

CRYSTALLOGRAPHIC APPROACHES TO OBTAIN STRUCTURAL SOLUTIONS FOR DYNAMIC CAPSID PROTEINS MEDIATING VIRAL ASSEMBLY

A Thesis By

AARON LIM

ORCID iD: 0000-0003-0982-0614

California State University, Fullerton
Spring, 2022

In partial fulfillment of the degree:

Master of Science in Chemistry (Biochemistry)

Department:

Department of Chemistry and Biochemistry

Committee:

Niroshika Keppetipola, Department of Chemistry and Biochemistry, Chair

Madeline Rasche, Department of Chemistry and Biochemistry

Stevan Pecic, Department of Chemistry and Biochemistry

DOI:

10.5281/zenodo.6567978

Keywords:

bacteriophage, virus, assembly, crystallography, XRD, structure solution

Abstract:

The assembly pathway of Bacteriophage λ is conserved in many eukaryotic and prokaryotic viruses. The Major Capsid Protein (gpE) is a subunit of the virus protein head that is assisted by an intrinsically disordered scaffold protein (gpNu3) to polymerize into an icosahedron shell. gpE forms hexamer and pentamer intermediates called capsomers; however, the initial steps of gpE forming these intermediate structures are unclear. Determining an atomic-resolution image via X-ray diffraction (XRD) can reveal details of unique protein-protein interactions that occur when these proteins modulate assembly pathways seen in eukaryotic and prokaryotic viruses. Structural determination of capsid and other viral proteins can serve as targets for drug discovery for treatment of viral infections or development of phage nanotechnologies as therapeutics to target diseases using phage-display conjugated ligand delivery systems. Atomic structural characterization of the gpE-gpNu3 complex (ENu3) is lacking due to the instability of the complex spontaneously self-assembling into the capsid which is influenced by the disordered folding. In this work, novel crystallographic approaches were utilized to improve crystal size and morphology which potentially can solve the ENu3 complex structure. The ENu3 protein complex was co-expressed and purified by anion-exchange chromatography and size-exclusion chromatography. We obtained one fraction that contained pure gpE and gpNu3 with sizes of 38 kDa and 18 kDa, respectively. The proteins in this fraction formed needle clusters during the purification process. Microscale Matrix Seeding (MMS) was applied to further increase the improvement of size of needle clusters. Multiple rounds of MMS yielded novel rectangular plate-like crystals that will be used to collect XRD data and solve the structure. In this report, we discuss our work related to improvements to the protein purification and crystallographic processes as a step towards obtaining the atomic structure of ENu3.

TABLE OF CONTENTS

LIST OF TABLES	iv
LIST OF FIGURES	v
ACKNOWLEDGMENTS	vi
1. INTRODUCTION.....	1
Viruses versus Cells	1
Bacteriophages as Model Systems	3
Phage Lambda	3
Molecular Genetics.....	6
Medicine and Therapeutics	7
Lambda Assembly and Packaging Reaction	9
Terminase.....	12
The Major Capsid Protein (gpE)	13
The Scaffolding Protein (gpNu3).....	14
gpE Interactions with Decoration Proteins.....	15
Structural Homology Modeling of the Capsid.....	16
Cryo-EM: Lambda Reconstruction.....	17
Crosslink Mass Spectrometry Modeling	19
gpE Characterization	21
Research Aim.....	24
2. METHODOLOGY	25
Materials and Chemicals	25
ENu3 Initial Crystal Optimizations	25
Growth and Expression of Recombinant pET7CAP[E _{W308A}].....	25
ENu3 Purification	26
Anion-Exchange Chromatography.....	26
SDS-PAGE	27
Dialysis Buffer Exchange	27
Size-Exclusion Chromatography (SEC) of ENu3	28
ENu3 Crystallization.....	28
Crystallization Screen.....	28
Optimization	29
Microscale Seeding Matrix (MMS)	29
Crystal Harvesting	30
Diffraction Processing	30
3. RESULTS.....	31
Phase One	31
Initial Crystallization Screen	31
ENu3 Expression and Purification	32
Phase Two	35
ENu3 Expression and Purification	35
ENu3 Crystallization	39
Phase Three.....	41

ENu3 Expression and Purification	41
ENu3 Crystallization	45
Microscale Seeding	46
4. DISCUSSION	50
5. CONCLUSION	62
REFERENCES	64

LIST OF TABLES

<u>Table</u>	<u>Page</u>
1. Crystallization Screen Conditions	29
2. Phase 2 SEC Stoichiometric Ratios of ENu3	53
3. Phase 3 SEC Stoichiometric Ratios of ENu3	58

LIST OF FIGURES

<u>Figure</u>	<u>Page</u>
1. Virus Infecting Bacterial Cell and Lambda Genome	5
2. Lambda Capsid Assembly Pathway	10
3. Sequence Alignment of Lambda gpE	17
4. Cryo-EM gpE and HK97 Crystal Structures of MCP	18
5. CXMS Homology Model	20
6. Cryo-EM Model of gpE Capsomers	23
7. Initial ENU3 Crystallization Optimizations	32
8. SEC Protein Standards Elution Profile	33
9. Phase 1 SEC Purification	35
10. Phase 2 Anion-Exchange Chromatography	36
11. Phase 2 SEC Purification	38
12. Phase 2 ENU3 Crystallization Optimization	40
13. Phase 3 Anion-Exchange Chromatography	42
14. Phase 3 SEC Purification	44
15. Phase 3 ENU3 Crystallization Optimization	46
16. Round 1 ENU3 Microscale Seeding	47
17. Round 2 ENU3 Microscale Seeding	49
18. Conserved Residues in the gpE+gpD Trimer Spike	51
19. Histidine 82	57

ACKNOWLEDGMENTS

I would like to acknowledge the Department of Chemistry and Biochemistry and my committee members (Dr. Niroshika Keppetipola, Dr. Madeline Rasche, and Dr. Stevan Pecic) for the tremendous amount of mentorship and guidance I have received throughout my academic career that led me to become a better scientist. Thank you to Dr. Marcos Ortega for allowing me to conduct research in this laboratory. Most importantly, thank you to my family and friends for the endless support I have received while working towards this Master's degree. This thesis is dedicated to my grandfather, Frankie Taitague Quintanilla.

CHAPTER 1

INTRODUCTION

Viruses versus Cells

Macromolecules such as nucleic acid, proteins, and lipids are the core foundation of cellular entities. A set of specialized microbes comprised of DNA, protein, and lipid molecules that do not align with the Cell Theory called Viruses function in some ways that differ from cellular organisms. The clear distinction of dissimilarity between life and viral forms is that cells are self-replicative and can transcribe proteins needed to carry out mechanical work to meet cellular needs. In contrast, viruses are not self-replicative and cannot synthesize proteins independently, requiring a host organism for their survival. The viral genome size is approximately less than 200 kbp and contains sequences to produce all proteins needed to strategically carry out the viral needs such as infection and production of progeny. Viral nucleic acids have short sequences that code for gene products that transcriptionally regulate infectious mechanisms which control the activation or repression of lysogeny or lytic cycles (Baltimore, 1971; Bertani & Weigle, 1953; Blasche et al., 2013; Casjens & Hendrix, 2015; Jacob et al., 1962; Jacob & Wollman, 1953; Kaiser, 1957; Kaiser & Jacob, 1957; Twort 1915, Yulinery et al., 2019)

The temperate lysogenic lifestyle of a virus allows it to become dormant in the host through integration of the viral genome into sites within the host genome, combined together as one in where the viral genetic material is called a prophage (Jacob & Wollman, 1954; Lwoff, 1953; Lederberg & Lederberg, 1953). Replication of the host genome with integrated prophage would occur normally and so viral genetic information is reproduced. The lytic cycle will occur in which the prophage becomes excised from the host genome and will be replicated by host polymerases while the nucleic acid simultaneously undergoes transcription and translation to produce proteins which precursors that take part in the assembly pathway to form an infectious viral particle. Accumulation of these viral particles will lead to cell lysis and burst the cell so that the virus can proceed to infect the next host (Alberts et al, 2002; Casjens & Hendrix, 2015; Hendrix, 2002; Lwoff, 1953; Sam et al., 2004). The significance to

understanding the lytic cycle can be beneficial in drug discovery as viral proteins that regulate the switch to turn on or initiate the assembly pathway into an infectious virus can be targets to stop the production of viral proteins before mature genome-packaged viral particles are formed.

A key feature that distinguishes viruses from cells is the structure composition. These microscopic entities are comprised of mainly nucleic acid and protein, where some viruses contain a lipid envelope. The structure consists of many protein units that associate in an orderly arrangement packing to produce either a rod-filamentous or spherical icosahedron shape. This protein coat provides the protective barrier to enclose their nucleic acid, similar to the purpose of a cell nucleus (Bawden & Pirie, 1938; Bernal & Fankuchen, 1941; Casjens, 1985; Caspar & Klug, 1962; Crick & Watson, 1956; Franklin, 1955; Hohn & Katsura, 1977; Louten, 2016; Stanley & Loring, 1936). Cells have phospholipid membranes that enclose the genome and other content such as organelles and biomolecules needed for replication, energy production, and survival. Viruses have a simpler composition of protein or enveloped coat to protect their genome, yet the key feature that highlights the distinction between viruses and cells is the formation of an infectious virus which is only possible at the expense of a cellular host; thus, viral genetic material requires a host cell to produce gene products that dictate cell function to form progeny through unique mechanisms. The parasitic nature of viral existence historically has shaped the biosphere and continues to remain relevant if there is a biological cell host available.

Since viruses do not have the abilities of self-replication and self-producing as seen in prokaryotes and eukaryotes, these specialized forces violate the Cell Theory (Hooke) and are not considered to be classified under biological life. Although viruses contain the same components of cellular life forms such as nucleic acid and protein, viral nucleic acids are to be considered Selfish Genetic Elements (Forterre, 2005; Forterre, 2006; Koonin et al., 2015; Krupovic & Koonin, 2015; Louten, 2016; Owens et al., 2017; Rault & Forterre, 2015; Wolkowicz & Schaechter, 2008) as their genome requires external factors such as using host ribosomes to produce viral protein required to control subsequent steps involved in the virus pathogenicity. Virus genes dictate instruction and

benefit from cellular forms by hijacking cell machinery which increases viral transmission, integration, and fitness, thus being a selfish element.

Bacteriophages as Model Systems

Viruses will remain relevant as long as cell forms exist and present harm for the hierarchy of cellular life. Not only do viruses threaten large eukaryotes such as animals and plants, but microscopic prokaryotes are also subject to viral invasiveness. Microbiota that regulates ecosystems are host to viruses where infectious particles alter prokaryotic systems that can greatly affect higher life forms, thus supporting the general definition of a virus being a microscopic parasite (Ankrah et al., 2014; Batinovic et al., 2019; Hauser et al., 2012; Hendrix, 2002; Levinthal et al., 1967; Shao et al., 2017; Sharma et al., 2021; Suttle, C. A., 2007; Wolkowicz & Schaechter, 2008; Yulinery et al., 2019). Understanding the virulent characteristic nature can be explored using bacteriophages, which are viruses that infect bacteria and not eukaryotes. Many common characteristics can be observed in the diverse viral world, such as having small genomes, capsid and tail protein features, and the molecular mechanism of infection and transduction; thus, utilizing bacteria as a host for bacteriophages can be an effective tool to study the viral infectious nature to gain a deeper understanding of functional selfish elements (Fortier & Sekulovic, 2013).

Phage Lambda

Bacteriophages such as λ , T7, P22, and HK97 are few of many phage viruses that are commonly used to study virus function in bacteria (Campbell, 1986; Hauser et al., 2012; Hendrix, 2002; Hershey, 1952). Bacteriophages have been used as a subject model to better understand viral genetics, structure, and function. Large dsDNA bacteriophages such as Lambda and HK97 are of the prototype virus models that infect *Escherichia coli* (Effantin et al., 2010; Fokine et al., 2005; Hendrix, 2002; Louten, 2016; Terzi & Levinthal, 1967; Wikoff 1999) and phage P22 that infects *Salmonella typhimurium* (Bezdek & Amati, 1967; Kahmann & Prell, 1971; Zhang et al., 2000). Viruses are classified based on their genetic makeup, protein coat structure, and what host they infect (Blasche et al., 2013; Grose & Casjens, 2014; Koonin et al., 2021). Many viruses share common characteristics

like having a small genome that contain unique sequences to express regulating products and also structural proteins. More importantly, the viral genome has the ability to be integrated into the host genome or utilize host enzymes to increase viral fitness and pathogenicity; thus, studying bacteriophages allows to understand the parasitic mechanisms in any viruses that can support why a virus is a selfish genetic element.

One of the most well-studied prototype bacteriophages that possesses common viral features seen in almost all viruses that infect every life domain is Bacteriophage Lambda (Figure 1). Lambda is a tailed temperate phage that infects *E. coli* and has a double-stranded DNA (dsDNA) (Lederberg & Lederberg, 1953; Lederberg, 1954). Lambda takes part in lysogenic and lytic cycles, where following infection of the host the viral genome immediately undergoes tight transcriptional regulation that controls whether the prophage will become temperate or to grow lytic and burst the cell (Bertani & Weigle, 1952; Casjens & Hendrix, 2015; Kaiser, 1957; Lwoff, 1953). Early regulating genes are transcribed and translated into products that are involved in transcriptional regulation and are involved in antitermination mechanisms that control mRNA synthesis of particular genes on the small 48 kbp genome. The regulation of lysogeny is complex but these early gene products prevent transcribing late structural genes so that way the prophage can be integrated into the host genome through integrase enzymes (Casjens & Hendrix, 2015; Echols, 1971). In the lytic cycle, a switch of this complex regulation occurs where now lytic genes are activated to repress lysogenic genes from being expressed which allows late structural genes to be transcribed and translated into product (Sam et al., 2004). These structural genes are precursors required to assemble viral progeny and to become infectious particles (Casjens, 2005; Caspar & Klug, 1962; Lwoff, 1953; Moisan et al., 2010). Proteins and enzymes that regulate the lysogeny or lytic cycles and the assembly pathway can be ideal targets to inhibit production of an assembled virion and reduce any viral genetic recombination. Structural analysis of bacteriophage viruses are ideal model for drug discovery to inhibit viral infection.

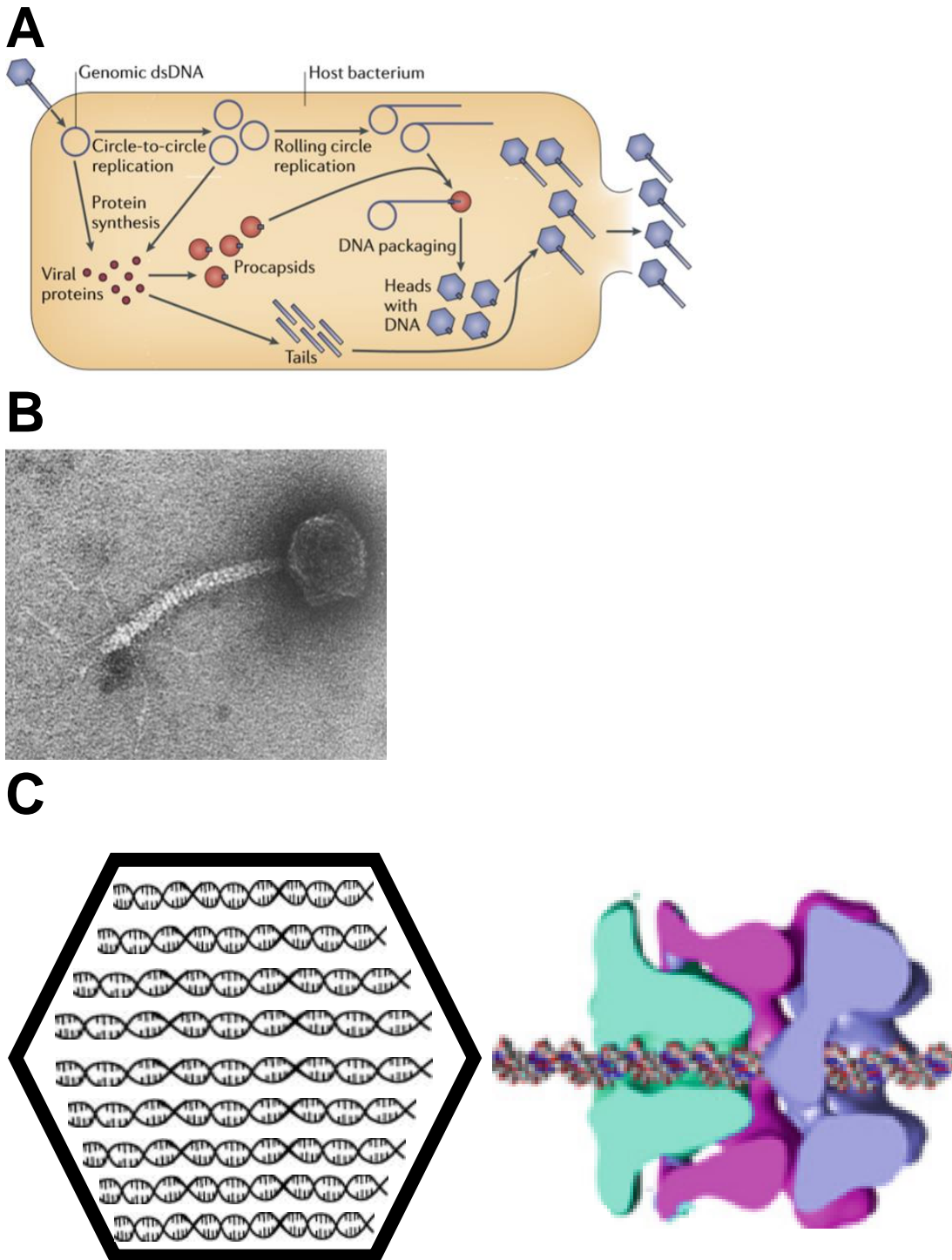


Figure 1. Virus Infecting Bacterial Cell and Lambda Genome. A) Replication of a virus occurs by infecting genetic material into host cell which a virus's DNA or RNA is replicated while simultaneously undergoing protein synthesis using host enzymes to form new viral particles. B) Electron microscope of Lambda Virus with body containing only head and tail proteins. C) Basic model of Lambda and representation of the DNA packaging reaction. Terminase motor (violet, purple) inserts the translocated lambda DNA through the portal protein (teal) and packages efficiently in the preformed capsid shell head (Casjens & Hendrix, 2015).

Much information has been gathered on the assembly pathway of an infectious virus using Lambda as a subject (Becker & Murialdo, 1977-1978, Casjens & Hendrix, 2015, 1974; Dokland &

Murialdo, 1978; Effantin et al., 2010; Hendrix, 2002; Hohn & Katsura 1997). The formation of the virus shell head consists multiple proteins which interact with each other in a novel oligomerization reaction that assembles into the virus capsid head which is proceeded by to the genome packaging event. In Lambda, late structural genes are transcribed and translated by host enzymes which produce precursors for the polymerization reaction of the virion capsid head. The viral protein head is required for genome packaging to occur as it is the precursor to the overall mature infectious virion that will protect and enclose the viral genome. The packaging enzyme called Terminase acts as a motor to translocate the DNA into the icosahedron shell to form a mature infectious virus (Casjens, 2011; Catalano, 2000; Feiss & Catalano, 2005; Feiss & Rao, 2012; Hohn et. al, 1974; Maluf et al., 2006; Yang & Catalano, 2003). The physical characteristics of lambda such as the protein coat shell are seen in almost all viruses that infect both eukaryotes and prokaryotes, thus it serves as an ideal model system for structure studies. Although there is much variability among the diverse world of viruses, the capsid plays a fundamental role to provide structure and function to enclose genetic material. Studying phages in bacteria is a useful tool that have allowed to better understand how selfish viral selfish genetic elements dictate the mechanisms of infection and their parasitic nature. Lambda and many other bacteriophages can serve as models to understand bacterial-viral interaction that can be applied to the broader sense of viral interaction with the surrounding world and overall understanding viral host dependency.

Molecular Genetics

The lambda virus genome become a tool for the investigation of viral molecular genetics as it demonstrates general features representative of other diverse viruses. Key discoveries were gaining insights of bacterial endonuclease activity on viral DNA, the understanding of unique gene regulation patterns outlined by the mapping of the genome, and the overall understanding of the mechanisms involved in lysogenic-lytic conversion leading to the pathway of viral assembly. The discoveries from Lederberg (1953-1954) of the lambda prophage existing as recombined viral DNA in host genomes led to more investigations of viral genetic function as the lambda genome appeared to be under

complex regulation not well understood at the time. The 48 kbp lambda genome provided a template for nucleotide sequencing that allowed to identify site specific regions on the DNA where endonucleases would recognize and cleave nucleotides in a particular arrangement (Sanger et al., 1982), which became an ideal cloning vector template as multiple sites on the genome can be recognized by restriction enzymes that enabled to clone genes of interest. Many sites of the lambda genome are included in various cloning and expression vectors (Arber, 1971; Cameron et al., 1975; Hohn & Murray, 1977; Loenen & Murray, 1986; Mertz & Davis, 1972; Sanger et al., 1982). Bacterial restriction endonucleases are commonly used in molecular DNA and recombinant technologies, where these enzymes have been characterized and understood through the use of lambda. The lambda genome contains specific recognition sites which are targeting by bacterial endonucleases that will cleave that segment of the viral genome, forming a nick in the DNA (Arber, 1971). Bacterial endonuclease enzymes act as a defense mechanism to protect against foreign molecule such viral DNA. Understanding the activity of restriction enzyme mechanisms using viral genomes played a critical role in advancing the field of molecular genetics (Arber, 1971; Echols, 1978; Halford et al., 1980). Lambda utilizes both lytic and temperate lysogenic lifestyles in host cell where the genome is controlled through unique regulation patterns (Bertani & Weigle, 1952; Campbell, 1986; Casjens & Hendrix, 1974; Casjens & Hendrix, 1992a; Kaiser, 1957; Kaiser & Jacob, 1957; Jacob et al., 1953; Lwoff, 1953). The regulation of lysogeny and lytic cycles outlines a defying example of how the lambda genome is tightly controlled which led to insights of repression or activation of certain genes that dictates the fate of a prophage (Jacob & Campbell, 1959; Thomas, 1971). These discoveries using the lambda as a prototypical virus as subject for investigation contributed to our fundamental understanding of other viral systems and the advancement of molecular biology.

Medicine and Therapeutics

Phage display is a relatively new technology where phages can be used as nano systems to deliver various specific antigens to target within a cell system, inducing a specific cellular response (Asadi-Ghalehni et al., 2015; Beghetto & Gargano, 2011; Chang et al., 2014; Huang et al., 2011;

Jepson & March, 2004; Johnson, 2010; Kanekiyo et al., 2015; Koudelka et al., 2013; Marintcheva, 2018; Nicastro et al., 2014; Saeedi et al., 2014; Sternberg & Hoess, 1995). These advancements have been seen to target various diseases that are not well understood and difficult to address such as cancer. Oncogenesis and growth of cancerous cells are controlled by many factors that still remain unclear, although the overexpression of certain genes can be markers for cancer cells (Asadi-Ghalehni et al., 2015; Huang et al., 2011; Kanekiyo et al., 2015; Marintcheva, 2018). Overexpression of particular proteins are an apparent indication of a cancerous cell, although the underlying mechanisms causing the overexpression is not well understood in many cancers. Development of alternative new methods using phage systems to target cancer growth factors such as overexpressed receptors may inhibit tumor progression and have been shown to be effective as an alternative therapeutic (Asadi-Ghalehni, 2015; Nicastro 2014). Phages are nanoparticles that can be utilized as a system for site-specific delivery of genetic material that can trigger host immune responses which in turn can target these growth factors seen in cancers. The premise behind these nanotechnology advancements deals with the phage structure and the ability for them to enclose a specific piece of DNA within the protein shell. The phage contains genes that encode for external decoration proteins producing a phenotype that extrudes on the coat surface, overall providing viral shell stability. Genes encoding for exterior decoration proteins can be fused with a gene for a ligand of interest; this fused gene becomes expressed on the exterior virus head fused to the virus decoration protein and the expressed fused ligand can specifically recognize various cell receptors. Due to the endocytosis nature of a virus, genes encoding for the antigen or ligand fused to a decoration proteins become displayed on the capsid surface to target the receptor intended to stimulate a host immune response. Phage display technologies that target cancerous growth factor receptors are to be inactivated by this process, ideally to inhibit tumor function and growth. A gene of interest can be fused to capsid decoration genes and will become expressed as a fusion protein with the decoration protein that is to be displayed on the exterior shell (Chang, et al. 2015; Koudelka et al., 2013; Nicastro, 2014; Yang et al., 2000).

