5-7-2016

Study Of The DNA Packaging Protein From A Phage That Replicates Under Extreme Conditions

Philip Lessans
plessans1@gsu.edu

STUDY OF THE DNA PACKAGING PROTEIN FROM A PHAGE THAT REPLICATES UNDER EXTREME CONDITIONS

by

PHILIP LESSANS

Under the Direction of Ming Luo, PhD

Abstract

Bacteriophages are viruses that infect bacteria. Phage DNA packaging is one process in the assembly of mature phages that is not well characterized. Elucidating the mechanism of this process would enhance our understanding of the biological significance of the machinery. A previous study suggested that the protein gp74 from bacteriophage HK97 functions as an HNH endonuclease and is required for phage DNA packaging. In this thesis the functionality was assessed for a protein from a phage that replicates under extreme conditions. Our data suggest that the endonuclease activity may not be essential for the role of gp74 in phage DNA packaging.

INDEX WORDS: Bacteriophage, HNH endonuclease, gp74

STUDY OF THE DNA PACKAGING PROTEIN FROM A PHAGE THAT REPLICATES UNDER EXTREME CONDITIONS

by

PHILIP LESSANS

A Thesis Submitted in Partial Fulfillment of the Requirements of the Degree of

Master of Science

in the College of Arts and Sciences

Georgia State University

2016

STUDY OF THE DNA PACKAGING PROTEIN FROM A PHAGE THAT REPLICATES UNDER EXTREME CONDITIONS

by

PHILIP LESSANS

Committee Chair: Ming Luo

Committee: Kathryn Grant

Ming Luo

Jenny Yang

Electronic Version Approved:

College of Graduate Studies

College of Arts and Sciences

Georgia State University

May 2016

DEDICATION

I would like to dedicate my thesis to my family and friends who are supportive and help me through thick and thin. Thank you.

ACKNOWLEDGEMENTS

I would like to acknowledge the members of my lab group Ryan Gumpper, Dr. Weike Li, Dr. Tales Rocha de Moura, Chelsea Severin, and James Terrell, who have contributed insight and support in my research projects. Without their help it would have taken longer to overcome hurdles.

I would also like to thank my advisor Dr. Ming Luo who has guided me and allowed me the opportunity to be a part of his group. It has been a privilege and a great experience to learn from you and explore these projects.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

LIST OF ABBREVIATIONS

ATP-Adenosine triphosphate

Bis-Tris- 2-[Bis(2-hydroxyethyl)amino]-2-(hydroxymethyl)propane-1,3-diol

βME-β-mercaptoethanol

cDNA-Complimentary DNA

DNA-Deoxyribonucleic acid

E. coli-Escherichia coli

EDTA-Ethylenediaminetetraacetic acid

His-Histidine

IPTG-Isopropyl-β-D-1-thiogalactopyranoside

LB-Lysogeny broth

MPD- 2-Methyl-2,4-pentanediol

ORF-Open reading frame

Psymv2-temperate bacteriophage Psychrobacter sp. MV2

SDS-PAGE-Sodium dodecyl sulfate polyacrylamide gel electrophoresis

ts49-phage

49T*^{P}-partially glucosylated mature ts49-tDNA

1 INTRODUCTION

1.1 Replication Cycle of a Tailed Phage

Bacteriophages are viruses that infect and reproduce in bacteria with two distinct life cycles: lytic and lysogenic. The type of life cycle that the phage uses depends on the genetic make-up and interactions with the host bacteria. After infection by lytic phages, the phage genome is replicated and assembled into progeny phages. These progenies are released through lysis (Figure 1A). Temperate phages are phages that proceed through a lysogenic cycle, in which the phage genome becomes a part of the host chromosome, thus becoming a prophage. Prophages remain in either a latent or dormant state in which the prophage does not promote cell death or production of progeny phages (Fig. 1B). The lysogenic state of prophages is sustained by having phage lytic genes repressed. When stressful conditions arise, such as DNA damage, the phage genome is excised from the host cell to replicate the DNA, assemble new phage particles, package the DNA and lyse the cell. The initiation of the lysogenic and lytic cycles are dictated by excision and integration of the phage genome into the host chromosome. The attachment of the phage genome into the host chromosome occurs at a specific site in the host genome (*attB*) and connects with a specific site in the phage genome (*attP*). Some phages integrate randomly within the host genome and therefore increase their variability and can induce mutations in a population. A less common phage life cycle is known as pseudolysogeny in which replication of the phage genome does not occur compared to the lytic cycle or lysogenic cycle (Fig. 1C).[1]

