monomeric (Fig. 1.23). It was determined that monomeric TF-Gam1 would elute at ~185 mL. Three distinct peaks are observed in the SEC profiles (Fig 1.23) where in the cold temperature optimized purification a larger peak is observed at the monomeric size of TF-Gam1, as compared to the room temperature pH'ed buffers. The fraction corresponding to the monomeric TF-Gam1 was collected and utilized for biochemical characterization. To determine the purity of TF-Gam1 after SEC, the samples were run in a SDS PAGE (Fig 1.22 & 1.23), demonstrating protein purity above 85%. Gam1 was further purified utilizing ion exchange chromatography (IEX) on HiTrap Q HP (GE), which produced a protein with purity exceeding 90% (Fig 1.24).

1.4.4 CHARACTERIZATION

DLS measurements were utilized to track the forms of recombinant Gam1 obtained from purification techniques. These measurements proved to be complimentary to the SEC profiles in helping assess the quality and status of recombinant TF-Gam1. The first elutions from SEC represent protein that elute in the exclusion volume of the column. These represent aggregates of recombinant Gam1, and when measured using DLS display a hydrodynamic diameter greater than 40 nm is size with an average typically 53 nm. The dimeric form of Gam1 generally gave measurements between 16-30 nm, with an average approximately at 23 nm. Finally, the monomeric form of TF-Gam1 would be about 9-13 nm, with the highest representation being around 10 nm. Based on SEC and DLS profiles, three factors have been found to be critical to increase the level of monomeric Gam1. First, it is critical to adjust buffers' pH at 4°C because the small pH difference caused by temperature displayed a dramatic

effect on obtaining monomeric TF-Gam1. Secondly, the addition of DTT to the purification buffers, clearly demonstrates a shift towards monomeric Gam1 while the dimeric and aggregate forms are decreased (Fig. 1.23,1.25). Zeta potential measurements were taken to assess the charge on the surface of TF-Gam1, which further verifying that the overall charge was indeed negative (Fig 1.26). Reducing environment could have decreased the formation of false intermolecular disulfide bonds that can cause dimerization or aggregation [69, 70]. A dramatic effect concerning DLS measurements at different temperatures was also observed when characterizing the sizing profiles of Gam1 samples at room temperature as compared to 4°C, with TF-Gam1 increasing in size over time at room temperature. The effect of salt on polydispersity was also tested utilizing different salt concentrations ranging from 100-1 M NaCl. The sizing profiles saw negligible differences after the addition of 400 mM NaCl, where the peak intensity shifted to a higher hydrodynamic measurement (Fig 1.27).

1.4.5 DISCUSSION

1.4.5.1 CLONING

Given that earlier trials of PCR amplification were carried out utilizing Long Amp PCR kits (NEB; based on the Taq DNA polymerase), mutation rates are more frequent than other polymerases and have been shown to also cause frame-shifts [71, 72]. To avoid these issues, the Phusion High Fidelity (Finnzymes), which has a mutation rate fifty times lower than Taq [73, 74]. Moreover, cloning issues coupled with the inability to produce viable protein for structural studies lead the decision to clone using a novel system know as Invitrogen's Gateway ® Directional Cloning system. This system was advertised to have a cloning efficiency higher than 90%, without the need for traditional Restriction enzyme digestions and ligations. In practice however, over 80 clones were screened using this system and the success rate was determined to be less than 10%. Additionally, verified clones that were later expressed never gave a soluble protein. Due to the difficulties encountered, coupled with the expense

associated with the system it was decided to move away from directional cloning. Gam1 was codon optimized for bacterial expression, and cloned into an alternative vector system, the pCold Trigger Factor (TF) that has a bacterial chaperone TF to assist the folding. The first sequence verified and viable protein expression was in the pCold-TF vector systems. This is a result of the fact that the vector contains a fusion protein that aids in the folding of Gam1.

1.4.5.2 PROTEIN EXPRESSION

Protein expression in early pET constructs had two key problems: 1) there was little protein in comparison to other major contaminants, and 2) protein expression was not observed at all. In regards to the first issue, of protein expressed in *E.coli* is quite common [75, 76]. There are various factors that lead to the insolubility of the recombinant protein. One of which could be the fact that the Gam1 construct provided by Dr. Chiocca was produced from Wild type Gam1, meaning it has codons that are rarely seen when trying to express in *E.coli.*. Secondly, and most importantly, Gam1 carries out it's biological function by interacting with proteins in the SUMOylation and ubiquitination pathways and could result in its interaction with proteins produced in bacteria. This would account for the inability to separate Gam1 from contaminants. As mentioned above, Gam1 was expressed in different strains of E. Coli to try to relieve expression issues, which have been shown to improve protein production [77, 78]. The pET vector systems utilized were under the control of T7 promoter systems. One of the major limitations using these strong promoter systems is the fact that recombinant protein produced may not be in its native conformation, which will shift its susceptibility to be isolated into the insoluble pellets (inclusion bodies) [79]. Furthermore, high levels of protein production can at times account for more than 40-50% of the total cellular proteins. This will call for high quantities of mRNA, which can lead to elimination of the ribosomes and ultimately lead to cell death. These strong promoter systems also have been known to have leaky expression [80] that may lead to the instability of protein expression [81].

Also, Gam1 contains a hydrophobic region at the C-terminus, which is purported to be the location where EloB/C binds. Without this interaction, Gam1 could potentially aggregate to protect that region. Another possibility is that protein expression with IPTG can produce large quantities of exogenous protein (toxic to the bacteria) and may be targeted for degradation. In tune with this, the large amounts of protein generated could also outpace the native folding process, leading to accumulation of unfolded protein as aggregates. Therefore, co-expression with solubility enhancers was utilized to resolve these issues. Molecular chaperones or solubility enhancers have been thoroughly investigated [82] and have resulted in the enhanced expression of soluble target protein. Moreover, the expression of protein at lower temperatures [83] can improve solubility of *E. Coli* produced proteins, however the behavior of proteins various from case to case and remain insoluble [84] or difficult to purify such as with Gam1. It is also common to express in systems that utilize promoters that generate protein at lower temperatures [85]. One of the well-characterized and utilized cold shock promoters is the CspA from *E. Coli* [86-88]. Many times the problem is not actual expression of target protein, albeit the toxicity that the exogenous protein is causing to the bacterial cell. Therefore, over production of a protein will harm the bacterial system and be targeted for degradation. pET vectors (pET 30b(+) and pET DEST 42) never successfully produced protein. This could be because no solubility factors were present and the gene was not codon optimized for bacterial expression. Due to these limitations, Gam1 was codon optimized for bacterial expression, and cloned into an alternative vector system that will also assist in the folding of the protein. The pCold Trigger Factor (TF) was selected to express Gam1. TF itself is a prokaryotic ribosome associated cold shock chaperone of 48kDa that helps in protein folding under low temperature. TF facilitates correct protein folding thus enabling efficient soluble protein production. In addition, at lower temperature, basal expression of endogenous bacterial proteins is reduced, resulting in overexpression of recombinant Gam1 in the total protein. Usage of expression tags, even as small as His-tags, however, does have some drawbacks to have possibilities to interfere with the conformation

and biological functions of proteins [89, 90]. For these reasons, Gam1 was cloned into pCold1 without the usage of TF. However, when induction trials were completed (Fig. 1.13), recombinant Gam1 was only detected in the insoluble fraction. Lower temperatures have been shown to destabilize hydrophobic interactions [91]. Gam1 contains a hydrophobic region at the C-terminus, which is purported to be the location where EloB/C (components of the ubiquitination pathway) binds [24]. Without a chaperone[92] to protect that hydrophobic region, Gam1 could potentially aggregate, which may attribute to mis-folded pCold1-Gam1 that will go to the inclusion bodies. This demonstrates that cold temperatures alone did not solve the problem of producing soluble protein; TF is playing a major role to protect the hydrophobic region in the proper folding of monomeric Gam1, especially at low temperature. Removal of TF by protease digestion was unsuccessful, further supporting that TF and Gam1 interact closely burying the linker region between them.

1.4.5.3 PURIFICATION AND *IN VITRO* SUMOYLATION

Purification of pET49b+-GST-Gam1 never reached homogeneity, even after a series of chromatographic techniques and optimization trials; contaminants could not be removed (Fig 1.14-17). As abovementioned, cooling to 16°C before induction is crucial for increased production of monomeric Gam1 (Fig. 1.23), suggesting that the cold temperature facilitated protein folding and increased the production of soluble TF-Gam1, and facilitated protein purification. Optimization of TF-Gam1 purification using affinity chromatography initially produced highly homogenous protein with a purity exceeding 90%, however it was determined that using high concentrations of imidazole resulted in large protein aggregates [93], which lead to the reduction of imidazole to obtain monomeric Gam1. Also, as mentioned above cold temperature purification carried out at 4°C coupled with buffers pH'ed for these temperatures greatly improved the production of monomeric Gam1, likely slowing the activity of protease activity and slowed the course of protein degradation [94]. Previously, Gam1 has been

demonstrated to inhibit global SUMOylation [11]; here, we want to verify that recombinant TF-Gam1 maintains its biological function. Experiments here showed that when incubated with recombinant TF-Gam1, NS1 SUMOylation was abridged (lanes 5- 10 of Fig. 1.28), when compared to the positive and negative controls (lanes 1 and 3, Fig. 1.28). These results are consistent with previous results demonstrating Gam1's function to reduce SUMOylation [9]. However, SUMOylation was not completely eliminated by TF-Gam1. Based on quantified data (Fig 1.29), it was estimated that approximately 40% reduction of NS1 SUMOylation by TF-Gam1 compared to the controls. As mentioned in the previous discussion, TF plays a vital role in the folding of Gam1, probably protecting the C-terminus hydrophobic region and potentially hiding the contact region that binds to the E1 heterodimer. This might reduce recombinant Gam1's SUMOylation inhibition efficiency as previously demonstrated *in vivo* [24]. Moreover, the conformational heterogeneity [95] caused by the additional TF could be another reason Gam1 cannot freely interact with E1 and completely abolish SUMOylation. To a lesser extent, during the SEC purification of monomeric Gam1, dimeric form cannot be completely separated (Fig. 1.23). Furthermore, it was observed that after SEC, storage of monomeric TF-Gam1 at 4°C for a prolong period of time gradually led to aggregation. This problem became more severe when sample are more concentrated. This suggests that small percentage of dimer and aggregates could also account for the reduction of Gam1's activity.

1.5 SUMMARY AND FUTURE DIRECTIONS

Gam1 is the first viral protein found to globally inhibit cellular SUMOylation by directly interfering with SUMO E1 activity. However, the structure of Gam1 is not available and detailed data regarding the interface between Gam1 and the first enzyme in SUMOylation remain unclear. Thus, recombinant Gam1 was cloned and expressed in various bacterial expression systems to obtain pure and soluble Gam1 protein for further *in vitro* functional and structural studies. Most expression systems

produced insoluble Gam1. The most soluble protein was produced from a system with the fusion tag of Trigger Factor (TF), a cold temperature ribosome associated bacterial chaperone. TF-Gam1 was expressed under low temperature and purified to high homogeneity through sequential chromatography techniques. The purified protein was characterized by dynamic light scattering and by its ability to inhibit *in vitro* SUMOylation reaction of target substrates. Recombinant Gam1 maintained it's biological function to a certain extent, TF dramatically helped the expression and purification of Gam1, it might interfere with the function. Nevertheless, the current TF-Gam1 constructs are still useful for *in vitro* functional and structural studies. Therefore, pure TF-Gam1 will be pre-screened for crystallization trials. If a crystal is obtained, structure determination will be carried out and given that TF's structure has been determined [51], molecular replacement [96] can be utilized for structural calculations. Future experiments will target co-expression of Gam1 with its native cellular counterparts to stabilize the hydrophobic interface. Another future endeavor will include cloning into a vector with novel solubility enhancers that can be removed during purification, or that do not affect biological activity. As a last resort, if protein quality is not suitable for structure determination eukaryotic expression such as insect baculovirus [97] cells will be utilized. The result from each of the cloning and expression vectors selected is summarized in table 1.2.

These results will also establish a foundation for the future structural determination of Gam1 with or without the binding components to investigate the interactions between Gam1 and its cellular target proteins, which will deepen our understanding of how Gam1 impinges on the SUMOylation pathway as well as its roles within viral replication.

TABLE 1.2 SUMMARY OF GAM1 VECTOR OUTCOME

Vector System	Cloned	Expressed	Result
pET-30b(+)	Yes	Questionable	Expression Level too low, Construct has mutation
pET-49b(+)	Yes	Yes	Degradation, low purity
N-p-DEST42-Gam1	Yes	No	Mutation
C-p-DEST42-Gam1	Yes	No	Mutation, Expression not detected
N-pCold-TF	Yes	Yes	Degradation
C-pCold-TF	Yes	Yes	Testing functionality
pCold1-Gam1	Yes	Yes	Insoluble

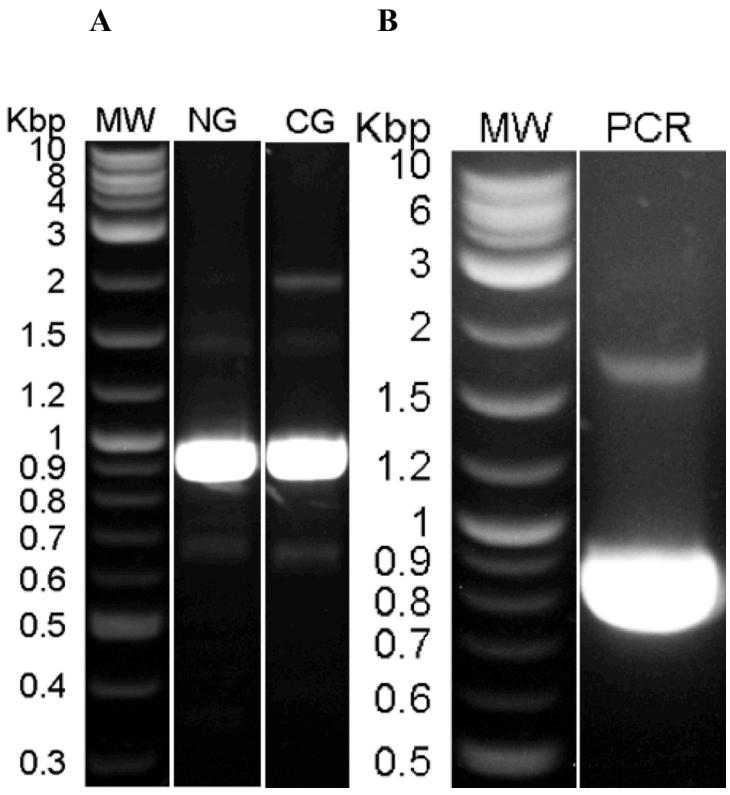

Fig 1.2 Agarose gel(s) of Gam1 PCR products for pET cloning. A) PCR products from pET-DEST42 cloning, NG is the N-terminal His tagged Gam1 and CG is C-terminal His tagged Gam1. B) PCR for pET49b(+). The bands above PCR products were removed by purifying via gel harvest. The molecular weight markers (MW) are in the first lane

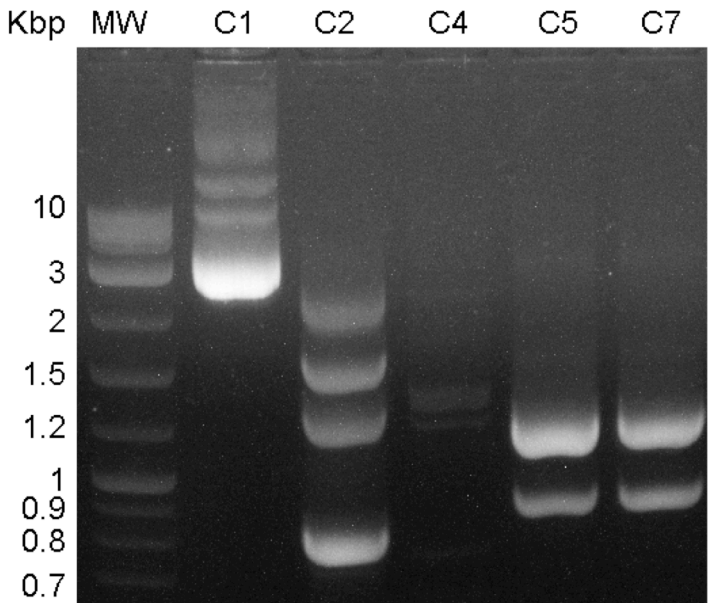

Fig 1.3 Agarose gel of Gam1 TOPO cloning screening. C stands for clone, and the wells above the panel denote the clone selected for screening. The molecular marker is in the first well and labeled to the left of the panel.

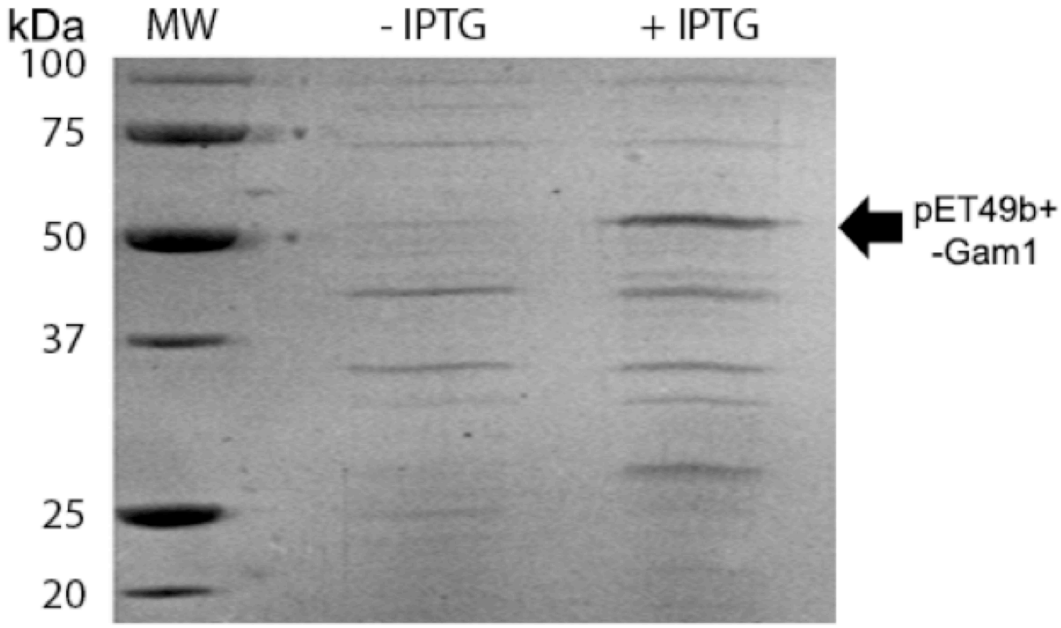

Fig 1.4 SDS-PAGE of small scale expression trials. pET49b+-GST-Gam1 supernatants with and without IPTG induction. Molecular weight (MW) markers are in the first lane of each SDS-PAGE with the corresponding size labeled on the left. Arrows on the right of each panel point to the expected size of recombinant Gam1.

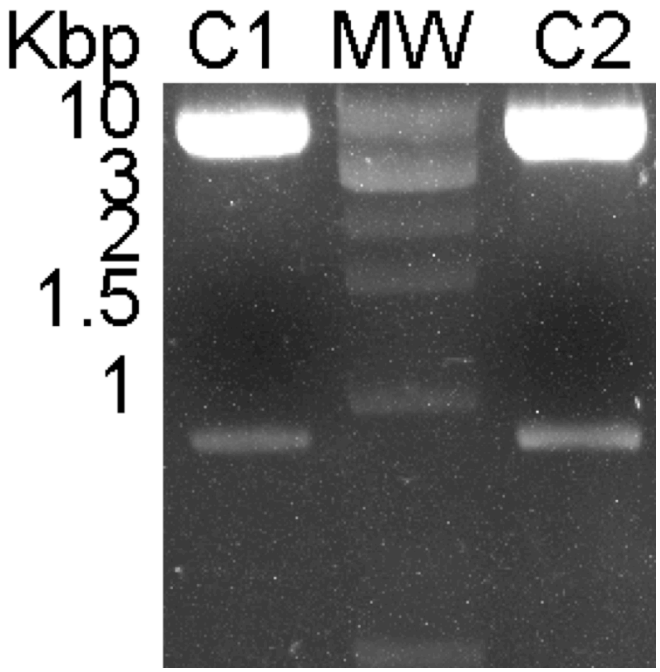

Fig 1.5 Agarose gel of pCold-TF-Gam1 dual restriction enzyme digestion. C stands for clone. The molecular marker is in the second well and labeled to the left of the panel.

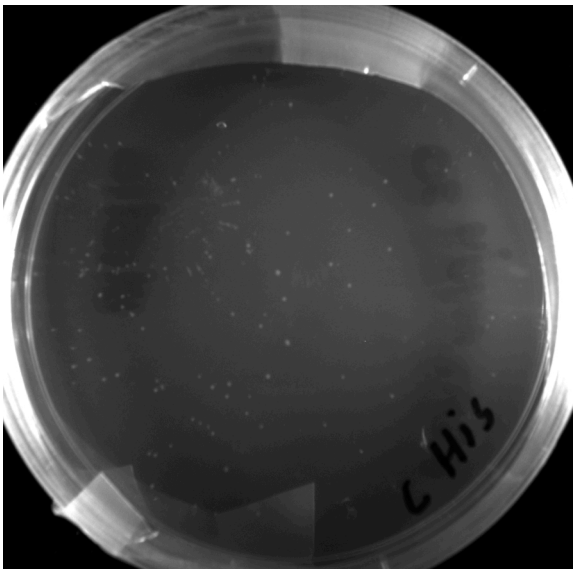

Fig 1.6 Selectivity plate of transformed Gam1 into bacterial competent cells.

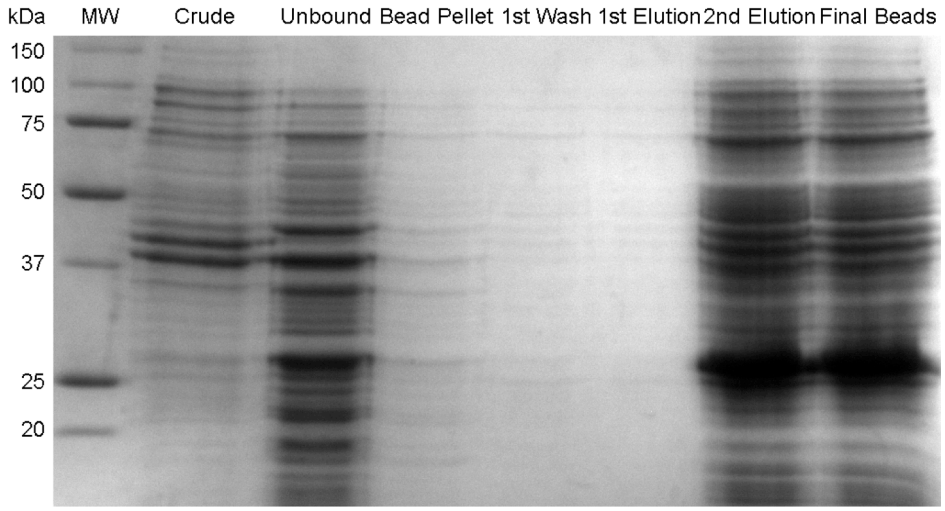

Fig 1.7 SDS-PAGE of pET30b+-Gam1. The lanes are labeled above the panel. The molecular weight marker is in the first lane and labeled to the left of the panel.

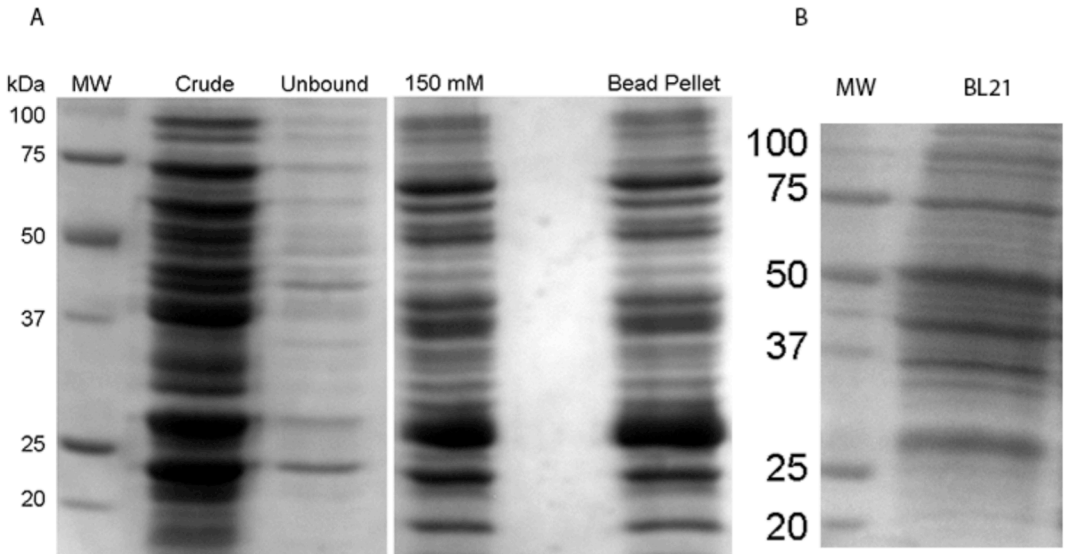

Fig 1.8 SDS-PAGE of Gam1 affinity pulldown/ trial expression. The image on the bottom right shows Empty BL21 (DE3) induced, as compared to pET-30b+-Gam1 on the left. The major expression could be non-specific. Molecular markers are in the first lane and labeled to the left of the panel.

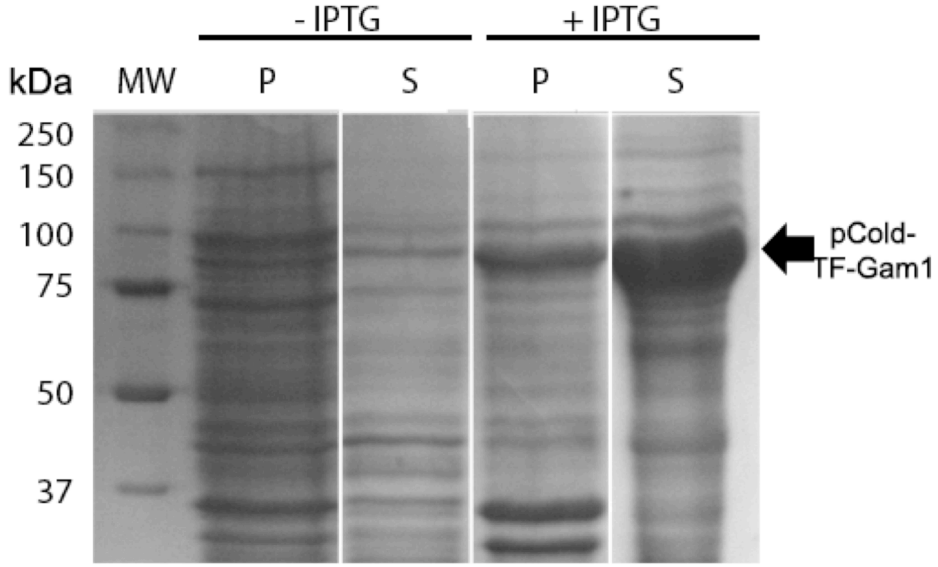

Fig 1.9 SDS-PAGE of small scale expression trials. pCold-His-TF-Gam1 pellet (P) and supernatant (S) samples with and without IPTG induction. Molecular weight (MW) markers are in the first lane of each SDS-PAGE with the corresponding size labeled on the left. Arrows on the right of each panel point to the expected size of recombinant Gam1.

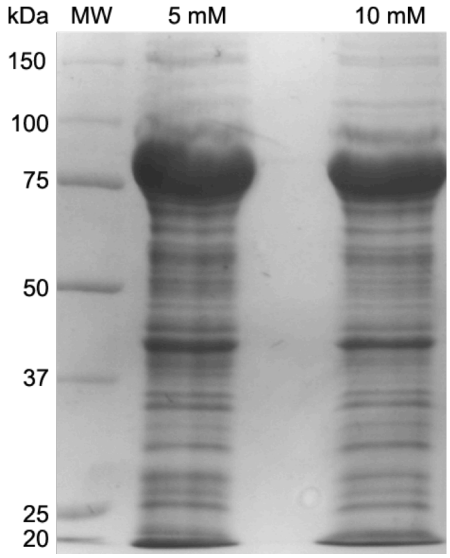

Fig 1.10 SDS-PAGE of IPTG induction optimization. pCold-His-TF-Gam1 was induced with the indicated concentrations of IPTG. Molecular weight (MW) marker is in the first lane and labeled to the left of the panel.

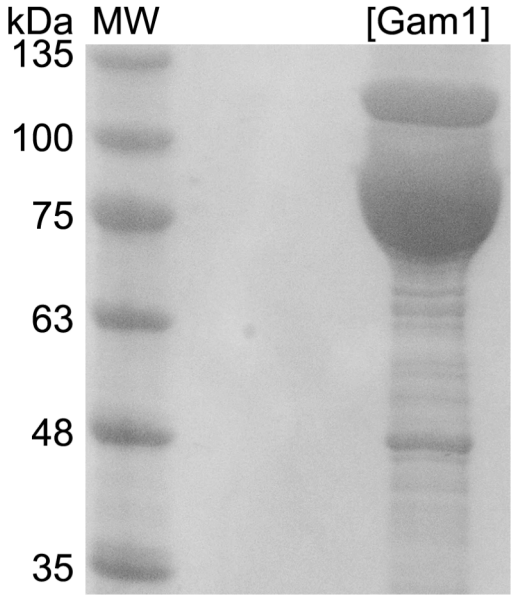

Fig 1.11 SDS-PAGE of pCold-His-TF-Gam1 concentrated after several chromatographic purification techniques. The molecular weight marker (MW) is in the first lane and labeled to the left of the panel.

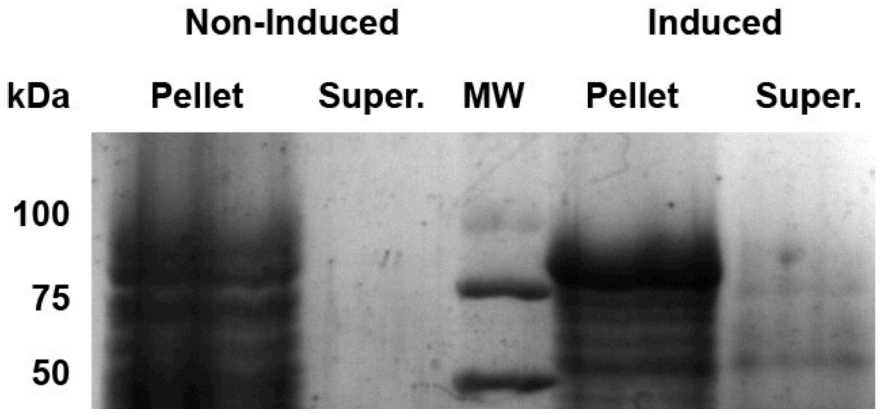

Fig 1.12 SDS-PAGE of small scale expression trials. pCold-TF-Gam1-His pellet (P) and supernatant (S) samples with and without IPTG induction. The molecular weight marker (MW) is in the first lane and labeled to the left of the panel

.

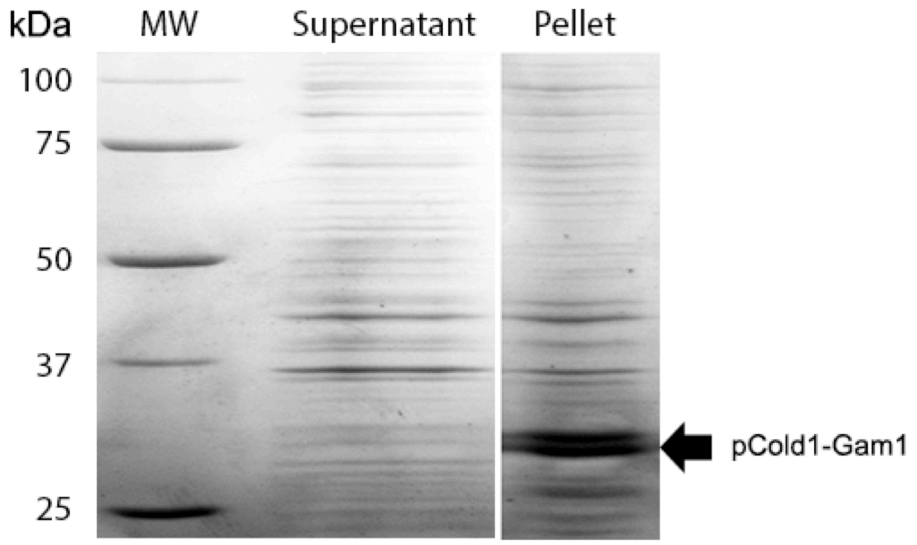

Fig. 1.13 SDS-PAGE of small scale expression trials. pCold1-Gam1 supernatant and pellet samples. Molecular weight (MW) markers are in the first lane of each SDS-PAGE with the corresponding size labeled on the left. Arrows on the right of each panel point to the expected size of recombinant Gam1.