Lambda Assembly and Packaging Reaction

Lambda assembly (Figure 2) outlines a distinct pathway that is highly conserved among many large double-stranded DNA (dsDNA) viruses such as Adenovirus, Herpesvirus, and other tailed-bacteriophage families (Casjens, 1985; Casjens et al., 1970; Dokland & Murialdo, 1993; Fane & Prevelige, 2003; Gertsman et al., 2009; Hauser et al., 2012; Hendrix & Casjens, 1974; Hohn & Katsura, 1977; Johnson, 2010; Johnson & Wikoff, 1998; Kobayashi, 1990; Lambert et al., 2017; Lander et al., 2008; Medina et al., 2012; Moisan et al., 2010; Murialdo, 1979; Murialdo & Becker, 1977-1978; Murialdo & Ray, 1975; Murialdo & Siminovich, 1972; Williams & Richards, 1974; Yang et al., 2000). While Adenovirus and Herpesvirus infect eukaryotes of higher order animals, bacteriophages specifically infect bacteria. Assembly, maturation, and DNA packaging is a highly coordinated process requiring many unique protein-protein interactions that are not observed in biological organisms. When viral proteins are synthesized using external cellular host enzymes and ribosomes, subsequent spontaneous polymerization of viral proteins cooperatively associate to begin the assembly process as novel reactions occur leading to formation of the mature icosahedron head.

All viruses are made up of a protein shell consisting of polymerized polypeptides that encloses the genetic material (Bawden & Pirie, 1938; Caspar & Klug, 1962; Crick & Watson, 1956; Forterre, 2005; Franklin, 1955; Stanley & Loring, 1936). The proteins involved in assembly appear to be disordered with unstructured domains that can be a factor of their flexibility to interact with one another as well as to undergo structural transitions during expansion into the final structure (Casjens, 1985; Dokland, 1999; Fane & Prevelige, 2003; Forkine et al., 2005; Hohn & Katsura, 1977; Hohn et al., 1974; Iwai et al., 2005; Lander et al., 2008; Medina et al. 2010-2011; Murialdo, 1979; Murialdo & Becker, 1977-1978; Murialdo & Siminovich, 1972; Ray & Murialdo, 1975; Singh et al., 2013; Wikoff et al., 2003; Wikoff et al., 1999). Due to the disorder seen within the structure of assembly proteins, this flexible characteristic introduces difficulty to express minor subunits as this instability is a result of the virus's natural ability to spontaneously oligomerize into higher-order capsid structures *in vivo* and *in vitro*. Capsids have been shown to tolerate stress internally such as withstanding DNA packaging as

the translocating mechanism produces high pressure within the interior as DNA becomes inserted into the preformed shell to a crystalline density (Casjens, 2011; Catalano, 2000; Feiss & Catalano, 2005; Feiss & Rao, 2012; Hohn et. al, 1974; Maluf et al., 2006; Yang & Catalano, 2003). In addition, the capsid proteins must tolerate the buildup of negative charge while the procapsid undergoes stable expansion.

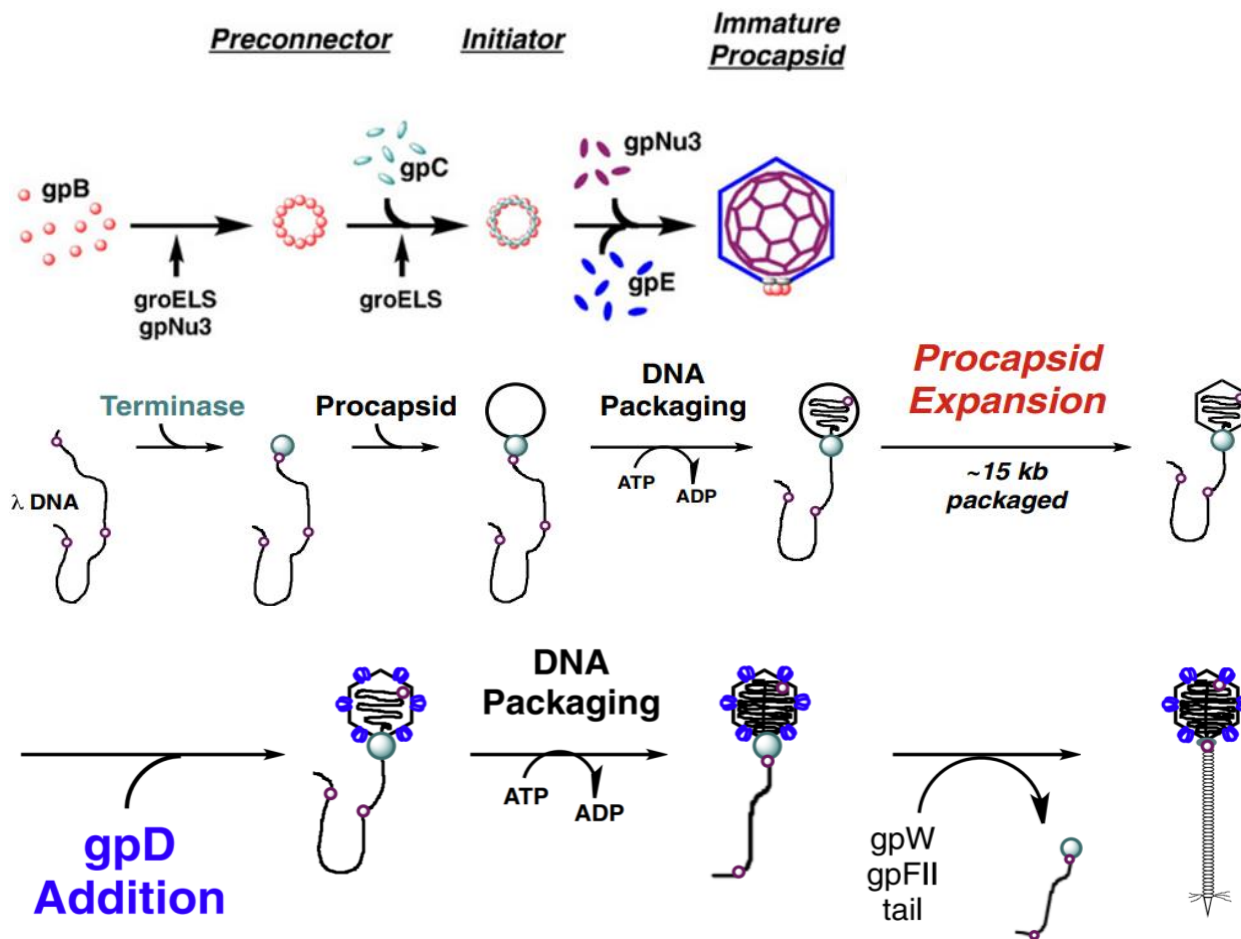


Figure 2. Lambda Capsid Assembly Pathway. 1) Initiation and procapsid assembly: the structural proteins involved in assembly into an infectious viral particle begins with gpB polymerization into the portal ring forming the preconnector, the nucleation site for assembly into the procapsid. Once viral protease gpC associates with gpB, the core proteins gpE (MCP) and gpNu3 (Scaffolding Protein) interact to form the immature procapsid with undefined circular shape. 2) The packaging event: Lambda DNA containing cos site is the recognition region for gpNu1 (Small Terminase) binding where gpA (Large Terminase) containing an ATPase domain powering the translocation reaction using DNA as the substrate. 3) Decoration and tail attachment: gpD (Decoration protein) forms a homotrimer on the capsid three-fold axis for shell stability. The packaging event and gpD addition occur simultaneously as the procapsid expands to a defined angular structure. Completing the process requires cleavage of portal and terminase proteins as tail proteins seal the capsid to form the infectious virus (Lambert et al., 2017).

The assembly pathway contains critical protein components that are responsible for providing the structural housing and protection of viral genome. Improperly assembled capsid and tail protein structures have been demonstrated to be inactive or fatal, which indicates that proteins involved in

the assembly pathway have an essential pivotal role in viral survival (Casjens & Hendrix, 1974; Lambert et al., 2017). Structural genes early in the lambda genome are labeled as *gp* (gene product) following a letter that denotes the protein. The process into forming the large protomeric complex begins the oligomerization of gene product *B* (gpB), known as the portal protein that forms a dodecameric O-shaped structure acting as a “portal” for DNA to pass through during the translocating event. The gpB initiator complex also forms the binding site for capsid and scaffolding proteins to bind, creating a nucleation core point for polymerization reactions to occur as it begins to form a higher-order structures. Once the procapsid is formed by the polymerization of gpE assisted by gpNu3, subsequent events leading to DNA packaging catalyzed by gene *A* called Terminase (gpA and gpNu1), will translocate the DNA into the immature procapsid simultaneously while the capsid proteins undergo structural transitions forming a more angular shell, a characteristic seen in many mature viruses. The DNA packaging event is modulated by terminase that binds 41 kbp lambda DNA and efficiently translocate the genome in an ordered arrangement into the preformed shell (Figure 2). Terminase mutants fail to package DNA into the capsid which leaves the capsid head empty and incomplete, therefore producing a failed assembly structure where now the virus is inactive (Bain et al., 2001; Casjens, 2011; Catalano, 2000; Feiss & Catalano, 2005; Feiss & Rao, 2012; Hohn & Hohn, 1974; Hohn et al., 1974; Maluf et al., 2006; Murialdo, 1979; Ortega & Catalano, 2006; Ortega et al., 2007; Yang & Catalano, 2003). The formation of a properly formed head is also critical, although mutating structural genes do not completely inhibit the formation of gpE protein assembly into the capsid shell; although these mutants (B, Nu3, D) do affect proper formation of the capsid shell, as defective and aberrant structures are observed but do not proceed onto the packaging event (Hendrix, 2012; Hohn & Katsura, 1977; Murialdo, 1979; Murialdo & Becker, 1977-1978; Murialdo & Siminovitch, 1972). Mutational studies of lambda structural genes have shown these components play a critical role in the assembly process of an infectious virus; therefore, the protein coat is an essential requirement for the DNA packaging event to take place.

Terminase

Terminase packaging motors are one of the most well studied viral enzymes and common to most dsDNA viruses (Becker & Murialdo, 1990; Catalano, 2000; Catalano et. al, 1995). This holoenzyme is essential for viral survival as its role is to insert the viral genome into the preformed capsid during expansion. The terminase motor is a critical component needed for viral fitness as the fate of the virus is dependent on the genome; failure to package DNA into the capsid is fatal. DNA packaging is a high energy process with high ATPase activity domain which this energy shuttling pumps the DNA into the preformed shell. The insertion of the genome can create an internal pressure up to 20 atm and packages DNA to a crystalline density (Catalano, 2000).

Terminase consists of two distinct domains that demonstrate cooperativity with one another: a large domain (Ter-L) consisting of 641 amino acids and 73.3 kDa, and small domain (Ter-S) consisting of 181 amino acids and 20.4 kDa. The small domain consists of a dimer that is responsible for recognizing and binding DNA from the pool of replicated genome, forming a ribonucleoprotein. The large domain known as the translocating motor, is fueled by ATP hydrolysis that physically inserts the DNA by working with the DBD in which DNA is transferred to ter-S to ter-L that inserts the genome into the preformed capsid shell (Andrews & Catalano, 2013; Bain et al., 2001; de Beer et al., 2002; Gaussier et al., 2005; Gaussier et al., 2006; Catalano, 2000; Maluf et al., 2006; Ortega & Catalano, 2006; Ortega et al., 2007) . Packaging triggers expansion of the shell and once the full viral genome is inserted through the portal into the preformed capsid, viral maturation occurs through addition of tail and decoration proteins that now enable virions to invade the next cellular host. Ter-S is a product of gene gpNu1 that purifies as a homodimer consisting of the DNA-binding domain (DBD) of the N-terminus and the ATPase site residing in the C-terminus. The DBD has high affinity to bind specific sequences in the early λ genome named *Cos*, containing repeating recognition elements called *R-sites* (R1-R3), while the C-terminus possesses catalytic properties of nicking fragmented DNA and maturing the genome through interaction of the large subunit gpA (Catalano, 2000). Mutational analysis and fluorescent quenching studies demonstrated that gpNu1-A55 containing the

Wing-Helix-Turn-Helix motif is the minimal domain needed for nucleotide binding that may be influencing conformational changes of the small subunits. Addition of ATP and ADP decrease fluorescence of tryptophan which suggest nucleotide binding induces conformational changes of the WTH motif (Gaussier et al., 2005; Gaussier et al., 2006; Catalano, 2000; Maluf et al., 2006; Ortega & Catalano, 2006; Ortega et al., 2007). Once ATP and ADP binds to this motif, the wing goes from a closed to open conformation exposing the helix that allows recognition of site-specific elements within the R-sites. This nucleotide binding that activates a more favorable conformation that increase the affinity for gpNu1 and DNA. This ribonucleoprotein complex formation of DNA and gpNu1 DBD is a crucial intermediate step that enables the N-terminus to transfer the genome to the C-terminal domain of Ter-S, which is followed by an interaction with the N-terminal domain of gpA (Ter-L) to begin the translocating step.

Major Capsid Protein (gpE)

Early genes within 1-7 kbp of the lambda genome contain the structural head genes *B*, *C*, *Nu3*, *E* (Casjens, 1985; Echols & Murialdo, 1978; Murialdo, 1979; Rajagopala et al., 2011). The gene *E* expresses the protein gpE known as the Major Capsid Protein (MCP) to form a polymerized structure containing 420 copies of gpE that is recapitulated across the entire virus head. The MCP monomer size is 38 kDa that appears to be disordered and flexible, consisting of unique secondary structure motifs involved in protein-protein interactions with other gpE proteins to form a procapsid (Lander, 2008). The primary sequence of gpE contains hydrophobic and negative charged amino acids. The hydrophobic content suggests these intramolecular interactions among gpE units are a significant factor in capsid-association stability, while negative residues create the slight repulsion to the negative charged DNA within the capsid interior. gpE forms two common hexamer and pentamer domains called capsomers, which are intermediate precursor structures of the mature head (Dokland & Murialdo 1993; Murialdo, 1979; Murialdo & Becker, 1977-1978; Murialdo & Siminovitch, 1972; Williams & Richards, 1974). The fully assembled lambda protein head consists of 420 individual gpE monomers in total, which form the core of the capsomers that takes part in novel reactions to

polymerize into the geometric arrangement of an icosahedron shell, stabilized by protein-protein interactions between gpE and other structural genes. The viral head consists of a symmetrical fold influenced by the protein arrangement making up the icosahedron. Lambda head displays a $T=7$ triangulation number where six monomer units form the basic hexamer unit and the seventh monomer gpE unit forms the site to construct the pentamer. The gpE a-domain belonging to three different hexamers connect together to form tri-fold vertices as maturation occurs, a critical factor during expansion as the corner formed by the three connected domains serve the putative binding site for decoration proteins (gpD) that increases capsid stability (Singh et al., 2013; Lander, 2008; Lambert et al., 2017). The preformed capsomers being a larger precursor structure than the monomer unit can act as a nucleation site for the novel polymerization reaction event to occur. The gene products involved in the assembly and morphogenesis into the mature head are well understood, although the mechanism of protein association in the initial steps to nucleate the polymerization reaction of gpE are unclear. MCP's are flexible and undergo different structural changes upon head expansion, thus the primary sequence of gpE is a factor of the overall fold that contributes to the protein-protein interactions in self-association as well as interactions with other structural proteins such as gpB, gpD and gpNu3 (Blasche et al., 2013; Casjens et al., 1970; Dockland & Murialdo, 1993; Fane & Prevelige, 2004; Hauser et al., 2012; Hendrix & Casjens, 1974; Hohn & Katsura, 1977; Lambert et al., 2017; Murialdo & Ray, 1975; Ray & Murialdo, 1975). The intra- and intermolecular forces these proteins exhibit greatly contribute to overall capsid stability and integrity.

Scaffolding Protein (gpNu3)

A small disordered 18 kDa peptide gpNu3, the product of gene *Nu3*, is the scaffolding protein that aids gpE into oligomerizing to the final form. Assembly of lambda particles containing gene *Nu3* mutations fail to assemble the procapsid properly, forming disrupted poorly formed virus shells in smaller amounts compared to WT lambda capsids (Casjens & Hendrix, 1974; Murialdo & Ray, 1975; Ray & Murialdo, 1975). Although constructs containing WT-gpE with *gpNu3* mutants were able to self-assemble to form aberrant unstructured forms, this indicates gpNu3 is not essential in capsid

oligomerization, although the absence of gpNu3 induces gpE to form improper capsids that is less stable and is morphologically different from the WT-*gpE* (Casjens & Hendrix 1974, Hohn et al., 1976; Dockland & Murialdo, 1993; Lambert et al., 2017). The scaffolding interactions are not exclusive to the MCP, Nu3 is responsible in assisting the formation of heterocomplexes with gpB creating an initiator core for nucleation of gpE subunits to build upon (Casjens & Hendrix, 1974; Murialdo & Siminovitch, 1971). The similar observations are seen with gpE-gpNu3⁻ mutations, where the absence or mutated gpNu3 affects gpB formation and its interaction with other proteins (Ray & Murialdo, 1975). gpB is directed to follow a different series of events in the absence of gpNu3 that can be explained based on the analytical ultra-centrifugation data. Expression of gpB with defective gpNu3 sediments the portal protein into a different fraction, which normally gpB has an S value of 25 and 30; suggesting gpNu3 is involved in a specific role to initiate the gpB portal. The absence of gpB shows that it becomes degraded into various cleavage products by protease gpC, further suggesting gpNu3 has some role in stabilizing the gpB complex which prevents premature degradation (Murialdo, 1979; Ray & Murialdo, 1975). The involvement of gpNu3 appears to be essential in forming the precursors responsible for proper and efficient virus assembly. The scaffold must be involved in modulating the various associations between the capsid and portal proteins that form large heterocomplexes; therefore, the morphogenesis into the procapsid structure requires the presence of Nu3. Although aberrant structures can be formed in the absence of functional gpNu3, this leads to inactive forms of the prohead that will be unable to take part in the DNA packaging reaction, thus becoming inactive viral particles.

gpE interactions with Decoration Proteins

In addition to the capsid proteins which are the main counterpart of the head, additional proteins called decoration proteins are bound to the exterior surface of gpE, stabilizing the viral head and increasing providing rigidity to prevent any external force that can impose any stress leading to disruption or degradation. The lambda decoration protein gpD is an 11 kDa monomer that is the product of gene *D* and naturally forms a trimer once bound to the each of the three-fold vertices

formed by gpE monomers from three different capsomers, where 140 copies of trimeric gpD are bound across each three-fold axis points of adjacent capsomers (Lambert et al., 2017; Singh et al., 2013; Yang & Catalano 2000).

This quasi three-fold axis serves as putative binding sites for gpD to cement itself which is stabilized by hydrophobic and electrostatic interactions (Lambert et. al, 2017). Docking the 1.1 Å crystal structure of gpD into the 6 Å cryo-EM gpE structure show that the three-fold axis is stabilized by His-Asp electrostatic interactions and Trp-Pro hydrophobic stacking interactions; where gpD contains His19 and Pro17 and gpE contains Asp292 and Trp308 (Lander et al., 2008; Lambert et al., 2017; Wikoff et al., 1999). The atomic resolved gpD crystal structure was used in-tandem with the high resolution gpE cryo-EM that was used to further characterize the MCP as this gave potential insights to details involving capsid and decoration protein interaction (Yang et al., 2000). To improve the resolution of the 6 Å cryo-EM structure, X-ray crystallography should be implemented for the analysis to determine the monomer gpE protein structure, enabling to fully distinguish what residues on the capsid protein are involved in these types of interactions.

Structural Homology Modeling of the Capsid

Examining viral protein structures are essential to learning the interactions that define viral integrity and understanding the nature of their lifestyle. Structural characterization is a process that is important to understand these characteristics. While Terminase and decoration protein crystal structures have been solved to an atomic-resolution, capsid proteins gpE and gpNu3 have no known crystal structure. There has been some insight on capsid structural details from cryo-EM and CXMS, but these structure models fail to provide atomic resolution which limits distinguishing the side chains of the amino acids involved in protein-protein interaction. As the current models present a snapshot of capsomers and shows interactions with gpD, obtaining a crystal structure provides potential for determining an atomic resolution that can reveal details of the capsid and scaffolding protein and their roles in interaction and catalysis. Solving these proteins on a model virus will aid in providing protein

targets for drug discovery for treatment of viral infections or using capsid structures to develop phage display systems.

Cryo-Electron Microscopy: Lambda Reconstruction

A relative of lambda is dsDNA bacteriophage HK97. This virus also infects *E. coli* and contains a capsid protein gp5. The crystal structure of gp5 was determined and resolved to 3.5 Å, providing a foundation model for lambda structural determination (Wikoff et al., 1999). Lambda and HK97 capsid proteins share a sequence of 29% similarity and 14% identity, although similarity may be low due to additional 44 amino acids located in different areas of the C-terminus (Figure 3) that are involved in cross linking domains with no requirement of a decoration protein (Johnson & Wikoff, 1998; Wikoff et al., 1999; Wikoff et al., 2000).

1	MSMYT-----TAQLLAANEQKFK-----FDPLFLRL----FFRESYPFTT	36
1	MSELALIQKAIEESSQQKMTQLFDAQKAEIESTGVSKQLQSDLMKVQEELTKSGTRLFDL	60
	** : : ** : * : : : : : : : : : : *	
37	EKVYL--SQIPGLVNMALYVSPIVSGEVIRSRGGSTSEFT-----	74
61	EQKLASGAENPGEK---KSFSERAAEELIKSWDGKQGTFGAKTFNKSLSGSDADSAGSLIQ	117
	* : : : ** : * : : * : * : * : . . *	
75	----PGYVVKPKHEVNPQMTLRRLPDEDPQNLADPAYRRRIIMQNMR-----DEELAI	123
118	PMQIPGIIMPGLR---RLTIRDLLAQGRSSNALEYVREEVFTNNADVVAEKALKPESDI	174
	** : * . : : * : * : . . * * . : : *	
124	A---QVEEM-----QAVSAVLKGYTMTGEAFDPVEVDMGRSEENITQSGGT-----	168
175	TFSKQTANVKTIAHWVOASRQVMDAPMLQSYINNRLMYGLALKEEGQLLNGDGTGDNLE	234
	: * . : : ** * : . : : : : . ** : : : **	
169	-----EWSKRDKSTYDPTDDIEA---YALNA-SGVVNIIVFDPKGWALFRSEFKAVKEK	217
235	GLNKVATAYDTSLNATGDTADIIAHAIYQVTESEFSASGIVLNPRDWHNIALKDNENGR	294
	: . : : * * : * : * : * : . . ** : : * : : *	
218	LDTRRGSNSLELETAVKDLGKAVSYKMGYGDVAIVVYSGQYVENGVKKNFLPDNTMVLGNT	277
295	YIF--GGPQ-----AFTSNIMWGL-----PVVPTKA	318
	* . . : : : * : *	
278	QARGLRITYGCIQDADAQREGINASARYPKNWVTTGDPAREFTMIQSAPLMLLADPDEFVS	337
319	QAAGTFTTVGGFDMA-----QVWDRMD-----ATVEVSREDRDNFVK	355
	** * * * : : * : : * : : : : : * * : * : *	
338	VQLA-----	341
356	NMLTILCEERLALAHYRPTAIKGTFFSSGS	385
	* :	

Figure 3. Sequence Alignment of Lambda gpE. The 341 amino acid sequence of Lambda (first row) is aligned with relative phage HK97 (second row). Purpose indicates hydrophobic residues and red indicates negative residues. Comparing the two phages there is a 13.7% identity and 99 similar residues (*) Identical, (:) Similarity.

Despite the lack of similarity, the crystal structure of HK97 gp5 that was used to map in the sequence of lambda gp5 shows similarly structural phenotypes between the two capsid proteins, indicating these capsids have a degree of homology and thus having similar function. Lambda and

HK97 monomer capsid proteins share striking similar features within the folds of the domains (Wikoff et. al, 1999; Lander et. al, 2008). The N-terminus, E and P loops, and C-terminus of lambda gpE appear to have disordered strands that may play a role in capsid flexibility to interact with neighboring subunits (Figure 4). Within the center of the structure lies the prominent spine α -helix measuring approximately 40 Å in length which this helix may be a critical factor of gpE flexible dynamic as it can act as a spring to extend the α -domain to bind with other gpE units. Additionally, the helix can act as a backbone that extends the other unstructured regions during capsid expansion to form the angular three-fold axis gpD binding site.

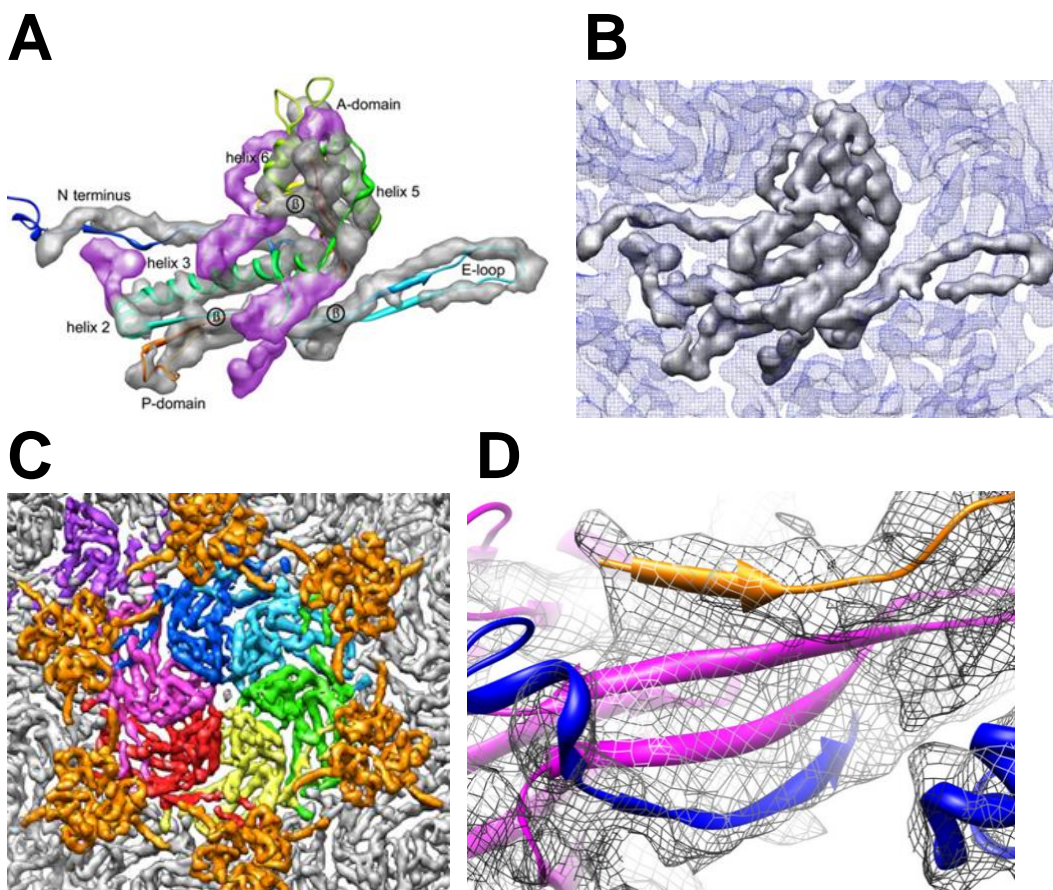


Figure 4. Cryo-EM gpE and HK97 Crystal Structures of MCP. A) Crystal structure of HK97 MCP B) Lambda MCP cryo-EM. C) the T7 symmetry of Lambda capsomer unit. Hexamer unit is shown in (blue, cyan, green, yellow, red, pink) where gpD sits on the axis bound by three gpE capsomer units. Purple indicates the adjacent pentamer unit. D) Lambda cryo-EM p-domain/e-domain was used to fit the docking of gpD (orange) crystal structure (Lander et al., 2008).

Cryo Electron Microscopy (EM) and Chemical Cross-Linking Homology Modeling of Lambda MCP have been used to characterize the structure which was made possible through usage of the crystal structure of relative phage HK97 capsid gp5 (Lander et. al, 2008. Singh et. al, 2013). By

threading in the lambda gpE sequence into the HK97 gp5 crystal structure along with fitting in the cryo-EM gpE densities allowed to better distinguish the gpE capsid structure. The lambda gpE “reconstruction” provided a snapshot of the mature capsid shell which contain the hexamer and pentamer units, but more importantly the integrated gpE cryo-EM model into the HK97 crystal structure provided the basis of the gpE monomer fold. The unstructured domains contain α -helical character within the interior fold while hydrophobic β -strands are exposed on the surface such as the E-loop and P-loop domains. While this model at 6 Å cannot infer distinction of the position of amino acids within the domains, the gpE fold being similar to gp5 model suggest the disordered monomer gpE serves a purpose to undergo various structural transitions as expansion occurs. The lambda reconstruction of the monomer unit using the models mentioned provided the snapshot of the general capsid fold which shares common characteristics seen within the HK97 capsid fold as well as other similar phages such as T4 and P22 (Forkine et al., 2005; Zhang et al., 2000).

Crosslink Mass Spectrometry Modeling

A novel method was implemented to better resolve the gpE structure using Chemical Crosslinking Mass Spectrometry (CXMS), which generated a homology model that aligns with previous structural characterization of the capsid proteins (Singh et al., 2008; Singh et al., 2013). CXMS reveals the intra- and intermolecular protein interactions gpE has with itself and other structural proteins. CX reactions were applied to form distances of the CX amino acids on gpE followed by proteolytic digestion and mass spectrometry for analysis of fragments to create a structural model (Figure 5A). For example, Lysine conjugated residues on gpE were measured within acceptable range of a particular CX-reagent (BS2, 21 Å) and threading in the gpE monomer sequence that matched the proteolytic fragments enabled the creation of a structural homology model that gave a better understanding of not only the intramolecular interactions within the gpE monomer fold, but also the intermolecular interactions gpE has within the capsomer structures (Singh et al., 2013; Lander et al., 2008; Wikoff et al., 1999)

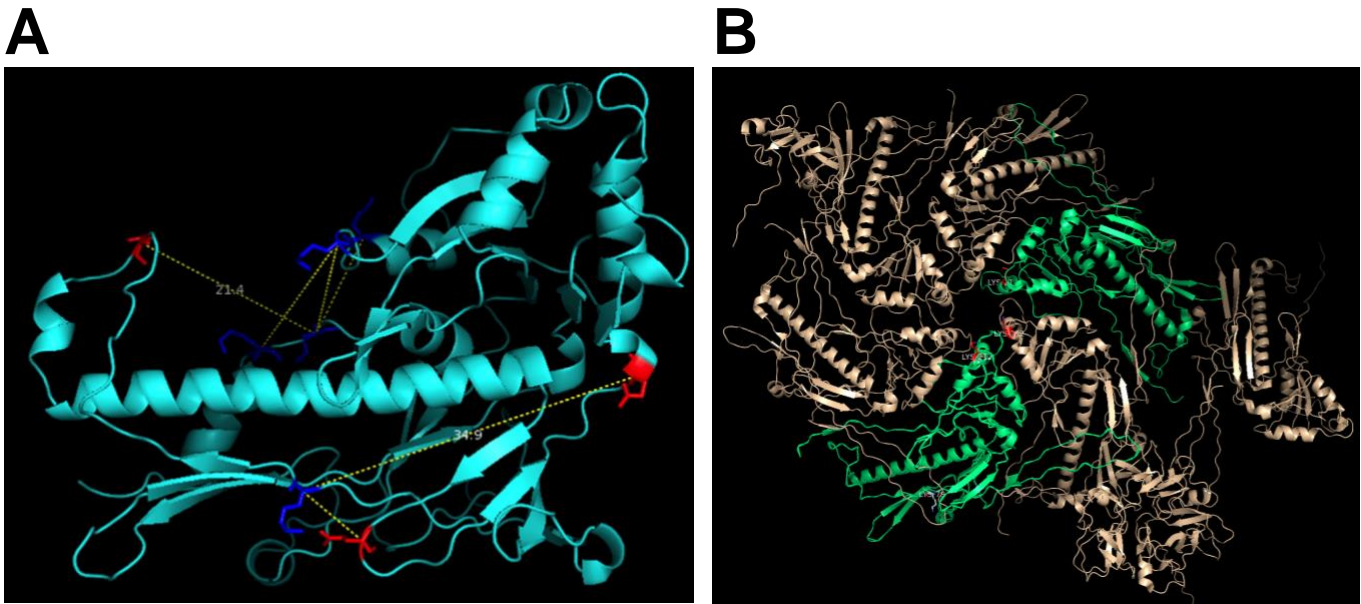


Figure 5. CXMS Homology Model of Monomer gpE and Capsomer Units. A) gpE-CX (indicating crosslinked residue) formed a conjugate with other tagged residues via carboxylation or amide reactions. The products were analyzed via MS and CX fragments used to determine m/w of cleaved peptide (Singh et al., 2013). The monomer unit formed CX distances with Lys (blue) and Glu (red). B) Two monomer units (green) associated with capsomer; the residues labeled in red correspond with the CXMS homology model (Singh et al., 2013). The acceptable distance formed by the BS2 reagent is 21 Å, where fragments of CX bound K212-K212 (red) formed an approximate distance of 15 Å in the adjacent hexamer units (19 Å in pentamer not shown). The single gpE monomer on the right extrudes out to form the pentameric unit having a T=7 symmetry fold. The bottom left (ii) is gpD binding to the site of the three-fold axis connect to two other gpE units in differing capsomers.

Structural analysis of the assembly proteins using the presented experimentation provided insights to viral capsid structure and function. While series of events utilizing the proteins involved in the directional assembly pathway are generally understood, the initial formation steps are not well defined as it is unclear to distinguish the residues involved in these steps. Analyzing the gpE sequence in-tandem with the solved homology models can infer putative residues that take part in capsid self-association in which the disordered character serves a role in the interacting process due to the flexibility. Much information has been gathered on gpE and gpNu3 proteins and their association together, yet the initial mechanisms of Nu3 and how they direct E and the polymerization reactions with other structural proteins are unclear. With the resolution of gpE determined from the past models are not atomically resolved, the amino acids that modulate the protein-protein interactions are indistinguishable. It is still unclear what the directional process is of how capsomers are formed and directed into positions so that polymerization, expansion, and maturation occur.

gpE Characterization

Capsids are known to spontaneously oligomerize into the preformed structure which is a critical step to proceed onto DNA packaging. Isolation of structural genes allowed expression lambda particles *in-vitro* which led to an understanding of proteins involvement that are essential for assembly to occur. The most abundant protein in the shell, gpE (MCP), can be directed into various pathways that form intermediates smaller than the capsomer, such as minor-precursors containing gpE-gpNu3 with a ratio not exactly known as very little is known about this initial complex formation which leads into capsomer oligomerization. gpE becomes polymerized through reactions which are also not well understood, but it is observed that gpE undergoes structural transitions from individual units to capsomer units, oligomerizing into a protomeric structure as the overall shape transforms from a circular form to an angular icosahedron. Individual capsid protein units naturally interact with one another to form higher-order structures assisted by gpNu3 and to understand the mechanics of gpE interactions with itself and other proteins, the individual monomer subunit should be examined more closely.

The characteristic of gpE wanting to naturally form higher-order complexes makes it difficult to isolate the monomer unit, as expression of constructs containing gene *E* would immediately produce procapsid structures. Studies on the capsid proteins also suggest that this process does not require gpB, gpD, and gpNu3, since expression of E in the absence of these proteins (B, D, Nu3) can still form procapsids, although these structures appear disordered and aberrant that differs from the wildtype final capsid structure (Lambert et al., 2017; Medina et al., 2010; Medina et al., 2011; Murialdo, 1979; Murialdo & Ray, 1975; Ray & Murialdo, 1975). Therefore, this must mean gpE requires these proteins to properly form into mature capsid structures. Expression of gpE constructs along with co-expression with or without structural genes enabled formation of gpE into procapsid structures that shows there must be a self-association event occurring for gpE to do so in the absence of structural proteins. To isolate the monomer gpE structure, site-directed mutagenesis of

critical residues was applied that enabled to successfully isolate gpE monomers (Lambert et al., 2017).

The high resolution cryo-EM models Lander et al. (2008) show at the three-fold axis formed by gpE units, W308 and D292 are positioned here within the A-domain which forms the three-fold, indicating these are putative residues involved in the interaction with gpD. The gpD crystal structure electron density was fit into the lambda cryo-EM density and the two structures fit into a position giving potential for a novel binding site. The 1.1 Å gpD crystal structure positioning onto the lambda models can suggest that residues H19 and P17 are positioned within this density (Lander et al., 2008; Lambert et al., 2017; Yang & Catalano, 2003). The possible interactions formed between W308 and H19 suggest hydrophobic stabilization due to the side chains, while D292 and P17 form electrostatic interactions. The W308 residue was of interest as it is conserved among other phages such as HK97, thus it was hypothesized that this aromatic residue is critical component in procapsid expansion as it's positioned at the three-fold axis which is important for interactions with stabilizing gpD. The side chain of W308 may potentially promote the hydrophobic interactions of pi-stacking and Van der Waal forces to bind with H19 gpD. Mutating W308A was predicted to abrogate the interactions between the capsid and decoration proteins, which would prevent higher-order structure formation. When locating the position on the monomer structure as seen in the Cryo-EM models show it is positioned on the outer-edge of the p-loop domain, which can be predicted to form the three-fold corner for which gpD binds to (Figure 6). In the hexamer and pentamer, W308 positions on the exterior edge of the capsomer structure which can be hypothesized that mutation of this residue should not affect capsomer formation, as the goal was to inhibit higher-order complexes from forming due to the three-fold axis where W308 positions at.

The gpE-W308A enabled isolation of the monomer unit, which can indicate that W308 must be critical in forming a stabilizing interaction involved in the polymerization reactions into the mature capsid. W308 must be essential in associating with gpD. Electron micrographs with this mutation failed to assemble higher order capsids structures (data not shown). The W308 residue (Figure 6)

positions at the exterior of the capsomer structure limiting the interactions with other preformed capsomers that inhibits formation of the three-fold axis; this greatly abrogated the formation larger quaternary structures.

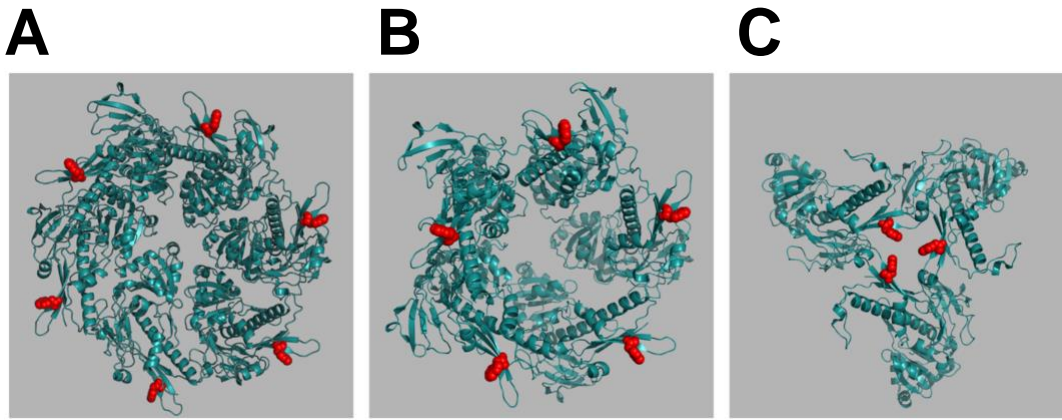


Figure 6. Cryo-EM Model of gpE Capsomers. Left is the basic hexamer unit, middle is basic pentamer unit, right is the three-fold trimer spike which is formed from three capsomer units where gpD cements to. Red indicates a Tryptophan residue that postulated to be involved with hydrophobic interactions with gpD (Lambert et al., 2017).

It is interesting to note that a W308F mutation did not fail to produce capsid structures, although these structures were deformed in comparison to the wildtype, which suggests that tryptophan must play more important roles beyond the hydrophobic interactions it exhibits (data not shown). It is not completely clear as to what the native gpE native structure is due to the different structures it can form; the native structure may be the monomer unit of a single peptide chain, the hexamer and pentamer capsomers, or the fully complete multidomain protomer complex. As the final result from the assembly pathway produces a shell with packaged DNA, this large protomer structure may be the native state gpE forms. The ability for gpE to undergo many structural transitions make it unclear what defines the native protein state, thus determination to resolve the details of capsid proteins are still under investigation as an atomic resolution may unravel the initial interactions gpE forms and the novel reactions taking place during oligomerization.

More importantly, site-directed mutagenesis of a gpE putative residue enabled the monomer unit to be soluble in the supernatant and to become purified. These findings show promising results that increases the probability to further characterize and purify a stable ENu3 complex to potentially

solve the structure. Structural analysis through X-ray crystallography can provide these atomic details to further inspect the subunits and amino acids involved in protein-protein interactions.

Research Aim

The ENu3 protein complex from bacteriophage lambda is investigated to isolate a stable form that facilitates structure determination by X-ray crystallography. Viral proteins are intrinsically disordered which play unique and specific roles in assembly. The natural ability for capsid proteins to self-assemble makes it difficult to isolate the monomer subunits for gpE and gpNu3, where often times purifying these proteins leads to a heterogenous population with various protein complex species, which in turn affects the crystallization.

The main objectives of this research project are as follows:

1. Obtain a highly pure ENu3 complex in a stable form desired for crystallization
2. Grow crystals to a maximum size with defined morphologies for stronger diffraction
3. If poor crystal morphology is produced, then apply a new method of Microscale Matrix Seeding (MMS) to improve the crystals size and morphology.

The end goal to meet the defining objectives of this research are to send crystals to be X-ray diffracted (XRD) at the ALS Synchrotron Facility (Berkeley, CA) and collect the data to attain an atomic resolution; thus, giving potential to solve the structures of the proteins presented. The long-term goal of this research project can provide insight regarding structural details of the lambda virus enabling a better understanding of the dynamic characteristics and function of ENu3 and in turn, provide novel therapeutic targets and approaches to treat virus-mediated diseases.

.

CHAPTER 2

METHODOLOGY

Materials and Chemicals

Chemicals were purchased from Sigma Aldrich, Research Products International, and Biorad. Crystallization screens, optimization trays, reagents, and tools were purchased from Hampton Research and Molecular Dimensions. Image J was used for densitometry analysis. Pymol was used to create the homology model structures.

Initial Crystallization Optimization of ENu3

Initial Crystallization Optimization of ENu3. Capsid and scaffolding proteins previously purified were provided by Dr. Marcos Ortega. and optimized using experimentally known crystallization conditions (Zhu, 2011). Co-crystallization of ENu3 was optimized via the hanging drop method using 24-well trays. Each well in all optimization trays contained 800 μ l of mother liquor buffer in the reservoir well. Each drop contained a 1:1 ratio of protein to crystallization reagent, where various drops had diluted 11 mg/ml protein sample (1:2, 1:3, 1:5, 1:7, and 1:10), forming a final drop volume of 2 μ l that was pipetted onto a glass cover slip set up in a triangular array. The glass cover slip was sealed upside down for vapor diffusion to occur and induce crystal growth. The plates were incubated for 1-2 weeks at 25° C or until any sufficient crystal growth was observed.

Growth and Expression of Recombinant pET7CAP[E_{W308A}]

Media containing tryptone, yeast extract, NaCl were autoclaved using (Priorclave 150L/320L). A frozen glycerol stock of cell line BL21(DE3)-pET7CAP[E_{W308A}] was used to inoculate 10 ml Luria Bertani (LB) broth medium containing 50 μ g/ml Ampicillin and incubated shaking overnight at 37° C. The overnight starter culture was then added to 1 L LB broth containing 100 μ g/ μ l ampicillin and maintained shaking at 37° C until an absorbance OD₆₀₀ of 0.4-0.5 was reached. Expression was induced with 300 μ l of 0.5 M IPTG and maintained shaking incubated at 25° C for 2-3 hours. Cells were harvested by centrifugation with a J-14 rotor 5000 RPM for 10 minutes (Beckman-3000). The pellet was obtained and resuspended in 200 ml lysis buffer (20 mM Tris pH 7.4, 10 mM NaCl, 10%

glycerol) and homogenized using a glass dounce until sample was completely resuspended. Cells were lysed using a N₂ high pressure homogenizer (Dyhydromatics). Prior to lysis, all buffers were chilled on ice and the homogenizer was washed with 50 ml 5% ethanol, following equilibration with 50 ml lysis buffer. Resuspended pellet was homogenized twice, following addition of 50 ml lysis buffer to flush out excess sample to collect approximately 150 ml of lysis. Sonication was used as an alternative lysis method, where 1.4 ml of 0.4 mg/ml lysozyme was added to 200 ml of the resuspended pellet and incubated at 4° C for 30 minutes. The lysis was sonicated for 10 second bursts and placed on ice for 1 minute and repeated for a total of 5 times. For both lysates collected by high pressure homogenizer and sonication, 100 µl of 10 µg/ml DNase was added following centrifugation using a J-10 rotor at 10K RPM to collect the soluble supernatant.

ENu3 Purification

Anion-Exchange Chromatography of ENu3

Supernatant containing expressed capsid proteins were purified by anion-exchange chromatography using a UNO-Q Ion-Exchanger column (Sigma Aldrich) and the elution was monitored by Biologic LP Data Flow Program. Prior to addition of the supernatant, the column was first flushed with elution buffer (20 mM Tris pH 7.4, 1 M NaCl, 1 mM EDTA) for 10 minutes to remove any remaining content within column. The lines and column were cleaned with 1 M NaOH for 10 minutes following charging the column with 1 M NaCl for 10 minutes. The column was then equilibrated with binding buffer (20 mM Tris pH 7.4, 1 mM EDTA, no salt) for 3 ml/min for 20 minutes until the conductivity reached about 5 mS/cm. The lysis supernatant was loaded to the column with a flow rate of 2 ml/min. 20 ml of the load flow-through was collected followed by the isocratic flow of elution buffer from 0-50% to elute 2 ml fractions that were automatically collected for a total of 180 ml (Biorad BioFrac Fraction Collector). Peaks were produced by measuring the elution at A₂₈₀ versus elution time. Fractions that corresponded to the peaks on the chromatogram were analyzed via SDS-PAGE.

SDS-PAGE

All denaturing polyacrylamide gels used contained 15% acrylamide. The acrylamide solution was prepared with 60 g (30% acrylamide), 1.6 g bis-acrylamide, 300 ml nanopure water. The resolving gel was prepared using 2 ml of 5x running buffer (10% SDS, 1.8 M Tris-HCl pH 8.8), 2.8 ml nanopure water, 10 μ l TEMED, and 0.2 ml ammonium persulfate which was added to the casting gel apparatus until solidified. The stacking gel was prepared using 1 ml of 5x stacking buffer (10% SDS, 0.5 M Tris-HCl pH 6.8), 2.9 ml nanopure water, 10 μ l TEMED, and 0.1 ml ammonium persulfate which added to the top of the casting gel apparatus. All gels were run using 2 μ l of 2X loading buffer (500 μ l 1 M Tris-HCl, 100 μ l 500 mM EDTA, 0.5 g 10% SDS, 2.5 ml 7 mM β -ME, 5 ml 25% glycerol, 1% bromophenol blue) that was added with 8 μ l protein and loaded into the wells and separated with a running current of 230 amps for 30-45 minutes. The gels were disassembled and stained with Coomassie blue (10% acetic acid, 25% methanol, 0.05% Coomassie blue) and microwaved for 10 seconds and stained for 30 minutes, following de-staining in nanopure water until all Coomassie blue was removed. The gel was analyzed using Omega-Lum imager.

Dialysis Buffer Exchange

All anion-exchange chromatography samples were dialyzed the same, except minor differences in buffer or salt concentrations as stated. Phase 1 and phase 2 anion-exchange fractions were dialyzed overnight in 2 L dialysis buffer (20 mM Tris pH 7.4, 100 mM NaCl, 1 mM EDTA, 5% glycerol). Phase 3 anion-exchange fractions were dialyzed overnight in 2 L buffer (20 mM Tris pH 8.0, 100 mM NaCl, 1 mM EDTA). Dialysate was pooled and concentrated by centrifugation at 3K RPM in 1-hour increments in 4° C with 2-minute rest intervals (Eppendorf 5804R). A pipet was used to aspirate the sample to prevent any unwanted clumping. The sample continued to spin at the indicated speed until a final volume of 2 ml was obtained. Once the concentrated sample reached 2 ml, a second centrifugation step was carried out and the sample was spun at 12K RPM for 10 minutes at 4° C. The 2 ml supernatant was collected and further purified by various SEC columns, the pellet was discarded.

Size-Exclusion Chromatography

Phase 1 used the HiLoad® 16/600 Superdex® 200 pg column (Sigma Aldrich) for gel-filtration size-exclusion chromatography to purify ENu3, operated by high pressured FPLC (Bio-rad Biologic Duo-Flow). The column was first equilibrated with equilibrating buffer (20 mM Tris pH 7.4, 100 mM NaCl and 1 mM EDTA, 5% glycerol) for 75-80 minutes with a flow rate of 0.3 ml/min. The 2 ml concentrated ENu3 protein sample was loaded to the injection loop line with a syringe and the sample eluted through the column with a flow rate of 0.3 ml/min. Once 30 ml of the load flow-through were collected together, the protein eluted was collected in 1.5 ml fractions using an automated fraction collector (Bio-Rad BioFrac Fraction Collector), and a chromatogram was produced that followed the elution process measuring A_{280} over elution time. Peak fractions were analyzed via SDS-PAGE (previously described) and fractions that contained pure amounts of protein were concentrated by centrifugation (previously described) to obtain an approximate volume of 3 ml. The absorbance was measured at 280 nm to determine the concentration using a Nanodrop, and centrifugation continued until a final concentration of 10 mg/ml, or greater was reached. Phase 2 SEC column, preparation, and FPLC elution parameters were the same as in phase 1, except the equilibrating buffer used (20 mM Tris pH 8.0). Phase 3 SEC used the HiPrep™ 16/60 Sephacryl® S-300 HR column (Sigma-Aldrich) that was chosen instead of the S-200 column due to the wider separation m/w range (10-1500 kDa). The FPLC elution parameters were changed to have an isocratic elution flow rate of 0.1 ml/min collecting 2 ml fractions. SEC protein standards (Bio-Rad Gel Filtration Standards) were ran using a Superose-6 Increase 10/300 GL column (Sigma-Aldrich) and prepared similarly as described above. The elution flow rate was 0.3 ml/min and collected 2 ml fractions.