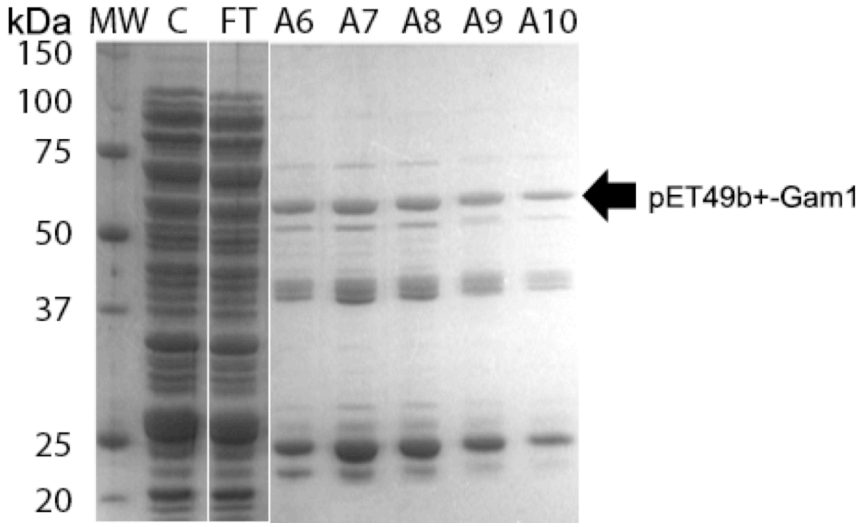

Fig 1.14 SDS-PAGE of pET49b+-GST-Gam1 samples from GST affinity column: crude (CR), flow through (FT), and elution fractions (A6-A10). Molecular weight (MW) markers are in the first lane of each SDS-PAGE with the corresponding size labeled on the left. Arrows on the right of each panel point to the expected size of recombinant Gam1.

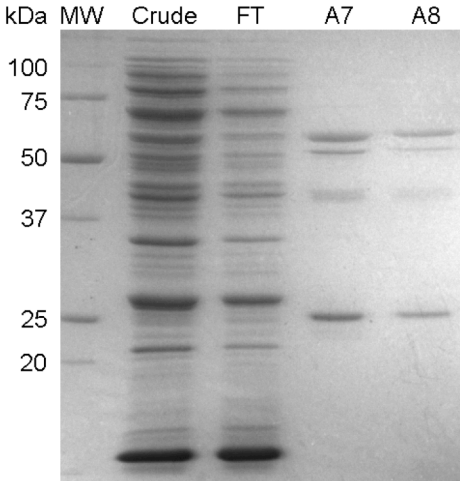

Fig 1.15 SDS-PAGE of pET49b+-GST-Gam1 samples from GST affinity column with reduced flow-rate (0.1 mL/min): crude (CR), flow through (FT), and elution fractions (A7, A8). Molecular weight (MW) markers are in the first lane of each SDS-PAGE with the corresponding size labeled on the left. Arrows on the right of each panel point to the expected size of recombinant Gam1.

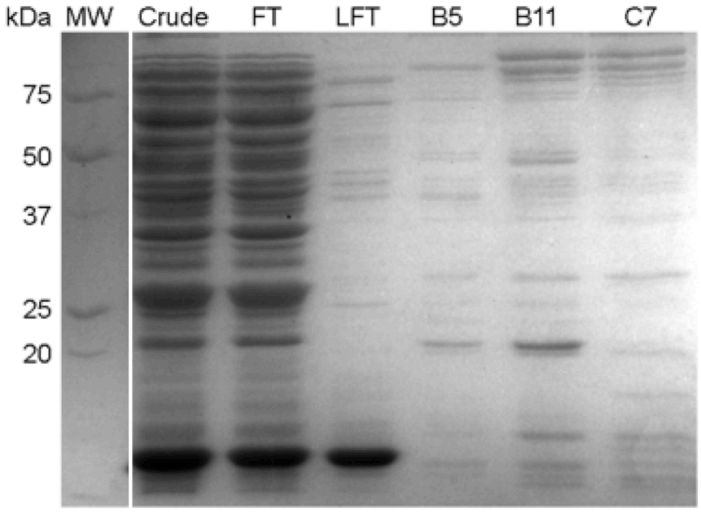

Fig 1.16 SDS-PAGE of pET49b+-GST-Gam1 using HiTrap SP FF Cation exchange chromatography. FT and LFT stands for Flowthough and late flowthrough, respectively. B5, B11, and C7 are elutions. Molecular weight (MW) markers are in the first lane of each SDS-PAGE with the corresponding size labeled on the left.

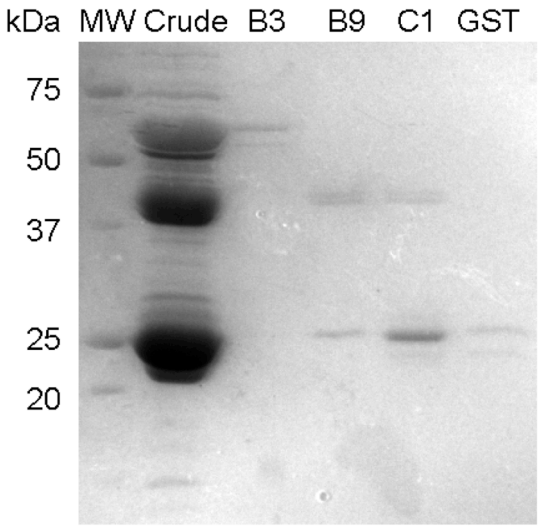

Fig 1.17 SDS-PAGE of pET49b+-GST-Gam1 using size exclusion chromatography. B3, B9, and C1 are elution fractions from the column. Molecular weight (MW) markers are in the first lane of each SDS-PAGE with the corresponding size labeled on the left.

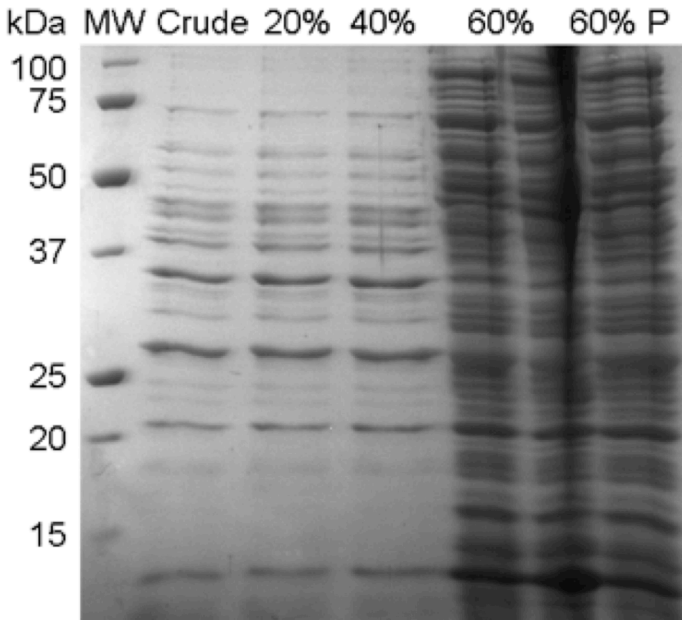

Figure 1.18 SDS-PAGE of pET30b+-Gam1 Ammonium sulfate precipitation trials. Different concentrations of saturated ammonium sulfate (as indicated above the lanes) were used to remove contaminants from the target Gam1 protein. The P stands for final pellet obtained. The molecular weight marker is in the first lane and labeled to the left of the panel.

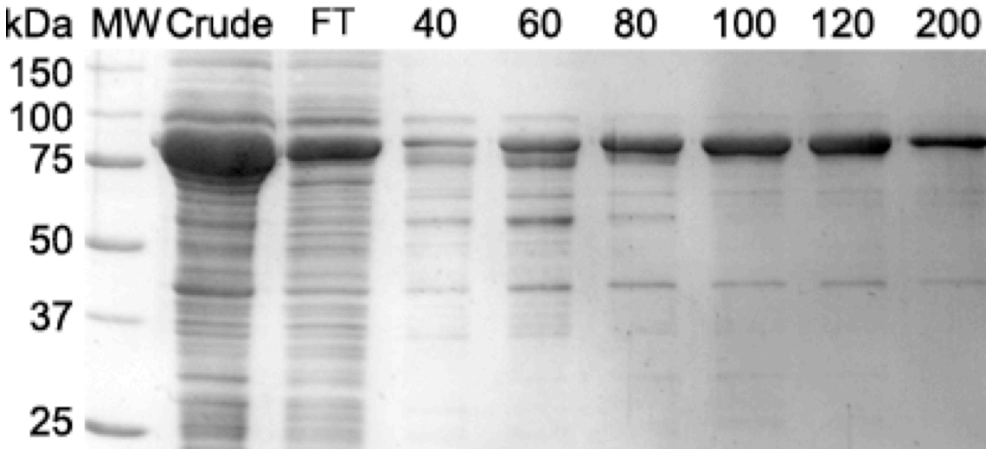

Fig. 1.19 SDS-PAGE of TF-Gam1 purification**.** Gravity flow affinity chromatography of TF-Gam1 using gradient concentrations (40 mM to 200 mM) of imidazole to determine the optimized wash and elution conditions, FT stands for flow through. Molecular weight marker is in the first lane and labeled to the left of the panel.

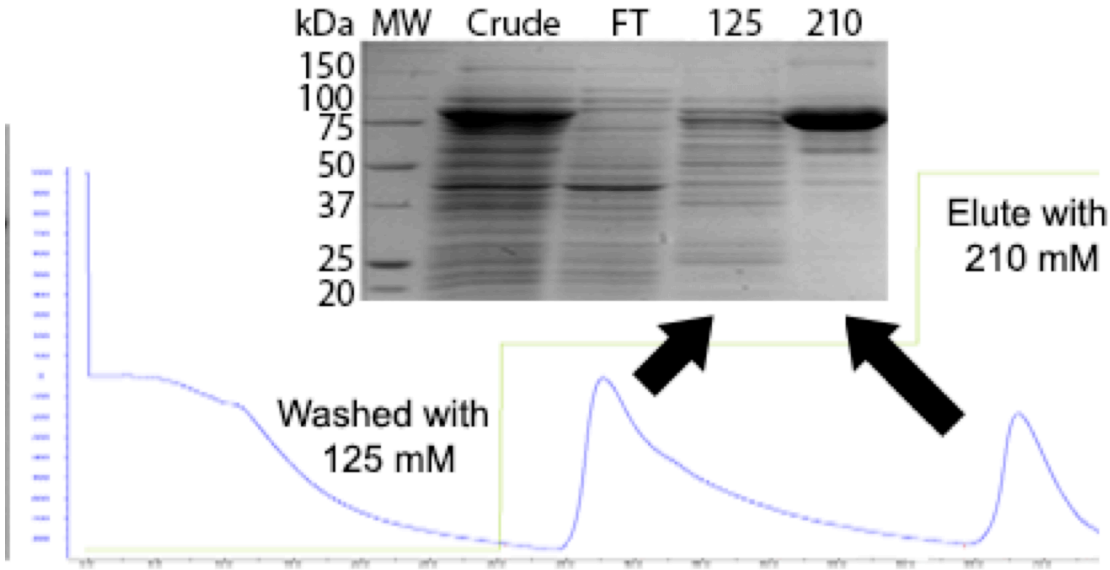

Fig 1.20 SDS-PAGE and chromatograms of TF-Gam1 purification. TF-Gam1 affinity chromatography using GE HisTrap HP with optimized wash and elution conditions determined in (A) and labeled on the chromatogram.

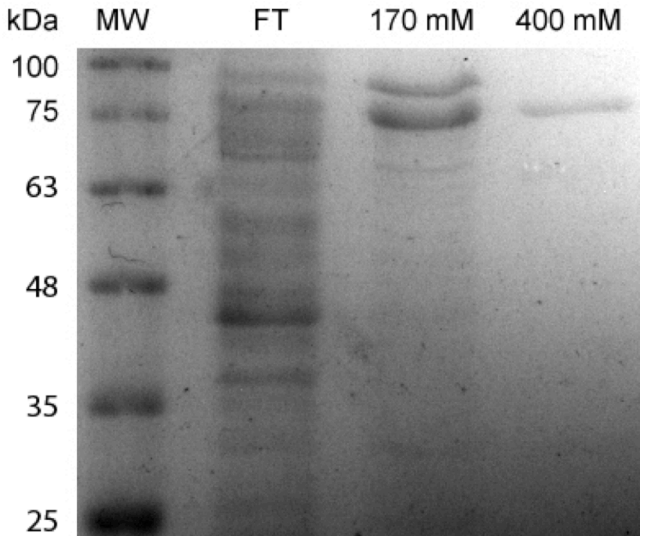

Fig 1.21 SDS-PAGE of TF-Gam1 purification. TF-Gam1 affinity chromatography using GE HisTrap HP with 170 and 400 mM wash and elution conditions for increased purity. FT stands for flow-through. The molecular weight marker is in the first lane and labeled to the left of the panel.

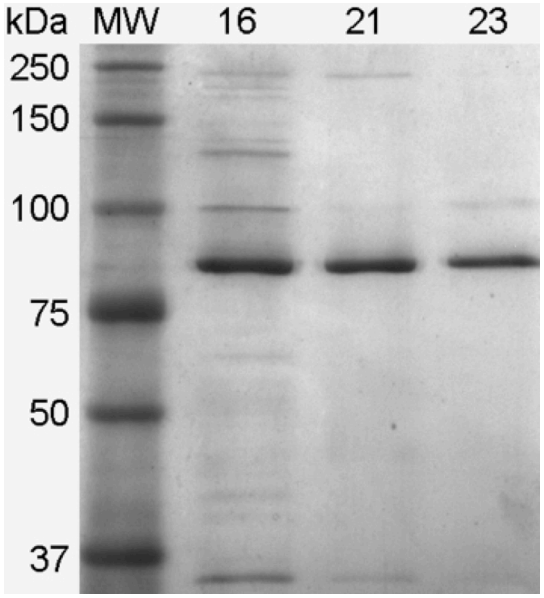

Fig 1.22 Size exclusion chromatography (SEC) purification of TF-Gam1. The three lanes correspond to the three states of TF-Gam1. The molecular weight marker is in the first lane and labeled to the left of the panel.

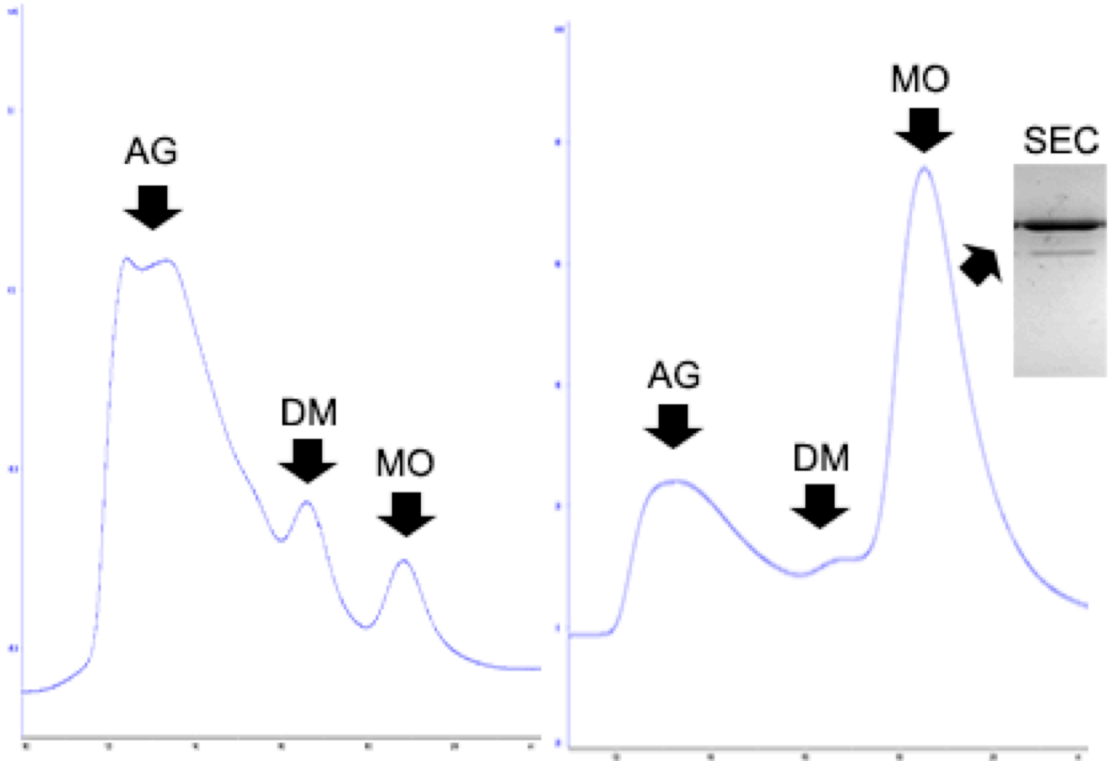

Fig 1.23 Size exclusion chromatography (SEC) purification of TF-Gam1. Left and right panel showing chromatograms before and after optimized purification conditions, respectively. Three peaks of aggregation (AG), dimeric (DM) and monomeric (MO) forms of TF-Gam1 are depicted by arrows. Lanes of an SDS-PAGE of TF-Gam1 are shown with arrows pointing to target peaks.

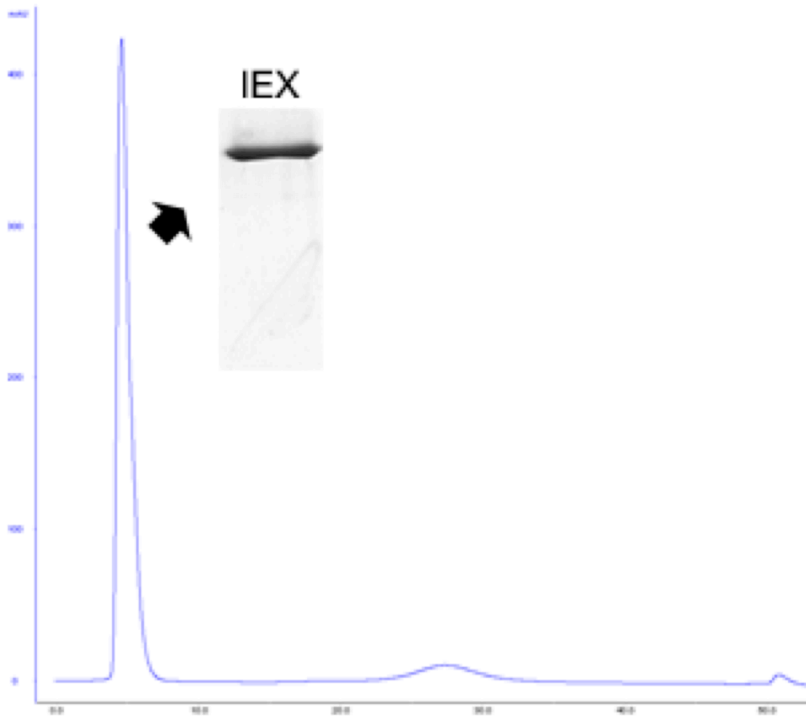

Fig 1.24 SDS-PAGE of Ion exchange chromatography (IEX) purification of TF-Gam1 after (SEC). Lane from an SDS-PAGE of TF-Gam1 with an arrow pointing to target peak.

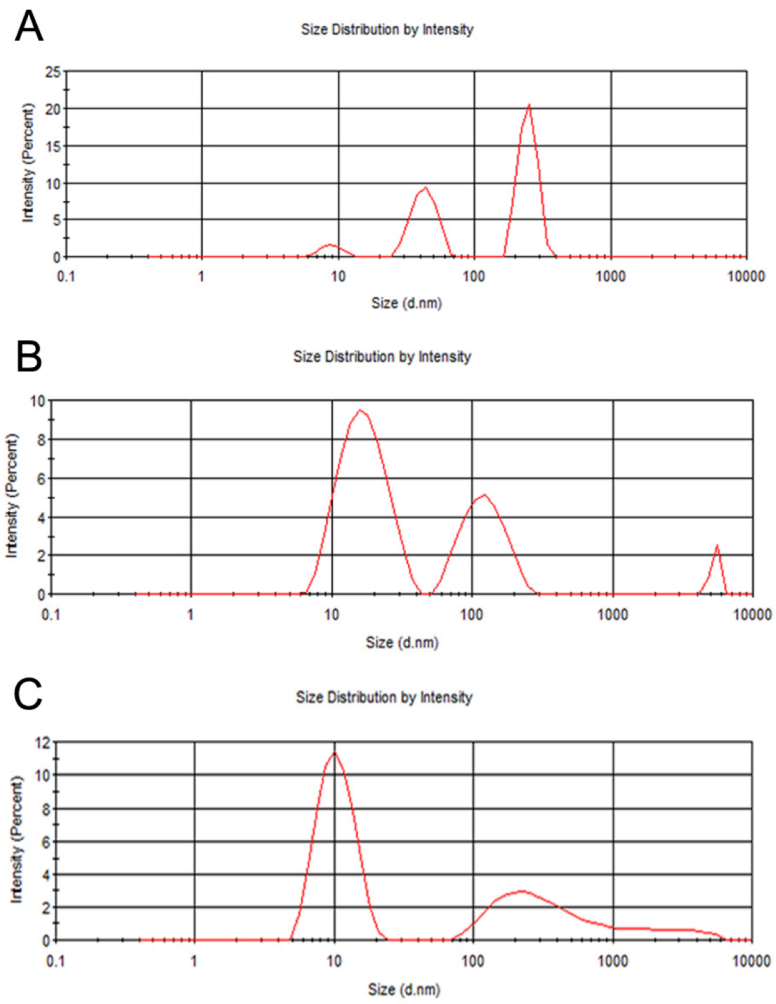

Fig. 1.25 The hydrodynamic size profile of TF-Gam1 characterized by dynamic light scattering (DLS). (A), (B), and (C) represent the aggregated, dimeric, and monomeric forms of TF-Gam1 eluted from SEC, respectively. DLS profiles can be used to determine TF-Gam1 forms.

Zeta Potential (mV): -13.8

Zeta Deviation (mV): 0.00

Conductivity (mS/cm): 12.3

Result quality : Good

Fig. 1.26 The Zeta potential surface charge measurement demonstrating TF-Gam1's overall negative charge.

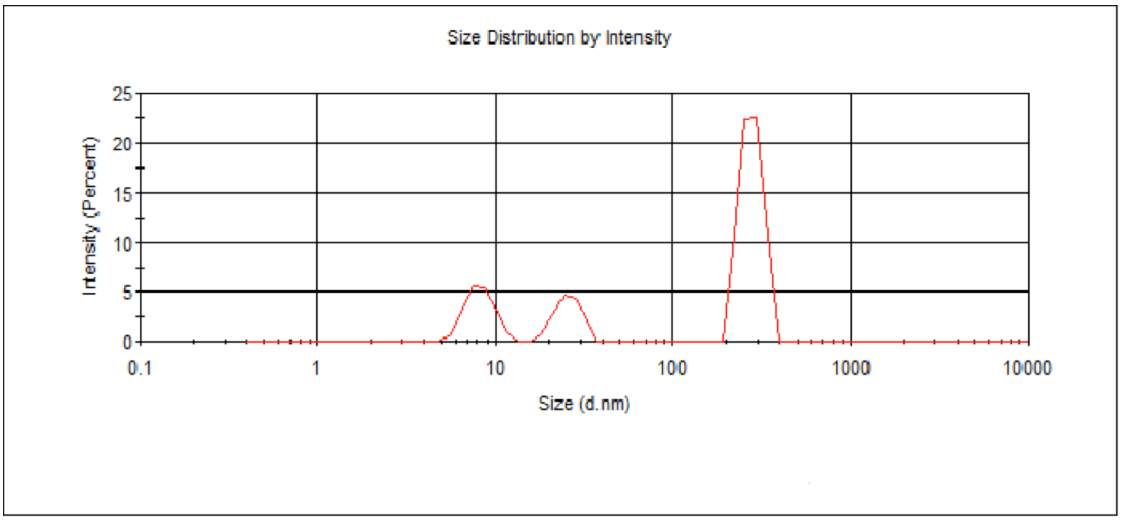

Fig 1.27 The hydrodynamic size profile of High salt TF-Gam1 characterized by dynamic light scattering (DLS). Polydispersity shifts to a higher hydrodynamic size with the addition of 500 mM NaCl.

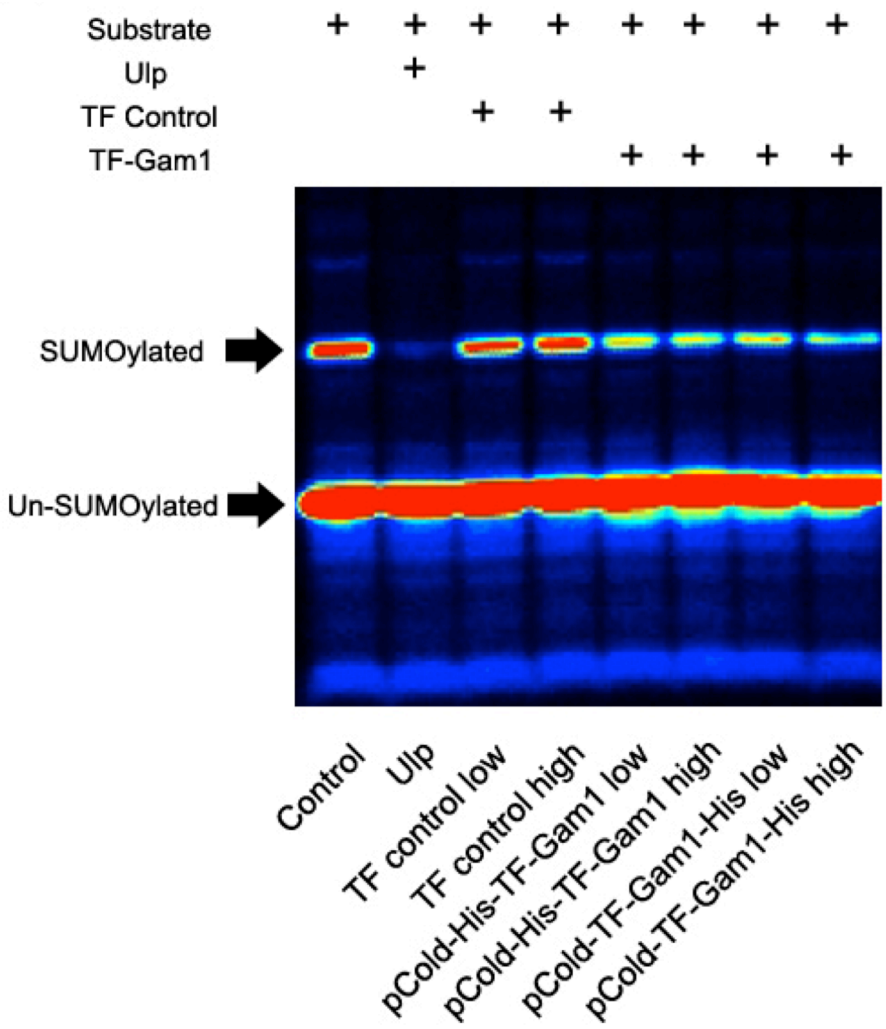

Fig 1.28 *In vitro* SUMOylation of TF-Gam1. NS1 (substrate) was incubated in the *in vitro* SUMOylation reaction mix comprising purified GST-SAE2/SAE1, GST-UBC9 in an ATP-containing buffer with increasing amounts of TF-Gam1 for 1.5 hr at 37°C. Reactions were stopped as described in Materials and Methods, loaded on SDS-PAGE, and scanned using autoradiography. Arrows on left point to the expected size of SUMOylated/Un-SUMOylated substrate. Samples are labeled below each lane in panel. For each protein sample tested, "low" (1.2 μg) and "high" (2.4 μg) quantities of purified Gam1 were used in the reaction to determine whether any dose dependent SUMOylation-inhibitory effects could be observed.

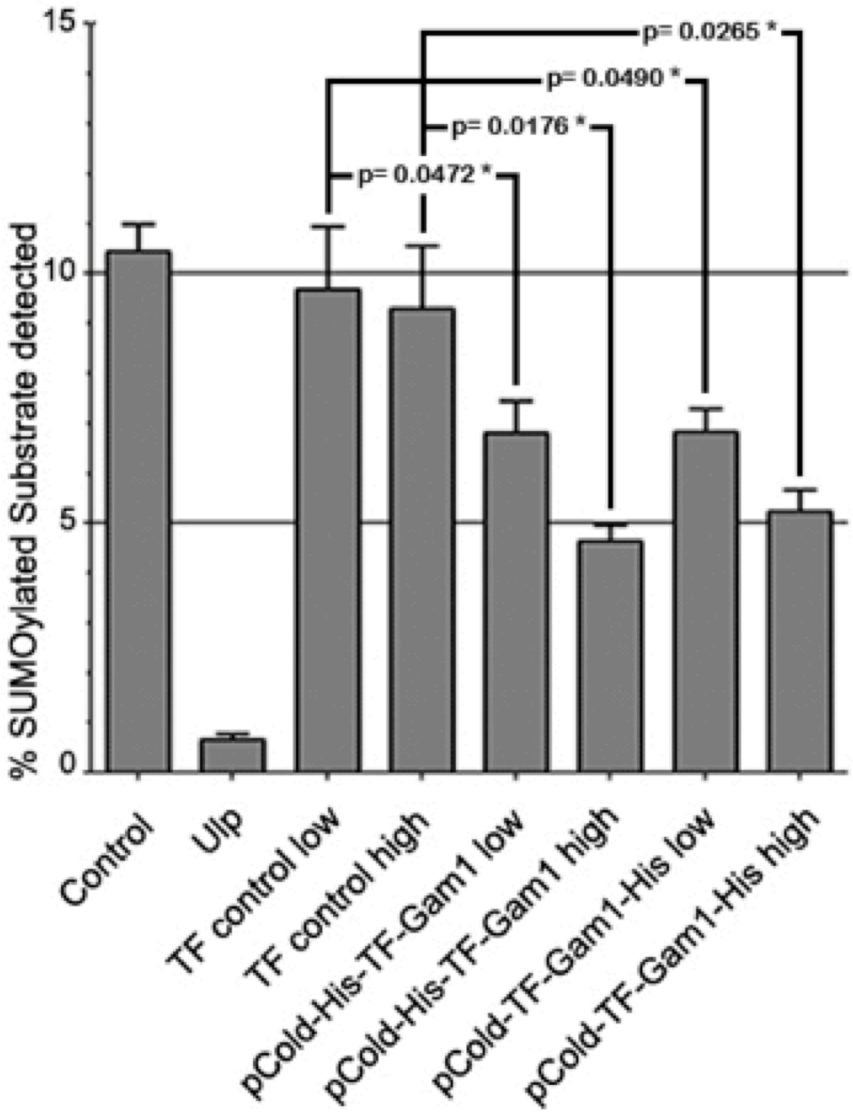

Fig 1.29 The *in vitro* SUMOylation efficiency was quantified by densiometric scans of (A) for the total amount of SUMO modified substrate using the un-SUMOylated substrate as internal standard. Three replicates were performed to estimate the standard errors. Statistics t-test was applied and p values were calculated and label accordingly above the bar graphs. * signifies statistical significance using conventionally accepted $p < 0.05$ criteria. Samples are labeled below each corresponding bar in panel.

CHAPTER 2:

STRUCTURAL INVESTIGATION OF THE ANTHRAX TOXIN RECEPTOR COMPLEXES

2.1 INTRODUCTION

In 2001, the importance of Anthrax was brought to light in the worst case of bioterrorism in the United States. Letters containing spore-contaminated correspondence were sent to political figures and media outlets, which resulted in a terrified nation as well as 22 infected with 5 deaths [98]. As a result, one billion dollars were spent for decontamination; new legislations were passed and research to find treatments thrived. The Anthrax disease is caused by the virulent strain of bacterium *Bacillus anthracis:* a gram-positive, endospore forming, rod-shaped bacterium [99]. Anthrax infection is typical in livestock namely: horses, sheep, cattle, and goats. The occurrence of Anthrax in humans is incidental when they come into contact with those afflicted animals, which accounts for more than 95 percent of Anthrax cases [100]. Intoxication with Anthrax can occur in three manners: Inhalation of the spore, through a cut in the skin, or intestinal [101]. The virulent properties of Anthrax were not discernable until the year 1954, when the toxin was discovered by Harry Smith in 1954 [99]. Anthrax toxin is a self-assembled complex, which is made up of three components: Protective Antigen (PA), Edema Factor (EF), and Lethal Factor (LF). Individually these components are non-toxic. They only become toxic once they form a complex after interacting with its respective cellular receptor [102]. Anthrax is classified as an A/B toxin, meaning it needs two components (a delivery vesicle PA, and the payload EF, and LF) [103-105]. PA and EF form Edema Toxin (ET), which cause non-lethal edema; PA and LF form Lethal Toxin (LT), which is lethal to the cell [101]. Protective Antigen is the delivery vehicle for EF and LF, and serves as the protein that recognizes one of the two known host cell-surface receptors: Anthrax toxin receptor 1 (ANTXR1) or Anthrax toxin receptor 2 (ANTXR2) [105, 106]. Only after PA (83 kDa) binds the receptor is it proteolytically processed by a cellular enzyme called Furin, leaving a

truncated PA monomer of about 63 kDa in size. The truncated PA begins the formation of a heptameric complex (pre-pore), which then attracts free (1 to 3) EF and LF proteins. The entire complex then undertakes endocytosis, where the pre-pore becomes a pore under low pH environment within the endosome after acidification and effectively releasing EF and LF into the cytoplasm [104, 107]. The mechanism is visualized in Fig 2.1.