ENu3 Crystallization

Crystallization Screen

The purified ENu3 from SEC (phase 3) was applied to a 96-well block screening matrix to observe if any initial crystal hits would form to potentially identify an optimal condition. The screening matrix plates (Molecular Dimensions) contained 96 different known crystallization reagents that have

crystallized various proteins (JCSJ, Structure 1&2, Proplex, Midas Plus, LMB). Approximately 80 μl of purified protein was used to screen each tray, where each well drop contained 0.8 μl mother liquor reservoir buffer and 0.8 μl protein that was pipetted onto a sitting drop well and allowed for crystals to grow via vapor diffusion. Screening trays were centrifuged at 1K RPM for 3 minutes to remove any unwanted air bubbles and incubated at 25° C for a week or until crystal hits had formed. Within 1-2 days, initial hits were recovered in the following mother liquor buffer seen in Table 1.

Table 1. Crystallization Screen Conditions

Screen Array	Buffer (M)	Salt (M)	Precipitant (% m/w)
JCSG+	0.1 Na Cacodylate pH 6.5	0.2 NaCl	2.0 M $(\text{NH}_4)_2 \text{SO}_4$
	0.1 NaKPO_3 pH 6.2	0.2 NaCl	50% PEG 200 m/w
LMB	0.1 NaKPO_3 pH 6.2	-	14% PEG 4000 m/w
	0.1 Na Cacodylate pH 6.2, pH 6.5	0.1-0.4 M NaCl	11%, 12% PEG 8000 m/w
	0.1 MES pH 6.0	0.25 $(\text{NH}_4)_2 \text{SO}_4$	11% PEG 8000 m/w

Optimization

The conditions from the screening matrices that produced initial crystal hits were investigated for optimization and prepared similar as previously stated in *Initial Crystallization Optimization of ENu3*.

Microscale Matrix Seeding (MMS)

This method was applied to improve the crystal optimization hits, where the two major consistent conditions were further explored and set over various seeding trays (see Table 1). The micro-needle crystals initially formed were used to create a “seed stock” that would be set over a wide range over new optimization trays to identify potential new conditions. In the well containing the best micro-needle crystals, 10 μl of reservoir well buffer was pipetted onto the cover slip and the crystal seed was selected with forceps and placed into the drop. The seed crystal was crushed into fine pieces using a crushing spade tool and aspirated with a pipet until mostly broken up. The drop seed was added to 100 μl of freshly prepared mother liquor that corresponded to the optimization condition

in the well, now called the working seed stock. The solution was vortexed for 30 seconds, with 15 second intervals cooled on ice that was repeated 2x or more to ensure crystals were broken up into finer pieces. Seed crystals were not completely solubilized but rather broken into very fine nano particles which followed diluting to 1:100, 1:500, and 1:1000 with the same reservoir buffer. Re-optimization using the prepared working seed stock would be applied to the optimization trays creating a drop containing equal parts of seed stock, fresh protein, reservoir buffer. The optimization would be carried out the same way except now with the addition of the seed stock.

Crystal Harvesting

Selected crystal hits from the optimization seeding matrices were harvested and added to a cryo-protecting buffer agent containing 2.5 M ammonium sulfate and the same mother liquor buffer crystals were grown in. The harvesting drop was placed on a micro-mount needle and placed in a storage puck containing liquid nitrogen that would flash-freeze the drop. Harvested crystals were sent to the Lawrence Berkely Berkley National Laboratory ALS Facility for XRD using a synchrotron and detection scans from the scattering light were taken at 0° and 90° and data was collected to analyze the diffraction results.

Diffraction Processing

Data obtained from the ALS Facility was processed by Dr. Marcos Ortega. All detected image frames were processed using HKL-2000. A peak search was first applied to scan for sharp peak spots on the diffraction pattern over the entire data set. Defined peak spots were identified and refined to fit into the best possible Bravais Lattice space group. The triclinic space group P1 was chosen to identify the orientation symmetry of the lattice. All image frames were integrated and continued to be refined until the best space group was identified. Phase determination by molecular replacement using Phenix/Phaser will be used in the future to determine structure.

CHAPTER 3

RESULTS

Phase One

Initial ENu3 Crystallization

The crystallization of ENu3 complex was previously investigated in the laboratory. A purified 18 mg/ml ENu3 sample stored frozen at -80° C was provided to continue the research; the objective was to optimize the crystallization of ENu3 complex that would give potential to produce reliable diffraction data. The previously identified optimization conditions that would yield any slight formation of crystalline solid were investigated and conditions were modified to identify the condition that provided an environment suitable for crystal growth. The optimization trays yielded similar results seen in previous work in this laboratory. The optimization results that were commonly observed were formation of microneedle crystals that were too small, disordered crystal particulates, formation of precipitate, or clear drops indicating no nucleation and growth. The initial micro-needle clusters formed were too small to be used for XRD and if remained incubated to allow growth, over time this would form disordered structures with varied appearances.

Figure 7 displays the similarities and differences between the current ENu3 crystallization work (a) and the previous optimizations (b). The shape and morphologies are similar in size and shape but differ in the crystallization buffers pH, where both drops had same concentration 0.1 mM NaKPO₃ but differed in a pH of (a) 5.6 and (b) 6.2. Micro-needles clusters formed are the most common result seen in the optimization screens. The protein appeared to grow from a center nucleation point that branched off in different directions, thus calling the form a needle cluster “asteroid”. There also appears to be slight precipitation in the drop forming, which is a result of longer incubation times. Drops not shown would form the needle clusters and then transition to precipitate over time, where some drops would contain a heterogenous structure containing needles or particulates which cannot be identified as true precipitate.

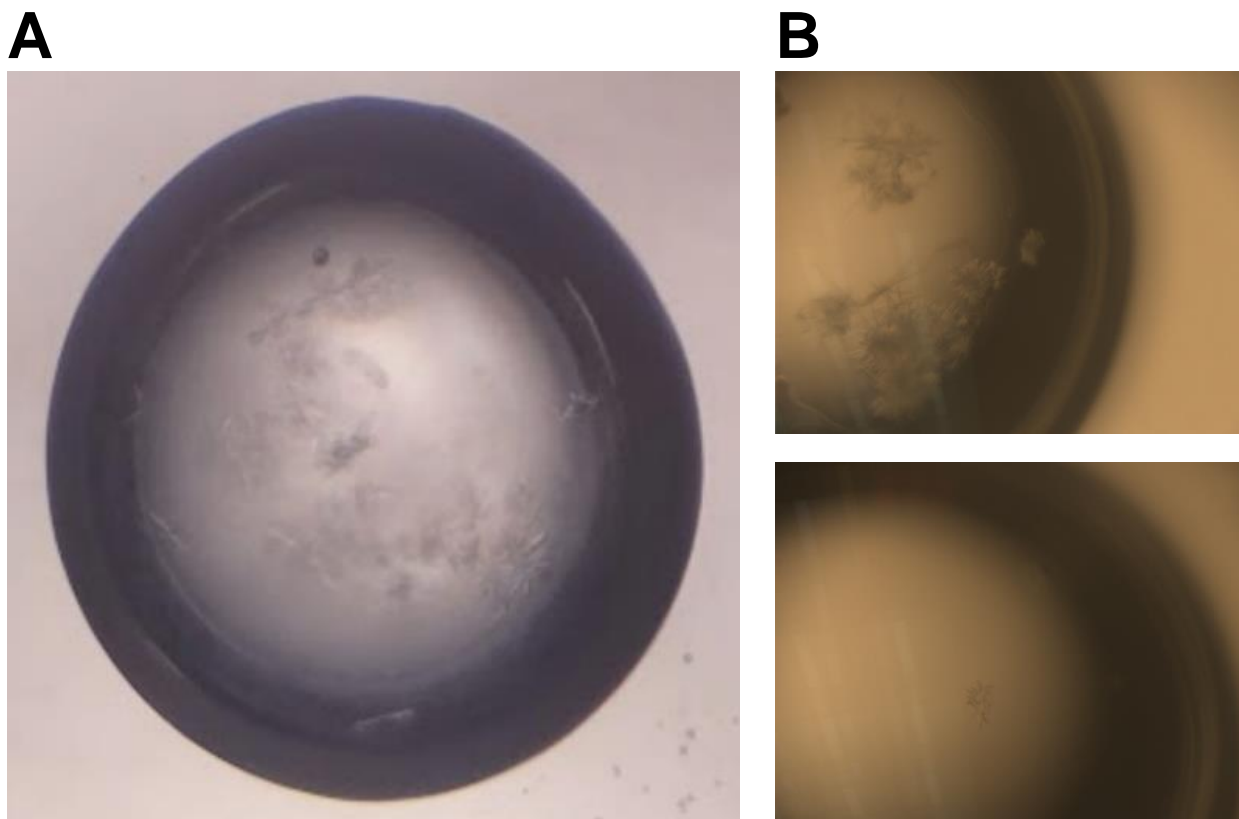


Figure 7. Initial ENu3 Crystallization Optimization. A) Current work using previously purified ENu3 was set using the hanging drop method for crystals to grow by vapor diffusion. Well drop contained 1:5 of diluted protein to a concentration of 2.2 mg/ml and crystallized in 0.1 mM NaKPO₃ pH 5.6, 10 % PEG 8K M/w, 0.2 M NaCl. The cover slip contained three dilutions for a single condition, 1:5 (shown), and 1:3, 1:7 (not shown). B) Previous crystallization studies under the Ortega Lab research at Macalester College (St. Paul, MN). The ENu3 protein complex was optimized with conditions 0.1 mM NaKPO₃ pH 6.2, 0.2 M NaCl, 10% PEG 8 K (above). Drops remained clear (below) using 1:2 diluted protein of 9 mg/ml ENu3 set in the same conditions.

Table 1 shows the screening wells that yielded the best initial crystal hits. These conditions were identified in JSCG+ and LMB screening matrices and were optimized using varying amounts of each reagent. The optimal condition that would consistently yield crystal needles were using 0.1 M pH 6.0-6.2 NaKPO₃, 0.1-0.4 M NaCl, and 10%-15% of PEG 8000 m/w.

ENu3 Expression and Purification

Previously transformed BL21(DE3) pET7CAP[EW308A] cell line containing lambda structural genes (B, C, Nu3, D, EW308A), was used for a new growth to isolate gpE+gpNu3 for structural characterization. Following expression, the lysate was collected and purified using anion-exchange column (Mono-Q 5/50 GL) which was further purified by SEC using a S-200 column. A Sup-6 column was used elute protein standards by SEC that was used as a size reference to identify the ENu3 complex native size (Figure 8A). The chromatogram produced in Figure 8A was used to create a

standard curve by taking the elution volume (ml) versus log molecular weight (Figure 8B). The separation enabled the understanding of what the native size and molecular weight of ENu3 complex is, but also provided insight to the various stoichiometric ratios it forms which can be estimated based off the SEC and using SDS-PAGE to verify the monomer subunit.

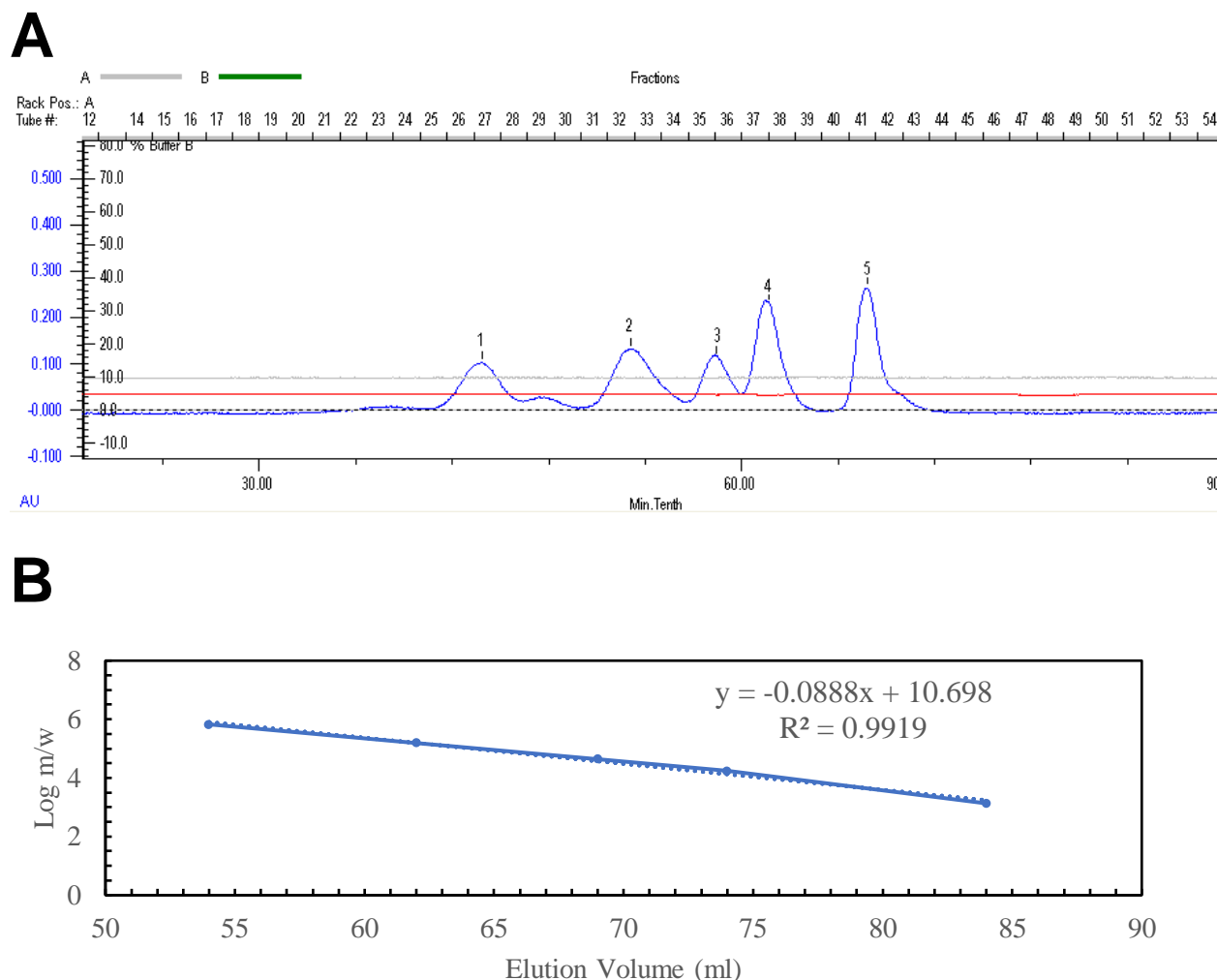


Figure 8. SEC Protein Standards Elution Profile. A) SEC protein standards (Bio-Rad) were eluted using a Sup-6 SEC column and sizes are indicated 1) Thyroglobulin (bovine): 670 kDa, 2) γ -globulin (bovine): 158 kDa, 3) Ovalbumin (chicken): 44 kDa, 4) Myoglobin (horse) 17 kDa, 5) Vitamin B12: 13.5 kDa. B) SEC standard curve of elution volume (ml) versus log m/w corresponding to a).

In Figure 9A, the chromatogram elution appeared to have a single broad flat peak which can indicate a few things, either a single species is present or the formation of heterogeneous small intermediate complexes that are relatively the same size. The standard size marker of 44 kDa is indicated which might suggest this protein eluted at 60-120 minutes with an elution volume 40 ml corresponding to the peak, is a single subunit of gpE. The monomer unit of gpE is 38 kDa and can be

verified by the SDS-PAGE in Figure 9B which shows moderately pure gpE at the 38 kDa size marker that correlates with the peak fraction size. Fractions 29-34 may have contained small amounts of gpNu3, indicated by the 17 kDa standard marker (myoglobin), although the gel shows most gpNu3 is absent which raises the question as to what form of gpNu3 is present in these fractions. The SDS-PAGE shows a faint band at the 18 kDa marker indicating minor presence of gpNu3, although most fractions did not contain gpNu3 as the absence of a band at this marker indicates gpNu3 was lost. There appears to be multiple fragments near the 18 kDa marker which may be a result of degradation and these bands represent possible degradation products. Since gpNu3 is mainly absent as seen in the gel, this may also indicate dissociation between scaffold and capsid protein, which may have caused gpNu3 to elute in later fractions due to the smaller size. As these fractions majority contained gpE correlating with the gel (Figure 9B), the absence of a band at the 18 kDa marker on the gel indicates little to no gpNu3 is present and thus dissociated from the ENu3 complex form. The broad wide peak may represent partial to full assembled capsid where gpNu3 is shown to not be present in the final structure, thus and not present in the final capsid structure, thus the 38 kDa band represents gpE dominates at this column fraction and can be in higher-order forms assembled in various subunit ratios. In early (22-23) and later (31-34) fractions very little gpNu3 appears to be present on the gel (Figure 9B), although the minor double bands under the 18 kDa marker suggests degradation of the scaffold which can correlate with it being cleaved by viral proteases thus not being in the final capsid structure. These results suggest gpNu3 was lost during the purification as it appeared to be dissociated from gpE. Previous experimentation in attempts to crystallize gpE alone without the gpNu3 scaffold have failed to crystallize numerous times; yet in the presence of gpNu3, this allows for crystallization to occur as a particular stoichiometric ENu3 complex forms a stabilized structure that may favor lattice nucleation and growth.

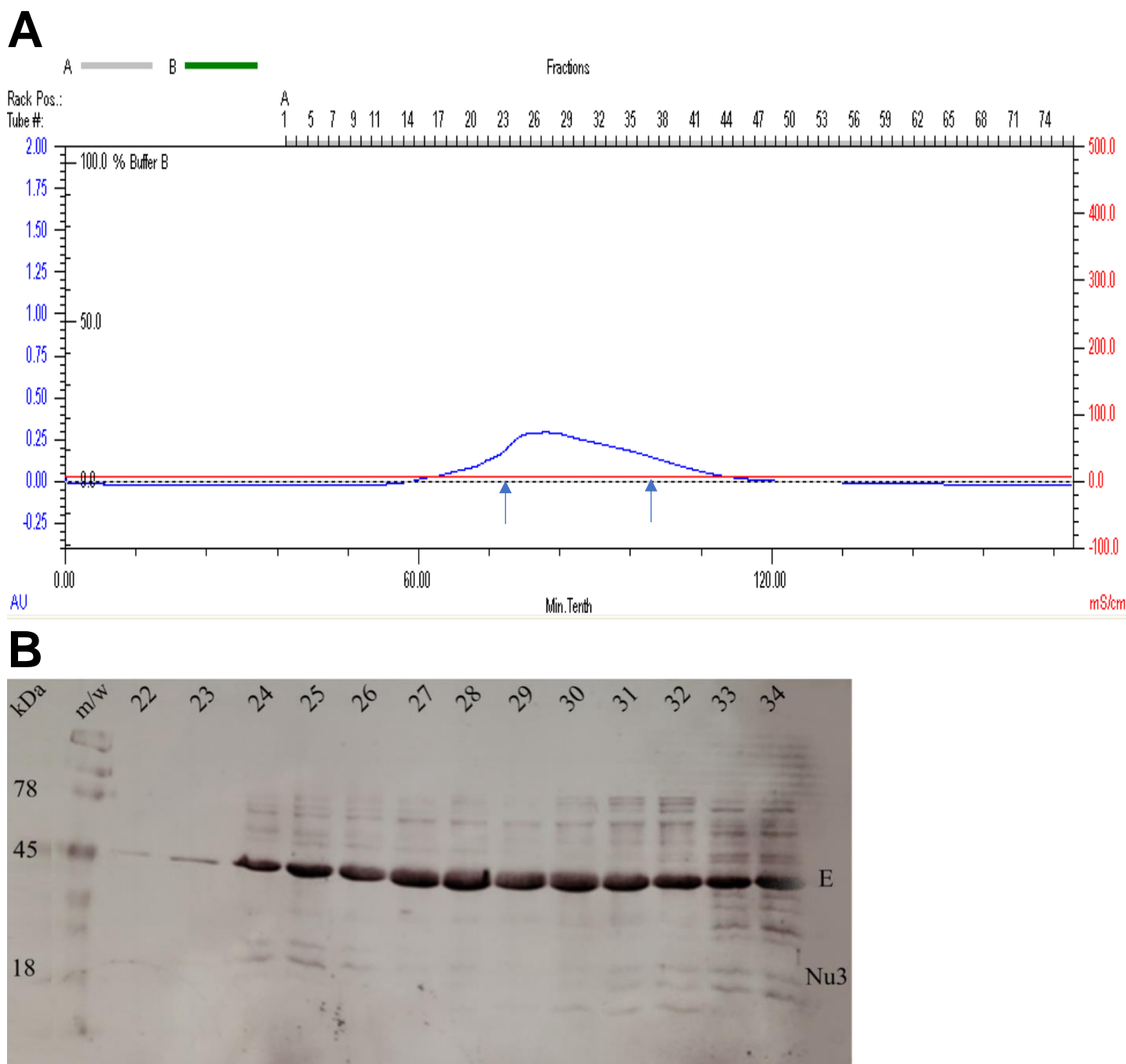


Figure 9. Phase 1 SEC Purification of ENU3. A) Chromatogram elution of ENU3 shows a single broad peak with a measured A_{280} of 0.25. The left arrow indicates the 44 kDa standard (Ovalbumin) and the right arrow indicates 17 kDa standard (Myoglobin). B) Fractions 22-34 corresponding to the single peak in (a) were analyzed via SDS-PAGE. At 38 kDa, the gpE monomer predominates being the most abundant protein separated in this purification, with little to no presence of gpN3 at 18 Kda marker.

Phase Two

ENU3 Expression and Purification

A glycerol stock containing BL21-DE3 pET7CAP[E_{W308A}] cell line was used to start a new growth and expression of capsid proteins as phase 1 expression failed to purify the gpE+gpNu3

complex. In phase 1 the Mono-Q column that is a strong anion-exchanger might have separated Nu3 from E, and to observe whether the column was a factor that affected the purification, a weaker anion-exchange column was used Diethylaminoethyl (DEAE) column and produced a chromatogram as fractions were collected and measured at A_{280} .

Peak fractions were analyzed to determine the presence of the ENu3 complex, where the fractions that corresponded to the second peak was appeared to have majority ENu3 (Figure 10A). The fractions collected within 160-190 minutes were analyzed via SDS-PAGE (Figure 10B).

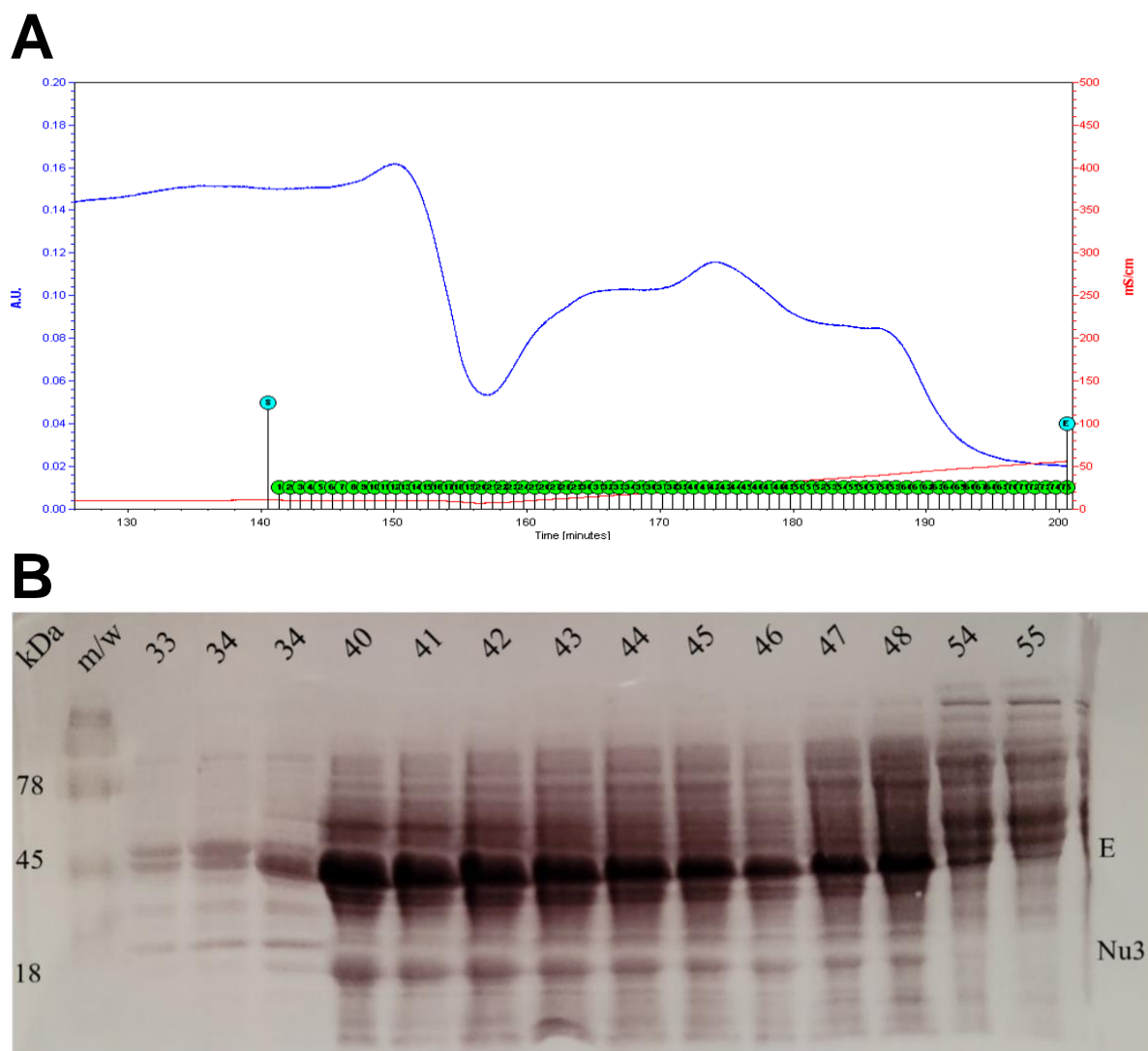


Figure 10. Phase 2 Anion-Exchange Chromatography of ENu3. A) Chromatogram following the elution of expressed ENu3 using a DEAE anion-exchange column. Peaks are measured at A_{280} versus elution time, fractions are labeled in green. The purification of Enu3 complex is suspected to be in fractions that correspond with the peaks on the chromatogram, indicated by red arrows. B) Peak fractions 33-35, 40-48, 54-55 in (a) were analyzed by SDS-PAGE. Majority of Enu3 elutes at fractions 40-48 indicated by the dark bands at 38 kDa (E) and 18 kDa (Nu3).