2.1.1 PA AND ATXR2

PA contains four domains, of which domains 2 and 4 have been implicated in receptor binding[108]. Domain 2 is involved in heptamer formation and is found along the inside of the heptamer. Domain 4 interacts with the receptors binding domain and has been suggested to play a role in the pH dependent pore formation [108, 109]. Despite the high sequence homology between ATXR1 and 2, binding activities with PA are quite different [110]. Mutagenesis studies performed by Heather Scobie, et. al in 2006 demonstrated the physiological importance of ANTXR2 in the disease progression, showing that a mutated form of PA could still interact and cause a lethal infection, while ANTXR1 played a smaller role [110]. ANTXR2, the one of primary interest to this research has four components. It is characterized as a type I transmembrane protein, with the primary interaction with PA relegated to the von Willebrand Factor A domain (VWA) [111]. It has what has been called a stalk region connected to a transmembrane domain, and finally a cytosolic domain. It is well known that the VWA domain is crucial for Anthrax toxin infectivity, but recently the stalk region has been implicated as an immunoglobulin-like (Ig) domain; which has demonstrated the importance of disulfide bonds for infection [112, 113]. In this research, structural studies utilizing Cryo and Negative Stain Electron Microscopy were performed on the Protective Antigen in complex with ANTXR2. It is expected that there will be a vast difference in the Ig domain with the WT and a Mutant version where disulfides have

been mutated, making this the first characterization of WT PA and a Mutant in which the disulfide linkage has been altered.

2.1.2 ELECTRON MICROSCOPY

Cryo Electron Microscopy (Cryo-EM) of biological specimens namely for single particle reconstructions utilizes samples that have been prepared and frozen in an aqueous environment to mimic their native state. In the past decade, electron microscopy of biological samples has come to the forefront of structural biology. There exist many obstacles that need to be resolved when handling biological samples in Cryo-EM such as: limited dosage of electrons due to potential radiation damage, the need for high vacuum, and the resulting low contrast due to the fact that higher dosage of electrons will damage the sample under investigation [114]. One of the steps taken to minimize these issues is vitrification of the sample, which means to freeze the sample in a very rapid manner as to produce ice that will not impede with data collection [115]. Another step is the advent of cryo-protecting the sample by introducing cold stages and as well as maintaining the sample at cryogenic (cryo) temperatures while collecting data; which allows for slightly higher dosage of electrons, while reducing radiation damage incurred [114]. A collection of individual macromolecules can be observed using the above-mentioned techniques, which is now approaching atomic resolution [116-118]. Once data have been collected, individual micrographs will be processed. This process is ubiquitous regardless of the software utilized. In this case, EMAN, and FREALIGN were used for the image processing and reconstructions. The reconstruction procedure begins by either digitizing the films, such as scanning the micrographs with a Nikon coolscan 9000, or transferring data from the CCD to a portable hard-drive. The image processing will commence by isolating single particles from the hundreds of other particles and background within the micrograph. Once all the particles have been "boxed", they undergo a series of pre-processing steps including: image normalization, defocus level determination, and correction of the contrast transfer

function (an artifact that is a byproduct of the microscope). Utilizing the pre-processed images from the previous steps, an initial model was generated for the following reconstruction cycles [116]. Within the iterative process, there are two major tasks to be conducted: orientation determination and three-dimensional reconstruction. In the orientation determination, using the initial model or the reconstructed model from the previous cycle, the individual particles' information regarding orientation and center is calculated. After this refinement cycles with the entire dataset is performed where the angles and orientation information will be used to reconstruct a new model for the ensuing cycles. This iteration cycle will be continued until the highest possible resolution from the dataset is obtained. In this chapter, we will compare the reconstruction between the wild type complex and a mutant in an effort to understand how the efficient pore formation has been disrupted by removal of the disulfide bonds in pore formation. The results obtained will guide the development of novel treatments for Anthrax infection.

2.1.3 SUMMARY

Overall, the key aspect covered in this introduction pertains to the reason for the boom in Anthrax investigation, namely the mechanism by which the toxins work and are shuttled into the cell. Inhalation of Anthrax spores poses a serious threat, and developing novel treatments that can halt the illness are crucial. Progression of the disease has mainly focused on the first domain of the human receptor (ANTXR 2), however, recent evidence points to a major role undertaken by the second Ig-like domain in the delivery of the toxin. As a result, structural studies utilizing transmission electron microscopy were performed. Cryo-electron microscopy is now at the forefront of macromolecular reconstruction, with many advances made in data collection, which helps to preserve biological samples in their native environment, while getting closer to atomic resolution. Data processing for single particle reconstructions is divided into three phases: preprocessing the images, building the initial model, and an

iterative refinement cycle to obtain the best resolution reconstruction from the data. In the remainder of this chapter, we will discuss how the reconstructions of Anthrax toxin complexes were carried out and comparing the structural differences in the Mutant (second domain disulfide mutation) with the WT in effort to locate the position of the second domain of the ANTXR 2 in the complexes.

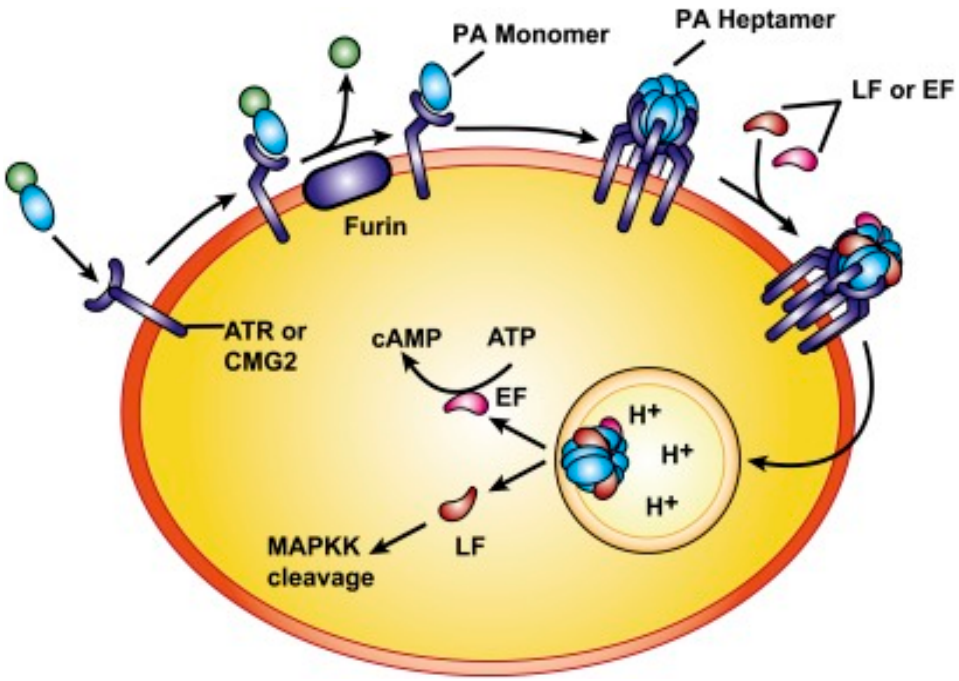

Fig 2.1 Model of Anthrax toxin receptor. Adapted from [101]

2.2 LONG TERM GOALS AND SPECIFIC AIMS

As mentioned in the introduction, multiple biochemical data as well as structural data have been collected for the PA and domain one (VWA) of ANTXR 2. However, not until recently has domain two been implicated in the overall course of infection. The **long-term goal** of this research is to determine high-resolution structures of PA and VWA and Ig-like domains of the ANTXR 2 receptor in complex. Information obtained would lay the foundation in understanding the role domain two of the receptor plays in toxin delivery. In this dissertation, a total of six Negative stain as well as Cryo-EM datasets of

the above mentioned mutant and WT were processed and final models were generated. One specific aim is proposed, which will test the hypothesis that *the complex structure of PA with domains one and two of ANTXR2 will maintain different conformations with respect to WT and the domain 2 (Ig-like) Mutant.*

Specific Aim: To generate a three-dimensional (3-D) model of the WT and MUT datasets and highlight the key structural differences in ANTXR 2 Ig-like domain. Cryo-electron Microscopy as well as negative stain datasets of the abovementioned complexes was collected and three-dimensional models were generated to investigate the structural details pertinent to Ig-like doamin of the ANTRX 2 receptor. Results obtained from these experiments will show that coupled with the biochemical analyses (Pedro Jacquez/Dr. Sun) Ig-like domain organizations within the macromolecular assembly plays a critical role in the efficient delivery of the toxins into the cell. Previous research has focused on the interaction of PA with VWA of the ANTRX 2 receptor; here we are interested in observing the conformation of the Ig-like domain and how the mutation of the disulfide bonds abrogates the function of toxin delivery.

2.3 MATERIALS AND METHODS

2.3.1 EM SPECIMEN PREPARATION AND DATA COLLECTION

The PA/receptor complexes were negatively stained on ultrathin carbon film (Ted Pella #01824) using 2% sodium phospho-tungstate. Negative stain data were collected on JEOL 3200FS transmission electron microscope equipped with field emission gun and in-column omega filter (University of Texas at El Paso). 3.5 uL sample aliquots of purified protein complexes were blotted onto plasma cleaned thin holey carbon grid (C-flat CF-1.2/1.3-2C) and manually plunged into ethane at liquid nitrogen temperatures. Cryo-EM data were recorded on a Gatan 4k X 4k CCD using a JEOL 2200FS transmission electron microscope equipped with field emission gun and in-column omega filter

(University of Texas Medical Branch) operating at 200 kV nominal magnification of 60,000x. The sampling rate was at 15 micron/pixel, generating a pixel size of 1.97 A at the sample.

2.3.2 Image Processing, 3D Reconstruction, and Docking of Atomic Models

Single particle reconstructions depend on high-end microscopes to collect the data, coupled with multicore computers for high throughput calculations. Multiple platforms to process the data exist, which may use different but very similar approaches. Many of the features from individual programs have over-lapping functionality, and many times certain aspects of the software can be combined from multiple programs to carry out the desired calculation. The major differences observed are relegated to the manner in which classification, CTF correction, and model refinement are performed.

2.3.2.1 Particle Selection and Preprocessing

Image analyses and three-dimensional reconstructions were carried out using the EMAN and FREALIGN [119] imaging suites. Particles were manually selected using EMAN1 software suite [120]. Data were linearized and normalized using EMAN2.1 [121]. Before image selection, the micrographs will be normalized with proc2d and the command *edgenorm*, which adjusts the means and standard deviation of the images. In order to accommodate the individual particles, a box size ~ 1.5-2 times the size of the longest particle dimension (160) was selected [122]. Only particles selected (BOXER) from micrographs demonstrating no visible signs of drift and minimal astigmatism were used in the reconstruction. Images demonstrating sample that is mono-disperse were targeted, and those demonstrating too high a sample concentration were avoided. This program displays two windows when in use; one for the entire micrograph, and another boxed multiple times that is updated once a selection (particle) has been made on the micrograph. EM datasets were boxed multiple times, to include more particles and remove aggregated particles. Individual micrographs were visually inspected

to ensure the CTF was concentric; all images not displaying this quality were not utilized in further processing. To save out the boxed particles, save boxed particles (IMAGIC format .img) for further EMAN processing and save box DB (Coordinate file) for re-boxing purposes were both selected. The defocus of each micrograph was determined by *CTFit*, verified and applied by *e2ctf.py*. *Proc2d* was utilized to append all the particles from each respective dataset (Cryo-EM) and generate a start file. The start file was then used to generate reference-free 2-D class averages from the particles in the dataset; *refine2d.py or* e2refine2d.py (negative stain) was run using 8-10 iterative cycles and a number of classes to contain approximately 30-40 images representing a class. The program will compare the similarity of a particle to others in the stack, and then place into an appropriate class based on cross-common line schemes. Particles were centered using the *Cenalignint* command or within the *e2refine2d.py* command. Bad class averages were manually removed and the good classes extracted for usage in the refinement steps. Phase flipping was applied during the reconstruction to correct for the contrast transfer function (CTF).

2.3.2.2 Initial Models and Reconstruction

The reference-free initial model was generated using EMAN2.1 *e2initialmodel.py*, using the processed stack from the previous steps. Multiple initial models were utilized for the refinement procedure. Maps were calculated from the dataset itself, or the available crystal structure with ANTRX 2 VWA domain (filtered to a resolution of 20 Å). The maps were iteratively refined using EMAN2.1 using the processed stack coupled with the respective initial model using a 90% classification selection and 30-70% of the classes to be used in generating the next model. In this step, once again particles can be eliminated from the reconstruction process, allowing only a percentage of the class averages to be included in the model generation. In each cycle of refinement, the particle center and the orientation information of the projected image were cross-correlated with projections (Fig 2.3) of the current 3D

model and subsequently utilized for the next round of refinement. The process was repeated over various cycles in an effort to improve the structural features that could be distinguished. The final 3D reconstruction for each dataset was computed from a number of particles listed in table 2.1. Fourier shell correlation analysis was used to assess the quality and convergence of the refinement cycles. The resolution of each map (Table 2.1) was determined Fourier shell correlation (FSC) at a cutoff of 0.5 (EMAN2.1). Cryo-EM maps were docked with the X-ray crystal structure of PA complexed with Anthrax Toxin Human Receptor VWA (PDB 1TZN), using UCSF Chimera [123]. All illustrations of structures were rendered using UCSF Chimera.

2.3.2.3 Docking of Atomic Structure Into Electron Density Maps

The docking of the atomic models was either performed by visual inspection and or utilizing the "Fit in map" function from Chimera under the menu of Tool/Volume/Data, which has the capability to fit the atomic coordinates into a density map, or one map into another for comparison [124].

2.3.2.4 Tables and Figures

Table 2.1 Dataset Specifics of Reconstructions.

Dataset	Pixel Size (A/pix)	Magnification	Defocus range	Micrographs taken	Micrographs used	Total Particles	Total Particles used
Cryo-EM TF_WT	1.97	74627	0.8-3.5	465	206	52186	29490
Cryo-EM TF_MUT	1.97	76074	0.87-2.236	272	83	32808	15134
Cryo-EM WT	1.97	74627	1.026-3.38	345	90	28484	10104
Negative Stain NS TF_WT	1.02	146785	0.3-1.45	31	30	10582	7995
NS_ TF_MUT	1.02	146785	0.72-1.59	76	71	16947	10972
NS_WT	1.02	146785	0.42-1.48	91	76	16901	11932

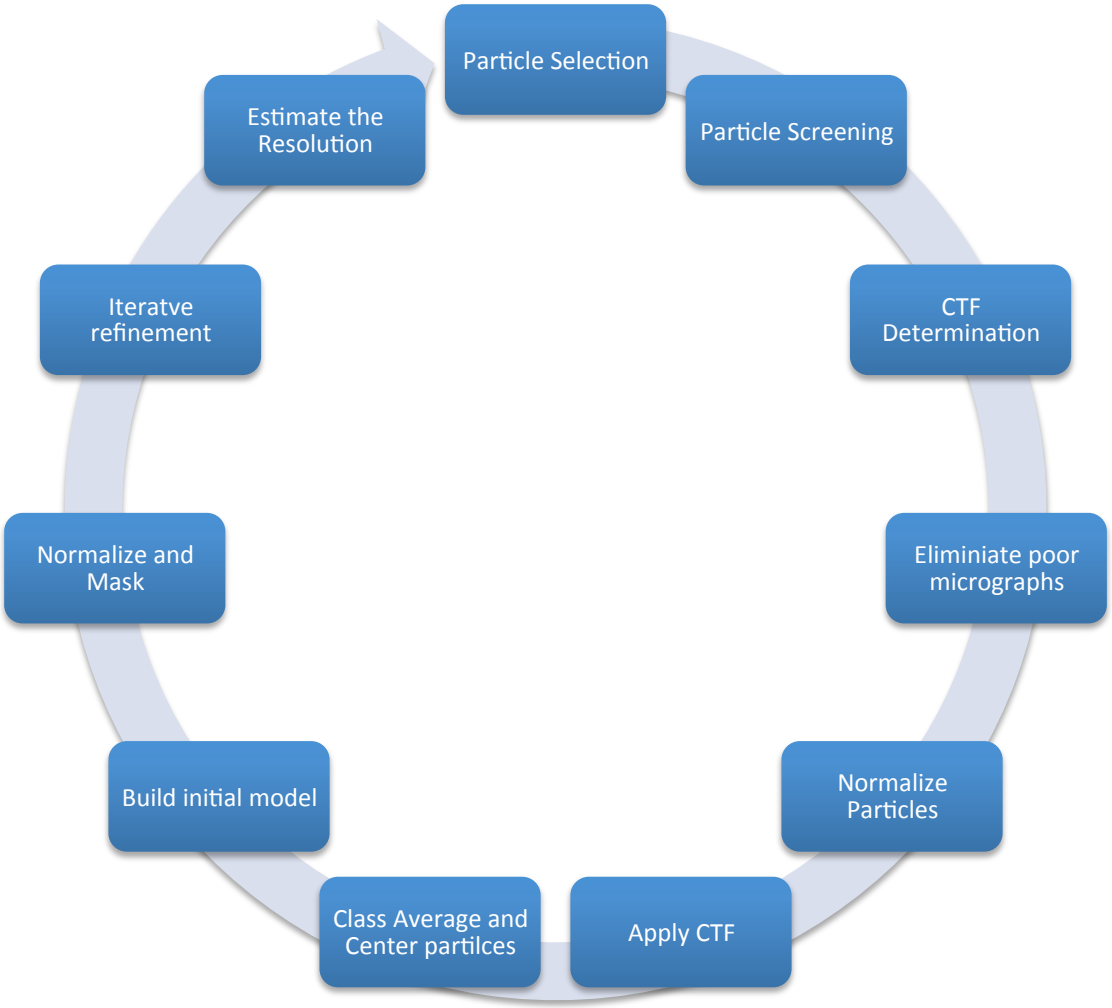

Fig 2.2 Schematic diagram of the EMAN reconstruction process.

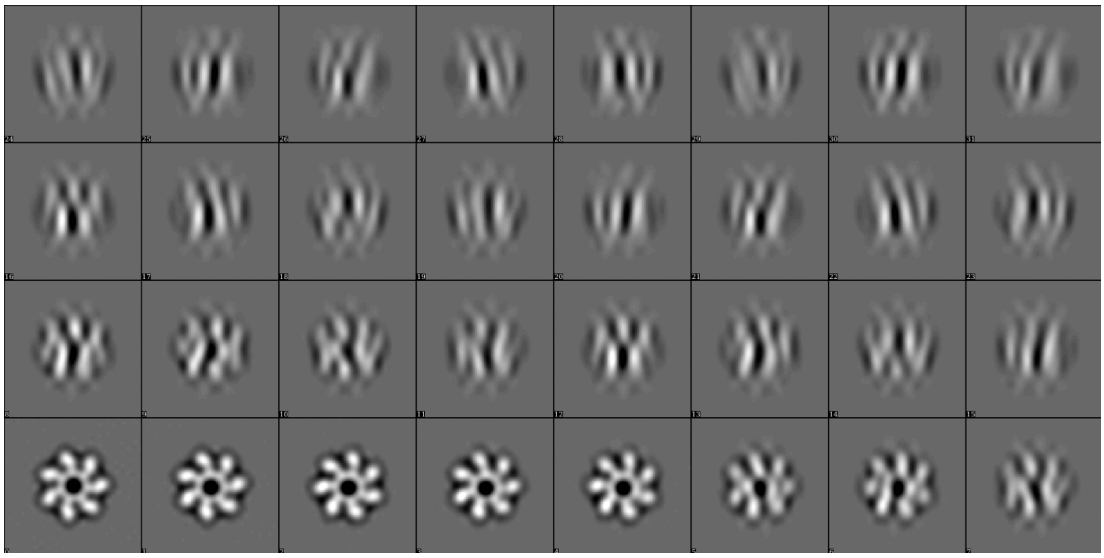

Fig 2.3 Projections from EMAN2.1 generated during iterative reconstruction.

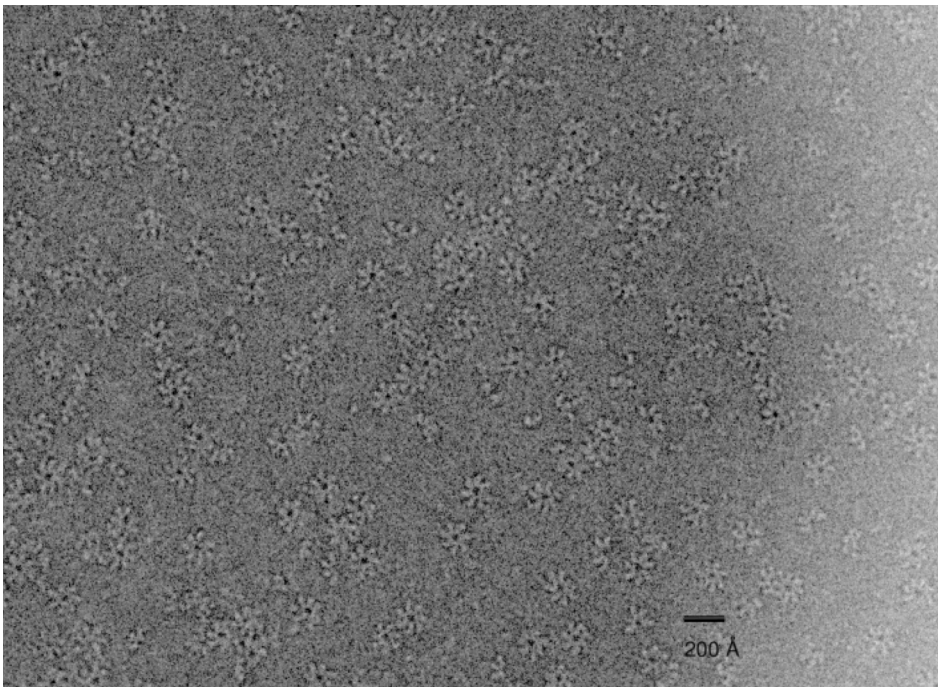

Fig 2.4 Raw image of an EM micrograph depicting disperse particles. A 200 A scale bar is in the bottom right corner.

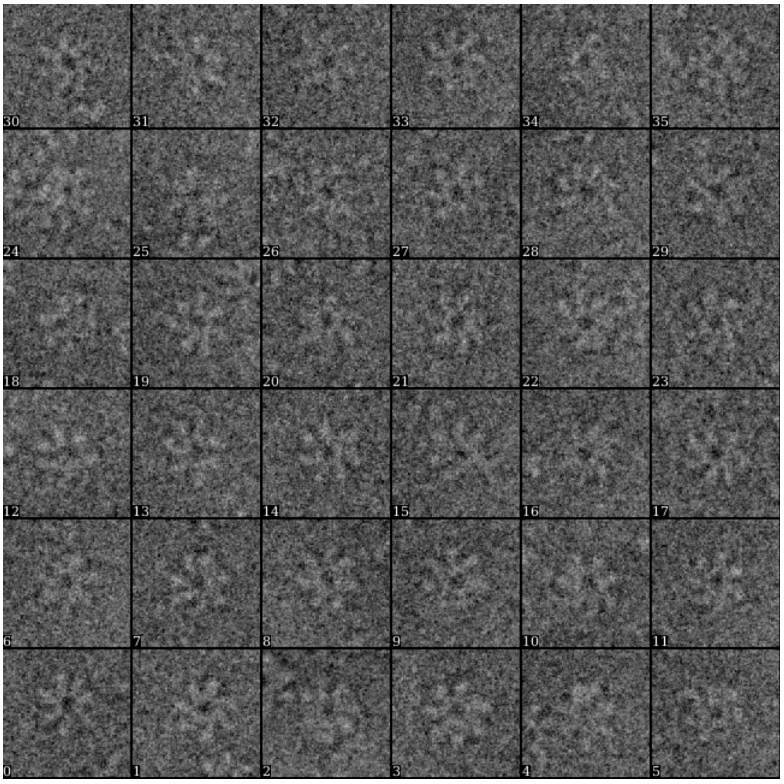

Fig. 2.5 Particles that have been separated from the micrograph into individual "boxes".

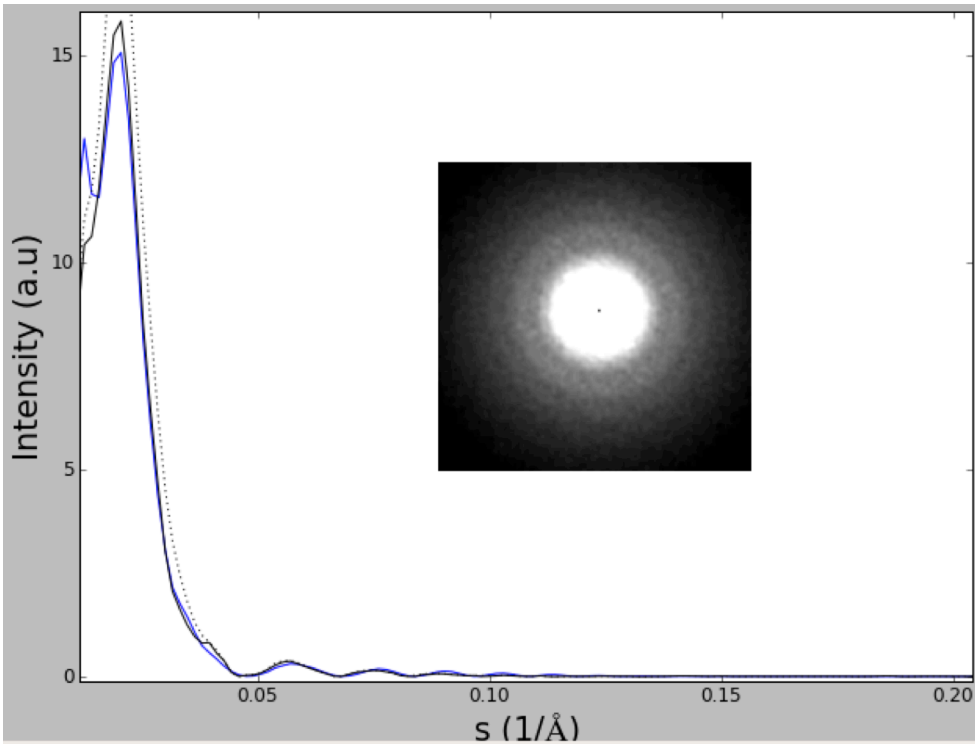

Fig 2.6 2-D CTF power spectrum of a micrograph, and the fit line to correct the CTF.

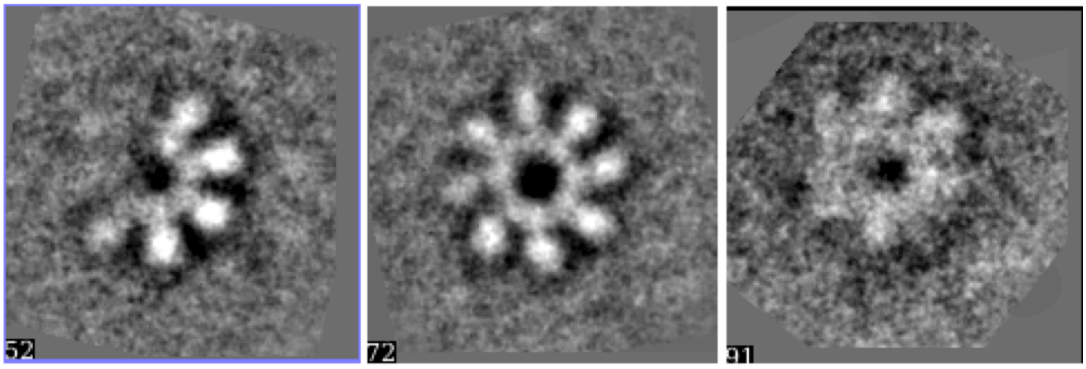

Fig 2.7 Classes demonstrating broken particles, octomers and hexamers when performing 2-D classification.

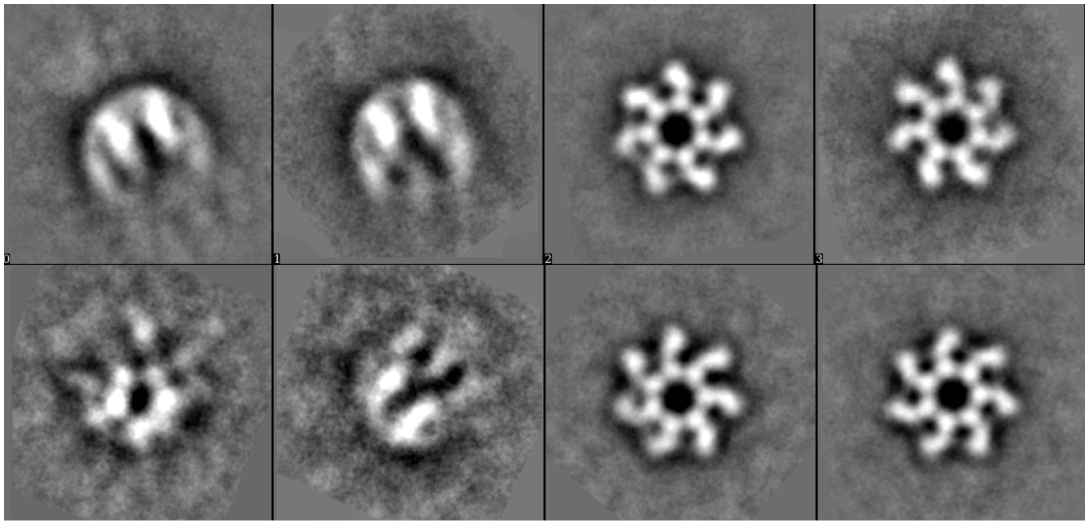

Fig 2.8 Negative Stain class averages demonstrating top, side and tilt views.

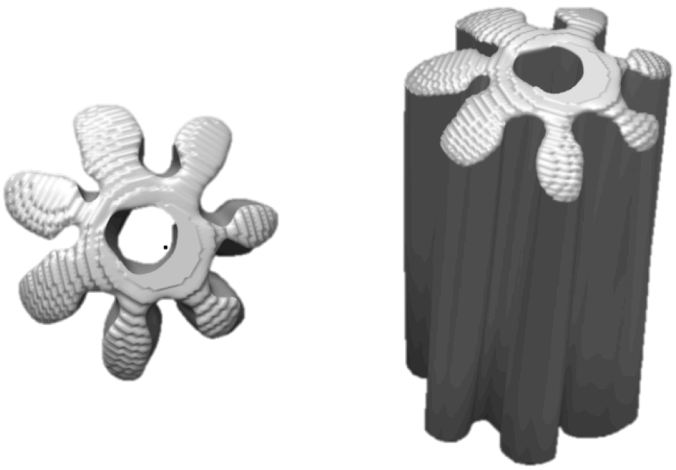

Fig 2.9 Initial models from cryo-EM reconstructions (EMAN1). Top and side views from left to right.

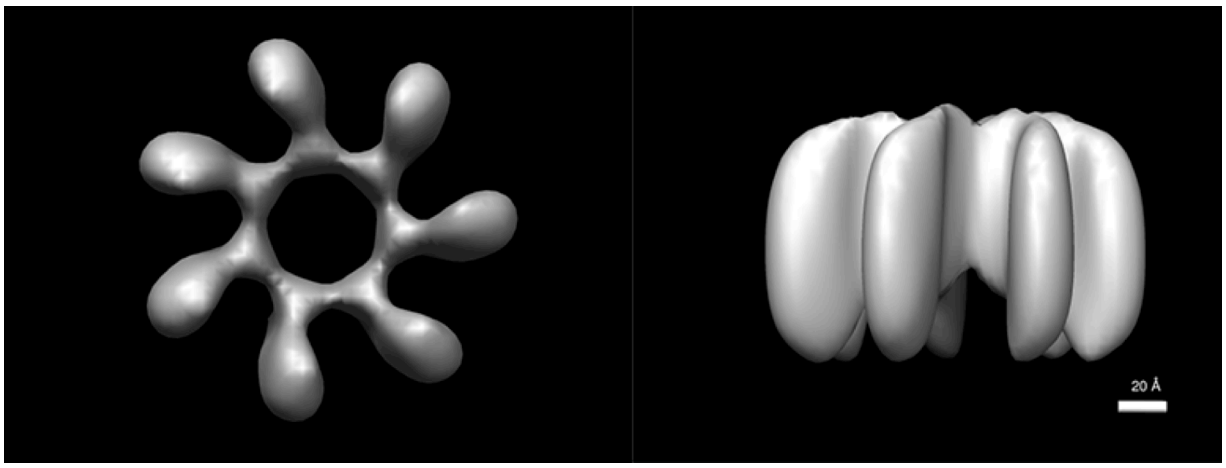

Fig 2.10 Initial models from Negative stain EM reconstructions (EMAN2.1). Top and side views from left to right.

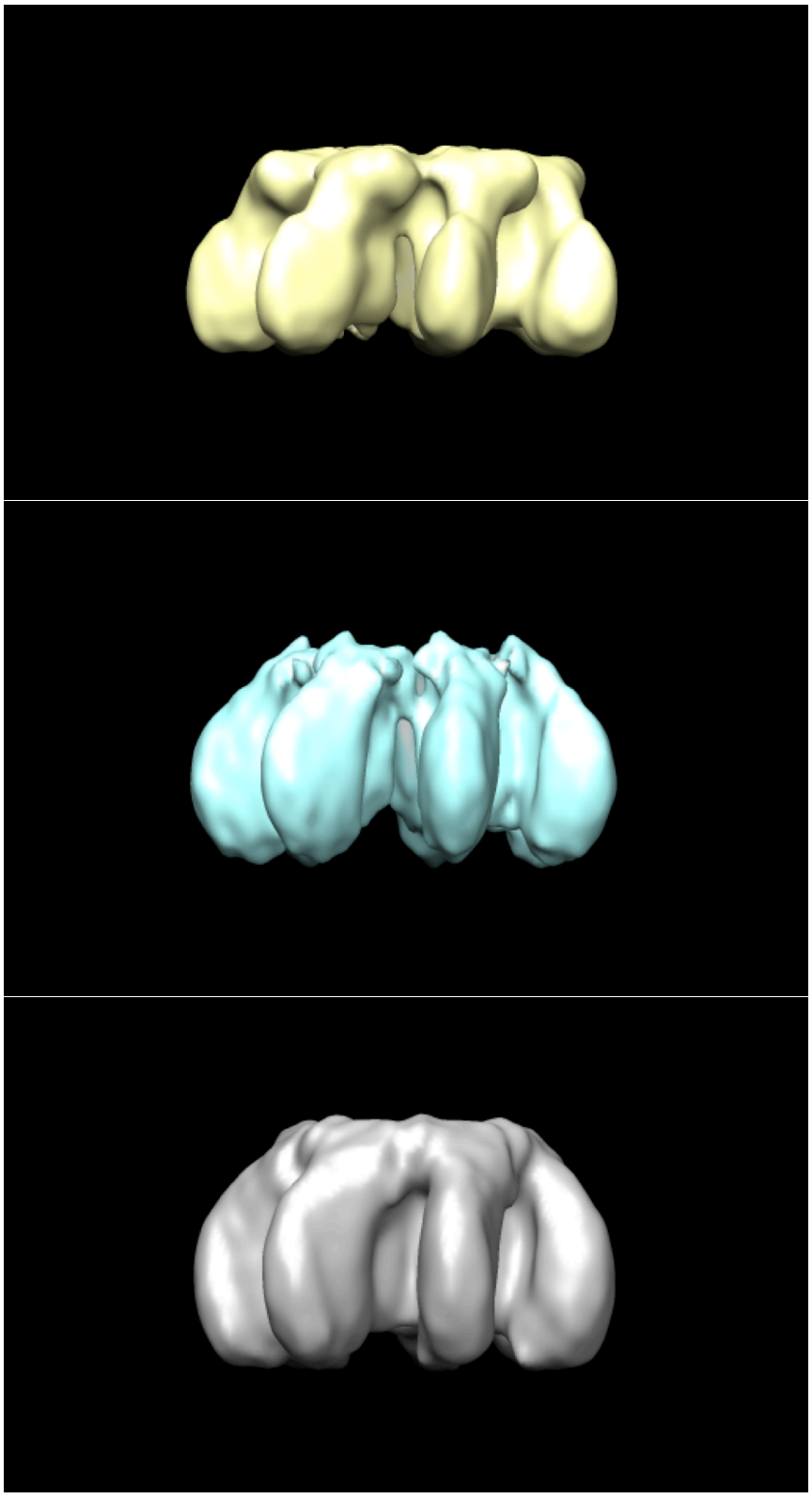

Fig 2.11 Cryo-EM reconstructions of Anthrax toxin receptor complexes using EMAN2.1, from top to bottom TF-WT, TF-MUT, and WT.

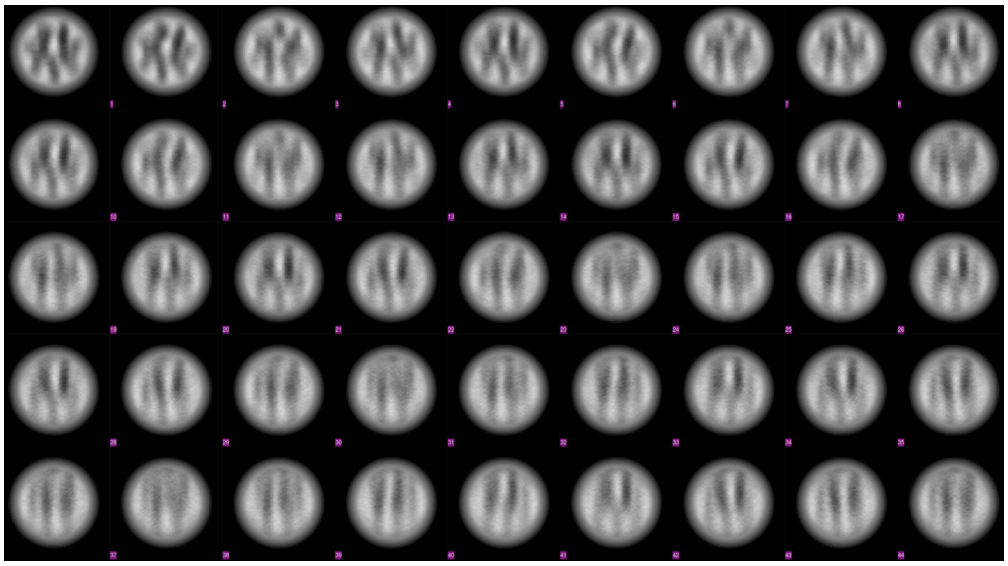

Fig 2.12 Side view projections from reconstructed map to be used as a reference for particle selection.

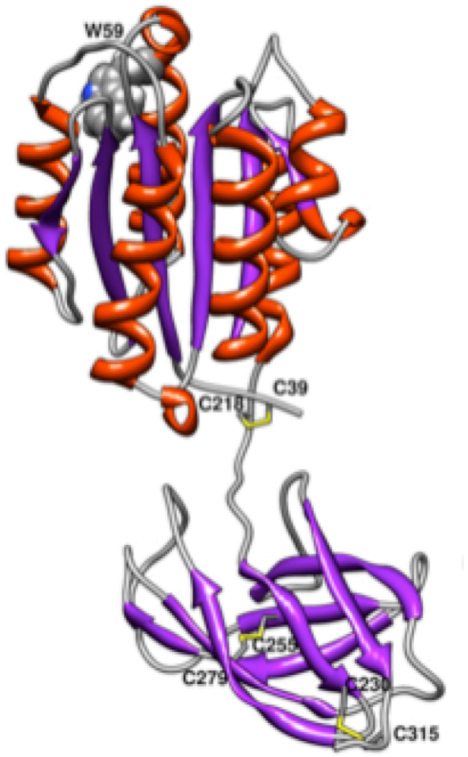

Fig 2.14 Crystal structure of VWA domain of the Anthrax toxin receptor with homology modeled Ig-like domain. Secondary elements have been colored as red (helix), green (beta-strand), and blue (coil). Disulfide bonds denoted at C279-C255 and C230-C315.

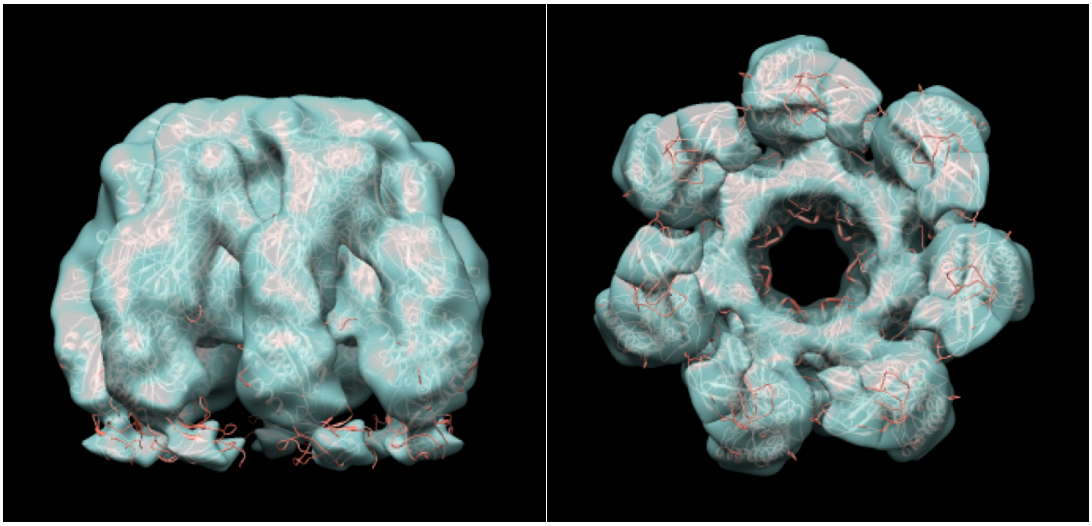

Fig 2.14 NS-WT reconstruction superimposed on the PA-VWA receptor complex. The second modeled domain (Ig-like) was located at the base of the structure.

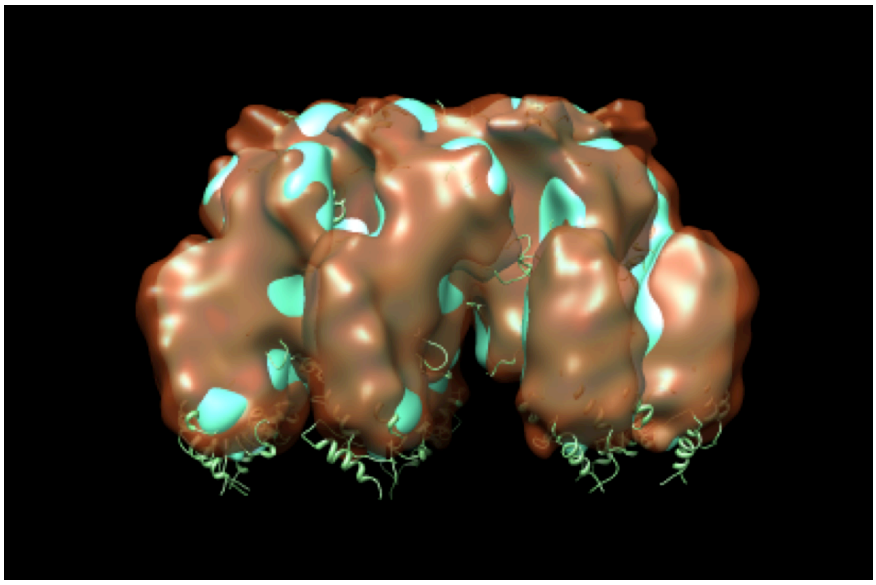

Fig 2.15 Negative stain reconstructions of TF-MUT (Green) and TF-WT (Orange) superimposed on the available PA-VWA receptor complex crystal structure.

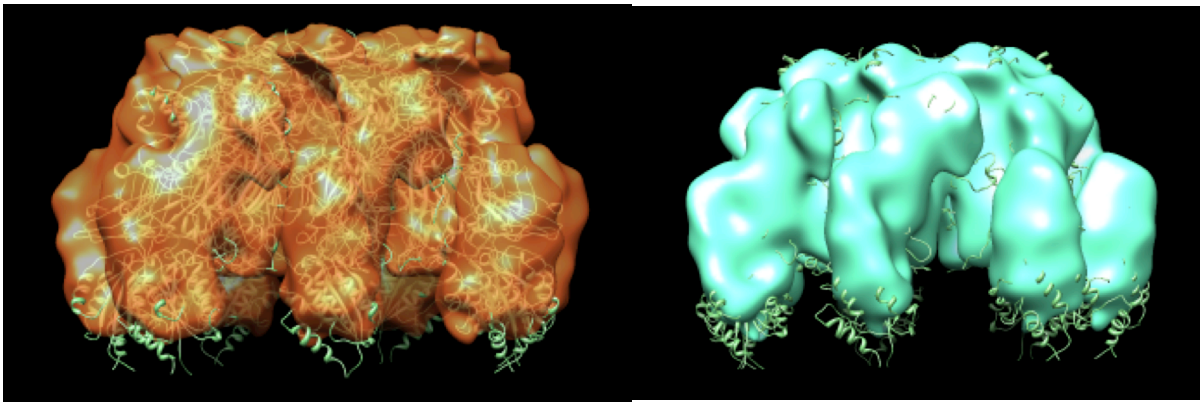

Fig 2.16 Negative stain reconstructions of TF-WT (Orange) and TF-MUT (Green) superimposed on the available PA-VWA receptor complex crystal structure. TF-WT clearly has more density where the VWA domain of the receptor is located as compared to the disulfide TF-MUT.

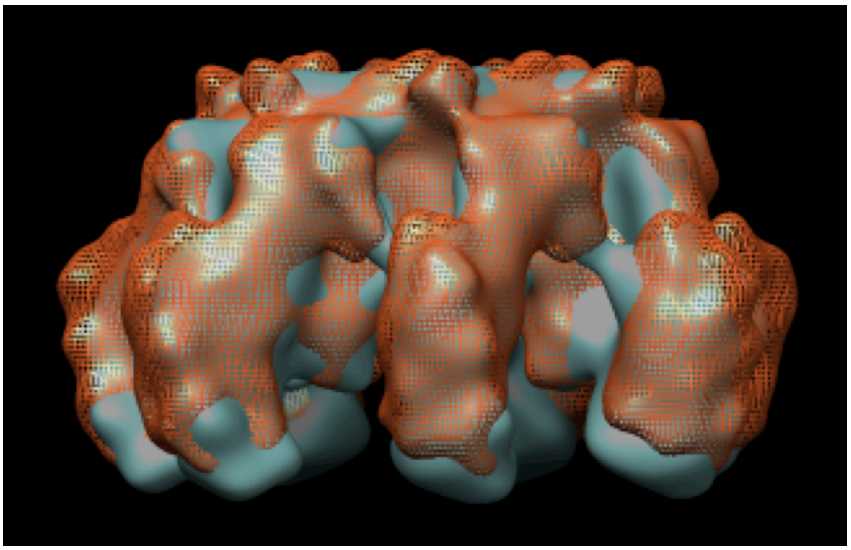

Fig 2.17 Negative stain reconstructions of TF-WT (orange mesh) superimposed on the WT (light blue). The density correspondent to the location of VWA domain of the receptor match well, as compared to the TF-MUT.

2.4 Results and Discussion

2.4.1 Data Processing

All data processing steps were performed on three multicore workstations (Xi Computer Corp.). Initially BOXER from EMAN1 was utilized to individually select out particles (Fig 2.4-2.5) from micrographs into boxes of 160 X 160 square pixels, with a pixel size of 1.97 Å. Images demonstrating too much aggregation and/or ice contamination were eliminated from further calculations. Particles were normalized and CTF defocus (Fig 2.6) calculations utilizing CTFit were performed. Manual screening of micrographs' CTF were analyzed to ensure it was concentric and not astigmatic. This assessment would typically eliminate more than 30 percent of the original data collection. The resulting

corrected particles were appended into a single stack file "start.hed" for further processing. Results from Cryo-EM datasets encountered many problems in the initial particle selection stages. It was observed that the majority of densities boxed out were predominantly top views. It was very difficult to distinguish side views from damaged particles or other artifacts. Heterogeneity of the sample could be observed when generating 2-D class averages, demonstrating octomers and hexamers and broken particles (Fig 2.7). Bad classifications were removed and the new stack was utilized to center and align the images (Cenalignint). Reclassification was once again repeated. The same processes were carried out with Negative stain samples, however the EMAN2 software suite was utilized. Particle orientations were more uniform, albeit, viable particles were significantly reduced.

2.4.2 Initial Model Building and Iterative Refinement

The 10-20 best representations of top, side, and tilt views were utilized to generate an initial model (Fig 2.8). Initial models using EMAN1 software would always generate an elongated cylindrical model that would be refined with the dataset (Fig 2.9), and then used to execute iterative refinement cycles with the entire datasets. The exact premise for model generation was utilized for the datasets utilizing EMAN2.1, however initial models were much closer in shape (Fig 2.10) to the PA/ANTXR 2 crystal structure. Collectively, reconstructed models in the cryo-EM datasets had an elongated side (Fig 2.11) view, suggesting lack of side views represented in the actual density map. Side view projections (Fig 2.12) were generated as a visual guide to conduct re-boxing, building a new initial model and refinement. This only marginally improved the quality of the data, and the density maps all appeared featureless when viewing from the side views (Fig 2.11). Multiple model refinement cycles were computed, and three-dimensional density maps with resolutions at 13.3, 11, and 13.1 (FSC criterion of 0.5) for the TF-WT, TF-MUT, and WT respectively. All reconstructed models fit the available crystal structure of PA with VWA of the ANTXR 2 receptor. Key differences from the reconstructions of WT

and TF-MUT were observed at the interface between domain 4 of PA and VWA relative to the actual conformation at the base of the complex (Fig 2.11). The WT reconstructions (TF-WT and WT) were longer when compared to the TF-MUT along the side view. Furthermore, the localization of VWA at the base of the PA, representing ANTXR 2 VWA maintains an inward type curvature towards the inside of complex for the TF-MUT. In the WT reconstructions, this density was extended outward and the contour of the density was more prominent as compared to the TF-MUT (Fig 2.11).

The negative stain (NS) reconstruction was able to locate the Ig-like domain in the PA/ANTXR2 complex from the NS-WT reconstruction, and from the NS-TF-MUT it was possible to see conformational differences in the VWA domain of the ANTXR 2 receptor. In silico analyses of domain 2 or the Ig-like domain of the ANTXR 2 receptor have been reported to form an Ig fold [112, 125], however the position within the complex and structure remain unknown. Multiple iterative refinements were calculated and a low resolution ~14 Å NS-WT was generated (Fig 2.13). The Ig-like domain of ANTXR 2 was also homology modeled (I-TASSER, Fig 2.14) and subsequently fit into the density map (Fig 2.13). The modeled domain fit the density at the base of the NS-WT complex, which is just beneath the first domain of the ANTXR 2. An important observation pertinent to the density is that it is relatively weak when taking the molecular weight into account. The density only appeared after the standard deviation was one deviation above the average, whereas the remainder of the map fit the available crystal structure quite well at three and a half deviations above the normal. This could suggest that the Ig-like domain is very flexible in solution as opposed to the PA and the VWA domain of ANTXR 2. To that end, the structural model did demonstrate a connecting hinge region between domains one and two of the receptor, which would allow for flexibility between the domains. Biochemical analyses performed previously (Dr. Sun) have demonstrated no effect on PA pore formation with the addition of the TF tag, and in terms of data collection appeared to improve the stability, contrast, and orientation of the particles within the grid. This however, did not assist in

locating the Ig-like domain in the reconstructed maps, as it appeared that both the TF and Ig-like domain flexibility made it difficult to map those regions. There is a long linker region between the TF tag and Ig-like domain, which could account for the flexibility. Density maps are shown at a contour level three and a half standard deviations above the average, in order to remove the unstructured densities below VWA of the ANTXR 2.

2.4.2.1 DOCKING OF ATOMIC STRUCTURE INTO ELECTRON DENSITY MAPS

The overall shape of all cryo and negative stained reconstruction maps matched that of the PA-VWA-D1 heptamer crystal structure (PDB 1TZN). After manually translocating the PDB coordinates into the map roughly by visual inspection, the Chimera "fit in map" tool can dock the structure with numeric calculations with a much higher accuracy. Particularly, the PA structure docks into all the maps very well, indicating no major conformational changes occurred in that area. The differences observed do appear between ANTXR2 VWA and Ig-like domains.

Given that all the reconstruction contain the heptameric PA part, in the discussion below, the complexes will be named according to the different ANTXR2 components. For instance, the reconstruction of TF-WT means the reconstruction of heptameric PA complexed with wild type ANTXR2 and a TF tag at the N-terminus of the receptor. Only negatively stained WT reconstruction shows partial density of ANTXR2 Ig-like domain (Fig 2.13), indicating the hinge region between the VWA domain and Ig-like domain are very flexible. This hinge region also is where the N-terminus of VWA domain located due to a disulfide bond linking the N- and C-termini of the VWA domain. Therefore, when the TF tag is linked to the same hinge region, the TF will interfere with the already flexible Ig-like domain, preventing it from being visible in all TF-tagged reconstructions that used symmetry averaging. Ig-like domain structure was homology modeled and docked into the partial density observed in the WT reconstruction. No literature has previously reported the position of the Ig-like domain and this is the first time that the Ig-like domain's relative location to the VWA domain is determined.

Comparison between negatively stained TF-WT and TF-MUT reconstructions show that the TF-MUT reconstruction has much less density where the VWA domain is located. The difference can potentially be attributed to the double disulfide mutations in the Ig-like domain, thus affecting the VWA domain's folding, leading to a more flexible and/or partially unstructured VWA domain. In order to assess whether the structural differences observed between TF-WT and TF-MUT were not attributable to the TF tag, negatively stained TF-WT reconstruction was superimposed on that of WT (Fig 2.17). No obvious differences in the densities correspondent to the VWA domain of ANTXR 2 were observed. This further demonstrates that the TF tag plays no role in affecting the conformations of PA and VWA domain of ANTXR2.

2.4.3 DISCUSSION

The cryo-EM datasets did not generate models with useful structural information. The reasons for the poor data quality are as follows. One caveat of EM reconstructions is the potential for sample heterogeneity. One of the major expectations when collecting cryo-EM data is that the macromolecules are "identical" in structure. Meaning sample heterogeneity [126-128] is not observed, as was the case in our reconstruction. All densities pertaining to a plausible particle were selected, and when 2-D class averaging was performed, we found hexameric and octameric complexes as well as broken particles. All of these factors could have contributed to the lack of structural details from the cryo-EM datasets. It also becomes challenging to produce a monodisperse specimen for cryo-EM data collection, relegating you to add components that could drastically reduce image contrast. To this end, cryo-EM intrinsically has poor contrast, which can make it difficult to discern a good particle from a denatured or broken particle. If most of the particles are not representative of the intact target structure, then you can expect the final model to be representative of that poor data. From our data collection utilizing cryo-EM, it was observed that many of the particles were mixed quite ubiquitously with these unwanted particles. To solve these issues, the complexed macromolecular assembly needs to be freshly purified and

immediately used for data collection. This alone will not completely remove the sample heterogeneity, however programs in the data processing packages will be utilized to remove those particles from the reconstruction. Another major issue with single particle reconstructions is the possibility that the macromolecular complex maintains a preferred orientation [129, 130] on the grid. As the name implies, this happens when the molecule absorbs on the carbon grid in one or a few orientations. This was one of the issues seen in our cryo-EM datasets, in which the top "flower" shape was predominant. Therefore, our cryo-EM data was unable to reach higher resolutions was the lack of side views, which could account for the elongation observed in the models. Modifying the method of data collection by tilting the specimen can solve this; however, that introduces new problems of electrical charges on the vitreous ice thus leading to image drift [131]. Another solution could be to utilize detergents or vesicles to disrupt the orientations with the grid. To obtain a high-resolution reconstruction, implementation of these solutions will be carried out and new data will be collected.

The negative stain reconstructed models showed more structural features, namely the visibility of the Ig-like domain in NS-WT. This might be attributed to the higher contrast inherent to negative stain EM technique. The negative stain datasets also contained more side views as compared to the cryo-EM datasets, perhaps due to two possible reasons. One is the fact that it was easier to locate side views due to the high contrast of negative stain. Secondly, the supporting carbon film might orient more complexes towards side views. As a result, not only the Ig-like domain of the ANTXR 2 receptor was visible, the negative stain reconstructions denote key structural differences in the VWA domains of NS-TF-WT and NS-TF-MUT related to the disulfide bond mutation in domain two of ANTXR 2. The fact that toxin translocation is inhibited as shown in biochemical data (Jacquez, P. et. al.) [113], it is plausible that the C255-C279 disulfide bond within the Ig fold maintains a pivotal role in the stabilization of the Ig domain. When this critical interaction is abridged, conformational or even unfolding events may occur in the Ig-like domain. Consequently, this structural change could potentially alter the functioning of the

PA. Moreover, based on negative stain reconstruction of PA with GroEL [132] and lipid nanodiscs [133], there is a 50-60 Å gap that is between the membrane and the cap of the PA. The height of our modeled receptor VWA plus Ig-like domains (NS-WT) is 60 Å. It could be the case that the receptor ANTXR2 is still associated with the complex after pore formation, surrounding the pore to stabilize the PA. Previous studies have shown a critical loop near residue F427 in domain 4 of PA involved in both pore formation and translocation [134, 135]. Docking of crystal structures into negative stain EM reconstructions allowed the positioning of VWA domain at the bottom of the PA pore cap within reach of F427 flexible loop. Previous electron paramagnetic measurements of a spin probe attached to Cys in PA (F427C) suggested rearrangement of the residue towards the 7-fold axis in pre-pore to pore formation [135]. This could likely implicate the VWA domain influences the function of PA pore through the interaction with residues positioned outside the F427 loop. Therefore, if the double disulfide mutant alters the conformation of the VWA domain, efficient delivery of the toxin may be impinged. In summary, the negative stain reconstructions were able to provide models with more structural features, which may provide key insights to the potential mechanism that the Ig-like domain plays in the delivery of the toxin into the host.

2.4.4 Summary and Future Directions

Bio terroristic acts in 2001 brought Anthrax research to the forefront of scientific research looking for novel treatments and potential cures from the infection. In this research, the main objective was to determine the role of the Ig-like domain of ANTXR 2 in the progression towards toxin release into the afflicted cellular host. Previous research has focused mainly on the VWA domain of the human receptor (ANTXR 2); however, recent evidence suggests that the second domain (Ig-like) plays a pivotal role in the delivery of the toxin. As a result, structural studies utilizing transmission electron microscopy were performed. Results obtained from negative stain electron microscopy located the

position of the Ig-like domain and demonstrated conformational differences in the VWA domain of the receptor complexes between the wild type and mutant. Cryo-EM datasets presented multiple issues, therefore, future work will aim to fix the orientation issues and obtain more side views using alternatives discussed above. In addition, recent EM studies of the PA pore structure were reconstructed bound to GroEL [132], yielding more side views of the pore formation. Future cryo-EM data could utilize GroEL to perhaps reduce the preferred orientation and yield more side views. Also, given the added density, the contrast of the particles will be improved and make it a little easier for particle selection. Furthermore, the PA pore has also been placed into nanodiscs and lipid vesicles [133, 136, 137], which could also be targeted in our reconstruction efforts as well. It is unclear as to the exact molecular details about how the Ig-like domain interferes with PA pore formation, whether by direct or indirect contact with the PA pore. Therefore, obtaining higher resolution models with more structural features will be imperative in deciphering the mechanism by which the Ig-like domain disulfide bond modulates PA pore function. Understanding the molecular details of the toxin delivery mechanism will aid in the development of novel therapeutics and potential treatments against the lethal Anthrax disease.

CHAPTER 3:

STRUCTURAL INVESTIGATIONS OF CORE CIRCADIAN REGULATORY COMPONENTS

3.1 INTRODUCTION

Approximately 50 to 70 million Americans suffer from sleeping disorders, which debilitate overall health and longevity [138]. Some of the risk factors include *hypertension, diabetes, obesity, depression, heart attack, and stroke* [138]. Much of these maladies are attributed to disruption of the circadian rhythm. Circadian rhythm is an intrinsic, roughly 24 hour timing mechanism whereupon oscillations of biochemical, physiological, and behavioral processes are synchronized with environmental elements such as light, nutrients, and temperature [139]. In mammals, this master biological "clock" is embedded in the hypothalamus that has control over clocks in peripheral tissues [140]. Circadian rhythm is found in most living organisms ranging from cyanobacteria to humans, and is a vital component for organisms to adjust their metabolism so as to use energy efficiently [139]. One example of circadian disruption in humans would be demonstrated by traversing various time zones after a long distance flight, resulting in what is known as "jet lag". Thus, resetting the "clock" or reducing the disruption period can provide benefits for patients with sleep disorders and for persons in high-stress occupations such as medical staff, long distance drivers, pilots and soldiers who need to maintain alertness under prolonged dangerous situations. For instance, accidents can be reduced if the circadian disruption can be modulated to decrease fatigue [141]. The strategy in developing new therapeutics requires a deeper understanding of the mechanisms in modulating the circadian rhythm, which has been investigated for over two decades, albeit the precise molecular mechanism by which the central components are regulated has yet to be determined. This can be partially attributed to the lack of structural information of core circadian regulatory molecules. In this proposal, structural biochemical techniques such as X-ray crystallography combined with traditional biochemical and biophysical

methods will be applied to investigate one core regulatory component, Cryptochrome (CRY), in the mammalian circadian clockwork. Cryptochromes are a set of blue-light sensitive flavoproteins, which have been found in plants, bacteria and animals [142-144]. Information obtained will challenge hypotheses about CRY and its usage of FAD as a cofactor, and will benefit the development of new therapeutic agents targeting circadian-related diseases [142-144].

3.2 BACKGROUND AND SIGNIFICANCE

Circadian related genes have been investigated since 1984, with the first observed genes discovered in drosophila mutants, then later in mammals [142-144]. It was found that the circadian rhythm was driven by a transcription-translational loop within the cell (Fig. 3.1). It has been shown [145] that two transcriptional factors CLOCK (Circadian Locomotor Output Cycle Kaput) and BMAL1 (brain and muscle aryl hydrocarbon receptor nuclear translocator like-1) form a heterodimer and bind a DNA transcriptional element known as E-box found in the promoter region of many circadian genes, including CRYs [143, 146-148]. The transcriptional activation complex serves as positive elements in establishing the circadian rhythm, and can be manipulated by most clock modifiers. For instance, once CRY accumulates in the cytoplasm (reaches a certain threshold), it will down-regulate its own production by inhibiting the transcriptional activity of the aforementioned activation complex. This will generate an inherent time-delayed negative feedback loop, controlling the rhythmic expression of various clock-controlled genes[149] (Fig. 3.1).

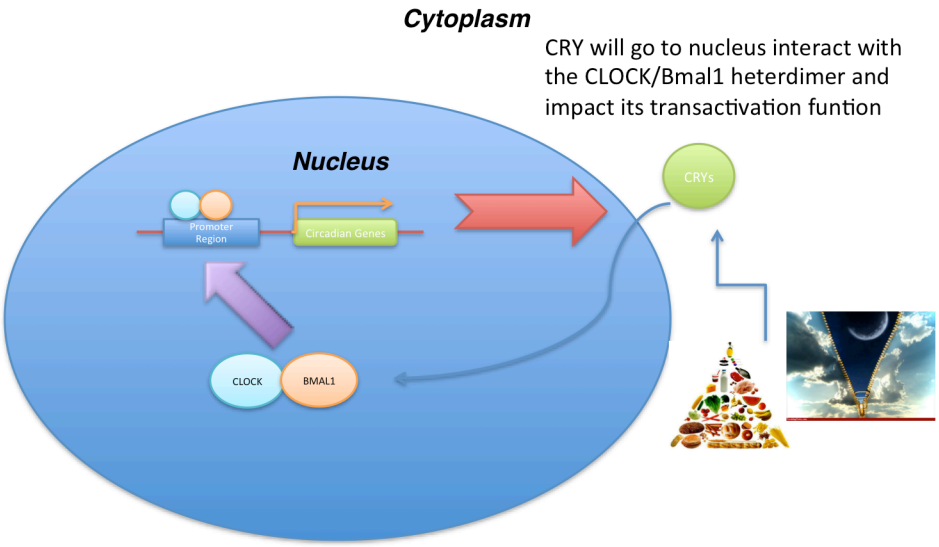

Fig 3.1 A schematic diagram of the transcriptional/translational negative feedback loop governing the circadian rhythm.

Cryptochromes are a class of photoreceptors maintaining the ability to modify the circadian rhythm in almost every species [150]. CRY's are multifunctional proteins, which have been determined to play viable roles in signal transduction important for growth, development, and magnetosensitivity[151-153]. Interestingly, CRYs are homologous to photolyases (PLs), both of which are classified in the same family of flavoprotein photoreceptors that mediate responses to ultraviolet and blue light exposure. Despite maintaining structurally conserved homology between species, CRYs and PLs maintain distinct functions[154]. PLs utilize visible light to repair DNA that has incurred ultraviolet light damage. CRYs on the other hand, act as blue-light photoreceptors in circadian systems [155]. To date, it is known that CRYs coupled with PLs preserve common structural motifs, the use of Flavins and intrinsic photoactivation mechanisms[156, 157]. It is possible to further separate animal CRYs into two distinct subclasses, which are based on phylogeny and dependence upon light. The CRY's that pertain to the type I subfamily pose as photoreceptors, and magnetoreceptors; whereas type II CRY's

(mammals) are non-light dependent and play major roles in repressing the negative feedback loop at the level of transcription. [158]

The differences that have been scrutinized between members of the CRY family types I and II are relegated to the substrates in which are employed and the manner in which activation mechanisms are brought forth. It remains a mystery how animal CRYs synchronize the circadian rhythm and whether this regulation requires any ligands. In 2011, the full-length atomic structure of drosophila CRY (dCRY) was determined using X-ray crystallography. It is the first animal CRY whose structure has been determined. This structure is composed of multiple alpha helices with many loops and approximately 13 beta sheet structures. dCRY also includes two domains: the amino-terminal photolyase homology domain (PHD), and an FAD binding domain conserved from photolyase. The structure has an additional C-terminal tail (Fig. 2), which has been proven to be a nuisance in obtaining the structure of mammalian CRYs.