The SDS-PAGE shows dark bands at the 38 kDa and 18 kDa marker, indicating presence of monomer gpE and gpNu3 forms, respectively. From the three peak fractions in the chromatogram (Figure 10A), the second peak corresponding to fractions 40-48 contained the highest amount of ENu3 seen in SDS-PAGE, which suggests that both proteins were eluted in this fraction volume and can be predicted that the complex was stable to remain associated. The first peak was separated on the gel has a faint band in fractions 33-35 and appears to contain little to no gpNu3 near the 18 kDa marker (Figure 10B), indicating that little to no gpNu3 was present in these fractions. The third peak containing fractions 54-55 also appeared to have a loss of gpNu3 as seen in the SDS-PAGE. Although contaminants are present in fractions 38-48 corresponding to the second peak, both gpE and gpNu3 are present in high amounts which might suggest that these proteins remained associated with one another.

The fractions containing the highest levels of ENu3 eluted from the DEAE column was further purified by SEC to remove unwanted contaminants and other proteins that were present. This separation differed from phase 1 purification as the peaks seen on the chromatogram are more resolved than before, rather than a flat broad peak; indicating ENu3 complexes with various stoichiometric ratios have formed (Figure 11A). The different peaks produced from SEC were investigated as an ideal stable ENu3 complex was isolated in the various fractions. The fractions that correspond to peak 2 on the chromatogram can suggest a single species of one protein complex was majority present and eluted at this column volume. The ENu3 complex ratio can be estimated based on the SEC protein standard curve, where peak 2 fraction volume of 58 ml at 70 minutes contained the presence of an approximate 318.5 kDa species based on the SEC chromatogram.

Fractions 19-21 appear to have the highest purity of ENu3 as seen on the SDS-PAGE (Figure 11B), where the 38 kDa marker represents gpE and 18 kDa marker represents gpNu3. There are minor contaminants present within these fractions, but nonetheless this purification contained the highest amount of pure ENu3 complex that can be compared with phase 1 SEC that contained majority E with no Nu3. Further purification using a SEC column with a wider range to separate the

unwanted contaminants is possible, although this presents the possibility of ENu3 dissociating or Nu3 forming degradation products over time, observed in the SDS-PAGE from phase 1 SEC fractions (Figure 9B).

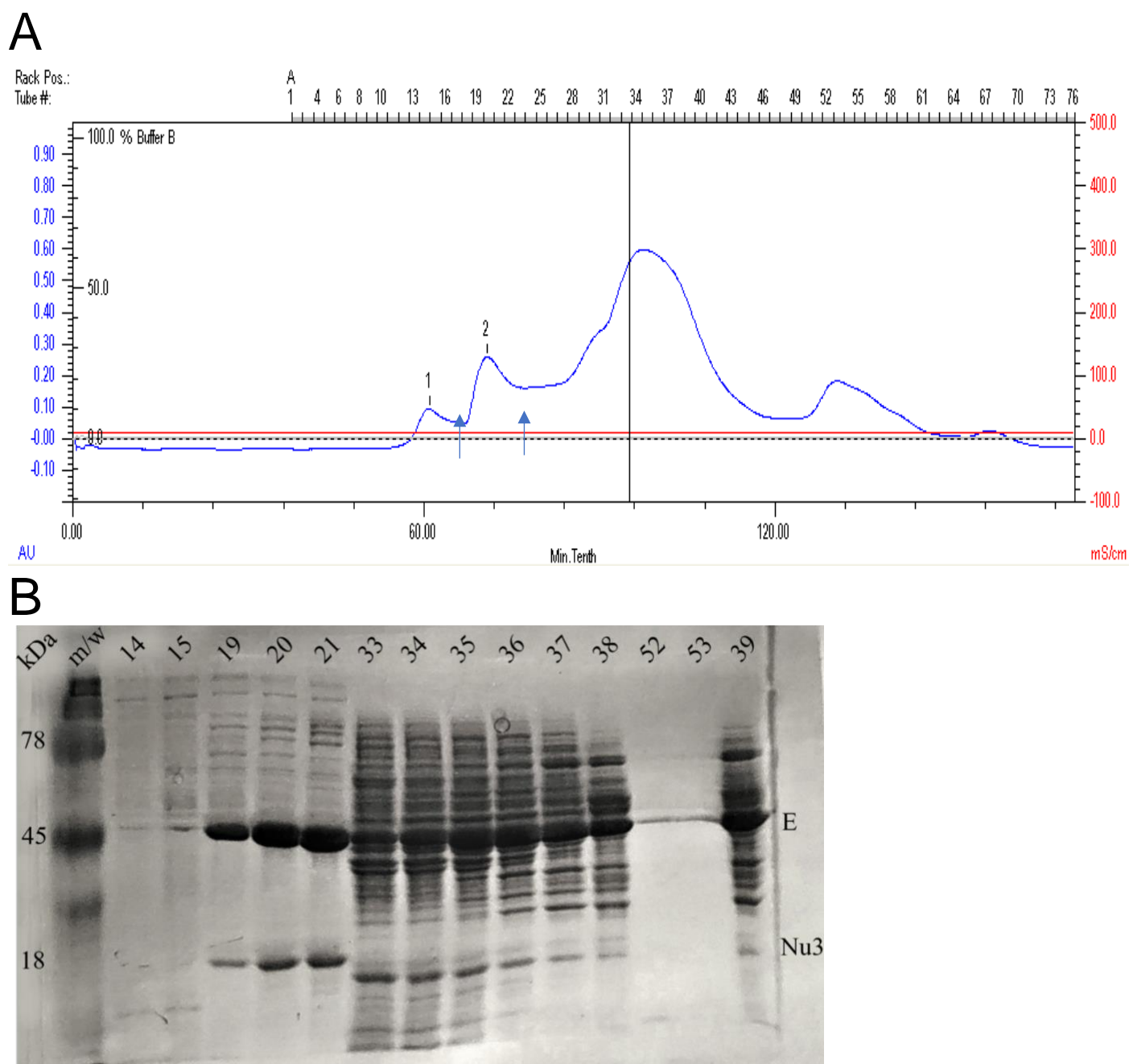


Figure 11. Phase 2 SEC Purification of ENu3. A) Chromatogram following the elution of anion-exchange purified ENu3. Peaks 1 and 2 are suspected to contain the desired ENu3 complex as seen in previous SEC purifications. Peak 1 starts to elute at a fraction volume of 49 ml at 60 minutes with a measured A_{280} of 0.15. Peak 2 starts to elute at a fraction volume of 58 ml at 70 minutes with a measured A_{280} at 0.3. The left arrow indicates the 670 kDa standard (Tyroglobulin), the right arrow indicates the 158 kDa γ -globulin. B) The fractions corresponding to peaks on the chromatogram in (a) were analyzed by SDS-PAGE to confirm the presence of ENu3 complex. Fractions 19-21 contained the highest purity of ENu3, indicated by dark bands at the 38 kDa (E) and 18 kDa (Nu3). The other peak fractions do contain sufficient ENu3 levels, although these fractions contain more contaminants than fractions 19-21.

Fractions 33-53 also contain high amounts of ENu3 complex but are contaminated with other proteins, where various bands correspond to different m/w sizes. The protein containing in fractions 19-21 was concentrated and absorbance was measured via Nanodrop with an A_{280} of 1.04 where the concentration was determined to be 10 mg/ml.

ENu3 Crystallization

The purified 10 mg/ml ENu3 sample was used to crystallize in the previous optimized conditions. It has yet to be explored what the ideal stable ratio of the ENu3 complex is that favors crystallization; this particular ratio of ENu3 has some relation that allows the state of stable nuclei to form in order for crystal grow to occur. Therefore, isolation of a homogenous single protein complex species can potentially form stable nuclei that will grow in the metastable state; this hypothesis can be based on the formation of a ENu3 association that maybe stable enough in a particular conformation to give rise to forming the crystal lattice. Similar to the initial crystal optimizations, the three main outcomes of this crystallization phase were needle cluster asteroids, no crystals, or formed precipitate. The common crystal morphology for ENu3 to form needle clusters is expected, as seen in previous crystallization studies. Although these needles appear to look slightly different in shape compared to phase 1 crystallization seen in Figure 7. In phase 1 optimization, the needles are in smaller clusters but have slightly thicker shard-like needles that extrude from the center. The phase 2 crystallization produced the common needle cluster or asteroids appeared to have a slight modification that differed from previous crystals. The needles are thinner and capillary-like extruding from the darkened center, possible implication of forming precipitate (Figure 12A). The rapid transition of stable nuclei forming to crystal growth produces disorderly crystals and as vapor diffusion continues some protein will precipitate. Comparing the conditions in this drop that formed crystals to a drop with no growth in (Figure 12B), the precipitant concentration may be the factor that initiates nucleation into growth. Higher m/w PEG may increase the rate of vapor diffusion that leads to fast forming disordered crystals and precipitate.

Over the time course it was observed that the S-200 purified ENu3 sample stored in -20° C

had changed from a clear to yellow color which indicated the sample became more concentrated; a result of aggregation to form larger structures that remained soluble. The -20° C stored ENu3 were set over an optimization tray where most drops precipitated which led to the decision to not move forward with this sample. Multiple cycles of ENu3 re-growth and expression were carried out, but numerous attempts failed to produce ENu3 in high amounts and to a high purity. Proteins expressed from these cycles were analyzed by SDS-PAGE which indicated poor expression of gpE and gpNu3; weak band signals or an absence of a band at their indicated size markers was observed after each attempt. The gel analysis also may indicate degradation of the protein complex had occurred as fragmented bands appear near the 38 kDa marker (gpE) and/or near the 18 kDa marker (gpNu3), thus coming to conclusion that degradation products of the complex were formed (data not shown). It was suspected that the glycerol stock containing the pET7CAP[EW308A] plasmid was not in high amount and may be due to a weakened BL21-DE3 cell line that resulted in poor expression or degradative cleavage.

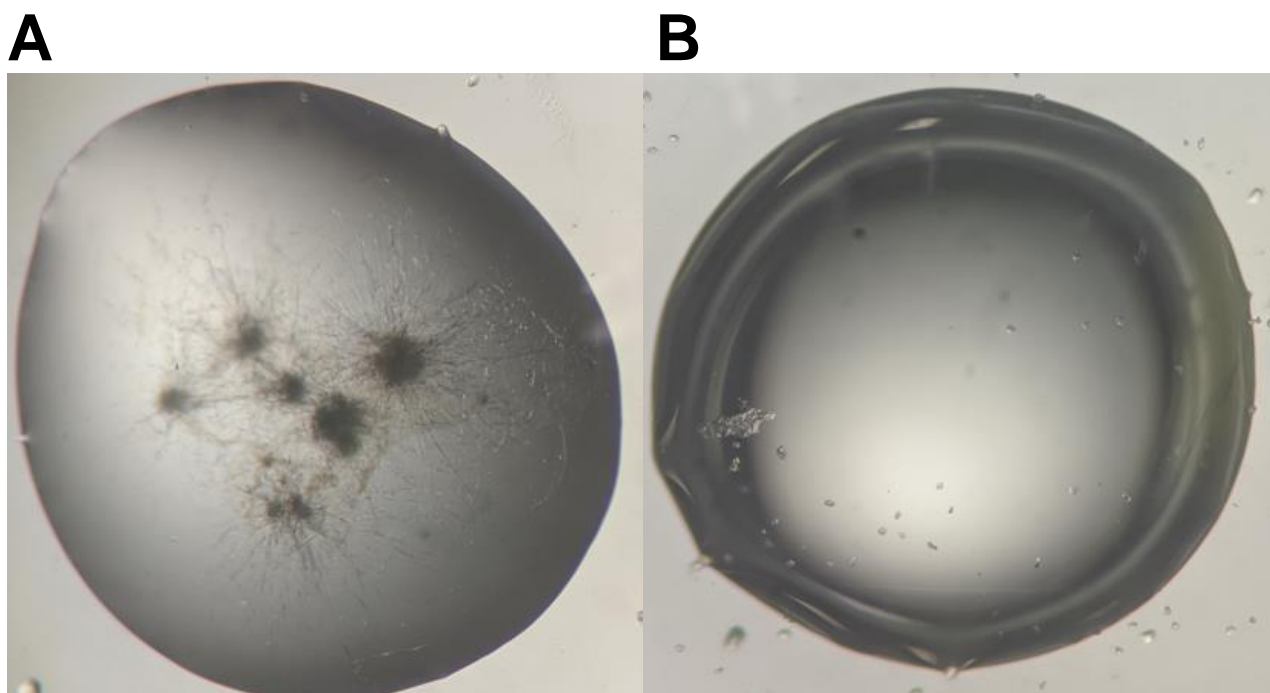


Figure 12. Phase 2 ENu3 Crystallization Optimization. The SEC fractions containing purified ENu3 was set over optimization trays, the product contained the following conditions: A) Needle cluster asteroids in 0.5 M MES pH 5.5, 200 mM Ammonium Sulfate, PEG 5% 10 K m/w. B) No crystals formed in 0.5 M MES pH 5.5, 200 mM Ammonium Sulfate, PEG 10% of 8K.

Phase Three

ENu3 Expression and Purification

To replicate the expression of ENu3 in high amounts as seen in the prior attempt (phase 2), a fresh transformation of the pET7CAP[EW308A] plasmid into competent DH5- α cells was intended to produce the plasmid efficiently in greater quantities, which was followed by isolation of the correct plasmid to be re-transformed into fresh competent BL21-DE3 cells. Phase 3 utilized the new transformed BL21-DE3 pETCAP[E_{W308A}] cell line that expressed ENu3. The protein was purified by anion-exchange chromatography to produce the chromatogram (Figure 13A) which measured A_{280} over the elution time. The fractions were analyzed by SDS-PAGE (Figure 13B).

Around 170-180 minutes with an A_{280} at 0.5, the peaks measured indicate presence of protein collected in these fractions, where fractions 32-36 correspond to peak 1 and fractions 39-44 correspond to peak 2 (Figure 13A). The fractions collected at the 170-180 minutes that corresponded to the peak absorbances were analyzed by SDS-PAGE and a presence of a band at the monomer size markers for ENu3 showed the protein complex had successfully eluted in these fractions (Figure 13B). While peak 1 and peak 2 fractions both contain high amounts of ENu3, it was decided to move forward using fractions 32-26 from peak 1 as there were less contaminants. The phase 2 anion-exchange purification investigated the second peak having the highest protein amount that was further purified by SEC, while phase 3 anion-exchange purification utilized the first peak instead, as these fractions related to peak 1 had less contaminants. The peak 1 fractions being purer than before eliminates the potential of protein contaminants to interfere with the stable ENu3 complex during any remaining purification step and additionally the crystallization process. It is interesting that phase 2 anion-exchange purification produced a widened broad peak that appeared to split into forming a triplet, while in this current phase the peak splits but appears to form a doublet.

This observation of different peak symmetries and elution times can be a factor of the difference in the proteins mass-to-charge ratio; agreeing with the natural ability of ENu3 to undergo

higher-order complex transitions to form various stoichiometric ratios resulting in different chromatogram profiles.

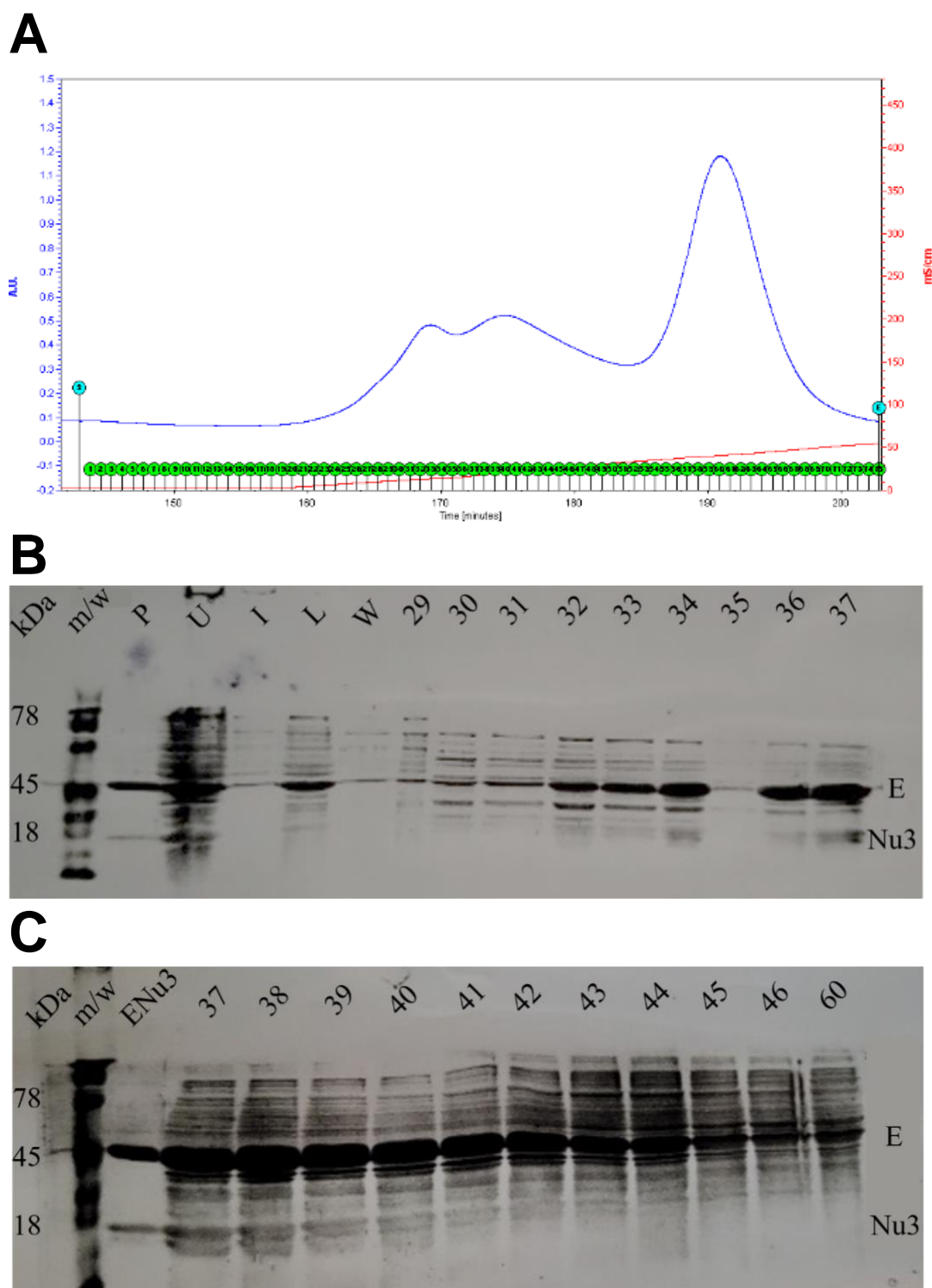


Figure 13. Phase 3 Anion-Exchange Chromatography of ENu3. A) Expressed ENu3 was applied to a DEAE-Q anion-exchange column and the elution was followed measuring A_{280} over time. The two most left peaks which are merged are suspected to contain ENu3, having an A_{280} of 0.5 with an elution time of 150-170 minutes. The tallest peak at elution time of 190 minutes corresponds to negatively charged DNA fragments. B) Peak 1 corresponds to fractions 29-36 analyzed by SDS-PAGE. Enu3: standard, P: pellet, U: uninduced, I: induced, L: load, W: wash, Fractions. C) Peak 2 corresponds to fractions 37-44 analyzed by SDS-PAGE. In both (a) and (b), a band at 38 Kda and 18 kDa indicate a presence of E and Nu3, respectively.

A difference between the previous growth and expression that may have affected these results were differences in buffers used. In phase 3, no salt was used during lysis and Tris pH 8.0 was used throughout purification, while phase 2 used Tris pH 7.4 and 100 mM NaCl. Previous ENu3 expression using Tris pH 7.4 may have enabled a protonated state of the side chains due to the lower pH and a higher salt concentration may affect the ionic strength of the protein that led to altered interactions with the protein and ion-exchange column. Lambert et al (2017) utilized Tris pH 8.0 throughout the expression and purification and so this was also used in phase 3 so it could be identified if pH was an issue that led to unwanted conformational states. The question of investigation now was determining how to selectively isolate a stable ENu3 ratio that is suitable for crystallization.

The fractions containing ENu3 from anion-exchange chromatography were eluted through a S-300 gel-filtration column. The S-300 column was used in this purification phase as it has a wider m/w separation range (10-1500 kDa), therefore it was hypothesized that the protein sample containing various unknown stoichiometric ratios of ENu3 complexes would partition into different fractions more efficiently; allowing for a better separation to isolate a single homogenous ENu3 complex.

Figure 14A shows the chromatogram following the elution of ENu3 which differs compared to previous SEC purifications. Peak 2 appears to be more sharply defined with an A_{280} of 0.5, compared to phase 2 SEC purification where peak 2 (Figure 11A) with an absorbance of 0.25. The fractions within peak 2 is expected to contain a desired stable ENu3 complex as seen previously, thus peak 2 being much more resolved than phase 1 and phase 2 was of interest as a greater absorbance with a sharper peak might suggest presence of a homogenous protein complex. This purification step yielded an approximate 100% pure separation of ENu3 consistent with the SDS-PAGE (Figure 14B), where fractions 13-15 are free of contaminants and contain high amounts of pure monomers indicated by solid bands at 38 kDa for gpE and 18 kDa for gpNu3. Later fractions had contaminants and other proteins in addition to moderate ENu3 levels. It is important to note that peak 2 has shifted into earlier fractions in comparison with phase 2 SEC purification (Figure 14A subset). This may

indicate ENU3 are associated with one another and formed a stable intermediate form that is most abundant within fractions 13-15.