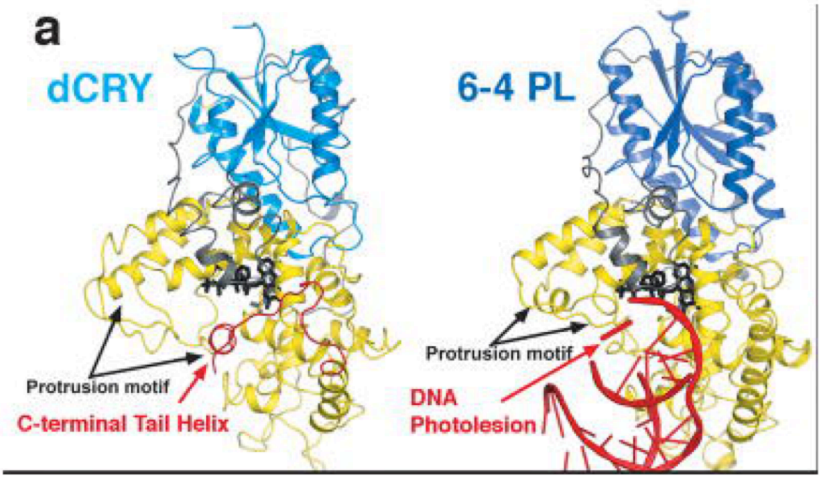

Fig 3.2 dCRY and PL demonstrating variable C-terminal CTT found in CRY's. Adapted from Zoltowski, 2011

Given the lack of structural information pertaining to mammalian circadian core components, scientists have based their assumptions utilizing the dCRY model to predict ligand recognition and

attempt to decipher the manner in which they interaction with other clock-controlled genes such as CLOCK and BMAL1. Zoltowski, et al observed that the composition of dCRYs C-terminal tail (type I) and that of PLs is conserved and serine rich, whereas type II CRYs have three Cysteine residues that are postulated to bind close to the active site of type II CRYs [155]. This could suggest a different purpose for the active site, which would shed light on human CRYs function. The variable C-terminal tail of dCRY was shown in Fig 3.2 to emulate the DNA binding position observed in PLs. Even though sequence homology is not preserved and DNA repair mechanisms are not undertaken by CRYs. The C-terminal region does however pose a physiologically important effector role in CRYs[159], namely could play a key role in the interaction with nucleic acids, flavin cofactors such as FAD. In 2013, Xing, et al compared the crystal structures of mammalian CRY2 in its free form, FAD bound, and in complex with ubiquitin ligases; suggesting regulation via ubiquitin degradation pathway [160]. Furthermore, it was observed that FAD binding to mCRY2 was exposed rather than buried within the binding pocket such as drosophila (6-4) PL (Fig 3.3). This could explain why expression models of human CRYs have shown to exhibit low levels of flavin binding [161-163] which could suggest the potential usage of the catalytic core for other functions as opposed to enzymatic functions such as DNA repair as seen in PLs.

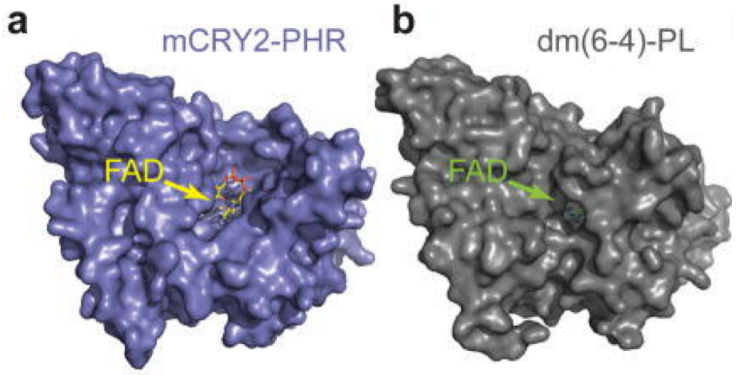

Fig. 3.3 Density map representation of mCRY2 compared to drosophila photolyase showing the comformation of the FAD binding pocket. Adapted from Xing, 2013

3.2.1 SIGNIFICANCE

Determining the crystal structure of the mammalian CRY will deepen our knowledge about the functions of CRYs. Much of what is known about the function of mammalian CRYs has been based on structures of related CRY models such as dCRY. However, the different mechanism used by type I CRYs (drosophila) to synchronize the circadian rhythm is light dependent and targets the protein "timeless" (TIM) for it subsequent degradation. Type II CRYs (mammals) conversely interact more directly with the heterodimer complex to inhibit its function, thus playing a more direct role in regulating the circadian rhythm [142-144]. Comparative protein sequence homology shows that dCRY and human Cry1 (hCRY1) only maintain approximately 40% overall homology, which could further indicate a distinctive function of hCRY1's variable C-terminal tail from that of dCRY. From the structure, it may be possible to predict the mechanism by which the protein functions and the manners in which it influences interacting components. By determining the full-length structure of hCRY1, it will be plausible to predict the manner in which it regulates the circadian rhythm and mechanisms by which it affects the transcriptional activation complex. This would shed light on the function of the catalytic core and the usage of the C-terminal tail not found in PL members.

3.2.2 LONG TERM GOALS AND SPECIFIC AIMS

As mentioned above, the full-length structure of mammalian CRY's have not been obtained. Moreover, the function of the Photolyase homology region (PHR) of hCRY1 is still unknown. The **long-term goal** of this research is to obtain a full-length atomic resolution structure that can provide molecular insights as to what these regions perform in the feedback cycle. Within this dissertation, recombinant hCRY1 has been cloned, expressed, and purification from bacterial expression system is being optimized. Protein obtained from optimized purification will be utilized for future structural

determination of Gam1 with or without the binding components. Two specific aims have been proposed in this research to test the *main hypothesis unlike dCRY (type I), hCRY1 (type II) utilize their Photolyase homology domain (PHD) catalytic core for DNA binding.*

In this research, two specific aims are proposed to test the hypothesis and examine the function of hCRY1.

Specific Aim One: To optimize purification hCRY1**.** The objective will be to obtain large quantities of pure protein for crystallization and for conducting biochemical and biophysical studies.

Specific Aim Two: To characterize the ligands binding to hCRY1. The pure protein obtained in specific aim one will be used to examine the binding of potential cofactors by various biochemical and biophysical methods.

The experimental aspects were designed in this matter so they could support or refute the hypothesis. Function can in part be explained with the structure, which requires a highly homogenous protein for crystallization carried out in Specific Aim One. Once the protein is crystallized and the atomic structure is determined, it will be possible to identify the two known structural domains. These domains display different functions as demonstrated in the structure from type I CRYs (dCRY), therefore the structure of hCRY1 can support or refute the stated hypothesis in part from the atomic structure if DNA is indeed found to be a ligand. Additionally, it will demonstrate any other type(s) of ligands they bind, and whether or not they serve common purposes from homologous structures. Also, the biochemical and biophysical experiments will corroborate the results from structural studies in terms of specific ligands namely DNA and their capacity to bind hCRY1.

3.3 MATERIALS AND METHODS

3.3.1 CLONING

The full-length gene sequence was: amplified via PCR, restriction endonuclease digested, and ligated into an expression vector(s) see table 3.1. During the PCR primer design, expression tags were added to either the N-terminal or C-terminus of hCRY1 gene. These tags are for affinity purification, enhancing solubility, and assisted folding. Following cloning, expression trials were carried out to optimize protein expression. See section 1.3.1 for detailed methods regarding directional cloning and into pCold-TF. hCRY1 was also obtained in a FLAG-tag (p3XFLAG-CMV-14) containing construct (Courtesy Dr. Erquan Zhang, Beijing School of Sciences) and also cloned into pMal-p4X with E. coli. maltose binding protein (MBP) at the N-terminus. The FLAG tag is a short 8 amino acid residue peptide utilized to purify via an Anti-FLAG antibody column. It has been utilized in a variety of expression systems from mammalian, yeast [164] and bacteria [165], and more importantly has been shown to not interfere with native protein function [162, 166]. Maltose binding protein has been used as a solubility enhancer when fused to proteins that are otherwise difficult to express [167-169]. Inducible expression with the tac promoter generates a fusion of MBP-hCRY1. Cloning was performed according to specification in table 3.1. E. coli strains as mentioned in section 1.3.1 were utilized, with the added strain BL21 Codon + (Agilent technologies). This strain similar to Rosetta2 (DE3)pLysS contain extra rare codons that are not as frequently utilized in bacterial expression [170].

3.3.1.1 EXPRESS TARGET PROTEIN:

Section 1.3.2 describes overlapping methods for hCRY1 expression, for more information reference to that section. For maltose binding protein construct culture was grown at 37°C agitating at 220 rpm until an OD_{600} of 0.5 is reached. Then the culture was immediately induced with 0.3 mM IPTG and incubated an additional three hours at 37°C agitating at 200 rpm. Culture was processed as

mentioned in section 1.3.2. For FLAG-tagged hCRY1, expression was carried out as mentioned in section 1.3.2.1.

3.3.2.2 COLD INDUCTION PROTEIN EXPRESSION

hCRY1 protein was produced in pCold-TF with a C and N-terminal His tag as detailed in section 1.3.2.2, however cultures were grown to an OD_{600} of 0.8, at which point they were removed from the incubator and maintained at room termperature induced with 0.1 or 1 mM IPTG for 30 minutes. Afterward, Culture was placed back into the incubator at 16°C for 12 hours with 200 rpm agitation. Cultures were harvested following the aforementioned section.

3.3.3 PROTEIN PURIFICATION TRIALS

3.3.3.1 WESTERN BLOT

10% acrylamide SDS-PAGE gels were transferred to polyvinylidene difluoride (PVDF) membranes utilizing the Bio-Rad Mini-PROTEAN 3 system and transfer buffer (25 mM Tris, 192 mM Glycine, and 20% methanol). Membranes were blocked overnight at 4°C with 5% nonfat dry milk reconstituted in TBST (50 mM Tris pH 7.4, 150 mM NaCl, 0.02% NaN_3, and 0.1% Tween-20). The next morning membrane was washed 3X with TBST and incubated with Anti-FLAG HRP conjugate (Courtesy of Dr. Sun; 1:1000) for 1 hour. Following incubation, membrane was washed as described above and processed with Immobilon Western Chemiluminescent HRP substrate (Millipore). After 5-minute incubation, image was exposed for 3-5 minutes and images were captured on X-ray films (Phenix Research Products).

3.3.4 Biochemical Characterization

3.3.4.1 DLS in hCRY1 Characterization

DLS of hCRY1 were collected following specifications mentioned in section 1.3.4.1. Briefly, dilute hCRY1 sample taken immediately following SEC fractionation was used for DLS measurement. Sample concentration of 0.1 mg/mL was used in characterization. Sample was maintained within the same buffer as mentioned in section 1.3.4.1.

3.4 Results and Discussion

Efficient cloning via the amplification of the hCRY1 gene encoding ~1.9 kbp (Fig 3.4) was performed using PCR and ligated into the vectors listed in Table 3.1. These constructs were further verified by DNA sequencing. The resultant constructs were transformed into BL21 (DE3), BL21 Codon+, and Rosetta 2 (DE3) pLysS E. coli (Fig 3.5). In some cases, dual restriction enzyme digestion was used to verify the insertion of the hCRY1 gene into various vectors (two shown in Fig 3.6-7). After induction trials, constructs with vectors pET-DEST42 (with or without N-terminal 3X-FLAG) and pMal-p4X (MBP-3XFLAG-hCRY1) produced either very little or highly degrading protein (Fig 3.8 A), even after pooling multiple batches of protein crude together for purification purposes. The protein was suspected to be unstable and degrade when expressed on its own, which is why the MBP vector was selected, as it is known to help solubilize challenging proteins[171]. His-affinity beads pull-down assays were used to access the expression of pET-DEST42 C-terminal 6X His construct, in which no band correspondent to hCRY1 at the expected size was detected (Fig 3.9). The major protein obtained by the pull-down assay was one that has molecular weight of slightly higher than 25 kDa. Since the His tag was on the C-terminal, this suggests an N-terminal degraded product of hCRY1. Western blot analyses were utilized to assess the expression of pET-DEST42-hCRY1-His with C-terminal 6X His and

pMal-p4X (MBP-FLAG-hCRY1) constructs (Fig 3.8 and 3.13). Results demonstrated that soluble protein was being obtained from these vectors. Nevertheless, expression quantities were quite limited and the mixture of extraneous proteins exceeded that of our target. More importantly, degradation of hCRY1 was observed (Fig 3.8 and 3.10). In addition, given that hCRY1 is a member of blue-light receptors, dark expression (no light exposure) trials were carried out for pMal-p4X (MBP-FLAG-hCRY1) and pET-DEST42-hCRY1-His constructs to see if this had an effect on hCRY1 production (Fig 3.10-11), but similar results were obtained as compared to the regular expression.

The hCRY1 is a eukaryotic gene with codon usages different to prokaryotic. The lack of viable hCRY1 produced in the aforementioned constructs led to having hCRY1 codon optimized and *de novo* synthesized (Genscript). In addition, the codon-optimized hCRY1 was cloned into pCold-TF (Fig. 3.12) that has several advantages for obtaining protein (see discussions in Chapter 1 section 1.3.1.2). As shown in (Fig 3.13), the crude sample of induced hCRY1 as compared to the non-induced sample displays an additional band of the correct molecular weight corresponding to hCRY1. Expression of hCRY1 in pCold-TF expressed a suitable amount of protein, which was then tested in small scale affinity purification trials to assess the binding to the column (Fig 3.14). The wash and elution conditions were optimized for large scale purification using HisTrap HP column (GE) as shown in (Fig 3.15, 3.17). Similar to pMal-p4X-MBP construct, light/dark expression tests of pCold-TF-hCRY1 did not show difference (data not shown). Affinity purified hCRY1 was then utilized for size exclusion chromatography (SEC) (Fig 3.16, 3.18). The hCRY1 protein was eluted in the void volume of the SEC suggesting that it was a large aggregate. Dynamic light scattering further confirmed that the only peak obtained (hCRY1) was indeed much larger than its monomeric or even dimeric state (Fig 3.19).

As mentioned in section 1.4.2, protein expression from bacterial expression systems encounters multiple problems. By far, mammalian cryptochromes were almost exclusively expressed in eukaryotic cell systems [160, 162, 172, 173], some of which only cloned a truncated portion of the entire gene [174-176]. To date, only two mammalian cryptochrome (mouse CRY1 and CRY2) structures have been determined [160, 175], both of which are truncated. This is likely attributable to the fact that the C-terminal region, which is highly flexible, exposed [177] and unstructured [178, 179] causes problems when attempting to produce a crystal. More importantly, the C-terminus has been found to easily degrade [160] as also seen in our results (Fig 3.8, and 3.14-16), making it quite difficult to obtain full-length hCRY1. Also, it was recently shown that mouse CRY1 (mCRY1) becomes phosphorylated at the C-terminus, which regulates the length of circadian period and serves to stabilize CRY1's activity [180]. Therefore, re-cloning hCRY1 into a vector with a C-terminal His tag was carried out, anticipating that only full-length protein would be purified because those truncated or degraded hCRY1 will not have the His tag to interact with affinity resin. Unfortunately, the C-terminal His-tag construct did not resolve the degradation issue, as demonstrated in Fig 3.15-16. Even after further purified by SEC, the aggregated form of hCRY1 still contains degradation (Fig 3.18). Nevertheless, the only elution profile observed from the C-terminal His construct corresponded to aggregates, similar to previous constructs.

3.5 SUMMARY AND FUTURE DIRECTIONS

Bacterial expression systems were first targeted due to their relative ease and scalability, therefore, hCRY1 will be re-cloned with only the PHR and a short extension after that, following the methods performed by Nangle et al for mCRY1 [175]. Given the current aggregated status of bacterial expressed recombinant full length hCRY1, denaturing and refolding in the presence of it's binding partner hPER1 could potentially yield a stable protein complex useful in structural studies. Along the same line, dual protein expression vector will be used in which hCRY1 and hPER1 will be co-expressed in hopes that they stabilize each other. Alternatively, it might be useful to stabilize hCRY1 with small

molecular compounds, which have been found to lengthen the circadian period by preventing mCRY1 degradation [175]. Given that hCRY1 is a eukaryotic gene, and solubility or folding issues cannot be overcome in prokaryotic systems, eukaryotic expression systems such as yeast, or the insect baculovirus system may be used [97]. The drawback of multiple steps for generation of recombinant clones; low protein yield, and high cost for media are reasons why the eukaryotic expression systems were not tried in the first place. Once highly homogenous hCRY1 is produced, crystallization trials will be performed. Members of the PL/CRY family have previously been crystallized [155, 172, 175, 181, 182]; therefore, conditions similar to those successful crystallizations can be screened. Subsequent to obtaining optimal crystallization conditions and suitable size crystals are produced, endeavors to determine the atomic structure will follow standard X-ray crystallography procedures [183, 184]. The resultant atomic structures will facilitate to generate a plausible regulatory mechanism how hCRY1 acts upon the transcriptional activation complex. Structural comparative analyses among hCRY1, dCRY, and E. coli Photolyases (PLs) will shed light on C-terminal tail regulatory function as well as the difference between PLs and CRYs. Moreover, structural analyses of the contact interface between hCRY1 and its potential binding components including ligands will reveal critical residues that are pertinent in the circadian rhythm. Information learned from these critical residues will be further compared with other CRY models and PLs to determine whether they are conserved between species, which will further prove or disprove the hypothesis that type 1 and 2 CRYs have different regulatory functions [156-158]. Having the atomic structure will aid our understanding of the mechanism with which CRY imparts to modulate the circadian rhythm. In addition, the available structures will show whether the catalytic core of hCRY1 binds FAD similar to the PL/CRYs in homologous species or binds other circadian components/DNA. The high resolution structures obtained from the research will deepen our understanding of the catalytic function of hCRY1 and its overall role in modulating the circadian rhythm.

Generating an atomic structure of hCRY1 will facilitate the long-term goal of conducting functional studies in terms of the entire transcriptional activation complex. Information obtained from the structure can help build new hypotheses pertinent to binding capabilities to the heterodimer complex,

which could be tested in future mutagenesis experiments at critical contact residues. The proteins of the heterodimer transcriptional activation complex CLOCK/BMAL1 can be purified and crystallized individually then assembled in the presence of hCRY1 and/or hPER1 and examined utilizing cryo-electron microscopy (cryo-EM). The individual atomic structures can then be fitted into the electron density map generated from cryo-EM to produce a pseudo-atomic structure of the complex. Overall, this complex structure would provide insight into how the transcription/translational negative feedback loop is regulated at the molecular level. Moreover, critical residues of the interfaces between these core regulatory components in the mammalian circadian clockwork can be targets for therapeutic agents that disrupt DNA binding or regulate the inhibitory function of hCRY1. The knowledge obtained will facilitate the development of treatments against circadian related diseases.

3.6 TABLES AND FIGURES

TABLE 3.1 hCRY1 gene amplification primers

Clone	Primer
pET-DEST42-hCRY1-His (the PCR was done in three steps)	Sense Primer Outside 5' GCTAGGATCCAAGGAGATATTCACCATGGGGGTGAACGCCGT BamHI Sense Primer Inside 5' CACCATGGGGGTGAACGCCGTG For TOPO cloning Antisense Primer 5'GGCACTCGAGCGTCCTTCAATATTAGTGCTCTGTCTCTGGACTTT XhoI
pET-DEST-42-3X-FLAG-hCRY1 (the PCR was done in two steps)	Sense Primer Outside 5'GCTAGAGCTCAAGGAGATATTCACCATGGGGGTGAACGCCGTGCA XhoI Sense Primer Inside 5' CACCATGGGGGTGAACGCCGTG For TOPO cloning Antisense Primer 5'CCGCGGATCCTCTTCCTTCAATATTAGTGCTCTGTCTCTGGACTTT BamHI

pMal-p4X-MBP-3XFLAG-hCRY1	Sense Primer Outside 5' GCTAGAGCTCAAGGAGATATTCACCATGGGGGTGAACGCCGTGCA SacI Sense Primer Inside 5’ CACCATGGGGGTGAACGCCGTG Antisense Primer 5’CCGCGGATCCTCTTCCTTCAATATTAGTGCTCTGTCTCTGGACTTT BamHI
pCold-TF-His-hCRY1 Codon Optimized	*De novo* synthesized construct (Genscript) in pUC57 digested with 5’-NdeI and 3’ HindIII
pCold-TF-hCRY1-His	Sense Primer 5' GCGCTACATATGGGTGTGAATGCGGTCCACTGGTTCCG… NdeI …TAAAGGTCTGCGTCTGCACGACAAT Antisense Primer 5’ GGTCGAATTCTTAATGATGATGATGATGATGGCGACC… EcoRI …TTCAATGTTCGTGCTCTGACGCTGGA

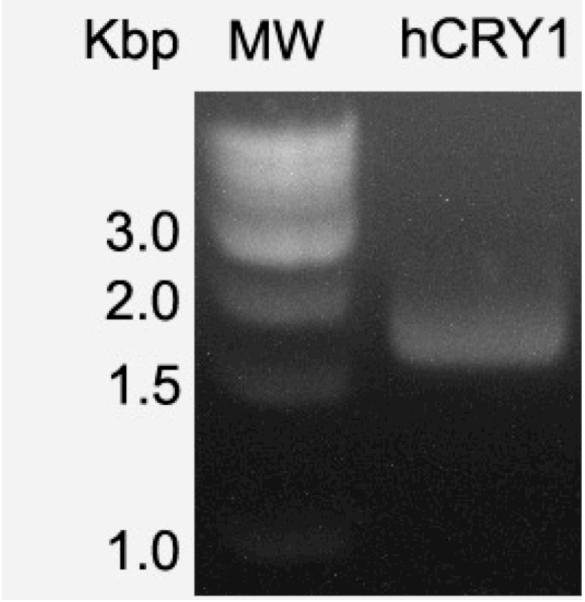

Fig 3.4 Agarose gel of hCRY1 PCR product. Molecular weight marker is in the first well, and labeled on the left of the panel.

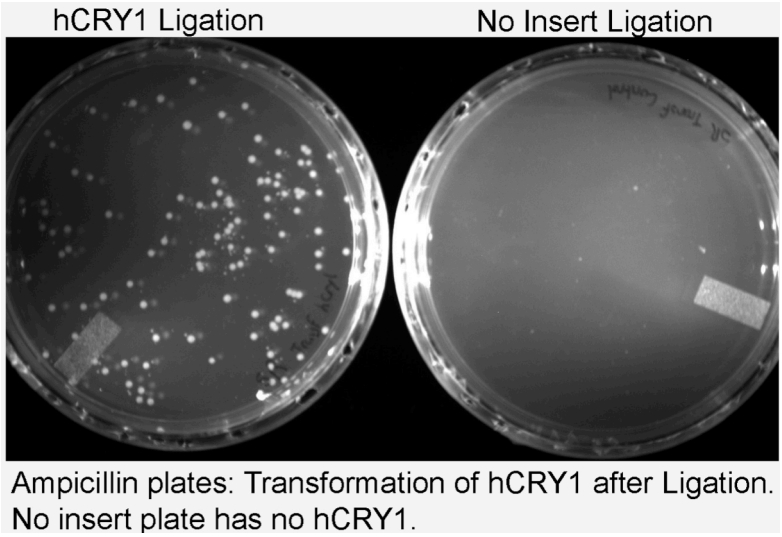

Fig 3.5 Ampicillin plates containing transformed hCRY1 after ligation reaction. The plate on the right denotes a negative control, while the plate on the left is the transformed ligation.

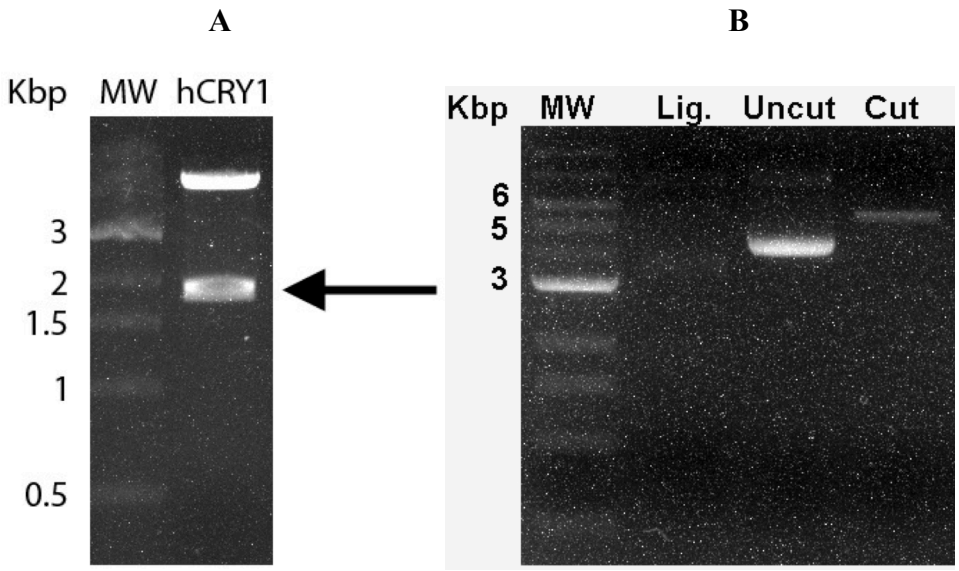

Fig 3.6 A) Agarose gel of hCRY1 dual digested ligation product. The black arrow denotes the location of hCRY1 at the correct size after the ligation and transformation reaction. B) Agarose gel of ligation trial of hCRY1 with a shift in plasmid position in the second lane. The adjacent lanes are uncut and cut plasmids as reference. The molecular weight marker is in the first well and labeled to the left of the panel.

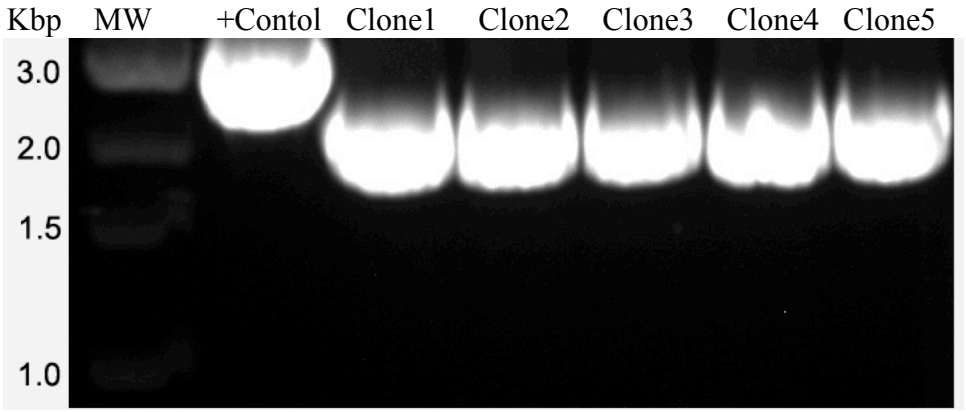

Fig 3.7 Agarose gel of hCRY1 PCR screening following ligation. Wells labeled as in figure, with the positive control in the second well and the molecular weight marker in the first well and labeled to the left of the panel.

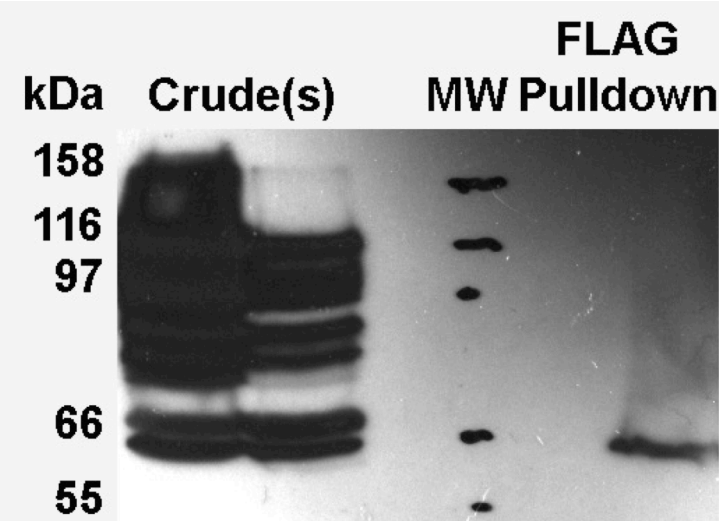

Fig 3.8 Western blot of hCRY1 FLAG-tag affinity pulldown in pET-DEST 42. The first and second lane represent crude samples, and the fourth lane is the pulldown hCRY1. The middle lane(s) contain the molecular weight marker, which is labeled to the left of the panel. As shown in the figure, hCRY1 is degrading, which is more severe in the SDS-PAGE gel (data not shown).

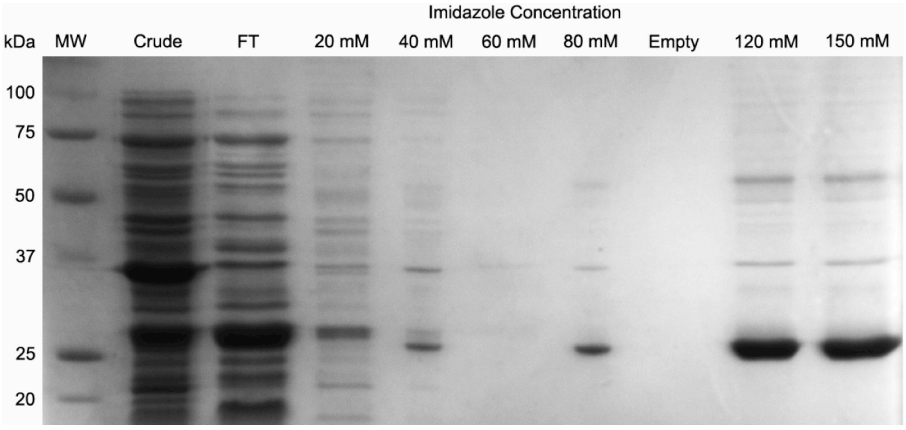

Fig 3.9 SDS-PAGE of hCRY1-pET-DEST42 affinity pulldown assays. Step elutions were performed at the indicated concentration of imidazole labeled above the lanes. FT stands for flow-through. The molecular weight marker is in the first lane and labeled to the left of the panel.

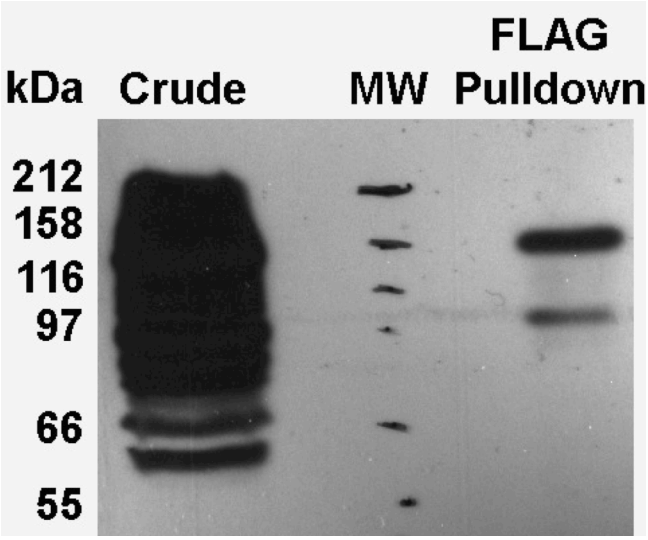

Fig 3.10 Western blot analysis of hCRY1-MBP affinity pulldown assay. The first lane is crude sample. The third lane is the pulldown product.

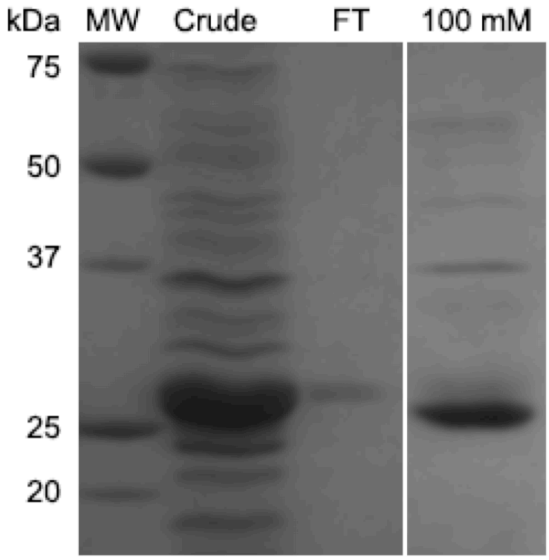

Fig 3.11 SDS-PAGE of hCRY1 dark expression pulldown assays. hCRY1 was applied to affinity column washed with binding buffer and eluted with 100 mM imidazole, to test binding of protein. FT stands for flow-through. The molecular weight marker is in the first lane and labeled to the left of the panel.

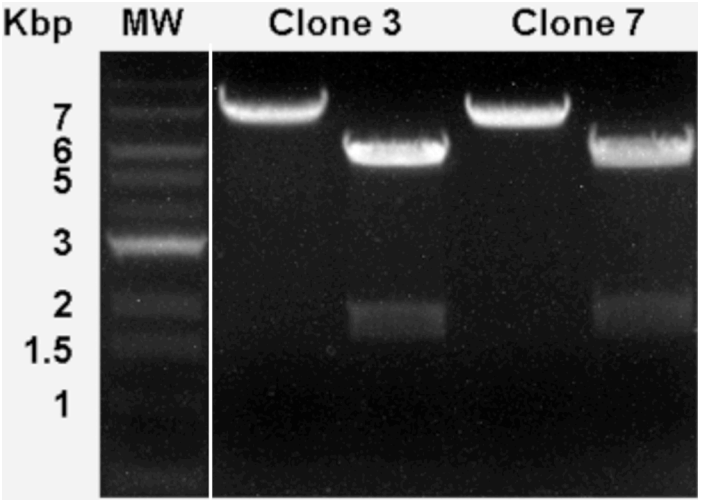

Fig 3.12 Agarose gel of hCRY1 after ligation into pCold-TF and dual restriction enzyme digested. wells two and four are uncut vector of clones 3 and 7 respectively. Wells three and five are dual digested hCRY1. The molecular weight marker is in the first well and labeled on the left of the panel.

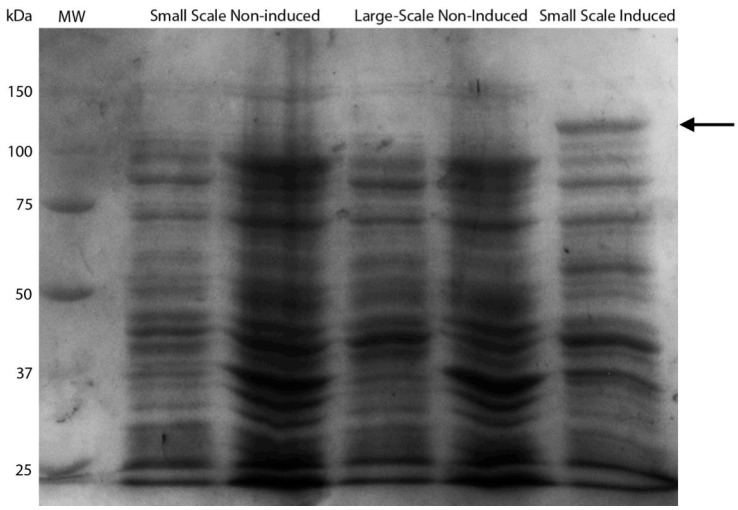

Fig 3.13 SDS-PAGE of hCRY1 induction trials. Lanes two and three are non-induced supernatants of small and large scale induction respectively. Lanes three and five are pellets of of small and large scale induction respectively. The sixth lane is induced hCRY1 at the correct fusion size as denoted by the black arrow to the right of the panel. The molecular weight marker is in the first lane and labeled to the left of the panel.

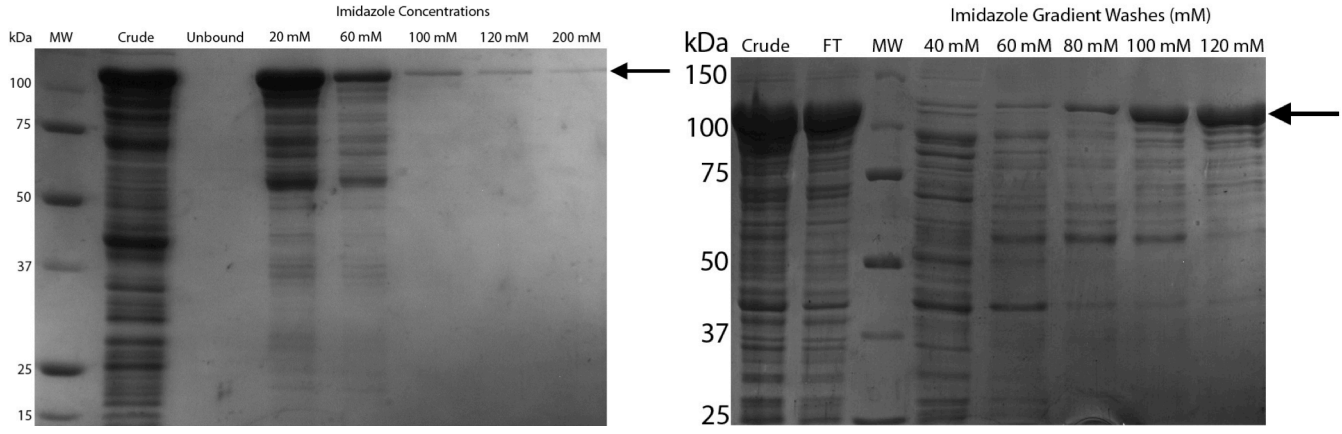

Fig 3.14 SDS-PAGE of hCRY1 small-scale affinity purification trials. Lanes are labeled above the figure, and the black arrow to the right of the panel denotes hCRY1-TF fusion size. The molecular weight marker is in the first lane and labeled to the left of the panel.

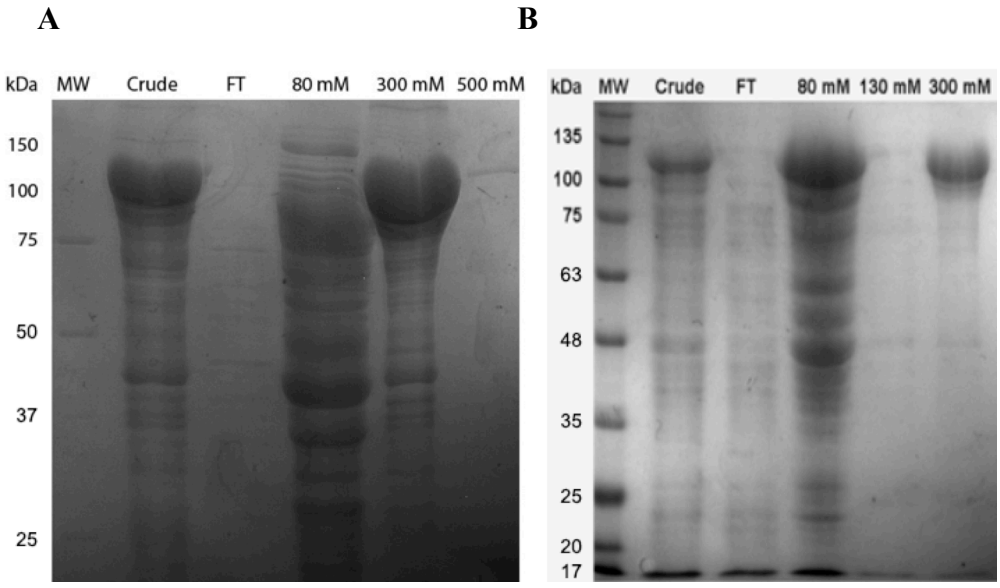

Fig 3.15 A) SDS-PAGE of hCRY1 affinity purification on HisTrap HP. Lanes are labeled above the figure with indicated imidazole concentrations. B) SDS-PAGE of hCRY1 (0.5L) on HisTrap HP using a higher gradient to wash out contaminants and purify hCRY1-TF fusion to higher homogeneity. FT stands for flow-through. The molecular weight markers are in the first lane(s) and labeled to the left of the panel.

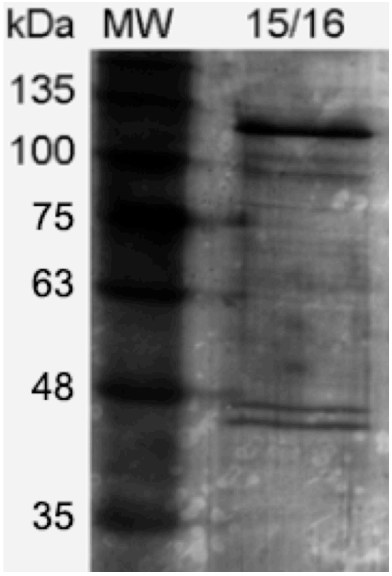

Fig 3.16 SDS-PAGE of hCRY1 (silver stained) after size-exclusion chromatography purification. The second lane contains the only elution fraction obtained from the column. The molecular weight marker is in the first lane and labeled to the left of the panel.

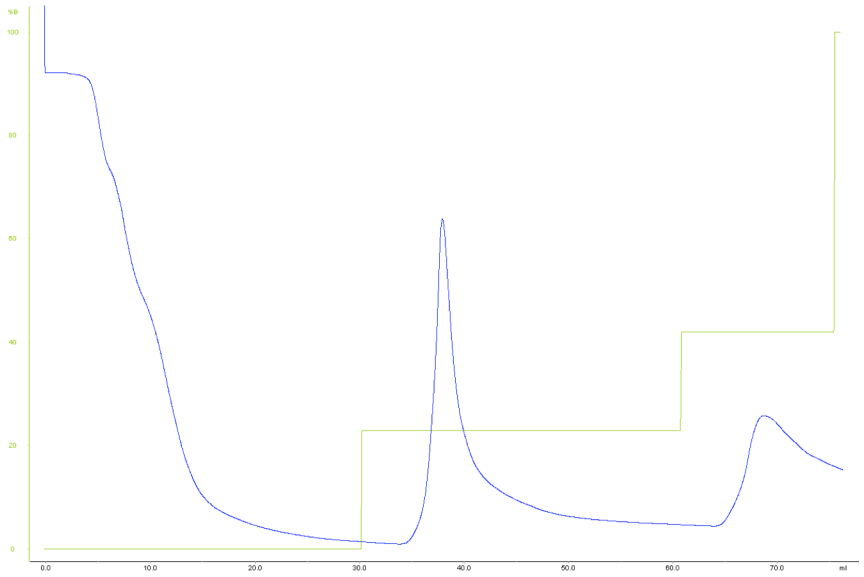

Fig 3.17 HisTrap HP (GE) affinity chromatogram of TF-hCRY1. The first peak is the wash at 110 mM and the second is elution at 200 mM imidazole concentration.

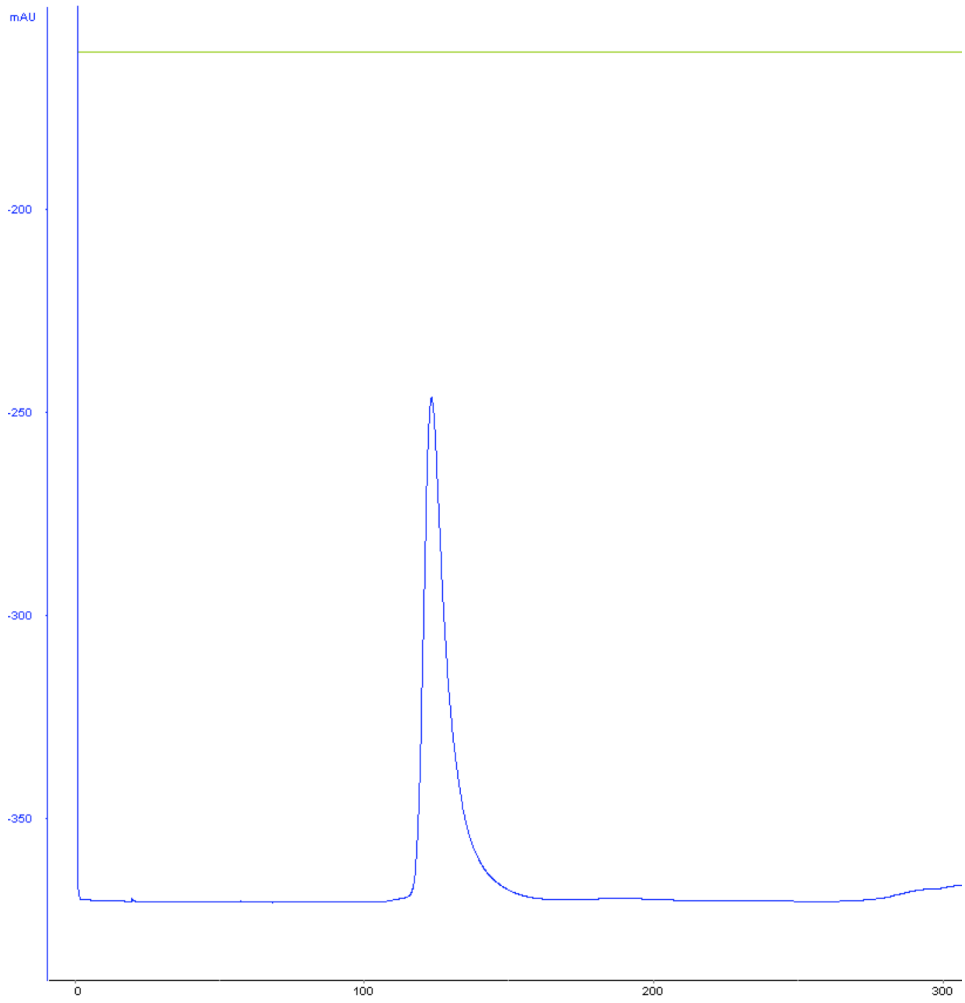

Fig 3.18 Size exclusion chromatogram of TF-hCRY1. The only peak observed corresponds to the elution of hCRY1.

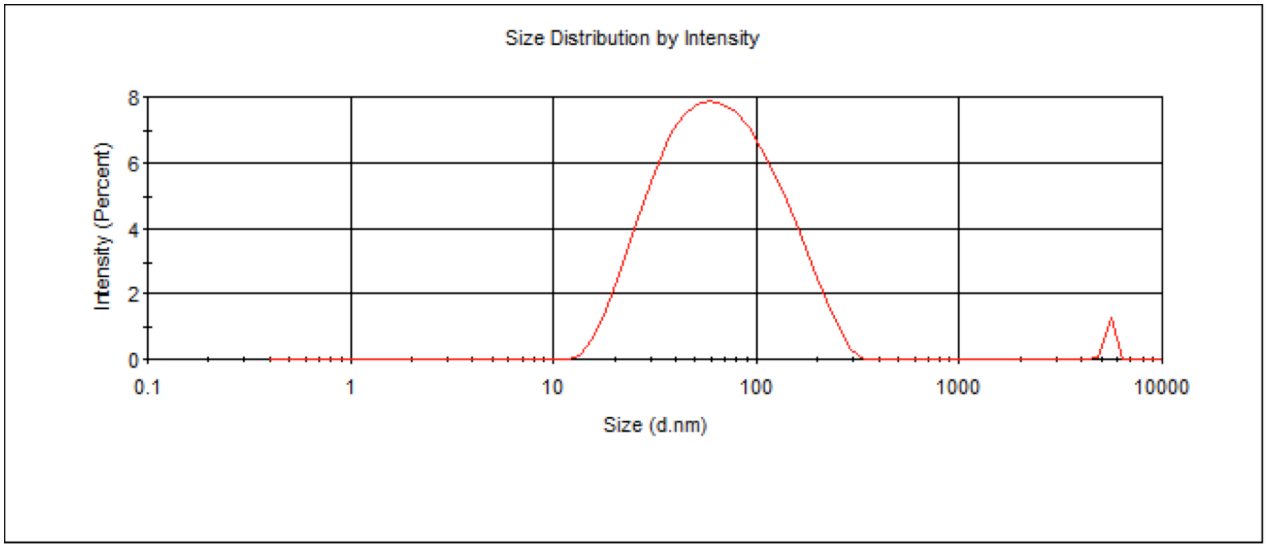

Fig 3.19 Dynamic light scatter profile of TF-hCRY1 purified via affinity and size exclusion chromatographies.

CHAPTER 4:

BIOCHEMICAL INVESTIGATION OF AN ANTIBODY AGAINST CANCER

4.1 INTRODUCTION

In 2013 alone, over 1.6 million new cases of cancer were reported, with an expected mortality of over half a million [185]. The importance of cancer in everyday life is clear, which necessitates the development of novel detection and treatment methods. Carbohydrates are vital macromolecules found within living organisms that play critical roles within the cell. Currently there are a handful of glycan-derived drugs on the market that have proven efficient in the treatment of the Flu, anticoagulation, and red blood cell production [186-189]. In addition many cell surface proteins are glycosylated and their glycan moiety serve as targets for microbes and viruses, cell adhesion and development, impact cancer spread, and regulate countless receptor:ligand interactions [187]. In the past, carbohydrates were believed to be poorly immunogenic or of little value as antigens for vaccine development [190]. However, this has changed after the discovery of cancer specific glycan immunogens [191, 192]. Even though the study of glycans as an immunogen is in its infancy with a little over 15 years, it is now clear that carbohydrate and glycoconjugates can serve as highly immunogenic B-cell epitopes (recognized by the immune system by antibodies) [191]. This discovery has vastly increased the efforts of glycobiologist to explore the glyco surface molecules as targets for cancer, used as therapeutics, vaccines, and biomarkers.

Antibodies are ubiquitously produced throughout vertebrates, and divided into five different classes (IgA, IgD, IgE, IgG, and IgM). Each class also has subclasses such as IgG1, IgG2, etc [193]. Antibodies work as “sticky” proteins as defensive molecules that recognize and attaches to its specific antigen (foreign substance). Once attached, they can recruit components of the immune system to eliminate the antigen [194, 195]. The most common Immunoglobulins (Ig), are IgG, which are found in

all body fluids and accounts for approximately 80% of all antibodies within the body [196]. IgG demonstrates a Y-shaped structure (Fig 4.1) with two identical arms representing Fab fragments (antigenic contact) connected to Fc through a flexible linker [197]. Antibodies are made up of two light chains linked by disulfides with two heavy chains; where the heavy chain (H) is a 50 kDa subunit and the light chain (L) is 23 kDa [198]. Each antibody has two copies of distinct antigen-binding site, located at the tip of the Y-shaped antibody. The antigen-binding site is comprised from variable domains of one light chain (V_L) and one heavy chain (V_H) [198]. The interaction between an antibody and its antigen directs the immune response and is indispensable for the development of novel vaccines for diseases [199].

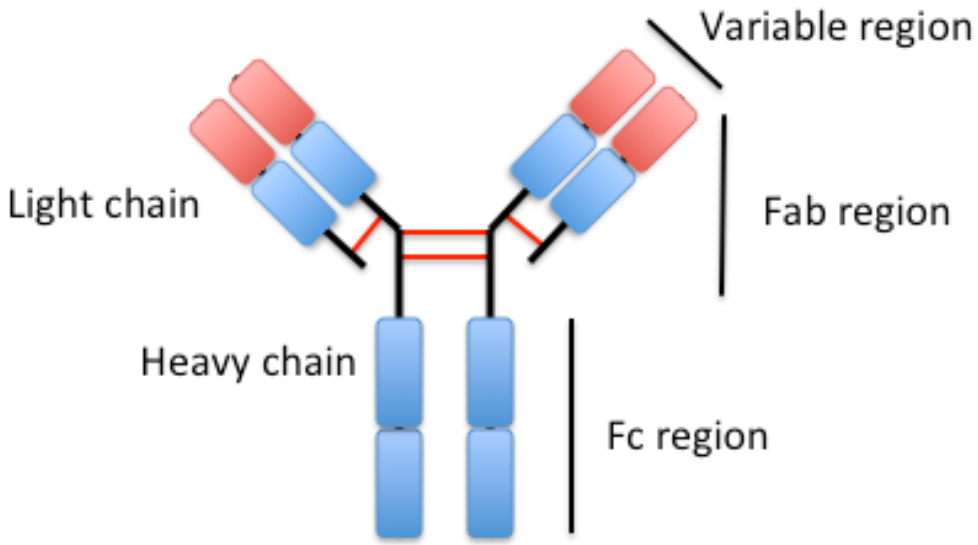

Fig. 4.1 Classic IgG Antibody structure depicting the different regions of the antibody. Adapted from [200].

Despite remarkable advances made in the field of glycobiology, progress has stalled due to the unknown mechanism by which a carbohydrate (antigen termed immunogen) is presented to elicit a response. As denoted by Gerstenbruch, S. *et. al.*, over 2000 antibody-carbohydrate interactions have been investigated and over half have shown cross-reactivity with other glycans [201]. The low affinity and molecular flexibility have deterred the usage of structural studies to examine these interactions. The

abovementioned cross-reactivity coupled with the lack of structural information has stagnated the successful development of carbohydrates as vaccine to target cancer. The interactions between protein and carbohydrates are still poorly understood. Due to the relatively weak interactions between proteins and monosaccharide, proteins require oligosaccharides for binding. The most well-characterized carbohydrate binding protein is lectin, a big family of surface proteins found throughout many species that serve a wide range of functions. One of the chief functions carried out by lectins involve the contact between two cells [202]. As for antibodies against carbohydrates, very few structures are available [191, 192, 203] due to the poor immunogenicity of carbohydrate for the above stated reasons. This has hindered the development of carbohydrates for cancer therapeutic or diagnostic agents. Therefore, it is imperative to bridge this need for more antibody-glycan complex structures. More antibody/glycol-antigen structures will contribute in general to unravel the interactions between protein and carbohydrate at the atomic level, which will facilitate the design and development of more effective carbohydrate-based immunogens.

16A, the antibody under investigation, is an IgG1a derived from mouse that targets the Tn antigen O-linked N-acetylgalactosamine (GalNac) epitope of MUC1 from cancer cells [204]. Antibodies (including 16A) toward glyco-peptides with authentic glycosylated forms of MUC1 were produced by Song, W. et. al. with 25 fold higher binding versus the non-glycosylated form. In specific, 16A antibody recognizes the glycopeptide sequences of RPAPGS(Tn)TAPPAHG and the Tn antigen is linked via O-glycosylation to the Ser residue of the peptide sequence. In order to characterizing the antibody-carbohydrate interactions, the biotinylated glycopeptides, RPAPGS(Ac3GalNAc)TAPPAHG-dPEG™11-Biotin (Fig. 4.2), and non-glycosylated form RPAPGTAPPAHG-dPEG™11-Biotin, were custom synthesized (Peptide International Inc., Louisville, KY). Subsequent deacetylation was performed to produce RPAPGS(GalNAc)TAPPAHG-dPEG™11-Biotin [204]. The binding characteristics between these glycopeptides and 16A were analyzed using ELISA and surface plasmon

resonance [204]. ELISA of glycopeptides RPAPGS(Ac3GalNAc)TAPPAHG and RPAPGS(GalNAc)TAPPAHG showed higher binding affinity (25 fold higher) as compared to the non-glycosylated RPAPGTAPPAHG peptide. However, surface plasmon resonance measurement of 16A demonstrated similar binding to both glycosylated and non-glycosylated forms. Nevertheless, 16A will provide us an interesting candidate to investigate the structural features that lead to a better understanding of the interaction between 16A and the abovementioned glycopeptides at the atomic level. In order to improve our understanding of how carbohydrate-antibody interacts, we must couple structural biology to determine a wide range of parameters. These include: how carbohydrates of a potential immunogen should be presented, the amount of carbohydrate epitopes present, their proximity to one another, the relative size and complexity, the portion(s) of the carbohydrate involved in the contact interface, and the role of the peptide backbone in terms of recognition.

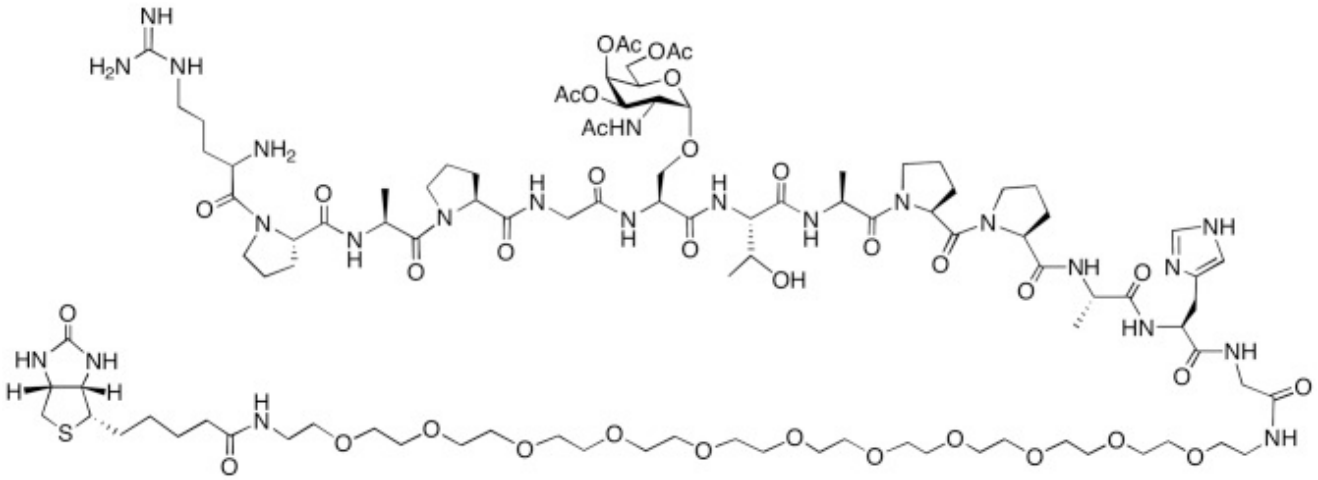

Fig. 4.2 Structure of biotinylated glycopeptide (RPAPGS(Ac3GalNAc)TAPPAHG) that is recognized by 16A antibody. Adapted from Song, W. et al. 2012

4.2 MATERIALS AND METHODS

Initial batches of the 16A antibody were purified by Dr. Dapeng Zhou at MD Anderson Cancer Center using the following protocol: mouse ascites were mixed 1:10 with binding buffer (1M Glycine pH8.6, 300 mM NaCl). Samples were then loaded into ProSep-vAUltra- (cat # 115115824-Millipore), subsequently washed with binding buffer and eluted with elution buffer (0.1M Citric Acid pH 3.0, 150 mM NaCl). Eluted antibodies were immediately neutralized with 1M Tris pH 9.0, and dialyzed in PBS buffer (10 mM Na_2HPO_4, 1.8 mM KH_2PO_4, 140 mM NaCl, 2.7 mM KCl pH7.4).

4.2.1 CELL CULTURE

Hybridoma cells (provided by Dr. Dapeng Zhou at MD Anderson Cancer Center) were grown at 37°C with 5% CO_2 in T-flasks. 16A was maintained in Hybridoma-SFM media (Gibco Cat# 12045) with 10% fetal calf serum and Penicillin-Streptomycin (Gibco Cat# 15140) at a final concentration of 100 ug/ml streptomycin, and 100 units/ml penicillin. Cell counts were performed daily to ensure cells were healthy and maintained between $2x10^5$-$1x10^6$ viable cells/ml, and split accordingly. Cultures were slowly adapted to the above media without fetal calf serum. Cell culture was then used for large-scale production of 16A in spinner flasks (Bellco Glass, Inc).

4.2.2 PROTEIN A AND PROTEIN G PURIFICATION.

Protein A is from the bacterium *Staphylococcus aureus* and has very strong affinity to IgGs [205, 206]. Due to its high affinity for IgG's Fc regions, Protein A has been utilized for purification of IgG's to high homogeneity [207, 208]. One issue is that Protein A has a weaker affinity towards mouse IgG1 [209]. Protein G, another protein from Streptococcal bacteria, also binds the Fc region of immunoglobulins. Protein G, however, binds IgG much stronger than Protein A [210, 211].

16A antibody was centrifuged at 5,500 xg for 10 minutes to remove cell debris and filtered by 0.45μm filter before loaded onto Hitrap HP Protein A column (GE), or onto Protein A resin (Goldbio) suspended in a 1mL gravity column and equilibrated with 10 mL of binding buffer (50 mM Tris pH 8.0, 2.7 mM KCl, 137 mM NaCl). Gravity chromatography was used to optimize the binding and eluting conditions. 50 mL of supernatant was run via gravity flow and any unbound sample was washed out with the same buffer, while allowing only the Antibody remains bound to the column. The flow through was collected in a separate sample tube for analysis. Elutions were tested using 1 M Glycine pH 2.8, or 0.1M Citric Acid pH 3.0, 150 mM NaCl. The optimal eluting conditions were found to be the 0.1M Citric buffer. Chromatographic techniques were performed in the same manner for HiTrap HP A and protein G (GE).

4.2.3 16A FAB FRAGMENTATION

Due to the flexible loop between Fab (variable region) and Fc (constant region), intact antibodies do not readily pack into crystal arrays. Fragmentation of full-length antibodies needs to be performed to obtain the Fab domains for crystallization. There are many proteases available for the generation of various fragments of Fab; one of the enzymes commonly utilized is Papain. Papain is a cysteine protease found in papaya (*Carica papaya*) [212]. Papain breaks the linkage of an antibody between the Fc and Fab regions generating three fragments: two copies of Fab (~50 kDa) and one copy of Fc (~25 kDa) [213, 214]. Ficin, Another enzyme similar to papain, is a non-specific sulfhydryl protease coming from the latex of the fig tree *Ficus glabrata* [215]. Fab obtained from ficin fragmentation has been found to have an improved reactivity and yield as compared to that from papain, specifically for mouse IgG1 [216, 217].

16A fragmentations were prepared by proteolytic digestion of the purified antibody with various enzymes including: immobilized papain (Pierce cat # 20341), immobilized ficin (Pierce; cat # 44881),

crystal papain (Sigma-Aldrich; cat # 76218), and mouse IgG1 Fab/F(ab')2 preparation kits (Pierce Ficin based; cat # 44980). Multiple trials were carried out to optimize the fragmentation of the antibody in order to obtain Fab fragments. Purified intact antibodies were buffer exchanged into the proper digestion buffer required for the specific enzyme. Small scale enzymatic digestion conditions were tested by: (1) changing the enzyme(s) to antibody ratio; (2) increasing the concentration of antibody; (3) increase or decreasing the total enzyme used; (4) changing digestion temperature; and (5) changing reaction receptacles. Samples were taken out of the digestion at different time points and quenched with the addition of SDS-PAGE buffer (reducing or non-reducing) and 20 mM iodoacetimide. SDS-PAGEs were used to analyze fragmentation results.

4.3 RESULTS AND DISCUSSION

16A received from Dr. Zhou as well as obtained from Hybridoma cells were successfully purified to homogeneity (Figures 4.3-6) using Protein A or Protein G chromatography, and the best yield was from Protein G. Purified 16A antibody was then utilized for subsequent enzymatic digestions. Small scale digestion trials following the conditions detailed with the kits or literature conditions were carried out first. A predominant 37 kDa fragment was observed on non-reducing SDS-PAGEs from three of the four digestion systems (Fig 4.7). Non-reducing SDS-PAGE gel of immobilized ficin kit digestion showed the intact antibody without fragmentation (Fig. 4.8), suggesting the antibody is resistant to ficin using the company suggested condition. The normal Fab fragment should be 50 kDa in non-reducing SDS-PAGEs. The 37 kDa fragments appear to be over digested fragments of the intact antibody. Approximately 50 enzyme digestions with various conditions (temperature, time, enzyme to substrate ratio, concentration, etc.) were carried out in order to obtain Fab fragments with little success. Results obtained in most cases were over digested with 37 kDa fragments (Fig. 4.9, 4.10, 4.11, 4.12, 4.13 with immobilized papain, immobilized ficin, crystal papain 1:200, crystal papain 1:80, and IgG1 mouse preparation kit 1:50; respectively). In relatively rare conditions, the antibody was resistant to the digestion (Fig 4.14 immobilized papain). When significantly reducing the enzyme to substrate ratio E:S (e.g. below 1:500), only partial digestion of the antibody was observed only yielding 37 kDa fragments.

None of the conditions completed yielded fragments at the expected 50 kDa size, pertaining to Fab. During the multiple trials, in order to have ample antibody samples for fragmentation, hybridoma 16A cell culture was utilized to produce the antibodies in house. More than 3 L of serum free media (SFM) with the 16A antibody were produced and purified to high homogeneity via protein G chromatography. Protein A has less specificity for mouse IgG1a as mentioned above, therefore protein G was utilized as the method to produce highly pure 16A. This yielded approximately 1.5-2 mg of antibody after this purification technique from 500 mL of media. The sample produced locally and that provided by Dr. Zhou showed the same over digestion fragments, suggesting the antibody source and purification procedure is irrelevant to the digestion. Moreover, 16A was sent to the MD Anderson Cancer Center to perform the Fab fragmentation with the same outcome.