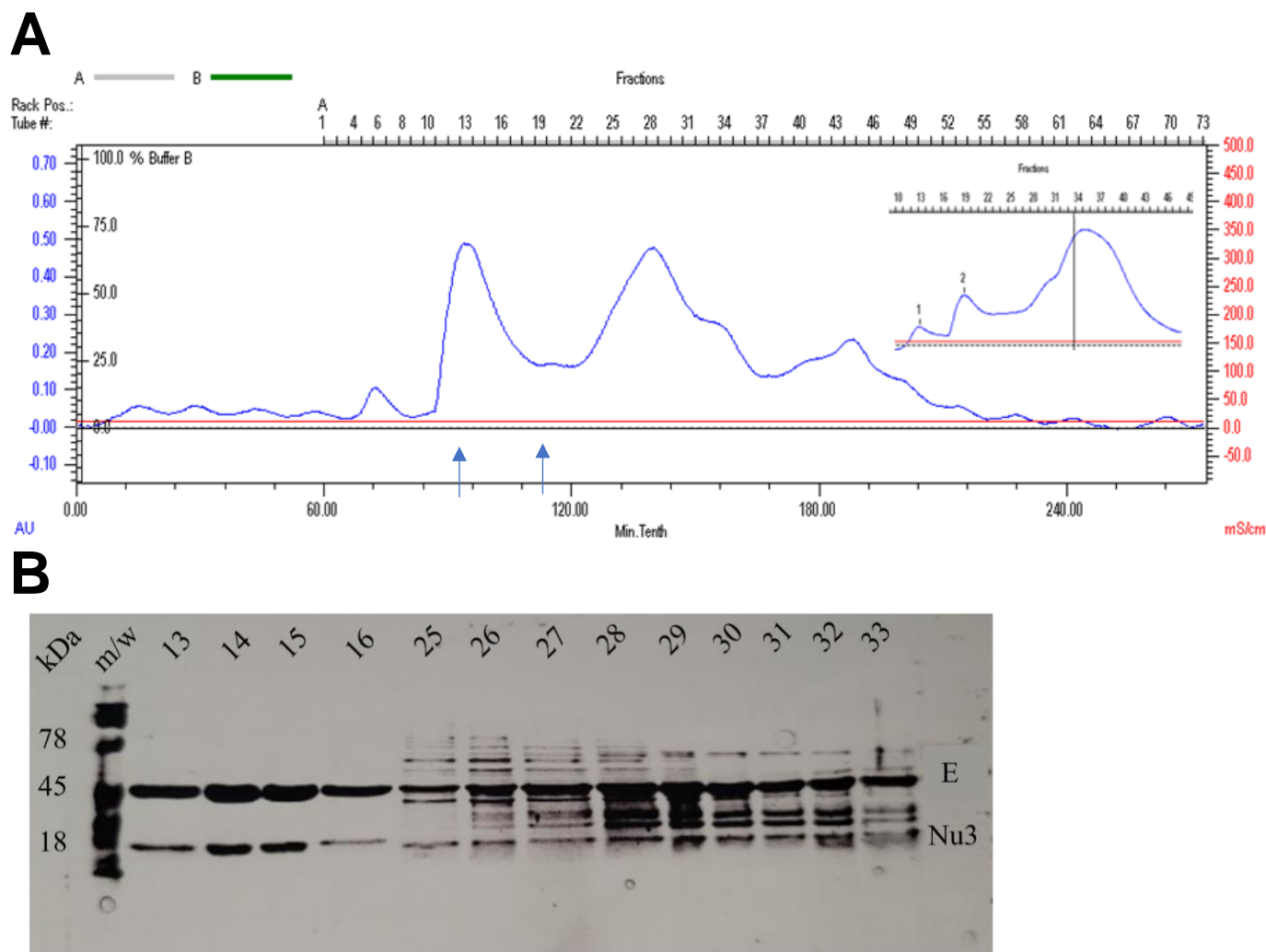


Figure 14. Phase 3 SEC Purification of ENU3 Using an S-300 Column. A) Chromatogram following the elution of expressed ENU3 from dialyzed and concentrated DEAE anion-exchange chromatography fractions 32-36 (peak 1, Figure 13A). The elution was measured at A_{280} to produce different peaks, where fractions that correspond with the peak absorbance indicate presence of protein. Fractions 13-15 were suspected to contain a stable ENU3 complex as it aligns with the most prominent defined peak at an elution time of 80-100 minutes. Subset shows phase 2 SEC purification for comparison, where the second peak differs in the fraction volume and overall elution time. The left arrow indicates the 670 kDa standard (Thyroglobulin), the right arrow indicates the 158 kDa γ -globulin B) The fractions corresponding to peak 2 and 3 on the chromatogram in (a) were analyzed via SDS-PAGE. Fractions 13-15 (peak 2) contained the highest purity of gpE at 38 kDa and gpNu3 at 18 kDa almost free of contaminants. Peak 3 fractions contain ENU3 in similar amounts but eluted with contaminants

This association complex having a particular stoichiometric ratio that is larger in size can be deduced by the peak shift into earlier fractions. Although the SEC purification can suggest a larger ENU3 complex has formed, the size and ratio can be calculated using SEC protein standards, but this only gives prediction as to what the complex ratio may be. The arrows indicate the protein standards

of 670 kDa (Thyroglobulin) and 158 kDa (γ -globulin) which can infer that this peak containing the fractions of eluted ENu3 has a protomeric size that is within this range. If a full constructed hexamer was formed, this would have a total of 6 gpE and 18 gpNu3, having a 1:3 ratio where the overall theoretical m/w is estimated to be 546 kDa, falling within the range of the protein standards on the chromatogram (Figure 8A). This observation is consistent with the SDS-PAGE in (b) that has a strong band at the 38 kDa and 18 kDa markers only, suggesting that majority of protein within fraction 13-15 is a homogenous single protein complex that aligns with the size estimate and sharp peak on the chromatogram. The outcome of phase 3 growth and expression was successful compared to phase 1 and phase 2 as the SDS-PAGE confirms the highest purity of ENu3 which is favorable for crystallography. Fractions 13-16 were concentrated to reach a final volume of 750 μ l and the concentration determined to be 10.5 mg/ml using the extinction coefficient of $36,330 \text{ M}^{-1} \text{ cm}^{-1}$.

ENu3 Crystallization

The 10.5 mg/ml ENu3 sample collected from SEC (phase 3) was re-screened to identify any other optimal condition that can be further assessed. The JSCG+ and LMB screening well-blocks produced the best potential hits in which these conditions were investigated for optimizing crystal growth of ENu3 (Table 1).

A 24-well optimization plate was used to set drops with the 10.5 mg/ml ENu3 sample to allow for crystal growth by vapor diffusion via the hanging drop method. As anticipated, formation of needle cluster asteroids was produced, a common crystal morphology seen in previous optimization screens. Figure 15A and Figure 15B displays the outcome of the optimization, which appeared to form needles that grow from a center core, more similar to the morphology from the initial optimizations (Figure 7). The crystal morphology seen in phase 2 optimizations (Figure 12) were not produced as this current phase 3 crystallization optimization rounds formed thicker needle shard-like morphologies similar to a snowflake, rather than the phase 2 optimizations needles that appeared to be much more thinner shards resembling that of a capillary system.

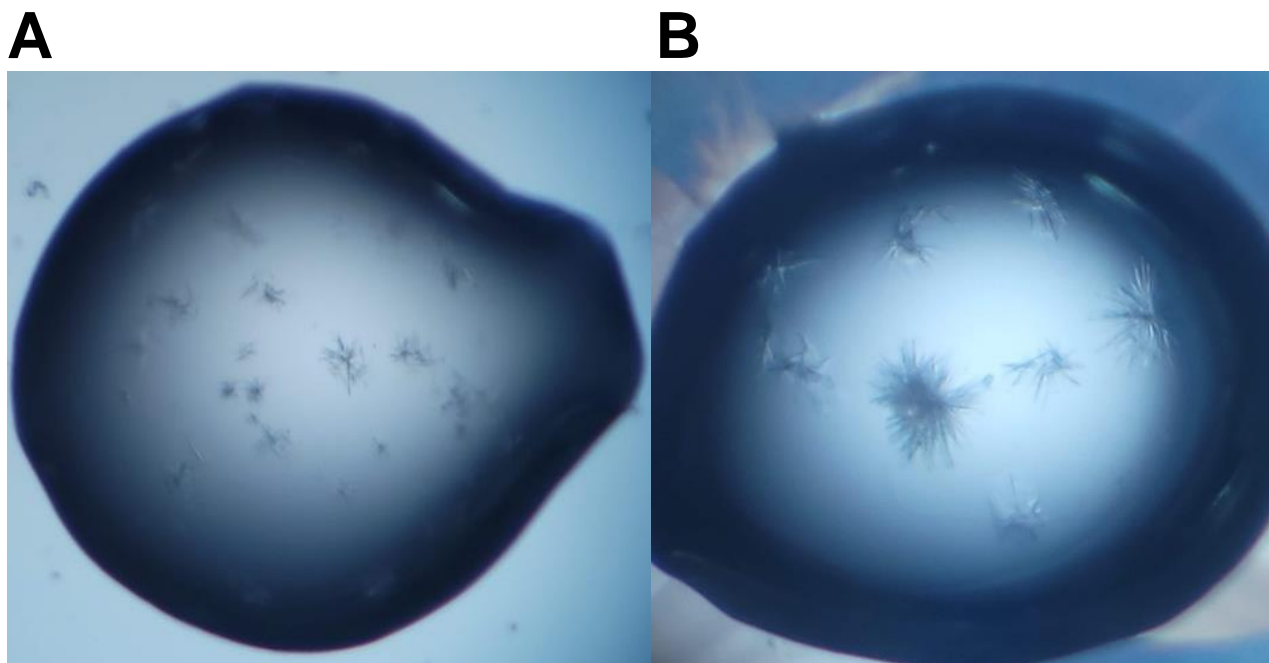


Figure 15. Initial Optimization of ENU3 from Phase 3 Purification. Purified SEC ENU3 (10 mg/ml) was used to set optimization trays in the following conditions: A) 3.3 mg/ml ENU3 (diluted 1:3) crystallized in 200 mM NaK Phosphate pH 6.2, 150 mM NaCl, 13% PEG 8K. B) 2 mg/ml ENU3 (diluted 1:5) crystallized in same condition above except 17% PEG 8K.

Microscale Seeding

The phase 3 crystallization optimization consistently yielded crystal morphologies that were the same as before. As the growth and purification cycles were carefully evaluated to produce a better crystallization result, needle clusters were the repeated outcome, thus the third objective was to be investigated next that would utilize the method of microscale-matrix seeding (MMS). The aim was to improve the crystal growth and morphology to attain the best possible maximum size that would give potential to provide stronger diffraction. MMS has been applied to improve the crystallization of many proteins having poor morphologies; thus, it was hypothesized that this technique would improve the size of these crystals.

The first round of seeding produced results that were promising as the growth differentiated greatly from thin needle cluster asteroids commonly seen to larger forming rectangular plate-like crystals shards that extend (Figure 16). The seed from phase 3 initial optimization crystal trays (Figure 15) was used to create a seed stock that would act as a scaffold to initiate stable nucleation using fresh protein sample to crystallize in the same conditions. As these crystals appeared to be

small, the goal was to improve poor quality crystals that could not grow to a maximum size without precipitate formed, therefore crystal seeding optimization was applied to improve the lattice growth to a maximum unit cell dimensions. The drops set for the first MMS round differed only in precipitant concentration of 11% PEG 8K (Figure 16A) and 13% PEG 8K (Figure 16B). The difference in precipitant concentration was minor but higher ENU3 concentration of 3.3 mg/ml (Figure 16A) favored lower PEG concentration of 11% while lower 2 mg/ml ENU3 concentration grew with higher PEG at 13%. No further nucleation or growth was observed in this seeding round. This initial outcome was a success as morphology improved from micro needles to a rigid shard-like rectangular symmetry which had not been observed previously.

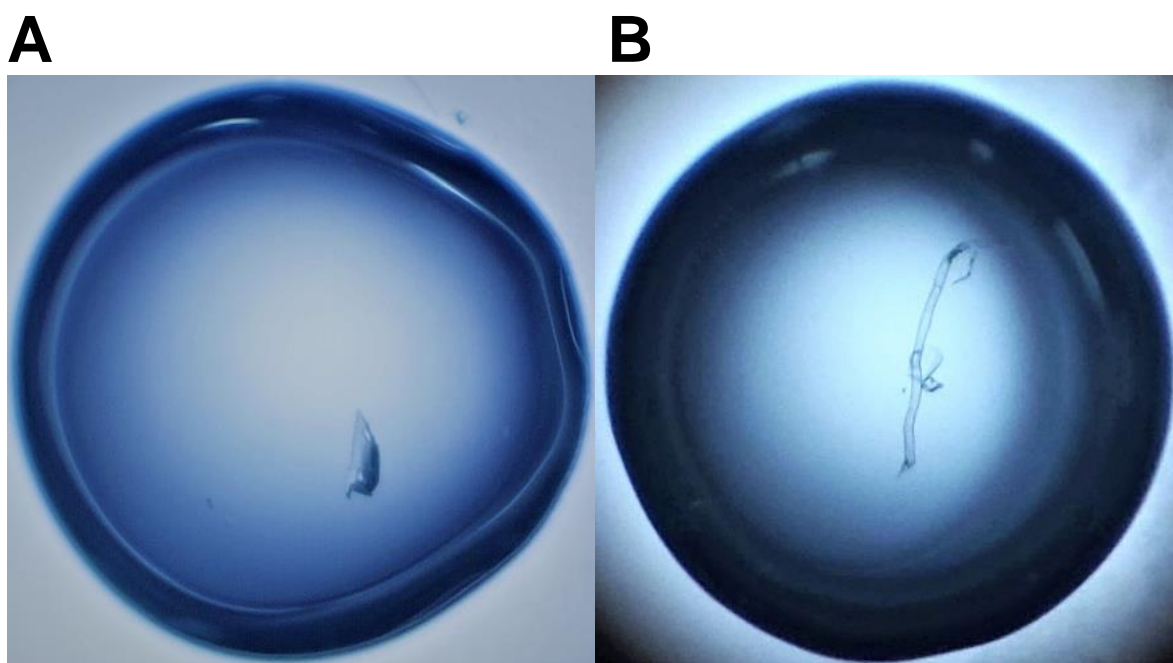


Figure 16. Round 1 MMS Optimization of ENU3 Initial Crystal Hits. Seed crystals from Figure 15 was used to prepare a seed stock that was set over new trays using the same optimization conditions and protein concentration as before. Wells that yielded hits were in conditions: A) 3.3 mg/ml ENU with 200 mM NaK Phosphate pH 6.2, 50 mM NaCl, 11% PEG 8k. B) 2 mg/ml ENU3 with same condition as drop (a) except 13% PEG 8K was used as the precipitant.

Additional rounds of MMS would be applied to improve the crystal yield to grow to a maximum size within unit cell dimensions. Larger crystals are a product of more atoms packed within the formed lattice, although the symmetry of the atom arrangement is dependent on formation of stable nuclei that serves a site for protein molecules to grow in an organized array to resemble symmetrical characteristics which can be correlate with geometrical lattice symmetry elements formed by the

atoms within the crystal lattice. To better improve the size of these crystals, a second round of seeding was carried out using the rectangular shard crystals (Figure 16) as a seed crystal to create new seed stock that would introduce more stable micro-nuclei to the drop to serve as a site for fresh protein to grow into the lattice.

The second round of microscale seeding yielded more desirable results that have not been observed in any previous crystallization experiment of ENu3. Figure 16 shows seeding had been successful in production of large three-dimensional crystals in greater quantity that formed larger shard-like rectangular symmetries. The crystals grew within 1-4 days though longer incubation times continued to diffuse the drop reagent concentration into the reservoir well which pushed the drop to a zone of supersaturation that led to formation precipitate. The drops shown in Figure 17 contained the same protein concentration and buffer pH but differed in salt and precipitant concentration. The lower salt and precipitant concentration may be more favorable for nucleation and growth (Figure 17A) compared to the drop with higher salt and precipitation concentration that might induce faster crystal formation (Figure 17B). The higher salt and precipitant concentration may dehydrate the drop quicker to induce an increased vapor diffusion rate and as a result began to start forming needles. Some crystals in this drop formed rectangular plate-like crystals, while other crystals appear to look more like the needle structures that extrude from the center. The lower concentration of salt and precipitant ratio may have allowed for steady-growth to occur, thus forming plate-like morphologies that stack on top of one another.

The previous needle crystals formed in phase 1 and 2 crystallizations were insufficient in size for XRD. The latest results were the outcome of the MMS method which produced viable crystals with a distinct rectangular plate-like morphology that can be used for XRD, giving potential to provide stronger and viable diffraction. To attain the best possible crystal, multiple rounds of seeding matrices over a wide range of optimization conditions were applied to increase the probability of a stable nucleation event to occur in order to crystals to efficiently grow in size.

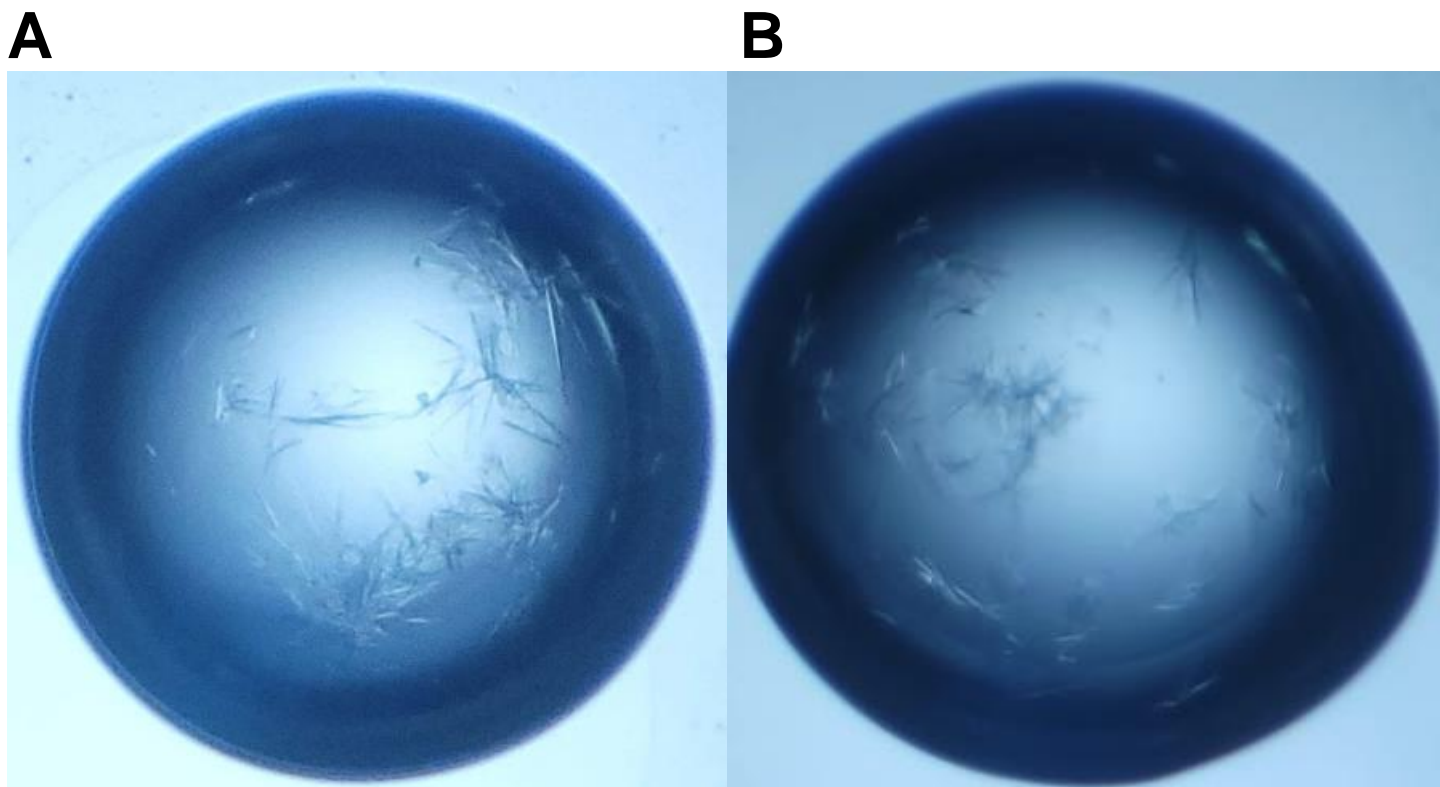


Figure 17. Round 2 MMS Optimization of ENU3. Seed stocks were prepared using the seed crystals from round 2 MMS (Figure 16) and diluted 1:50 and 1:100. 5 mg/ml ENU3 was used to set each drop and crystallized in the following conditions: A) 2.0 M NaK Phosphate pH 6.2, 40 mM NaCl, 12% PEG 8K B) 45 mM NaCl, 16% PEG 8K, same buffer and pH in (a).

CHAPTER 4

DISCUSSION

The results presented here outline the crystallization of gpE-gpNu3 protein complex from phage lambda. Information of the interatomic details of these proteins are lacking as no crystal structure has been determined to date. High-resolution electron microscopy studies have presented insights on the overall shape and configuration of gpE, although only a crystal structure solved to an atomic resolution will allow to distinguish the amino acids that are the basis of gpE interaction and function.

Previous attempts to crystallize gpE and gpNu3 have been partially successful in yielding micro needle crystals (Figure 7). In order to meet the research objective, various purifications were conducted and used to optimize the ENU3 protein complex in a wide range of different crystallization conditions, although despite the many attempts, the crystals produced were consistent in forming the same microcrystal needle morphology as seen previously. The needle crystals grew and formed a heterogenous population with undefined symmetries, or in longer incubation times led to formation of precipitate which limits the ability to undergo xrd to solve the structure (Figure 12). The root of this problem lies within the gpE structure itself as its shape and hydrophobic content might have an effect on the crystallization. Homology modeling and cryo-EM has shown that gpE has an unusual, disordered shape that serves a purpose in the protein-protein contacts it forms with other gpE molecules and interactions with gpNu3. Reconstructing lambda cryo-EM structure into homolog HK97 gp5 capsid crystal structure showed striking similarities, where the α -helix spine may be a factor of its flexible characteristics that allow to extend and interact with neighboring subunits; a feature that is hypothesized that might take part in the polymerization reaction into the full formed capsid shell. GpE naturally tends to spontaneously form higher-order structures to assemble the capsid, thus making it difficult to isolate the monomer form. Mutational studies of lambda capsid proteins have been employed to observe what factors may be critical in capsid formation, where it was concluded that the gpE-W308A mutation limited large capsid formation in comparison to the wildtype, indicating that

W308 plays a putative role in interactions with the decoration protein gpD (Figure 18). More importantly, gpE-W308A afforded monomer capsid subunits to be soluble which has not been seen before and thus allows to further characterize the structure via XRD (Lambert et al., 2017).

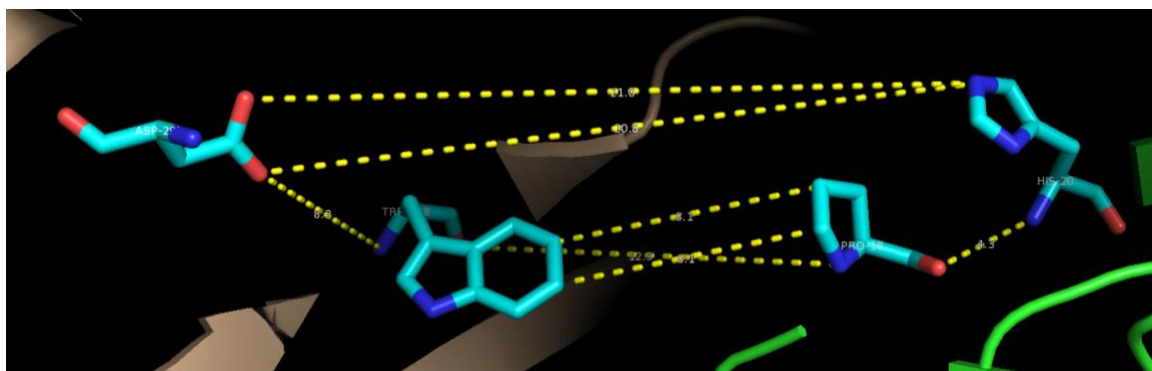


Figure 18. Conserved Residues in the gpE+gpD Trimer Spike. gpE (tan) p-domain containing W306 makes contacts with H19 on gpD (green). The W308A mutation abrogates the gpE-gpD association to prevent fully formed capsid structures.

Attempts of crystallizing a subunit of the capsid led to the hypothesis that crystal formation may only occur in the presence of Nu3; where in the absence of Nu3, E does not crystallize as it tends to form precipitate (data not shown). This observation does not imply that the E monomer is unable to crystallize, a suitable condition has yet to be identified that would allow this to occur. Since ENu3 proteins tend to crystallize as a complex when co-expressed, it was preferred to use purified ENu3 as the goal intended to examine the association between capsid and scaffold.

In phase one, the expressed ENu3 was purified by anion-exchange using a Mono-Q column which is a strong anion exchanger that resulted in a loss of the scaffold. Due to Nu3 having high acidic content and a low isoelectric point of 4.3, this may have caused the 18 kDa protein to bind tightly to the column and potentially dissociate from E, causing the proteins to elute in different fraction volumes. Further purified ENu3 by SEC, the SDS-PAGE verifies that most of Nu3 was lost in the purification, while E remains present (Figure 9B). The broad peak indicates that the corresponding fractions contain heterogeneous species that have protein forms with different stoichiometric ratio complexes that fall within the protein standard size marker of 44 kDa and 17 kDa. Due to the loss of gpNu3, the protein sample was of no use for crystallizing as E would not crystallize in the absence of Nu3.

The ENu3 complex associates in-vivo to assemble into the capsid and during final maturation steps Nu3 becomes cleaved and degraded to exit the structure mainly by gpC protease, although dissociation of Nu3 in the absence of gpC can still occur. While gpC may not be essential for Nu3 dissociation, a factor not yet observed may trigger Nu3 dissociation from the complex; thus, it was intended to purify the proteins in the associated state as Nu3 has the ability to associate-dissociate at any moment. A new growth and expression followed up phase one, where parameters such as pH and column were evaluated to ensure the complex would remain associated. The phase two anion-exchange purification showed an improvement from phase one that successfully isolated ENu3 (Figure 10A). Using a moderately weaker DEAE anion-exchange column allowed for Nu3 to elute with E in the same fraction as both monomers are present (Figure 10B). Further purification was followed by SEC to analyze the size of the complex and to remove excess proteins that would interfere with the crystallization.

Figure 11 shows the chromatogram elution of ENu3 purified by SEC that revealed different assembly states. Majority of pure ENu3 eluted in fractions 19-21 measured at 280 nm denoted by peak 2, confirmed the monomer sizes were present by SDS-PAGE (Figure 11B). The various peaks on the chromatogram suggest that E is shifting into different oligomeric states with different ratios as the protein elutes in earlier volume fractions. The fraction volume collected corresponding to peak 2 contain protein with an estimated size of 318 kDa, which is roughly between the size markers of 670 kDa and 158 kDa.

The estimated size ratio of the complex is an approximation as seen Table 2. The total size of the purified ENu3 complex can be estimated from protein standards separated by SEC with sizes that range from 318-172 kDa (Table 2). Fraction 20 elutes a protein with an estimated size of 234 kDa that correlates with the densitometry analysis of the phase 2 SEC SDS-PAGE (Figure 11B) with an approximate ratio of 3:1 E:Nu3. This fraction may contain 6 gpE to 2 gpNu3 with a calculated estimated weight of 264 kDa. It can be predicted that this fraction potentially elutes ENu3 in the hexamer conformation which is hypothesized to form a stable complex unit. The capsomer contains

six bound gpE monomers that are adjoined together by protein-protein contacts through the p-loop and e-loop stabilized by hydrophobic interactions. Based on homology modeling and cryo-EM structures of the gpE monomer, the spine helix that contributes to the p-loop and e-loop flexibility might serve a role in the polymerization reaction to higher-order structures mediated by gpNu3. The flexible attributions of gpE may be reduced when in the form of a capsomer as each of the bound subunits are held together by the hydrophobic forces invoked by p-loop and e-loop domains of each monomer unit. The ratios of ENu3 vary within these fractions which also suggests that capsid proteins are assembling into higher-order subunits assisted by gpNu3, although the mechanisms of gpNu3 involvement with gpE and subsequent steps that part in the polymerization reaction remain undefined.

Table 2. Phase 2 SEC Complex Size Estimation Based on SEC and Densitometry Analysis

SEC Total Size Standard (kDa)	Predicted Complex Size (kDa)	E	Nu3	Densitometry ENu3 Ratio
Fraction 19 (318)	340	8	2	4.1
Fraction 20 (234)	264	6	2	2.8
Fraction 21 (172)	168	3	3	1.1

The fractions pertaining to peak 2 from SEC (phase 2) that contained the ENu3 complex was optimized in previously known conditions. Most drops from the optimization screens produced two major outcomes which were the formation of needle clusters (Figure 7A and 12A) or no crystal growth at all (Figure 7B and Figure 12B). ENu3 crystals consistently formed disorderly undefined needle clusters, although this sample produced a morphology that slightly differed than before. These needle cluster asteroids appeared to form a capillary-like morphology with thin needles that differed than smaller needle clusters as seen in Figure 7. The morphology produced may be due to protein contaminants that remained present in the SEC purification (Figure 11B), which would interfere with the crystallization. Although these fractions contained purified ENu3 complex with bands present at their indicated size (Figure 11B), the presence of contaminants at larger m/w sizes indicate fractions

19-21 contain a population of various protein species that produced crystals not suitable for XRD analysis.

Needle crystals give potential to provide resolution if the lattice packed is organized and ordered in an array free of micro-heterogeneity, although this observed needle morphology with slight precipitation suggests the ENu3 protein sample contained various protein species other than the ENu3 complex which affected lattice arrangement. It is suspected that these crystals may fail to provide strong diffraction as micro-heterogeneity is present in the crystal drops (Figure 12). The crystal morphology seen from this phase is uncommon like previous ENu3 crystallizations, thus the possibility of disordered lattice organization is affected by the heterogenous protein sample (Krauss et al., 2013; McPherson, 2004; Pike et al., 2016).

New growth and expression of ENu3 was carried out in order to produce high amounts of ENu3 protein and to obtain a pure sample free of contaminants unlike the phase 2 SEC purification which had bands present at larger m/w indicating heterogeneity in this sample (Figure 11B). Multiple growths using the previous frozen glycerol stock of BL21-DE3 pET7CAP[E_{W308A}] cell line failed to reproduce the results that would yield a high purity of ENu3. Expression of ENu3 was low and many times would fail to express Nu3, like phase one (Figure 9), where the protein would be absent or degraded seen on SDS-PAGE. It was suspected that a weakened cell line that failed to produce the plasmid containing capsid genes in high amounts which would affect downstream expression as protein is reduced. The previous ENu3 growth experiments used a frozen glycerol stock and it was deduced that this stock underwent many freeze-thawing cycles that may have reduced the *E. coli* cells competency resulting in an unhealthy cell line affecting cellular basal-function. Altering cell function such as production of enzymes needed for the cells replicate will downstream affect the amount of plasmid availability. The gene or the plasmid may have been lost or removed from the cells by the freeze-thawing events and thus a new transformation was applied using competent DH5 α cells to ensure the plasmid was efficiently produced in high amounts and following a fresh transformation using healthy competent DE3 cells for expression of the ENu3 complex. This new transformation was

successful as this expression purified by anion-exchange and SEC removed majority contaminants (Figure 13 and Figure 14).

The protein was purified by anion-exchange chromatography and the eluted fractions were measured at 280 nm to provide the similar peak symmetry seen in previous anion-exchange purifications. The fractions containing the most purified ENu3 complex had eluted in fractions 32-36 that corresponds with the first peak on the chromatogram (Figure 13A). The first peak fractions were isolated and contained ENu3 and less contaminants than the fractions corresponding to the second peak. This peak also appears to be more prominent, and sharper compared to peak 1 anion-exchange chromatography elution from phase 2 in Figure 10A, which indicates a particular complex is present in higher amounts than before, verified by SDS-PAGE lanes 32-36 (Figure 13B). This observation can be based upon negatively charged gpNu3 and its ability to associate with gpE, where negative residues within the gpNu3 sequence bind to positive residues of gpE as electrostatic interaction of these salt bridges reduce the overall net negative charge of the protein complex, therefore eluting in earlier fractions. While the mechanism of gpNu3 binding and the roles it partakes in the polymerization reaction are not evident, gpNu3 association to gpE may be dictated by side-chains which function can be altered by change in pH. Phase 3 purification used Tris pH 8.0 while phases 1-2 used Tris pH 7.4, a β -strand containing a Histidine residue positioned at the C-terminus. The lower pH buffer may have kept protein side chains protonated while higher pH buffer caused side chains to deprotonate which would mainly affect His82 on gpE. Histidine has a side chain pK_a of 6.0 and a pI of 7.6 which this observation led to the hypothesis that Tris pH 8.0 would induce the zwitterionic form which would prevent any formation of salt bridges.

The gpE amino acid sequence contains only one histidine residue where one can assume the difference in pH would be negligible, although looking closer at the H82 position is oriented towards the N-terminal region of the spine helix which can imply that this single histidine residue might serve a multi-functional role dependent on its protonated or zwitterionic state (Figure 19). Higher pH of 8.0 deprotonates the H82 imidazole side-chain to become neutral reducing electrostatic interactions

formed in the protonated positively charged state with negatively charged residues in nearby positions. Acidic residues within the spine helix D118 and E119 are approximated to form distances of 4-6 Å with H83 to form putative electrostatic interactions in low pH conditions. The putative interaction that H82 forms with the spine helix may be linked to the assembly process and gpNu3 binding, where in acidic conditions, the positive H82 side-chain might form salt bridges with negative Asp and Glu positioned in the spine helix altering the conformation that may limit the association gpNu3 forms with gpE. Conversely, in higher pH H82 is deprotonated and remains in a neutral state that reduces any potential interaction it may form with the negative charged residues in the spine helix. As H82 cannot form charge-charge interactions, the β -strand becomes destabilized and free to move away from the spine helix that results in an opening between the folds of surrounding motifs. Thus, the intramolecular forces potentially formed by H82 and the spine helix influences the ability for gpNu3 to freely bind and associate with gpE within this pocket juxtaposed near helix-2/helix-3. The full-length gpE sequence contains only a single histidine residue which may be a critical factor that invokes tertiary structure interaction between the β -strand and spine helix in low pH that results in an open or closed conformation. The changes in pH buffer affects H82 side-chain which proposes a role in forming potential electrostatic interactions with the spine-helix which in turn alters Nu3 binding affinity to gpE.

To further analyze this proposed hypothesis, when comparing phase 2 (Figure 11B) and phase 3 (Figure 14B) SEC purified samples observed by SDS-PAGE, Nu3 appears to be present in higher amounts in phase 3 SEC purification which is confirmed by densitometry analysis that shows the 18 kDa band was measured to have greater integrated density in phase 3 gel than it is phase 2 gel (data not shown). The Nu3 band in the phase 3 gel (lanes 13-15) also appears to have a darker intensity present over phase 2 (lanes 19-21), suggesting that more Nu3 is present in these fractions which aligns with the hypothesis of the H82 role mediating potential binding of Nu3 in higher pH buffers, where it is seen that pH 8.0 altered the H82 side-chain to become neutral that limits potential interaction with the spine-helix allowing Nu3 to bind freely to gpE. These interaction between capsid

and scaffold are critical as Nu3 promotes shuttling of monomer gpE units to form higher-order domains such as capsomers.

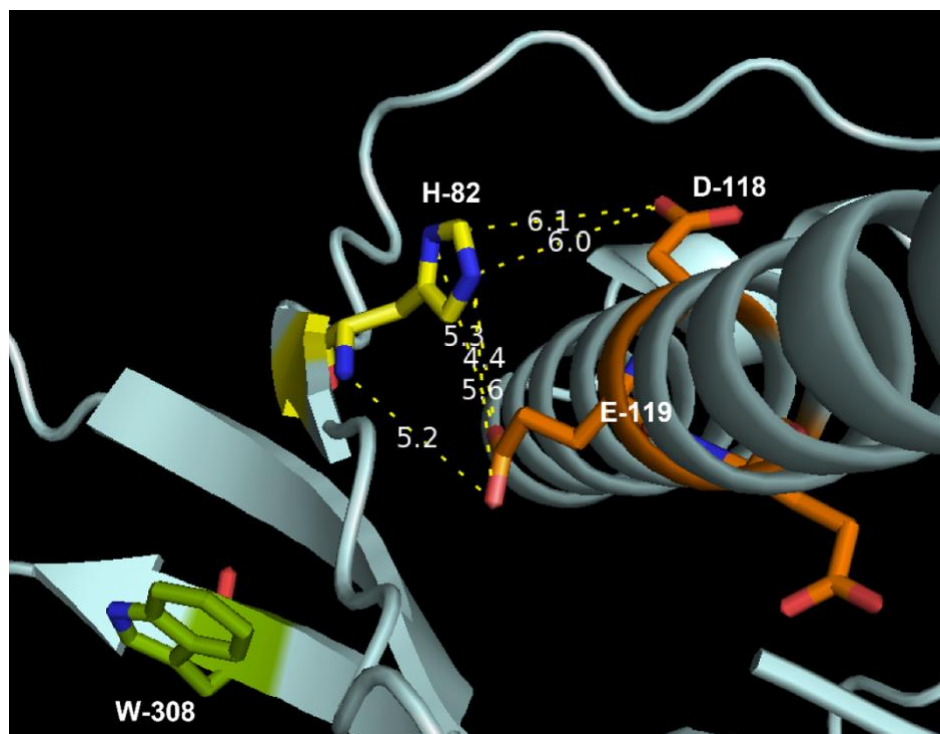


Figure 19. Histidine 82 positioned in the β -sheet N-terminal to helix 2 and 3 of the spine α -helix domain lies. The spine helix contains D118 and E119 in orange. W308 positioned in adjacent p-domain. Interatomic distances are indicated.

The crystals produced in phase 2 appeared to be poorly formed with a size not sufficient for XRD, which may be a result of lower pH used during purification which potentially affected interatomic forces within gpE. Reducing Nu3 binding to E might result in disordered formation of the capsid. Mutational studies of capsid genes show the knockout of the Nu3 gene affects gpE assembly into the mature capsid, although in Nu3's absence gpE can still form higher-order structures but cryo-EM shows these structures are aberrantly formed and fail to package DNA properly, thus suggesting the role of Nu3 is essential in capsid morphogenesis (Murialdo & Siminovitch 1972; Ray & Murialdo, 1975; Lambert et al., 2017). Limiting Nu3 interactions with E can be proposed to affect assembly as reduced binding of the scaffold leads to poorly formed capsid structures. These findings can correlate with the calculated densitometry ratios of the ENu3 complex (Table 2), implying that phase 2 SEC fractions 19-21 contained heterogenous protein complexes that affected the crystallization as protein molecules disorderly aggregated to form thin needle shards which led to undesirable precipitation.

A fresh transformation of the plasmid into competent *E. coli* expressing cells was essential to over-express ENu3 proteins and SEC purified showed to be more efficient than phase 2 with an approximate degree of purity of 100%, indicated by the dark bands are present in fractions 13-16 (Figure 14B). The chromatogram shows the fractions collected were measured at 280 nm and the second peak with an absorbance of 0.50 is suspected to contain a homogenous protein complex as this peak is sharply defined. This sharp peak was more resolved than previous SEC purifications and fractions 13-15 corresponding to the second peak were analyzed by SDS-PAGE (Figure 14B) and contained pure amounts of ENu3 showing a band at 38 kDa (E) and 18 kDa (Nu3). The fractions that corresponded to the second peak size were analyzed and compared to the protein standards which has a size between 670 kDa and 158 kDa (Figure 14A). The size of the protein complex collected in these fractions were calculated by taking the elution volume (56 ml) at the sharp peak and comparing it with the protein standard elution volume (Figure 8), where the complex size is approximately 531 kDa that falls within the size range seen on the chromatogram. This estimated size can consist of various ENu3 stoichiometric ratios. The complex size that would be most significant that elutes at this volume can be that of a capsomer, correlating with densitometry analysis of the phase 3 SEC SDS-PAGE (Figure 14B and Table 3).

Table 3. Phase 3 SEC Complex Size Estimation Based on SEC and Densitometry Analysis

SEC Total Size Standard (kDa)	Predicted Complex Size (kDa)	E	Nu3	Densitometry Ratio
Fraction 13 (531)	528	12	4	3.3
Fraction 14 (352)	376	8	4	2.0
Fraction 15 (234)	282	6	3	2.0

The size of the protein complex eluted in fractions 13-16 can be approximated to have a size complex of 531-234 kDa based off protein standards eluted by SEC. The ratios calculated in Table 3 suggest that these fractions contain sizes of aligning with the size of a hexamer. The densitometry ratio of fraction 13 and fraction 15 suggest the size of the quaternary structure may be a 6-fold domain, thus a larger size complex is expected to elute in earlier fractions. As the hexameric unit is

predicted to form a stable interaction with each other, this stable complex in turn may impact the crystallization. The hypothesis of the capsomer being a stable unit is based upon the limiting flexibilities of p-loop and e-loop domains of gpE, where the capsomer is bound together by the protein-protein contacts through the p-loop and e-loop domains which are stabilized by hydrophobic interactions.

The sharp peak can suggest that a single homogenous protein complex is present which can represent an ENu3 stable association, verified by the SDS-PAGE which these fractions contain highly pure ENu3. The sharp peak on the chromatogram can indicate majority of the sample may contain a homogenous complex which is desirable for crystallography. The gpE monomers connect via the e-loop and p-loop domains to form the 6-fold domain stabilized by hydrophobic forces. The size of the quaternary structure of the purified SEC protein in fractions 13-15 aligns with the size of a hexamer which appears to favor crystal nucleation and growth. The flexibility attributed by the domains of the gpE monomer fold is reduced in the capsomer form, where the binding of adjacent gpE units through p-loop and e-loop domains may lock in particular stable conformation; thus, capsomer units in this protein sample is predicted to be a stable higher-order structure that can influence the packing into an organized lattice arrangement to produce a symmetrical crystal structure.

In all previous ENu3 crystallization experiments, the production of needle clusters was commonly observed while longer incubation growth led to precipitation in the drops. Attaining highly purified ENu3 that can produce crystals with a larger size and symmetrical element was the objective as more atoms and electrons within the packed crystal lattice may result in stronger diffraction intensities which can be integrated to solve the structure (Krauss et al., 2013; McPherson, 2004; Pike et al., 2016).

The purified protein complex by phase 3 SEC was screened to identify new conditions that yield hits. Interestingly, the previously identified crystallizing conditions (Figure 7A) were used to further optimize the protein complex. The phase 3 optimization produced familiar needle crystals except these formed in approximately 2-3 days and were larger in size and more abundant than

before (Figure 15). The morphology of the crystal to form in a three-dimension symmetrical arrangement was aimed for improvement as this may provide unique diffraction intensities due to the particular arrangement of electrons present within the crystal lattice which may be needed to help solve the structure. Thus, formation of large isomorphous crystals free of micro-heterogeneity may increase the probability of strong diffraction due to its shape.

The third approach of the research aim used the method of crystal seeding which was a type of optimization to specifically improve reoccurring crystal forms that could not be optimized further. The crystals formed in Figure 15 would be used as a seed crystal to create a seed stock that would be applied to crystal drops set over iterative seeding rounds, where round one would produce the first crystal seeding product (Figure 16) and round two of the second seeding product (Figure 17).

The first round of seeding formed a single crystal with rare cubic symmetry which differed from any other morphology ENu3 ever formed (Figure 16). The goal of the crystal seed is to have a role in acting as a scaffold to promote crystal growth, where a fresh protein sample added to the drop regains equilibrium in the vapor diffusing environment to reach saturation, the seed crystal presents a nucleation site for fresh protein to grow in the metastable zone, bypassing the nucleation phase as seen in crystal drops without seed crystal. Subsequent rounds of seeding are needed to take place as conditions have yet to be identified that yield the best crystals, but also as some drop conditions may induce a supersaturated environment where further heterogenous nucleation occurs and eventually forms precipitate. More importantly, rounds of optimization took place to find a condition that would maintain the seed crystal in the metastable state where growth of new protein can occur, where identifying an optimal ratios of seed crystal to protein sample and crystallization reagent was applied to induce saturation in the drop to steadily form a crystal that bypasses *de-novo* nucleation. This first seeding round crystallized a 5 mg/ml protein sample in a 1:50 and 1:100 seed stock but could not further grow or produce excess crystals. This round of crystals was used to create another seed stock which produced the second seeding round crystals to the maximum size and morphology (Figure 17). Larger rectangular plates had formed while keeping the signature needle form as ENu3 always

produced. Round 2 crystals were used to create the second seed stock which was hypothesized that the crystal array formed produced interatomic forces that were much stronger and packed tightly where if broken up, nano size crystals retained a specific lattice that introduces stable nuclei for fresh protein sample to grow and arrange in an orderly array. This is observed in the second seeding round (Figure 17) and the larger crystals would be used for XRD analysis.

CHAPTER 5

CONCLUSION

The outcome of studies done by Lambert et al. (2017) discovered that the gpEW308A mutation afforded soluble gpE monomer subunits which became an advantage to analyze the capsid structure alone rather than in the assembled form. The stepwise process into forming these intermediate structures that take part in the polymerization reaction remain undefined, although these studies can imply that the capsomer is a prominent feature of the capsid and this structure may be a stable formed intermediate that has the ability to crystallize, unlike other various subunit stoichiometric ratios formed. The assembly reaction begins spontaneously to immediately form the mature capsid shell making it difficult to isolate a single stable E_{Nu3} protein complex. Purification by SEC yields different E_{Nu3} ratios implying heterogeneity of various assembled states that may impact the crystallization.

Crystallography is a meticulous process that requires a trial-and-error approach to identify reagent conditions that promotes nucleation needed for protein molecules to grow into an ordered crystal arrangement. Crystal formation is dependent on many factors such as expression of the protein in a specific conformation but also heavily dependent on the screening-optimization phases.

Using crystallography as a tool for structural analysis of lambda capsid and scaffolding proteins can provide structural details that will elucidate the roles of capsid and scaffolding protein interactions and their involvement in assembly and DNA packaging. To better understand the stepwise mechanisms and interactions between these proteins, a crystal structure of E_{Nu3} must be produced that gives potential to solve the structure that may answer the questions of what vital roles these proteins take part in capsid assembly and morphogenesis. The structural solutions of these proteins can be used in-tandem with other similar known viral structures similar to lambda, such as eukaryotic adenovirus, which will be a useful model for drug design to target viral infections. Additionally, the structure of lambda capsid proteins can be used as a subject for development and engineering of phage display systems for gene therapeutics.

Microscopic virus particles are composed of nucleic acid and protein which form an infectious parasite that affects all domains of life and has shaped the biosphere throughout history. Their pathogenic and virulent nature imposes a threat on cellular life as these machines require cellular organisms as a host to benefit their survival and production of progeny. Bacteriophages have provided an understanding of viral systems and viral selfish genetic elements that dictate their infectious nature.

GpE eluting in various volume fractions indicated that the protein complex will self-assemble into higher-order complex forms over time. The size of the quaternary structure that had the ability to crystallize was determined by SEC and the purity was confirmed by SDS-PAGE. A fresh transformation of the plasmid containing ENu3 into competent cells allowed for efficient expression that was able to be purified completely which removed 100% of contaminants. The size of the ENu3 quaternary structure that crystallized was predicted to be 531-234 kDa which correlates with densitometry ratios of ENu3 suggesting the capsomer structure form eluted in these fractions (Table 3).

In summary, the work described in this thesis report has contributed to the advancement of the crystallization of lambda capsid proteins as novel crystals with improved morphology were produced. The ENu3 crystals were sent to Berkley ALS facility and XRD data was collected which will be used for processing to solve the ENu3 crystal structure.

REFERENCES

- Andrews, B. T., & Catalano, C. E. (2013). Strong subunit coordination drives a powerful viral DNA packaging motor. *Proceedings of the National Academy of Sciences of the United States of America*, 110(15), 5909–5914.
- Ankrah, N.Y., May, A.L., Middleton, J.L., Jones, D.R., Hadden, M.K., Gooding, J.R., LeCleir, G.R., Wilhelm, S.W., Campagna, S.R., Buchan, A. (2014). Phage infection of an environmentally relevant marine bacterium alters host metabolism and lysate composition. *The ISME Journal*, 8, 1089–1100
- Asadi-Ghalehni, M., Ghaemmaghami, M., Klimka, A., Javanmardi, M., Navari, M., & Rasaei, M. J. (2015). Cancer immunotherapy by a recombinant phage vaccine displaying EGFR mimotope: an in vivo study. *Immunopharmacology Immunotoxicology*, 37(3), 274-279.
- Arber, W., (1971). Host-controlled variation. In Hershey, A. D. (Ed.), *The bacteriophage lambda* (pp. 83-96). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Bain, D. L., Berton, N., Ortega, M., Baran, J., Yang, Q., & Catalano, C. E. (2001). Biophysical characterization of the DNA binding domain of gpNu1, a viral DNA packaging protein. *Journal of Biological Chemistry*, 276(23), 20175-20181.
- de Beer, T., Fang, J., Ortega, M., Yang, Q., Maes, L., Duffy, C., Berton, N., Sippy, J., Overduin, M., Feiss, M., & Catalano, C. E. (2002). Insights into specific DNA recognition during the assembly of a viral genome packaging machine. *Molecular Cell*, 9(5), 981–991.
- Baltimore, D. (1971). Expression of animal virus genomes. *Bacteriological Reviews*, 35(3), 235-241.
- Batinovic, S., Wassef, F., Knowler, S. A., Rice, D. T. F., Stanton, C. R., Rose, J., Tucci, J., Nittami, T., Vinh, A., Drummond, G. R., Sobey, C. G., Chan, H. T., Seviour, R. J., Petrovski, S., & Franks, A. E. (2019). Bacteriophages in natural and artificial environments. *Pathogens*, 8(3), 100.
- Bawden, F. C., & Pirie, N. W. (1938). Crystalline preparations of tomato bushy stunt virus. *British Journal of Experimental Pathology*, 19(4), 251-263.
- Beghetto, E., & Gargano, N. (2011). Lambda-display: a powerful tool for antigen discovery. *Molecules*, 16(4), 3089-3105.
- Bernal, J. D., & Fankuchen, I. (1941). X-ray and crystallographic studies of plant virus preparations: i. Introduction and preparation of specimens ii. Modes of aggregation of the virus particles. *Journal of General Physiology*, 25(1), 111-146
- Bertani, G., & Weigle, J. J. (1953). Host controlled variation in bacterial viruses. *Journal of Bacteriology*, 65(2), 113-121.
- Bezdek M, Amati P (1967) Properties of P22 and a related Salmonella typhimurium phage. I. General features and host specificity. *Virology* 31:272–278
- Blasche, S., Wuchty, S., Rajagopala, S. V., & Uetz, P. (2013). The protein interaction network of bacteriophage lambda with its host, Escherichia coli. *Journal of Virology*, 87(23), 12745-12755.