4.4 Summary and Future Directions

As mentioned above, multiple efforts were made to optimize the enzyme and actual conditions to perform the digestion reactions. While viable Fab fragments were not produced, the protocol for producing, and purifying 16A to homogeneity is now established. Utilizing various protocols, suggested from the kit as well as exploring the literature, those conditions that do not generate Fab have been recognized. Suboptimum digestion conditions consistently produced an over digested antibody. Literature has rare instances where this piece maintained function [218], however, upon further characterization of 16A was shown not to be the Fab fragment. Therefore, future experiments will aim to crystallize the intact antibody. This however, might prove challenging given the flexibility [219] of the hinge region. Within the current protein database, only 4 full-length antibodies have been crystallized. The hinge regions of 2 of them were modified to be rigid. Another alternative would be to incubate 16A with the target glycopeptide to protect the Fab domain. It was shown in some cases that the 37 kDa fragment is due to the digestion of the flexible epitope-bind loops at the top of Fab [220]. Co-incubation with the antigen could potentially make those loops more rigid and resistant to enzymatic digestion [221]. Thirdly, the cDNA sequence encoding for 16A can be obtained and utilized to construct bacterially expressed single chain antibodies [222, 223]. These alternatives might help to obtain functional regions of 16A, open the gate towards the ultimate goal to determine its structure.

After obtaining large quantity of 16A variable domain, pre-crystallization screening kit (PCT™ Pre-Crystallization Test, Hampton Research) will be used to determine the ideal antibody concentrations for crystallization. Crystallization screening of 16A variable domain with or without its glycocojugates will ensue to obtain the crystals. The atomic structures will be determined using X-ray crystallography [190, 224, 225]. The structures determined will aid in the development of novel carbohydrate based vaccines to combat diseases such as cancer.

4.5 TABLES AND FIGURES

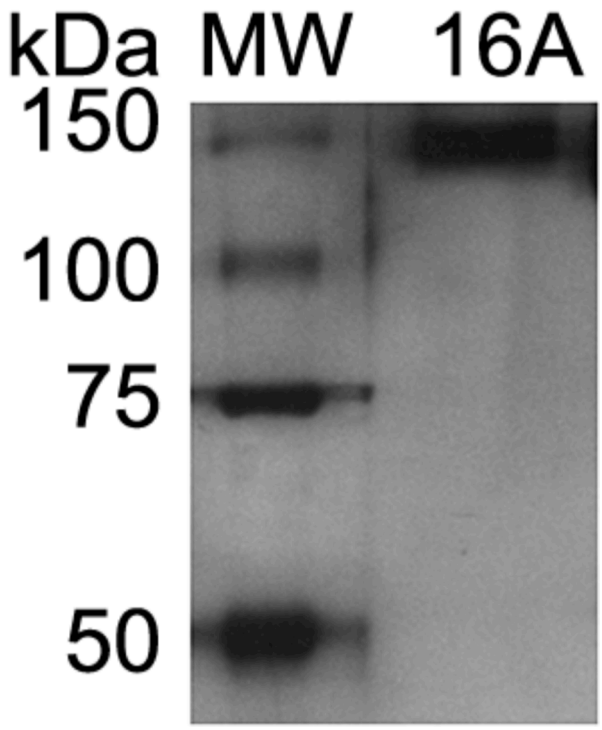

Fig. 4.3 SDS-PAGE (Non-reducing and silver stained) of 16A from Dr. Zhou to assess purity.

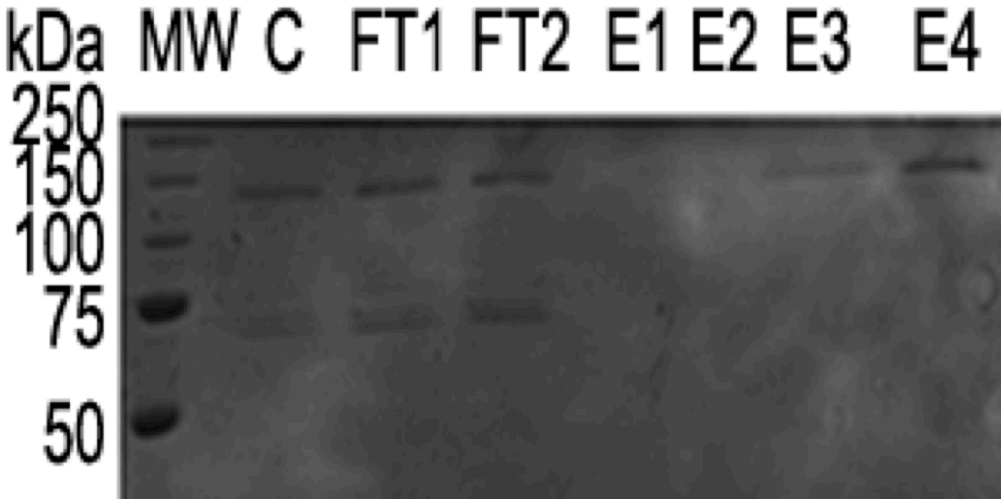

Fig. 4.4 Affinity chromatography on Protein A (GE) of 16A testing binding conditions, C stands for crude, FT stands for flow through, and E stands for elution. The lanes are labeled as in the figure, with the molecular weight marker to the left of the samples.

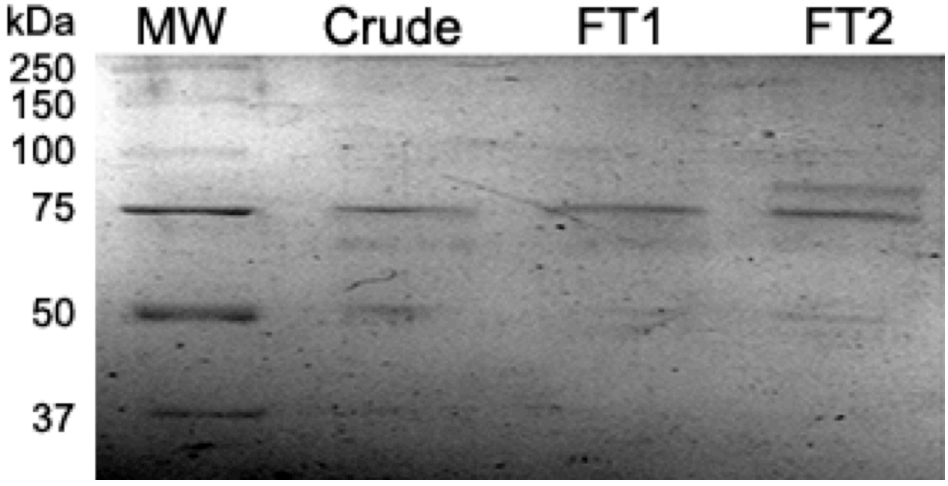

Fig. 4.5 Affinity chromatography on Protein G (GE) of 16A testing binding conditions, FT stands for flow through. The lanes are labeled as in the figure, with the molecular weight marker to the left of the samples.

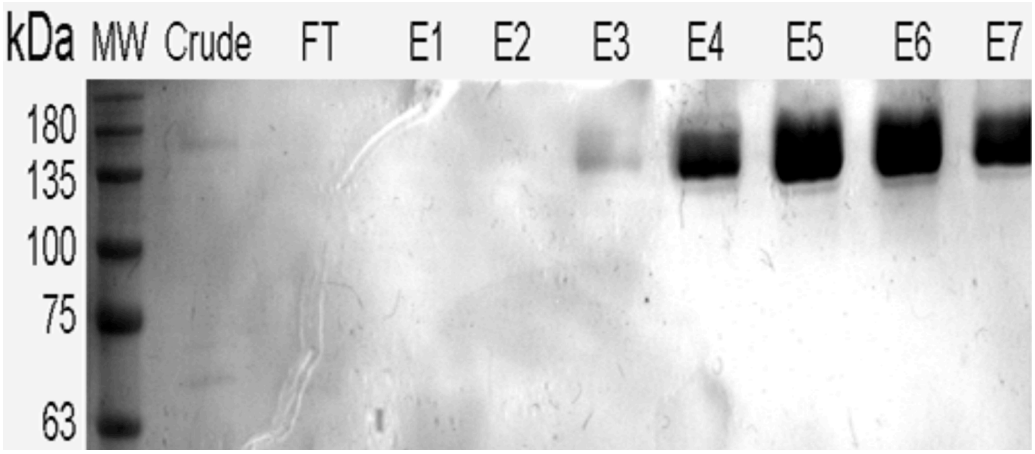

Fig. 4.6 Affinity chromatography on Protein G (GE) of 16A with optimal binding conditions, FT stands for flow through. E stands for elution. The lanes are labeled as in the figure, with the molecular weight marker to the left of the samples.

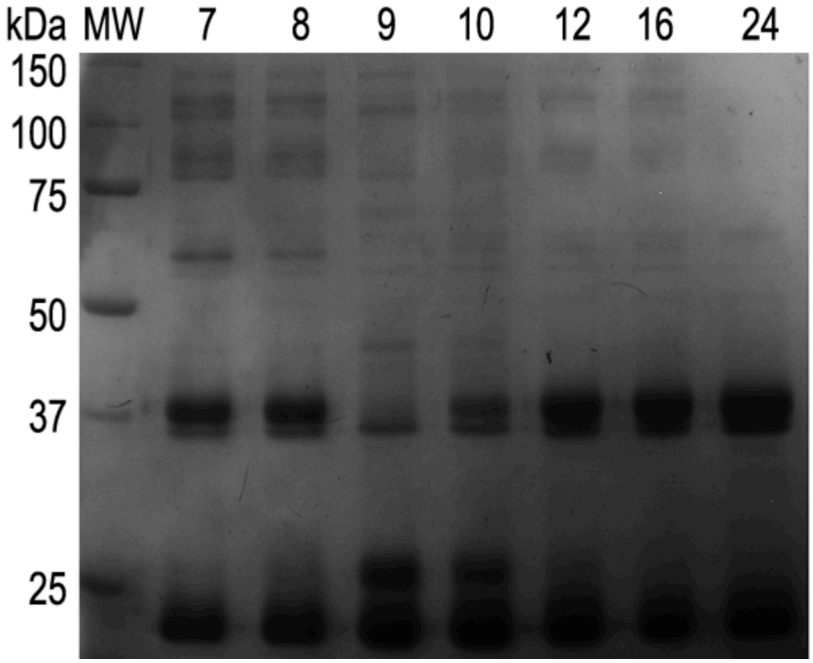

Fig. 4.7 SDS-PAGE of 16A fragmentation with IgG1 mouse preparation kit at an enzyme to substrate ratio of 1:200. Lane one is the molecular weight marker. The antibody was then incubated with kit enzyme and samples were collected at the indicated hours after enzyme incubation.

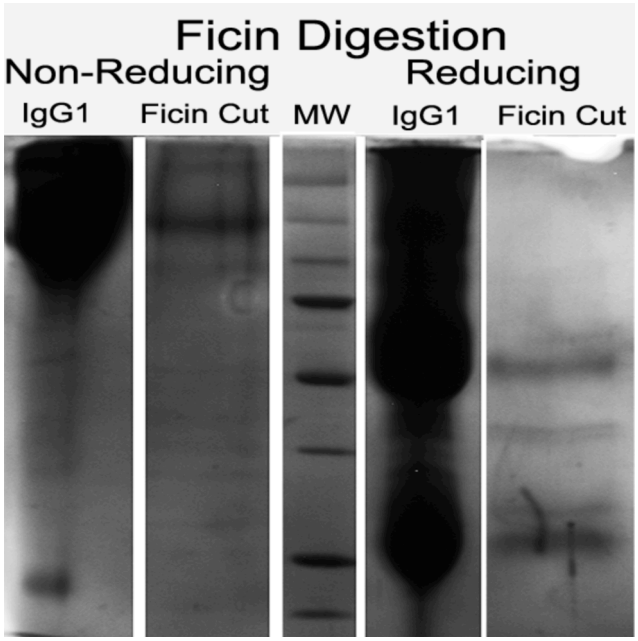

Fig. 4.8 SDS-PAGE of 16A fragmentation with immobilized ficin at an enzyme to substrate ratio of 1:33. The first lane and fourth lane are uncut antibody before digestion (non-reducing) and (reducing). The antibody was then incubated with immobilized ficin and samples were collected at 12 hours after enzyme incubation. The molecular weight marker is in the third lane.

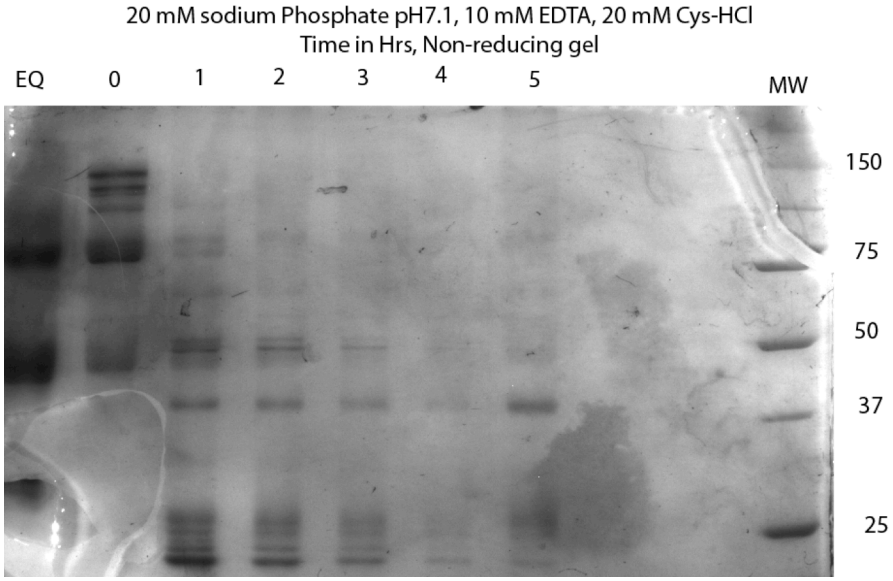

Fig. 4.9 SDS-PAGE of 16A fragmentation with immobilized papain at an enzyme to substrate ratio of 1:500. The first lane is equilibrated antibody before digestion (reducing). The antibody was then incubated with immobilized papain and samples were collected at the hours labeled above the lanes. The molecular weight marker is to the right of the samples.

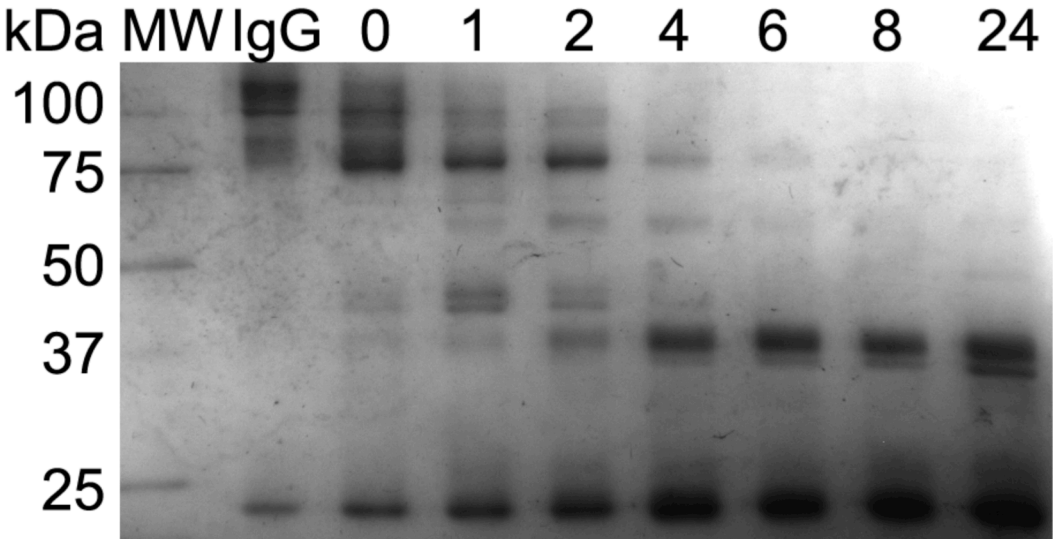

Fig. 4.10 SDS-PAGE of 16A fragmentation with immobilized ficin at an enzyme to substrate ratio of 1:7. Lane one is the molecular weight marker Lane two is uncut antibody. The antibody was then incubated with immobilized ficin and samples were collected at the indicated hours after enzyme incubation.

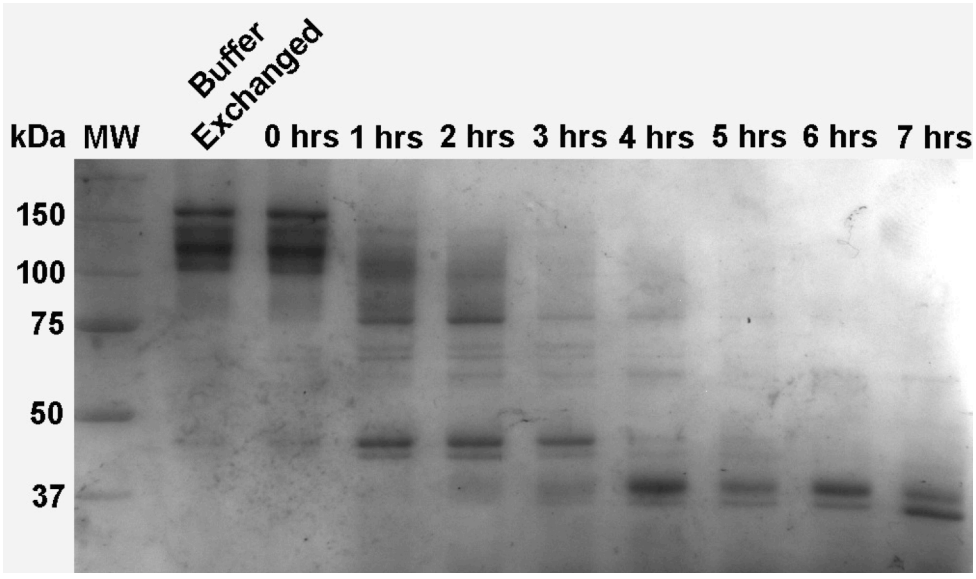

Fig. 4.11 SDS-PAGE of 16A fragmentation with crystal papain at an enzyme to substrate ratio of 1:200. Lane one is the molecular weight marker Lane two is uncut antibody. The antibody was then incubated with papain and samples were collected at the indicated hours after enzyme incubation.

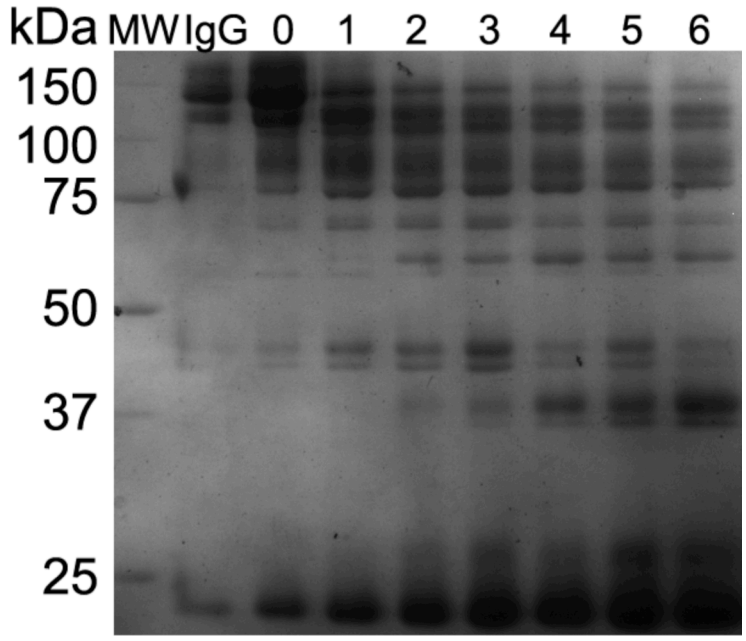

Fig. 4.12 SDS-PAGE of 16A fragmentation with crystal papain at an enzyme to substrate ratio of 1:80. Lane one is the molecular weight marker Lane two is uncut antibody. The antibody was then incubated with papain and samples were collected at the indicated hours after enzyme incubation.

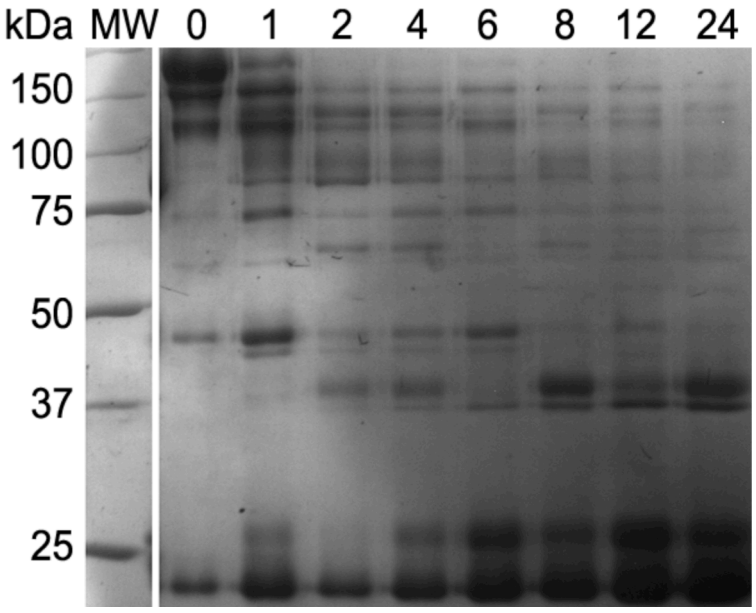

Fig. 4.13 SDS-PAGE of 16A fragmentation with IgG1 mouse preparation kit at an enzyme to substrate ratio of 1:50. Lane one is the molecular weight marker. The antibody was then incubated with kit enzyme and samples were collected at the indicated hours after enzyme incubation.

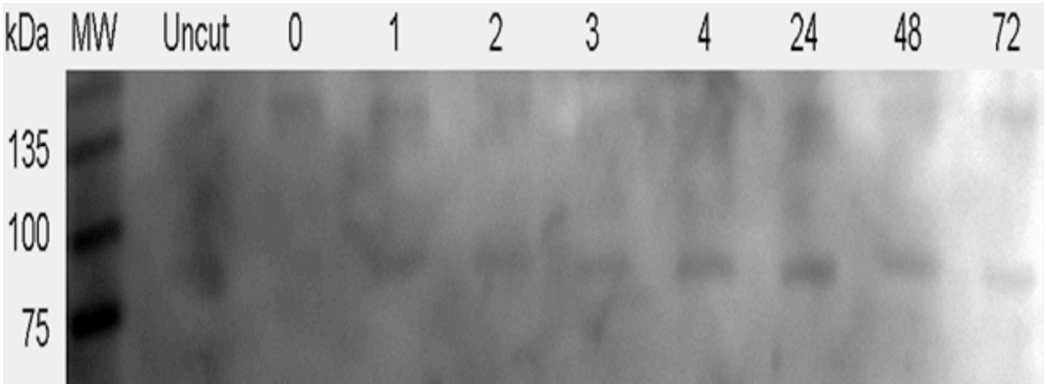

Fig. 4.14 SDS-PAGE of 16A fragmentation with immobilized papain at an enzyme to substrate ratio of 1:160. The first lane is uncut antibody before digestion (non-reducing). The antibody was then incubated with immobilized papain and samples were collected at the hours labeled above the lanes. The molecular weight marker is to the right of the samples.

REFERENCES

1. Seo, J. and K.J. Lee, *Post-translational modifications and their biological functions: proteomic analysis and systematic approaches.* J Biochem Mol Biol, 2004. **37**(1): p. 35-44.
2. Walsh, C.T., S. Garneau-Tsodikova, and G.J. Gatto, Jr., *Protein posttranslational modifications: the chemistry of proteome diversifications.* Angew Chem Int Ed Engl, 2005. **44**(45): p. 7342-72.
3. Gareau, J.R. and C.D. Lima, *The SUMO pathway: emerging mechanisms that shape specificity, conjugation and recognition.* Nat Rev Mol Cell Biol, 2010. **11**(12): p. 861-71.
4. Johnson, E.S., *Protein modification by SUMO.* Annu Rev Biochem, 2004. **73**: p. 355-82.
5. Zhao, J., *Sumoylation regulates diverse biological processes.* Cell Mol Life Sci, 2007. **64**(23): p. 3017-33.
6. Chiocca, S., A. Baker, and M. Cotten, *Identification of a novel antiapoptotic protein, GAM-1, encoded by the CELO adenovirus.* J Virol, 1997. **71**(4): p. 3168-77.
7. Glotzer, J.B., et al., *Activation of heat-shock response by an adenovirus is essential for virus replication.* Nature, 2000. **407**(6801): p. 207-11.
8. Chiocca, S., et al., *Histone deacetylase 1 inactivation by an adenovirus early gene product.* Curr Biol, 2002. **12**(7): p. 594-8.
9. Boggio, R., et al., *A mechanism for inhibiting the SUMO pathway.* Mol Cell, 2004. **16**(4): p. 549-61.
10. Boggio, R. and S. Chiocca, *Gam1 and the SUMO pathway.* Cell Cycle, 2005. **4**(4): p. 533-5.
11. Colombo, R., et al., *The adenovirus protein Gam1 interferes with sumoylation of histone deacetylase 1.* EMBO Rep, 2002. **3**(11): p. 1062-8.
12. Everett, R.D., C. Boutell, and B.G. Hale, *Interplay between viruses and host sumoylation pathways.* Nat Rev Microbiol, 2013. **11**(6): p. 400-11.
13. Everett, R.D., et al., *PML contributes to a cellular mechanism of repression of herpes simplex virus type 1 infection that is inactivated by ICP0.* J Virol, 2006. **80**(16): p. 7995-8005.
14. Moller, A. and M.L. Schmitz, *Viruses as hijackers of PML nuclear bodies.* Arch Immunol Ther Exp (Warsz), 2003. **51**(5): p. 295-300.
15. Mattoscio, D., C.V. Segre, and S. Chiocca, *Viral manipulation of cellular protein conjugation pathways: The SUMO lesson.* World J Virol, 2013. **2**(2): p. 79-90.
16. Pozzebon, M.E., et al., *BC-box protein domain-related mechanism for VHL protein degradation.* Proc Natl Acad Sci U S A, 2013. **110**(45): p. 18168-73.
17. Meluh, P.B. and D. Koshland, *Evidence that the MIF2 gene of Saccharomyces cerevisiae encodes a centromere protein with homology to the mammalian centromere protein CENP-C.* Mol Biol Cell, 1995. **6**(7): p. 793-807.
18. Shen, Z., et al., *UBL1, a human ubiquitin-like protein associating with human RAD51/RAD52 proteins.* Genomics, 1996. **36**(2): p. 271-9.
19. Geiss-Friedlander, R. and F. Melchior, *Concepts in sumoylation: a decade on.* Nat Rev Mol Cell Biol, 2007. **8**(12): p. 947-56.
20. Bayer, P., et al., *Structure determination of the small ubiquitin-related modifier SUMO-1.* J Mol Biol, 1998. **280**(2): p. 275-86.
21. Mossessova, E. and C.D. Lima, *Ulp1-SUMO crystal structure and genetic analysis reveal conserved interactions and a regulatory element essential for cell growth in yeast.* Mol Cell, 2000. **5**(5): p. 865-76.
22. Bernier-Villamor, V., et al., *Structural basis for E2-mediated SUMO conjugation revealed by a complex between ubiquitin-conjugating enzyme Ubc9 and RanGAP1.* Cell, 2002. **108**(3): p. 345-56.
23. Guo, D., et al., *A functional variant of SUMO4, a new I kappa B alpha modifier, is associated with type 1 diabetes.* Nat Genet, 2004. **36**(8): p. 837-41.

24. Boggio, R., A. Passafaro, and S. Chiocca, *Targeting SUMO E1 to ubiquitin ligases: a viral strategy to counteract sumoylation.* J Biol Chem, 2007. **282**(21): p. 15376-82.
25. Palacios, S., et al., *Quantitative SUMO-1 modification of a vaccinia virus protein is required for its specific localization and prevents its self-association.* Molecular Biology of the Cell, 2005. **16**(6): p. 2822-2835.
26. Rossman, J.S. and R.A. Lamb, *Influenza virus assembly and budding.* Virology, 2011. **411**(2): p. 229-236.
27. Rosas-Acosta, G.W.V.a.W., V.G. , *in Sumoylation: Molecular Biology and Biochemistry* (2004), Horizon Bioscience. pp. 331–377.
28. Chang, P.C., et al., *Kaposi's Sarcoma-associated Herpesvirus (KSHV) Encodes a SUMO E3 ligase That Is SIM-dependent and SUMO-2/3-specific.* Journal of Biological Chemistry, 2010. **285**(8): p. 5266-5273.
29. Chang, T.H., et al., *Ebola Zaire Virus Blocks Type I Interferon Production by Exploiting the Host SUMO Modification Machinery.* Plos Pathogens, 2009. **5**(6).
30. Petroski, M.D. and R.J. Deshaies, *Mechanism of lysine 48-linked ubiquitin-chain synthesis by the cullin-RING ubiquitin-ligase complex SCF-Cdc34.* Cell, 2005. **123**(6): p. 1107-20.
31. Johnson, E.S. and A.A. Gupta, *An E3-like factor that promotes SUMO conjugation to the yeast septins.* Cell, 2001. **106**(6): p. 735-44.
32. Kagey, M.H., T.A. Melhuish, and D. Wotton, *The polycomb protein Pc2 is a SUMO E3.* Cell, 2003. **113**(1): p. 127-37.
33. Kile, B.T., et al., *The SOCS box: a tale of destruction and degradation.* Trends Biochem Sci, 2002. **27**(5): p. 235-41.
34. Kamura, T., et al., *The Elongin BC complex interacts with the conserved SOCS-box motif present in members of the SOCS, ras, WD-40 repeat, and ankyrin repeat families.* Genes Dev, 1998. **12**(24): p. 3872-81.
35. Kamura, T., et al., *VHL-box and SOCS-box domains determine binding specificity for Cul2-Rbx1 and Cul5-Rbx2 modules of ubiquitin ligases.* Genes Dev, 2004. **18**(24): p. 3055-65.
36. Stebbins, C.E., W.G. Kaelin, and N.P. Pavletich, *Structure of the VHL-ElonginC-ElonginB complex: Implications for VHL tumor suppressor function.* Science, 1999. **284**(5413): p. 455-461.
37. Terpe, K., *Overview of bacterial expression systems for heterologous protein production: from molecular and biochemical fundamentals to commercial systems.* Appl Microbiol Biotechnol, 2006. **72**(2): p. 211-22.
38. Schmidt, F.R., *Recombinant expression systems in the pharmaceutical industry.* Appl Microbiol Biotechnol, 2004. **65**(4): p. 363-72.
39. Erlich, H.A., D. Gelfand, and J.J. Sninsky, *Recent advances in the polymerase chain reaction.* Science, 1991. **252**(5013): p. 1643-51.
40. Mullis, K.B., *Target amplification for DNA analysis by the polymerase chain reaction.* Ann Biol Clin (Paris), 1990. **48**(8): p. 579-82.
41. Chiocca, S., et al., *The complete DNA sequence and genomic organization of the avian adenovirus CELO.* J Virol, 1996. **70**(5): p. 2939-49.
42. Weiss, B., et al., *Enzymatic breakage and joining of deoxyribonucleic acid. VI. Further purification and properties of polynucleotide ligase from Escherichia coli infected with bacteriophage T4.* J Biol Chem, 1968. **243**(17): p. 4543-55.
43. Jesse, J., *Focus® 6:4.* 1984.
44. Sambrook, J.e.a., *Molecular Cloning: A Laboratory Manual.* 1989. **2nd**.
45. Shuman, S., *Novel approach to molecular cloning and polynucleotide synthesis using vaccinia DNA topoisomerase.* J Biol Chem, 1994. **269**(51): p. 32678-84.