- Cameron, J.R., Panasencko, S.M., Lehman, I.R., Davis, R.W. (1975). In vitro construction of bacteriophage lambda carrying segments of the *Escherichia coli* chromosome: selection of hybrids containing the gene for DNA ligase. *Proceedings of the National Academy of Sciences of the United States of America* 72, 3416–3420.
- Campbell, A. M. (1986). Bacteriophage lambda as a model system. *BioEssays: News and Reviews in Molecular, Cellular and Developmental Biology*, 5(6), 277-280.
- Casjens, S. R. (1985). Bacteriophage L: Chromosome physical map and structural proteins. *Virology*, 147, 431-440.
- Casjens, S. R. (2005). Comparative genomics and evolution of the tailed bacteriophages. *Current Opinion in Microbiology*, 8(4): 451-458.
- Casjens, S. R. (2011). The DNA-packaging nanomotor of tailed bacteriophages. *Nature Reviews Microbiology*, 9(9), 647-657.
- Casjens, S., & Hendrix, R. (1974). Comments on the arrangement of the morphogenetic genes of bacteriophage lambda. *Journal of Molecular Biology*, 90(1), 20-23.
- Casjens, S. R., & Hendrix, R. W. (2015). Bacteriophage lambda: Early pioneer and still relevant. *Virology*, 479-480, 310-330.
- Casjens, S., Hohn, T., & Kaiser, A. D. (1970). Morphological proteins of phage lambda: Identification of the major head protein as the product of gene E. *Virology*, 42(2), 496–507.
- Caspar, D. L., & Klug, A. (1962). Physical principles in the construction of regular viruses. *Cold Spring Harbor Symposia on Quantitative Biology*, 27, 1-24.
- Catalano, C. E. (2000). The terminase enzyme from bacteriophage lambda: A DNA-packaging machine. *Cellular and Molecular Life Sciences (CMLS)*, 57(1), 128-148.
- Chang, J. R., Song, E. H., Nakatani-Webster, E., Monkkonen, L., Ratner, D. M., & Catalano, C. E. (2014). Phage lambda capsids as tunable display nanoparticles. *Biomacromolecules*, 15(12), 4410-4419.
- Crick, F. H., & Watson, J. D. (1956). Structure of small viruses. *Nature*, 177(4506), 473–475.
- D'Arcy, A., Bergfors, T., Cowan-Jacob, S. W., & Marsh, M. (2014). Microseed matrix screening for optimization in protein crystallization: what have we learned? *Acta Crystallographica. Section F, Structural Biology Communications*, 70(Pt 9), 1117-1126.
- Derewenda, Z. S. (2004). The use of recombinant methods and molecular engineering in protein crystallization. *Methods*, 34(3), 354-363.
- Dokland, T. (1999). Scaffolding proteins and their role in viral assembly. *Cellular and Molecular Life Sciences (CMLS)*, 56, 580-603.
- Dokland, T., & Murialdo, H. (1993). Structural transitions during maturation of bacteriophage lambda capsids. *Journal of Molecular Biology*, 233(4), 682-694.
- Echols, H., & Murialdo, H. (1978). Genetic Map of Bacteriophage Lambda. *Microbiological Reviews*, 42(3), 577-591.

- Effantin, G., Figueroa-Bossi, N., Schoehn, G., Bossi, L., & Conway, J. F. (2010). The tripartite capsid gene of Salmonella phage Gifsy-2 yields a capsid assembly pathway engaging features from HK97 and lambda. *Virology*, 402(2), 355-365.
- Emsley, P., & Cowtan, K. (2004). Coot: Model-building tools for molecular graphics. *Acta Crystallographica. Section D, Biological Crystallography*, 60(Pt 12 Pt 1), 2126-2132.
- Evans, P. R., & Murshudov, G. N. (2013). How good are my data and what is the resolution? *Acta Crystallographica. Section D, Biological Crystallography*, 69(Pt 7), 1204-1214.
- Fane, B. A., & Prevelige, P. E., Jr. (2003). Mechanism of scaffolding-assisted viral assembly. *Advances in Protein Chemistry*, 64, 259-299.
- Feiss, M., & Catalano, C. E. (2005). *Bacteriophage lambda terminase and the mechanism of viral DNA packaging*. In C. Honour (Ed.), *Viral genome packaging machines: Genetics, structure, and mechanism. Molecular biology intelligence unit* (pp. 5-39). Boston, MA: Springer.
- Feiss, M., & Rao, V. B. (2012). The bacteriophage DNA packaging machine. *Advances in Experimental Medicine and Biology*, 726, 489-509.
- Flanigan, A., & Gardner, J. (2008). Structural prediction and mutational analysis of the Gifsy-I Xis protein. *BioMed Central (BMC) Microbiology*, 8, 199.
- Fokine, A., Leiman, P. G., Shneider, M. M., Ahvazi, B., Boeshans, K. M., Steven, A. C., Black, L. W., Mesyanzhinov, V. V., & Rossmann, M. G. (2005). Structural and functional similarities between the capsid proteins of bacteriophages T4 and HK97 point to a common ancestry. *Proceedings of the National Academy of Sciences of the United States of America*, 102(20), 7163-7168.
- Forterre, P. (2005). The two ages of the RNA world, and the transition to the DNA world: A story of viruses and cells. *Biochimie*, 87(9-10), 793-803.
- Forterre, P. (2006). The origin of viruses and their possible roles in major evolutionary transitions. *Virus Research*, 117(1), 5-16.
- Franklin, R. (1955) Structure of tobacco mosaic virus. *Nature* 175, 379–381
- Gaussier, H., Ortega, M. E., Maluf, N. K., & Catalano, C. E. (2005). Nucleotides Regulate the conformational state of the small terminase subunit from bacteriophage lambda: Implications for the assembly of a viral genome-packaging motor. *Biochemistry*, 44 (28), 9645-9656
- Gaussier, H., Yang, Q., & Catalano, C. E. (2006). Building a virus from scratch: Assembly of an infectious virus using purified components in a rigorously defined biochemical assay system. *Journal of Molecular Biology*, 357(4), 1154-1166.
- Gertsman, I., Gan, L., Guttman, M., Lee, K., Speir, J. A., Duda, R. L., Hendrix, R. W., Komives, E. A., & Johnson, J. E. (2009). An unexpected twist in viral capsid maturation. *Nature*, 458(7238), 646-650.
- Grose, J. H., & Casjens, S. R. (2014). Understanding the enormous diversity of bacteriophages: the tailed phages that infect the bacterial family Enterobacteriaceae. *Virology*, 468-470, 421–443.
- Halford, S. E., Johnson, N. P., & Grinsted, J. (1980). The EcoRI restriction endonuclease with bacteriophage lambda DNA. Kinetic studies. *The Biochemical Journal*, 191(2), 581–592.

- Häuser, R., Blasche, S., Dokland, T., Haggård-Ljungquist, E., von Brunn, A., Salas, M., Casjens, S., Molineux, I., & Uetz, P. (2012). Bacteriophage protein-protein interactions. *Advances in Virus Research*, 83, 219–298.
- Hendrix, R. W. (2002). Bacteriophage λ and its relatives. In U. N. Streips & R. E. Yasbin (Eds.), *Modern Microbial Genetics* (pp. 127-143). New York, NY: Wiley-Liss.
- Hendrix, R. W. (2002). Bacteriophages: Evolution of the majority. *Theoretical Population Biology*, 61(4), 471–480.
- Hendrix, R. W., & Casjens, S. R. (1974). Protein fusion: A novel reaction in bacteriophage lambda head assembly. *Proceedings of the National Academy of the Sciences of the United States of America*, 71(4), 1451-1455.
- Hershey, A. D. (1953). Inheritance in bacteriophage. *Advances in Genetics*, 5, 960–962.
- Hohn, B., & Hohn, T. (1974). Activity of empty, headlike particles for packaging of DNA of bacteriophage lambda in vitro. *Proceedings of the National Academy of Sciences*, 71(6), 2372-2376.
- Hohn, T., & Katsura, I. (1977). Structure and assembly of bacteriophage lambda. *Current Topics in Microbiology and Immunology*, 78, 69–110.
- Error! Hyperlink reference not valid.** Hohn, B., Wurtz, M., Klein, B., Lustig, A., & Hohn, T. (1974). Phage lambda DNA packaging, in vitro. *Journal of Supramolecular Structure*, 2(2-4), 302-317.
- Hohn, B., & Murray, K., (1977). Packaging recombinant DNA molecules into bacteriophage particles in vitro. *Proceedings of the National Academy of Sciences of the United States of America*, 74(8), 3259–326
- Huang, R. K., Steinmetz, N. F., Fu, C. Y., Manchester, M., & Johnson, J. E. (2011). Transferrin-mediated targeting of bacteriophage HK97 nanoparticles into tumor cells. *Nanomedicine*, 6(1), 55–68.
- Iwai, H., Forrer, P., Pluckthun, A., & Guntert, P. (2005). NMR solution structure of the monomeric form of the bacteriophage lambda capsid stabilizing protein gpD. *Journal of Biomolecular NMR*, 31(4), 351-356.
- Jacob, F., & Campbell, A. (1959). Sur le Systeme de Repression Assurant l'immunité Chez les Bacteries Lysogenes. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences*. 248, 3219–3221.
- Jacob, F., Sussman, R., & Monod, J. (1962). On the nature of the repressor ensuring the immunity of lysogenic bacteria. *Comptes Rendus Hebdomadaires des Seances de l'Academie des Sciences*. 254, 4214–4216
- Jacob, F., & Wollman, E. (1953). Induction of phage development in lysogenic bacteria. *Cold Spring Harbor Symposia on Quantitative Biology*. 18, 101–121.
- Jepson, C. D., & March, J. B. (2004). Bacteriophage lambda is a highly stable DNA vaccine delivery vehicle. *Vaccine*, 22(19), 2413-2419.
- Johnson, J. E. (2010). Virus particle maturation: Insights into elegantly programmed nanomachines. *Current Opinion in Structural Biology*, 20(2), 210-216.

- Johnson, J. E., & Wikoff, W. R. (1998). Macromolecular assembly: Chainmail stabilization of a viral capsid. *Current Biology: CB*, 8(25), R914–R917.
- Kabsch, W. (2010). Integration, scaling, space-group assignment and post-refinement. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 2), 133–144.
- Kahmann, R., & Prell, H. (1971). Complementation between P22 amber mutants and phage L. *Molecular & General Genetics*, 113(4), 363–366.
- Kaiser, A. D. (1957). Mutations in a temperate bacteriophage affecting its ability to lysogenize *Escherichia coli*. *Virology* 3, 42–49.
- Kaiser, A.D., & Jacob, F. (1957). Recombination between related bacteriophages and the genetic control of immunity and prophage localization. *Virology* 4, 509–517.
- Kanekiyo, M., Bu, W., Joyce, M. G., Meng, G., Whittle, J. R., Baxa, U., Yamamoto, T., Narpala, S., Todd, J. P., Rao, S. S., McDermott, A. B., Koup, R. A., Rossmann, M. G., Mascola, J. R., Graham, B. S., Cohen, J. I., & Nabel, G. J. (2015). Rational Design of an Epstein-Barr Virus Vaccine Targeting the Receptor-Binding Site. *Cell*, 162(5), 1090–1100.
- Kobayashi, I. K. a. H. (1990). Structure and inherent properties of the bacteriophage lambda head shell. *Journal of Molecular Biology*, 213, 503–511.
- Kobe, B., Guncar, G., Buchholz, R., Huber, T., Maco, B., Cowieson, N., Martin, J. L., Marfori, M., & Forwood, J. K. (2008). Crystallography and protein-protein interactions: biological interfaces and crystal contacts. *Biochemical Society transactions*, 36(Pt 6), 1438–1441.
- Koonin, E. V., Dolja, V. V., & Krupovic, M. (2015). Origins and evolution of viruses of eukaryotes: The ultimate modularity. *Virology*, 479–480, 2–25.
- Koonin, E. V., Krupovic, M., & Agol, V. I. (2021). The Baltimore classification of viruses 50 years later: How does it stand in the light of virus evolution?. *Microbiology and Molecular Biology Reviews: MMBR*, 85(3), e0005321.
- Koudelka, K. J., Ippoliti, S., Medina, E., Shriver, L. P., Trauger, S. A., Catalano, C. E., & Manchester, M. (2013). Lysine addressability and mammalian cell interactions of bacteriophage lambda procapsids. *Biomacromolecules*, 14(12), 4169–4176.
- Krauss, R. I., Merlino, A., Vergara, A., & Sica, F. (2013). An overview of biological macromolecule crystallization. *International Journal of Molecular Sciences*, 14(6), 11643–11691.
- Krupovic, M., & Koonin, E. V. (2015). Polintons: a hotbed of eukaryotic virus, transposon and plasmid evolution. *Nature Reviews. Microbiology*, 13(2), 105–115.
- Lambert, S., Yang, Q., De Angeles, R., Chang, J. R., Ortega, M., Davis, C., & Catalano, C. E. (2017). Molecular dissection of the forces responsible for viral capsid assembly and stabilization by decoration proteins. *Biochemistry*, 56(5), 767–778.
- Lander, G. C., Evilevitch, A., Jeembaeva, M., Potter, C. S., Carragher, B., & Johnson, J. E. (2008). Bacteriophage lambda stabilization by auxiliary protein gpD: Timing, location, and mechanism of attachment determined by cryo-EM. *Structure*, 16(9), 1399–1406.
- Lederberg, E.M., & Lederberg, J. (1953). Genetic studies of lysogenicity in *Escherichia Coli*. *Genetics*, 38, 51–64.

- Loenen, W. A., & Murray, N. E., (1986). Modification enhancement by the restriction alleviation protein (Ral) of bacteriophage lambda. *Journal of Molecular Biology*, 190, 11–22
- Louten, J. (2016). *Essential Human Virology*. Boston, MA: Academic Press.
- Lwoff, A. (1953). Lysogeny. *Bacteriological Reviews*, 17(4), 269–337
- Maluf, N. K. Gaussier, H., Bogner, E., Feiss, M., & Catalano, C. E. (2006). Assembly of bacteriophage lambda terminase into a viral DNA maturation and packaging machine. *Biochemistry* 45(51): 15259-15268.
- Marintcheva, B. (2018). *Harnessing the power of viruses*. London, England: Academic Press.
- Maruyama, I. N., Maruyama, H. I., & Brenner, S. (1994). Lambda foo: A lambda phage vector for the expression of foreign proteins. *Proceedings of the National Academy of Sciences of the United States of America*, 91(17), 8273–8277.
- McPherson, A. (2004). Introduction to protein crystallization. *Methods*, 34(3), 254-265.
- McPherson, A., & Cudney, B. (2014). Optimization of crystallization conditions for biological macromolecules. *Acta Crystallographica. Section F, Structural Biology Communications*, 70(Pt 11), 1445-1467.
- McPherson, A., & Gavira, J. A. (2014). Introduction to protein crystallization. *Acta Crystallographica Section F, Structural Biology Communications*, 70(Pt 1), 2-20.
- McPherson, A., & Larson, S. B. (2015). A guide to the crystallographic analysis of icosahedral viruses. *Crystallography Reviews*, 21(1-2), 3-56.
- Medina, E., Nakatani, E., Kruse, S., & Catalano, C. E. (2012). Thermodynamic characterization of viral procapsid expansion into a functional capsid shell. *Journal of Molecular Biology*, 418(3-4), 167-180.
- Medina, E., Wieczorek, D., Medina, E. M., Yang, Q., Feiss, M., & Catalano, C. E. (2010). Assembly and maturation of the bacteriophage lambda procapsid: GpC is the viral protease. *Journal of Molecular Biology*, 401(5), 813-830.
- Medina, E. M., Andrews, B. T., Nakatani, E., & Catalano, C. E. (2011). The bacteriophage lambda gpNu3 scaffolding protein is an intrinsically disordered and biologically functional procapsid assembly catalyst. *Journal of Molecular Biology*, 412(4), 723-736.
- Mertz, J. E., & Davis, R. W. (1972). Cleavage of DNA by R1 restriction endonuclease generates cohesive ends. *Proceedings of the National Academy of the Sciences of the United States of America*, 69(11), 3370–3374.
- Moisant, P., Neeman, H., & Zlotnick, A. (2010). Exploring the paths of (virus) assembly. *Biophysical Journal*, 99(5), 1350-1357.
- Murialdo, H. (1979). Early intermediates in bacteriophage lambda prohead assembly. *Virology*, 96(2), 341-367.
- Murialdo, H., & Becker, A. (1977). Assembly of biologically active proheads of bacteriophage lambda in vitro. *Proceedings of the National Academy of the Sciences of the United States of America*, 74(3), 906-910.

- Murialdo, H., & Becker, A. (1978). Head morphogenesis of complex double-stranded deoxyribonucleic acid bacteriophages. *Microbiological Reviews*, 42(3), 529-576.
- Murialdo, H., & Ray, P. N. (1975). Model for arrangement of minor structural proteins in head of bacteriophage lambda. *Nature*, 257(5529), 815-817.
- Murialdo, H., & Siminovitch, L. (1972). The morphogenesis of bacteriophage lambda. IV. Identification of gene products and control of the expression of the morphogenetic information. *Virology*, 48(3), 785-823.
- Nicastro, J., Sheldon, K., & Slavcev, R. A. (2014). Bacteriophage lambda display systems: Developments and applications. *Applied Microbiology and Biotechnology*, 98(7), 2853-2866.
- Ortega, M. E., & Catalano, C. E. (2006). Bacteriophage lambda gpNu1 and Escherichia coli IHF proteins cooperatively bind and bend viral DNA: Implications for the assembly of a aenome-packaging motor. *Biochemistry*, 45, 5180-5189.
- Ortega, M. E., Gaussier, H., & Catalano, C. E. (2007). The DNA maturation domain of gpA, the DNA packaging motor protein of bacteriophage lambda, contains an ATPase site associated with endonuclease activity. *Journal of Molecular Biology*, 373(4), 851-865.
- Owen, S. V., Wenner, N., Canals, R., Makumi, A., Hammarlof, D. L., Gordon, M. A., Aertsen, A., Feasey, N. A., & Hinton, J. C. (2017). Characterization of the prophage repertoire of african salmonella typhimurium st313 reveals high levels of spontaneous induction of novel phage btp1. *Frontiers in Microbiology*, 8, 235.
- Parera, M., Clotet, B., & Martinez, M. A. (2004). Genetic screen for monitoring severe acute respiratory syndrome coronavirus 3C-like protease. *Journal of Virology*, 78(24), 14057-14061.
- Pike, A. C., Garman, E. F., Krojer, T., von Delft, F., & Carpenter, E. P. (2016). An overview of heavy-atom derivatization of protein crystals. *Acta Crystallographica. Section D, Structural Biology*, 72(Pt 3), 303-318.
- Powell, H. R. (2017). X-ray data processing. *Bioscience Reports*, 37(5).
- Qiu, X. (2012). Heat induced capsid disassembly and DNA release of bacteriophage lambda. *Public Library of Science One*, 7(7), e39793.
- Rajagopala, S. V., Casjens, S., Uetz, P. (2011). The protein interaction map of bacteriophage lambda. *BioMed Central (BMC) Microbiology* 11(1): 213.
- Ray, P., & Murialdo, H. (1975). The role of gene Nu3 in bacteriophage lambda head morphogenesis. *Virology*, 64(1), 247-263.
- Saeedi, A., Ghaemi, A., Tabarraei, A., Moradi, A., Gorji, A., Semnani, S., Soleimanjahi, H., Adli, A. H., Hosseini, S. Y., & Vakili, M. A. (2014). Enhanced cell immune responses to hepatitis C virus core by novel heterologous DNA prime/lambda nanoparticles boost in mice. *Virus genes*, 49(1), 11-21.
- Sam, M. D., Cascio, D., Johnson, R. C., & Clubb, R. T. (2004). Crystal structure of the excisionase-DNA complex from bacteriophage lambda. *Journal of Molecular Biology*, 338(2), 229-240.
- Sanger, F., Coulson, A. R., Hong, G. F., Hill, D. F., & Petersen, G. B. (1982). Nucleotide sequence of bacteriophage λ DNA. *Journal of Molecular Biology*, 162(4), 729-773.

- Shao, W., Li, X., Goraya, M. U., Wang, S., & Chen, J. L. (2017). Evolution of influenza A virus by mutation and re-assortment. *International Journal of Molecular Sciences*, 18(8) 1650.
- Sharma, S., Datta, S., Chatterjee, S., Dutta, M., Samanta, J., Vairale, M. G., Gupta, R., Veer, V., & Dwivedi, S. K. (2021). Isolation and characterization of a lytic bacteriophage against *Pseudomonas aeruginosa*. *Scientific Reports*, 11(1), 19393. **Error! Hyperlink reference not valid.**
- Stanley, W. M., & Loring, H. S. (1936). The isolation of crystalline tobacco mosaic virus protein from diseased tomato plants. *Science*, 83(2143), 85.
- Singh, P., Nakatani, E., Goodlett, D. R., & Catalano, C. E. (2013). A pseudo-atomic model for the capsid shell of bacteriophage lambda using chemical cross-linking/mass spectrometry and molecular modeling. *Journal of Molecular Biology*, 425(18), 3378-3388.
- Singh, P., Shaffer, S. A., Scherl, A., Holman, C., Pfuetzner, R. A., Larson Freeman, T. J., Miller, S. I., Hernandez, P., Appel, R. D., & Goodlett, D. R. (2008). Characterization of protein cross-links via mass spectrometry and an open-modification search strategy. *Analytical Chemistry*, 80(22), 8799-8806.
- Steven, A. C. (1993). Conformational Change - An Alternative Energy Source? Exothermic Phase Transition in Phage Capsid Maturation. *Biophysical Journal*, 65(1), 5-6.
- Sternberg, N., & Hoess, R. H. (1995). Display of peptides and proteins on the surface of bacteriophage lambda. *Proceedings of the National Academy of Sciences of the United States of America*, 92(5), 1609–1613.
- Suttle, C.A. (2007). Marine viruses-major players in the global ecosystem. *Nature Reviews. Microbiology*, 5(10), 801–812
- Taylor, G. L. (2010). Introduction to phasing. *Acta Crystallographica. Section D, Biological Crystallography*, 66(Pt 4), 325-338.
- Terzi, M., & C. Levinthal (1967). Effects of λ -phage infection on bacterial synthesis. *Journal of Molecular Biology* 26(3): 525-535.
- Thomas, R. (1971). Control circuits. In: Hershey, A.D. (Ed.), *The Bacteriophage Lambda* (pp. 211-220). Cold Spring Harbor, NY: Cold Spring Harbor Laboratory.
- Tzamtzis, H. M. a. D. (1997). Mutations of the coat protein gene of bacteriophage Lambda that overcome the necessity for the FI gene; the EFi domain. *Molecular Microbiology*, 24(2), 341-353.
- Wendt, J. L., & Feiss, M. (2004). A fragile lattice: replacing bacteriophage lambda's head stability gene D with the shp gene of phage 21 generates the Mg²⁺-dependent virus, lambda shp. *Virology*, 326(1), 41-46.
- Williams, R. C., & Richards, K. E. (1974). Letter: Capsid structure of bacteriophage lambda. *Journal of Molecular Biology*, 88(2), 547-550.
- Wikoff, W. R., Che, Z., Duda, R. L., Hendrix, R. W., & Johnson, J. E. (2003). Crystallization and preliminary analysis of a dsDNA bacteriophage capsid intermediate: Prohead II of HK97. *Acta Crystallographica. Section D, Biological crystallography*, 59(Pt 12), 2060–2064.

- Wikoff, W. R., Liljas, L., Duda, R. L., Tsuruta, H., Hendrix, R. W., & Johnson, J. E. (2000). Topologically linked protein rings in the bacteriophage HK97 capsid. *Science*, 289(5487), 2129–2133.
- Wikoff, W. R., Duda, R. L., Hendrix, R. W., & Johnson, J. E. (1998). Crystallization and preliminary x-ray analysis of the dsDNA bacteriophage hk97 mature empty capsid. *Virology*, 243, 113–118.
- Wikoff, W. R., Duda, R. L., Hendrix, R. W., & Johnson, J. E. (1999). Crystallographic analysis of the dsDNA bacteriophage HK97 mature empty capsid. *Acta Crystallographica. Section D, Biological Crystallography*, 55(Pt 4), 763–771.
- Wolkowicz, R., & Schaechter, M. (2008). What makes a virus a virus? *Nature Reviews Microbiology*, 6(8), 643.
- Yang, Q., & Catalano, C. E. (2003). Biochemical characterization of bacteriophage lambda genome packaging in vitro. *Virology*, 305(2), 276–287.
- Yang, Q., Catalano, C. E., & Maluf, N. K. (2009). Kinetic analysis of the genome packaging reaction in bacteriophage lambda. *Biochemistry*, 48(45), 10705–10715.
- Yang, F., Forrer, P., Dauter, Z., Conway, J. F., Cheng, N., Cerritelli, M. E., Steven, A. C., Plückthun, A., & Wlodawer, A. (2000). Novel fold and capsid-binding properties of the lambda-phage display platform protein gpD. *Nature Structural Biology*, 7(3), 230–237.
- Yulinery, T., Triana, E., Suharna, N., & Nurhidayat, N. (2019). Isolation and anti-Escherichia coli biofilm activity of lytic bacteriophages isolated from water environment in vitro. *IOP Conference Series: Earth and Environmental Science*, 308(1), 012010.
- Zhang, Z., Greene, B., Thuman-Commike, P. A., Jakana, J., Prevelige, P. E., Jr, King, J., & Chiu, W. (2000). Visualization of the maturation transition in bacteriophage P22 by electron cryomicroscopy. *Journal of Molecular Biology*, 297(3), 615–626.
- Ziegelhoffer, T., Yau, P., Chandrasekhar, G. N., Kochan, J., Georgopoulos, C., & Murialdo, H. (1992). The purification and properties of the scaffolding protein of bacteriophage lambda. *The Journal of Biological Chemistry*, 267(1), 455–461.