46. Cheng, C. and S. Shuman, *Recombinogenic flap ligation pathway for intrinsic repair of topoisomerase IB-induced double-strand breaks.* Mol Cell Biol, 2000. **20**(21): p. 8059-68.
47. Shuman, S., *Recombination mediated by vaccinia virus DNA topoisomerase I in Escherichia coli is sequence specific.* Proc Natl Acad Sci U S A, 1991. **88**(22): p. 10104-8.
48. Landy, A., *Dynamic, structural, and regulatory aspects of lambda site-specific recombination.* Annu Rev Biochem, 1989. **58**: p. 913-49.
49. Oakley, A., *Glutathione transferases: a structural perspective.* Drug Metab Rev, 2011. **43**(2): p. 138-51.
50. Merz, F., et al., *Molecular mechanism and structure of Trigger Factor bound to the translating ribosome.* EMBO J, 2008. **27**(11): p. 1622-32.
51. Ferbitz, L., et al., *Trigger factor in complex with the ribosome forms a molecular cradle for nascent proteins.* Nature, 2004. **431**(7008): p. 590-596.
52. Patzelt, H., et al., *Three-state equilibrium of Escherichia coli trigger factor.* Biol Chem, 2002. **383**(10): p. 1611-9.
53. Maier, T., et al., *A cradle for new proteins: trigger factor at the ribosome.* Curr Opin Struct Biol, 2005. **15**(2): p. 204-12.
54. Jiang, X., et al., *Reduction of protein degradation by use of protease-deficient mutants in cell-free protein synthesis system of Escherichia coli.* J Biosci Bioeng, 2002. **93**(2): p. 151-6.
55. Seidel, H.M., D.L. Pompliano, and J.R. Knowles, *Phosphonate biosynthesis: molecular cloning of the gene for phosphoenolpyruvate mutase from Tetrahymena pyriformis and overexpression of the gene product in Escherichia coli.* Biochemistry, 1992. **31**(9): p. 2598-608.
56. Kane, J.F., *Effects of rare codon clusters on high-level expression of heterologous proteins in Escherichia coli.* Curr Opin Biotechnol, 1995. **6**(5): p. 494-500.
57. Bornhorst, J.A. and J.J. Falke, *Purification of proteins using polyhistidine affinity tags.* Methods Enzymol, 2000. **326**: p. 245-54.
58. Harper, S. and D.W. Speicher, *Purification of proteins fused to glutathione S-transferase.* Methods Mol Biol, 2011. **681**: p. 259-80.
59. Shuichi Yamamoto, e.a., *Ion-Exchange Chromatography of Proteins.* Chromatographic Science Series, 1988. **43**: p. 35-78.
60. Kostanski, L.K., D.M. Keller, and A.E. Hamielec, *Size-exclusion chromatography-a review of calibration methodologies.* J Biochem Biophys Methods, 2004. **58**(2): p. 159-86.
61. Rosas-Acosta, G., et al., *Proteins of the PIAS family enhance the sumoylation of the papillomavirus E1 protein.* Virology, 2005. **331**(1): p. 190-203.
62. Pal, S., J.M. Rosas, and G. Rosas-Acosta, *Identification of the non-structural influenza A viral protein NS1A as a bona fide target of the Small Ubiquitin-like MOdifier by the use of dicistronic expression constructs.* J Virol Methods, 2010. **163**(2): p. 498-504.
63. Pal, S., et al., *Influenza A virus interacts extensively with the cellular SUMOylation system during infection.* Virus Res, 2011. **158**(1-2): p. 12-27.
64. Santos, A., et al., *SUMOylation affects the interferon blocking activity of the influenza A nonstructural protein NS1 without affecting its stability or cellular localization.* J Virol, 2013. **87**(10): p. 5602-20.
65. Galluccio, M., et al., *Over-Expression in E. coli and Purification of the Human OCTN2 Transport Protein.* Molecular Biotechnology, 2012. **50**(1): p. 1-7.
66. Crowe, J., et al., *6xHis-Ni-NTA chromatography as a superior technique in recombinant protein expression/purification.* Methods Mol Biol, 1994. **31**: p. 371-87.
67. Schmitt, J., H. Hess, and H.G. Stunnenberg, *Affinity purification of histidine-tagged proteins.* Mol Biol Rep, 1993. **18**(3): p. 223-30.

68. Westra, D.F., et al., *Immobilised metal-ion affinity chromatography purification of histidine-tagged recombinant proteins: a wash step with a low concentration of EDTA.* J Chromatogr B Biomed Sci Appl, 2001. **760**(1): p. 129-36.
69. Baneyx, F. and M. Mujacic, *Recombinant protein folding and misfolding in Escherichia coli.* Nat Biotechnol, 2004. **22**(11): p. 1399-408.
70. Siurkus, J. and P. Neubauer, *Reducing conditions are the key for efficient production of active ribonuclease inhibitor in Escherichia coli.* Microb Cell Fact, 2011. **10**: p. 31.
71. Tindall, K.R. and T.A. Kunkel, *Fidelity of DNA synthesis by the Thermus aquaticus DNA polymerase.* Biochemistry, 1988. **27**(16): p. 6008-13.
72. Eckert, K.A. and T.A. Kunkel, *DNA polymerase fidelity and the polymerase chain reaction.* PCR Methods Appl, 1991. **1**(1): p. 17-24.
73. Frey, B.S., B., in *BioChemica*. 1995. p. 34-35.
74. Chester, N. and D.R. Marshak, *Dimethyl sulfoxide-mediated primer Tm reduction: a method for analyzing the role of renaturation temperature in the polymerase chain reaction.* Anal Biochem, 1993. **209**(2): p. 284-90.
75. Fahnert, B., H. Lilie, and P. Neubauer, *Inclusion bodies: formation and utilisation.* Adv Biochem Eng Biotechnol, 2004. **89**: p. 93-142.
76. Villaverde, A. and M.M. Carrio, *Protein aggregation in recombinant bacteria: biological role of inclusion bodies.* Biotechnol Lett, 2003. **25**(17): p. 1385-95.
77. Miroux, B. and J.E. Walker, *Over-production of proteins in Escherichia coli: mutant hosts that allow synthesis of some membrane proteins and globular proteins at high levels.* J Mol Biol, 1996. **260**(3): p. 289-98.
78. Rosano, G.L. and E.A. Ceccarelli, *Rare codon content affects the solubility of recombinant proteins in a codon bias-adjusted Escherichia coli strain.* Microb Cell Fact, 2009. **8**: p. 41.
79. Higgins, S.J. and B.D. Hames, *Protein expression : a practical approach*. The practical approach series. 1999, Oxford ; New York: Oxford University Press. xix, 282 p.
80. Studier, F.W., *Use of bacteriophage T7 lysozyme to improve an inducible T7 expression system.* J Mol Biol, 1991. **219**(1): p. 37-44.
81. Kang, Y., M.S. Son, and T.T. Hoang, *One step engineering of T7-expression strains for protein production: increasing the host-range of the T7-expression system.* Protein Expr Purif, 2007. **55**(2): p. 325-33.
82. de Marco, A. and V. De Marco, *Bacteria co-transformed with recombinant proteins and chaperones cloned in independent plasmids are suitable for expression tuning.* J Biotechnol, 2004. **109**(1-2): p. 45-52.
83. San-Miguel, T., P. Perez-Bermudez, and I. Gavidia, *Production of soluble eukaryotic recombinant proteins in is favoured in early log-phase cultures induced at low temperature.* Springerplus, 2013. **2**(1): p. 89.
84. Kataeva, I., et al., *Improving solubility of Shewanella oneidensis MR-1 and Clostridium thermocellum JW-20 proteins expressed into Esherichia coli.* J Proteome Res, 2005. **4**(6): p. 1942-51.
85. Qing, G., et al., *Cold-shock induced high-yield protein production in Escherichia coli.* Nat Biotechnol, 2004. **22**(7): p. 877-82.
86. Phadtare, S., J. Alsina, and M. Inouye, *Cold-shock response and cold-shock proteins.* Curr Opin Microbiol, 1999. **2**(2): p. 175-80.
87. Vasina, J.A. and F. Baneyx, *Expression of aggregation-prone recombinant proteins at low temperatures: a comparative study of the Escherichia coli cspA and tac promoter systems.* Protein Expr Purif, 1997. **9**(2): p. 211-8.

88. Vasina, J.A., M.S. Peterson, and F. Baneyx, *Scale-up and optimization of the low-temperature inducible cspA promoter system.* Biotechnol Prog, 1998. **14**(5): p. 714-21.
89. Chant, A., et al., *Attachment of a histidine tag to the minimal zinc finger protein of the Aspergillus nidulans gene regulatory protein AreA causes a conformational change at the DNA-binding site.* Protein Expr Purif, 2005. **39**(2): p. 152-9.
90. Fonda, I., et al., *Attachment of histidine tags to recombinant tumor necrosis factor-alpha drastically changes its properties.* ScientificWorldJournal, 2002. **2**: p. 1312-25.
91. Dias, C.L., et al., *The hydrophobic effect and its role in cold denaturation.* Cryobiology, 2010. **60**(1): p. 91-9.
92. Strocchi, M., et al., *Low temperature-induced systems failure in Escherichia coli: insights from rescue by cold-adapted chaperones.* Proteomics, 2006. **6**(1): p. 193-206.
93. Hefti, M.H., et al., *A novel purification method for histidine-tagged proteins containing a thrombin cleavage site.* Anal Biochem, 2001. **295**(2): p. 180-5.
94. Chi, E.Y., et al., *Physical stability of proteins in aqueous solution: mechanism and driving forces in nonnative protein aggregation.* Pharm Res, 2003. **20**(9): p. 1325-36.
95. Smyth, D.R., et al., *Crystal structures of fusion proteins with large-affinity tags.* Protein Sci, 2003. **12**(7): p. 1313-22.
96. Rossmann, M.G., et al., *Molecular Replacement Real-Space Averaging.* Journal of Applied Crystallography, 1992. **25**: p. 166-180.
97. Fraser, M.J., *The Baculovirus-Infected Insect Cell as a Eukaryotic Gene-Expression System.* Current Topics in Microbiology and Immunology, 1992. **158**: p. 131-172.
98. Imperiale, M.J. and A. Casadevall, *Bioterrorism: Lessons Learned Since the Anthrax Mailings.* Mbio, 2011. **2**(6).
99. Spencer, R.C., *Bacillus anthracis.* Journal of Clinical Pathology, 2003. **56**(3): p. 182-187.
100. Dixon, T.C., et al., *Anthrax.* New England Journal of Medicine, 1999. **341**(11): p. 815-826.
101. Sweeney, D.A., et al., *Anthrax Infection.* American Journal of Respiratory and Critical Care Medicine, 2011. **184**(12): p. 1333-1341.
102. Young, J.A. and R.J. Collier, *Anthrax toxin: receptor binding, internalization, pore formation, and translocation.* Annu Rev Biochem, 2007. **76**: p. 243-65.
103. Beauregard, K.E., et al., *Anthrax toxin entry into polarized epithelial cells.* Infection and Immunity, 1999. **67**(6): p. 3026-3030.
104. Sun, J., et al., *Insertion of anthrax protective antigen into liposomal membranes: effects of a receptor.* J Biol Chem, 2007. **282**(2): p. 1059-65.
105. Scobie, H.M., et al., *Human capillary morphogenesis protein 2 functions as an anthrax toxin receptor.* Proc Natl Acad Sci U S A, 2003. **100**(9): p. 5170-4.
106. Bradley, K.A., et al., *Identification of the cellular receptor for anthrax toxin.* Nature, 2001. **414**(6860): p. 225-9.
107. Zhang, S., A. Finkelstein, and R.J. Collier, *Evidence that translocation of anthrax toxin's lethal factor is initiated by entry of its N terminus into the protective antigen channel.* Proc Natl Acad Sci U S A, 2004. **101**(48): p. 16756-61.
108. Lacy, D.B., et al., *Structure of heptameric protective antigen bound to an anthrax toxin receptor: a role for receptor in pH-dependent pore formation.* Proc Natl Acad Sci U S A, 2004. **101**(36): p. 13147-51.
109. Williams, A.S., et al., *Domain 4 of the anthrax protective antigen maintains structure and binding to the host receptor CMG2 at low pH.* Protein Sci, 2009. **18**(11): p. 2277-86.
110. Scobie, H.M., et al., *Anthrax toxin receptor 2-dependent lethal toxin killing in vivo.* PLoS Pathog, 2006. **2**(10): p. e111.

111. Lacy, D.B., et al., *Crystal structure of the von Willebrand factor A domain of human capillary morphogenesis protein 2: an anthrax toxin receptor.* Proc Natl Acad Sci U S A, 2004. **101**(17): p. 6367-72.
112. Sun, J. and R.J. Collier, *Disulfide bonds in the ectodomain of anthrax toxin receptor 2 are required for the receptor-bound protective-antigen pore to function.* PLoS One, 2010. **5**(5): p. e10553.
113. Jacquez, P., et al., *Expression and purification of the functional ectodomain of human anthrax toxin receptor 2 in Escherichia coli Origami B cells with assistance of bacterial Trigger Factor.* Protein Expr Purif, 2013. **95C**: p. 149-155.
114. Kourkoutis, L.F., J.M. Plitzko, and W. Baumeister, *Electron Microscopy of Biological Materials at the Nanometer Scale.* Annual Review of Materials Research, Vol 42, 2012. **42**: p. 33-58.
115. Dubochet, J., et al., *Cryo-Electron Microscopy of Vitrified Specimens.* Quarterly Reviews of Biophysics, 1988. **21**(2): p. 129-228.
116. Baker, M.L., et al., *Cryo-EM of macromolecular assemblies at near-atomic resolution.* Nat Protoc, 2010. **5**(10): p. 1697-708.
117. Chang, J., et al., *Reconstructing virus structures from nanometer to near-atomic resolutions with cryo-electron microscopy and tomography.* Adv Exp Med Biol, 2012. **726**: p. 49-90.
118. Kostyuchenko, V.A., et al., *Near-atomic resolution cryo-electron microscopic structure of dengue serotype 4 virus.* J Virol, 2014. **88**(1): p. 477-82.
119. Grigorieff, N., *FREALIGN: high-resolution refinement of single particle structures.* J Struct Biol, 2007. **157**(1): p. 117-25.
120. Ludtke, S.J., P.R. Baldwin, and W. Chiu, *EMAN: semiautomated software for high-resolution single-particle reconstructions.* J Struct Biol, 1999. **128**(1): p. 82-97.
121. Tang, G., et al., *EMAN2: an extensible image processing suite for electron microscopy.* J Struct Biol, 2007. **157**(1): p. 38-46.
122. Ludtke, S.J., *3-D structures of macromolecules using single-particle analysis in EMAN.* Methods Mol Biol, 2010. **673**: p. 157-73.
123. Pettersen, E.F., et al., *UCSF Chimera--a visualization system for exploratory research and analysis.* J Comput Chem, 2004. **25**(13): p. 1605-12.
124. Goddard, T.D., C.C. Huang, and T.E. Ferrin, *Visualizing density maps with UCSF Chimera.* J Struct Biol, 2007. **157**(1): p. 281-7.
125. Deuquet, J., et al., *Hyaline fibromatosis syndrome inducing mutations in the ectodomain of anthrax toxin receptor 2 can be rescued by proteasome inhibitors.* EMBO Mol Med, 2011. **3**(4): p. 208-21.
126. Zhang, W., et al., *Heterogeneity of large macromolecular complexes revealed by 3D cryo-EM variance analysis.* Structure, 2008. **16**(12): p. 1770-6.
127. Orlova, E.V. and H.R. Saibil, *Structure determination of macromolecular assemblies by single-particle analysis of cryo-electron micrographs.* Curr Opin Struct Biol, 2004. **14**(5): p. 584-90.
128. Scheres, S.H., et al., *Disentangling conformational states of macromolecules in 3D-EM through likelihood optimization.* Nat Methods, 2007. **4**(1): p. 27-9.
129. Liu, Y., X. Meng, and Z. Liu, *Deformed grids for single-particle cryo-electron microscopy of specimens exhibiting a preferred orientation.* J Struct Biol, 2013. **182**(3): p. 255-8.
130. van Antwerpen, R., *Preferred orientations of LDL in vitreous ice indicate a discoid shape of the lipoprotein particle.* Archives of Biochemistry and Biophysics, 2004. **432**(1): p. 122-127.
131. Glaeser, R.M. and K.H. Downing, *Specimen charging on thin films with one conducting layer: discussion of physical principles.* Microsc Microanal, 2004. **10**(6): p. 790-6.
132. Katayama, H., et al., *GroEL as a molecular scaffold for structural analysis of the anthrax toxin pore.* Nat Struct Mol Biol, 2008. **15**(7): p. 754-60.

133. Katayama, H., et al., *Three-dimensional structure of the anthrax toxin pore inserted into lipid nanodiscs and lipid vesicles.* Proc Natl Acad Sci U S A, 2010. **107**(8): p. 3453-7.
134. Sun, J., et al., *Phenylalanine-427 of anthrax protective antigen functions in both pore formation and protein translocation.* Proc Natl Acad Sci U S A, 2008. **105**(11): p. 4346-51.
135. Krantz, B.A., et al., *A phenylalanine clamp catalyzes protein translocation through the anthrax toxin pore.* Science, 2005. **309**(5735): p. 777-81.
136. Akkaladevi, N., et al., *Assembly of anthrax toxin pore: lethal-factor complexes into lipid nanodiscs.* Protein Sci, 2013. **22**(4): p. 492-501.
137. Gogol, E.P., et al., *Three dimensional structure of the anthrax toxin translocon-lethal factor complex by cryo-electron microscopy.* Protein Sci, 2013. **22**(5): p. 586-94.
138. Colten, H. *Sleep Disorders and Sleep Deprivation: An Unmet Public Health Problem*. National Academies Press, 2006.
139. Green, C.B., J.S. Takahashi, and J. Bass, *The meter of metabolism.* Cell, 2008. **134**(5): p. 728-42.
140. Reppert, S.M. and D.R. Weaver, *Molecular analysis of mammalian circadian rhythms.* Annu Rev Physiol, 2001. **63**: p. 647-76.
141. Williams, G., et al., *Therapy of circadian rhythm disorders in chronic fatigue syndrome: no symptomatic improvement with melatonin or phototherapy.* Eur J Clin Invest, 2002. **32**(11): p. 831-7.
142. Zehring, W.A., et al., *P-element transformation with period locus DNA restores rhythmicity to mutant, arrhythmic Drosophila melanogaster.* Cell, 1984. **39**(2 Pt 1): p. 369-76.
143. Konopka, R.J. and S. Benzer, *Clock mutants of Drosophila melanogaster.* Proc Natl Acad Sci U S A, 1971. **68**(9): p. 2112-6.
144. Zhang, E.E. and S.A. Kay, *Clocks not winding down: unravelling circadian networks.* Nat Rev Mol Cell Biol, 2010. **11**(11): p. 764-76.
145. Panda, S., et al., *Coordinated transcription of key pathways in the mouse by the circadian clock.* Cell, 2002. **109**(3): p. 307-20.
146. Dunlap, J.C., *Molecular bases for circadian clocks.* Cell, 1999. **96**(2): p. 271-90.
147. Panda, S., J.B. Hogenesch, and S.A. Kay, *Circadian rhythms from flies to human.* Nature, 2002. **417**(6886): p. 329-35.
148. Gegear, R.J., et al., *Animal cryptochromes mediate magnetoreception by an unconventional photochemical mechanism.* Nature, 2010. **463**(7282): p. 804-7.
149. Zhang, E.E., et al., *Cryptochrome mediates circadian regulation of cAMP signaling and hepatic gluconeogenesis.* Nat Med, 2010. **16**(10): p. 1152-6.
150. Lin, C. and T. Todo, *The cryptochromes.* Genome Biol, 2005. **6**(5): p. 220.
151. Cashmore, A.R., *Cryptochromes: enabling plants and animals to determine circadian time.* Cell, 2003. **114**(5): p. 537-43.
152. Sancar, A., *Structure and function of DNA photolyase and cryptochrome blue-light photoreceptors.* Chem Rev, 2003. **103**(6): p. 2203-37.
153. Yoshii, T., M. Ahmad, and C. Helfrich-Forster, *Cryptochrome mediates light-dependent magnetosensitivity of Drosophila's circadian clock.* PLoS Biol, 2009. **7**(4): p. e1000086.
154. Vieira, J., et al., *Human cryptochrome-1 confers light independent biological activity in transgenic Drosophila correlated with flavin radical stability.* PLoS One, 2012. **7**(3): p. e31867.
155. Zoltowski, B.D., et al., *Structure of full-length Drosophila cryptochrome.* Nature, 2011. **480**(7377): p. 396-9.
156. Partch, C.L. and A. Sancar, *Photochemistry and photobiology of cryptochrome blue-light photopigments: the search for a photocycle.* Photochem Photobiol, 2005. **81**(6): p. 1291-304.

157. Hitomi, K., et al., *Functional motifs in the (6-4) photolyase crystal structure make a comparative framework for DNA repair photolyases and clock cryptochromes.* Proc Natl Acad Sci U S A, 2009. **106**(17): p. 6962-7.
158. Yu, X., et al., *The Cryptochrome Blue Light Receptors.* Arabidopsis Book, 2010. **8**: p. e0135.
159. Sancar, A., *Cryptochrome: the second photoactive pigment in the eye and its role in circadian photoreception.* Annu Rev Biochem, 2000. **69**: p. 31-67.
160. Xing, W., et al., *SCF(FBXL3) ubiquitin ligase targets cryptochromes at their cofactor pocket.* Nature, 2013. **496**(7443): p. 64-8.
161. Hsu, D.S., et al., *Putative human blue-light photoreceptors hCRY1 and hCRY2 are flavoproteins.* Biochemistry, 1996. **35**(44): p. 13871-7.
162. Ozgur, S. and A. Sancar, *Purification and properties of human blue-light photoreceptor cryptochrome 2.* Biochemistry, 2003. **42**(10): p. 2926-32.
163. Ozgur, S. and A. Sancar, *Analysis of autophosphorylating kinase activities of Arabidopsis and human cryptochromes.* Biochemistry, 2006. **45**(44): p. 13369-74.
164. Einhauer, A., et al., *Expression and purification of homogenous proteins in Saccharomyces cerevisiae based on ubiquitin-FLAG fusion.* Protein Expr Purif, 2002. **24**(3): p. 497-504.
165. Blanar, M.A. and W.J. Rutter, *Interaction cloning: identification of a helix-loop-helix zipper protein that interacts with c-Fos.* Science, 1992. **256**(5059): p. 1014-8.
166. Einhauer, A. and A. Jungbauer, *The FLAG peptide, a versatile fusion tag for the purification of recombinant proteins.* J Biochem Biophys Methods, 2001. **49**(1-3): p. 455-65.
167. Pryor, K.D. and B. Leiting, *High-level expression of soluble protein in Escherichia coli using a His(6)-tag and maltose-binding-protein double-affinity fusion system.* Protein Expression and Purification, 1997. **10**(3): p. 309-319.
168. Kapust, R.B. and D.S. Waugh, *Escherichia coli maltose-binding protein is uncommonly effective at promoting the solubility of polypeptides to which it is fused.* Protein Sci, 1999. **8**(8): p. 1668-74.
169. Riggs, P., *Expression and purification of recombinant proteins by fusion to maltose-binding protein.* Mol Biotechnol, 2000. **15**(1): p. 51-63.
170. Weiner, M.P.e.a., *Strategies*. 1994. p. 41-43.
171. Fox, J.D. and D.S. Waugh, *Maltose-binding protein as a solubility enhancer.* Methods Mol Biol, 2003. **205**: p. 99-117.
172. Czarna, A., et al., *Structures of Drosophila cryptochrome and mouse cryptochrome1 provide insight into circadian function.* Cell, 2013. **153**(6): p. 1394-405.
173. Lamia, K.A., et al., *AMPK regulates the circadian clock by cryptochrome phosphorylation and degradation.* Science, 2009. **326**(5951): p. 437-40.
174. Hirota, T., et al., *Identification of small molecule activators of cryptochrome.* Science, 2012. **337**(6098): p. 1094-7.
175. Nangle, S., W. Xing, and N. Zheng, *Crystal structure of mammalian cryptochrome in complex with a small molecule competitor of its ubiquitin ligase.* Cell Res, 2013. **23**(12): p. 1417-9.
176. Ozber, N., et al., *Identification of two amino acids in the C-terminal domain of mouse CRY2 essential for PER2 interaction.* BMC Mol Biol, 2010. **11**: p. 69.
177. Khan, S.K., et al., *Identification of a novel cryptochrome differentiating domain required for feedback repression in circadian clock function.* J Biol Chem, 2012. **287**(31): p. 25917-26.
178. Chaves, I., et al., *Functional evolution of the photolyase/cryptochrome protein family: importance of the C terminus of mammalian CRY1 for circadian core oscillator performance.* Mol Cell Biol, 2006. **26**(5): p. 1743-53.
179. Partch, C.L., et al., *Role of structural plasticity in signal transduction by the cryptochrome blue-light photoreceptor.* Biochemistry, 2005. **44**(10): p. 3795-805.

180. Gao, P., et al., *Phosphorylation of the cryptochrome 1 C-terminal tail regulates circadian period length.* J Biol Chem, 2013. **288**(49): p. 35277-86.
181. Huang, Y., et al., *Crystal structure of cryptochrome 3 from Arabidopsis thaliana and its implications for photolyase activity.* Proc Natl Acad Sci U S A, 2006. **103**(47): p. 17701-6.
182. Park, H.W., et al., *Crystal structure of DNA photolyase from Escherichia coli.* Science, 1995. **268**(5219): p. 1866-72.
183. Ladd, M.F.C. and R.A. Palmer, *Structure determination by X-ray crystallography*. 4th ed. 2003, New York: Kluwer Academic/Plenum Publishers. xlii, 819 p.
184. Rossmann, M.G.A., E.V., *International Tables for Crystallography*. Crystallography of Biological Macromolecules. Vol. F. 2001: Springer.
185. Institute, N.C., *Surveillance, Epidemiology, and End Results Program: Tuning Cancer Data Into Discovery.* 2013.
186. Liao, H.Y., et al., *Differential receptor binding affinities of influenza hemagglutinins on glycan arrays.* J Am Chem Soc, 2010. **132**(42): p. 14849-56.
187. Hart, G.W. and R.J. Copeland, *Glycomics hits the big time.* Cell, 2010. **143**(5): p. 672-6.
188. Jelkmann, W., *The enigma of the metabolic fate of circulating erythropoietin (Epo) in view of the pharmacokinetics of the recombinant drugs rhEpo and NESP.* Eur J Haematol, 2002. **69**(5-6): p. 265-74.
189. Lindahl, U., *'Heparin'--from anticoagulant drug into the new biology.* Glycoconj J, 2000. **17**(7-9): p. 597-605.
190. Evans, D.W., et al., *Structural insights into parallel strategies for germline antibody recognition of lipopolysaccharide from Chlamydia.* Glycobiology, 2011. **21**(8): p. 1049-59.
191. Wu, C.S., et al., *Cancer-Associated Carbohydrate Antigens as Potential Biomarkers for Hepatocellular Carcinoma.* Plos One, 2012. **7**(7).
192. Slovin, S.F., S.J. Keding, and G. Ragupathi, *Carbohydrate vaccines as immunotherapy for cancer.* Immunol Cell Biol, 2005. **83**(4): p. 418-28.
193. Schroeder, H.W., Jr. and L. Cavacini, *Structure and function of immunoglobulins.* J Allergy Clin Immunol, 2010. **125**(2 Suppl 2): p. S41-52.
194. Steplewski, Z., et al., *Hypothesis - Macrophages as Effector-Cells for Human-Tumor Destruction Mediated by Monoclonal-Antibody.* Hybridoma, 1983. **2**(1): p. 1-5.
195. Shields, R.L., et al., *High resolution mapping of the binding site on human IgG1 for Fc gamma RI, Fc gamma RII, Fc gamma RIII, and FcRn and design of IgG1 variants with improved binding to the Fc gamma R.* J Biol Chem, 2001. **276**(9): p. 6591-604.
196. Elgert, K., *Immunology: Understanding the Immune System.* 1998: p. 58-78.
197. Davies, D.R. and S. Chacko, *Antibody Structure.* Accounts of Chemical Research, 1993. **26**(8): p. 421-427.
198. Nelson, D.C., M, *principles of biochemistry.* 2005. **4**.
199. Davies, D.R., E.A. Padlan, and S. Sheriff, *Antibody-antigen complexes.* Annu Rev Biochem, 1990. **59**: p. 439-73.
200. Oliver, J., *Antibody Applications*. 2013.
201. Gerstenbruch, S., et al., *Analysis of cross-reactive and specific anti-carbohydrate antibodies against lipopolysaccharide from Chlamydophila psittaci.* Glycobiology, 2010. **20**(4): p. 461-72.
202. Lee, Y.C. and R.T. Lee, *Carbohydrate-Protein Interactions - Basis of Glycobiology.* Accounts of Chemical Research, 1995. **28**(8): p. 321-327.
203. Shibata, T.K., et al., *Identification of mono- and disulfated N-acetyl-lactosaminyl Oligosaccharide structures as epitopes specifically recognized by humanized monoclonal antibody HMOCC-1 raised against ovarian cancer.* J Biol Chem, 2012. **287**(9): p. 6592-602.

204. Song, W., et al., *MUC1 glycopeptide epitopes predicted by computational glycomics.* Int J Oncol, 2012. **41**(6): p. 1977-84.
205. Deisenhofer, J., *Crystallographic refinement and atomic models of a human Fc fragment and its complex with fragment B of protein A from Staphylococcus aureus at 2.9- and 2.8-A resolution.* Biochemistry, 1981. **20**(9): p. 2361-70.
206. Nardella, F.A., et al., *IgG rheumatoid factors and staphylococcal protein A bind to a common molecular site on IgG.* J Exp Med, 1985. **162**(6): p. 1811-24.
207. Boyle, M.D.P. and K.J. Reis, *Bacterial Fc-Receptors.* Bio-Technology, 1987. **5**(7): p. 697-703.
208. Hjelm, H., K. Hjelm, and J. Sjoquist, *Protein A from Staphylococcus aureus. Its isolation by affinity chromatography and its use as an immunosorbent for isolation of immunoglobulins.* FEBS Lett, 1972. **28**(1): p. 73-6.
209. Ey, P.L., S.J. Prowse, and C.R. Jenkin, *Isolation of pure IgG1, IgG2a and IgG2b immunoglobulins from mouse serum using protein A-sepharose.* Immunochemistry, 1978. **15**(7): p. 429-36.
210. Akerstrom, B., et al., *Protein G: a powerful tool for binding and detection of monoclonal and polyclonal antibodies.* J Immunol, 1985. **135**(4): p. 2589-92.
211. Akerstrom, B. and L. Bjorck, *A physicochemical study of protein G, a molecule with unique immunoglobulin G-binding properties.* J Biol Chem, 1986. **261**(22): p. 10240-7.
212. Balls, A.K. and H. Lineweaver, *Isolation and properties of crystalline papain.* Journal of Biological Chemistry, 1939. **130**(2): p. 669-686.
213. Nisonoff, A., F.C. Wissler, and L.N. Lipman, *Properties of the major component of a peptic digest of rabbit antibody.* Science, 1960. **132**(3441): p. 1770-1.
214. Putnam, F.W., Y.S. Liu, and T.L. Low, *Primary structure of a human IgA1 immunoglobulin. IV. Streptococcal IgA1 protease, digestion, Fab and Fc fragments, and the complete amino acid sequence of the alpha 1 heavy chain.* J Biol Chem, 1979. **254**(8): p. 2865-74.
215. Leiner, I.E.a.F., B., *Ficin.* Methods Enzymol, 1970. **19**: p. 261-273.
216. Kurkela, R., L. Vuolas, and P. Vihko, *Preparation of F(ab')2 fragments from monoclonal mouse IgG1 suitable for use in radioimaging.* J Immunol Methods, 1988. **110**(2): p. 229-36.
217. Mariani, M., et al., *A new enzymatic method to obtain high-yield F(ab)2 suitable for clinical use from mouse IgGl.* Mol Immunol, 1991. **28**(1-2): p. 69-77.
218. Tsai, W.C. and P.J.R. Pai, *Surface plasmon resonance-based immunosensor with oriented immobilized antibody fragments on a mixed self-assembled monolayer for the determination of staphylococcal enterotoxin B.* Microchimica Acta, 2009. **166**(1-2): p. 115-122.
219. Saphire, E.O., et al., *Contrasting IgG structures reveal extreme asymmetry and flexibility.* J Mol Biol, 2002. **319**(1): p. 9-18.
220. Sourial, S. and C. Nilsson, *HIV-2 neutralization by intact V3-specific Fab fragments.* Virol J, 2008. **5**: p. 96.
221. Eriksson, A. and M. Norgren, *Cleavage of antigen-bound immunoglobulin G by SpeB contributes to streptococcal persistence in opsonizing blood.* Infect Immun, 2003. **71**(1): p. 211-7.
222. Colcher, D., et al., *In vivo tumor targeting of a recombinant single-chain antigen-binding protein.* J Natl Cancer Inst, 1990. **82**(14): p. 1191-7.
223. Worn, A. and A. Pluckthun, *Stability engineering of antibody single-chain Fv fragments.* J Mol Biol, 2001. **305**(5): p. 989-1010.
224. Calarese, D.A., et al., *Dissection of the carbohydrate specificity of the broadly neutralizing anti-HIV-1 antibody 2G12.* Proc Natl Acad Sci U S A, 2005. **102**(38): p. 13372-7.
225. Pejchal, R., et al., *A potent and broad neutralizing antibody recognizes and penetrates the HIV glycan shield.* Science, 2011. **334**(6059): p. 1097-103.

Vita

Gustavo A. Avila, Son to Manuel and Leticia Avila is from El Paso, TX. He received his Bachelors of Science Degree at the University of Texas at El Paso, where he continued his academic endeavors under the mentorship of Dr. Chuan Xiao. While pursuing his degree, he had the privilege of serving as teaching assistant to over 300 students in chemistry lab, as well as teaching over 15 undergraduate and graduate students laboratory techniques. During his tenure as PhD, he has accrued 2 first author publications and one co-author. Gustavo is fascinated by the intricate nature of the cell and how proteins can impinge on cellular functions disturbing its homeostasis and potentially resulting in deleterious effects on the host. In order to better understand the events transpiring within the cell, it is important to investigate both the structure and the function of the proteins causing said abuses to develop novel treatments and therapeutic agents to combat the illnesses brought about by these imbalances.

Gustavo currently resides in El Paso, where he will seek employment opportunities in the area.

Permanent address: 3322 Wedgewood
El Paso, TX, 79925

This dissertation was typed by Gustavo A. Avila.