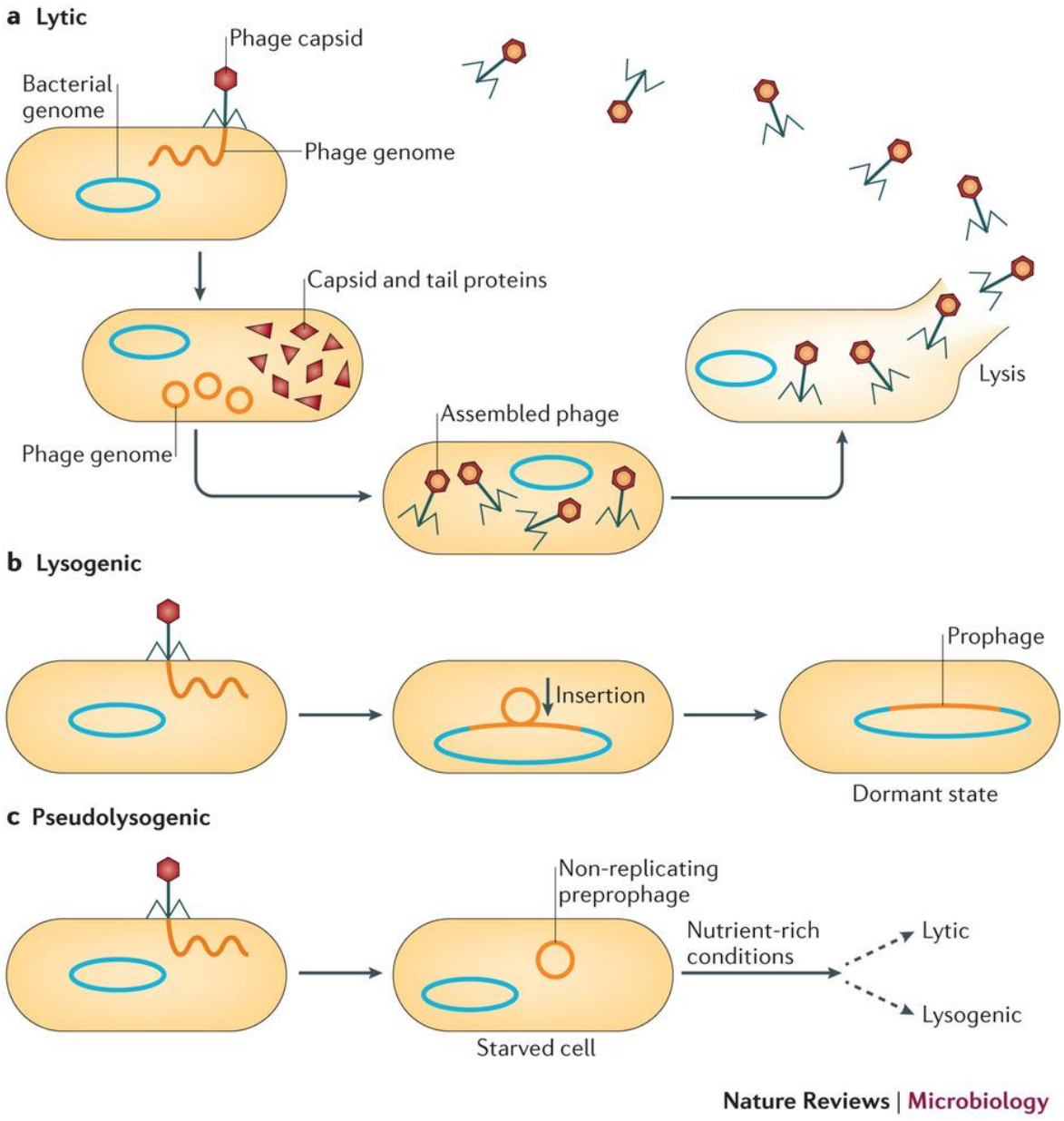

Figure 1 Different Replication Cycles of Phages.

Figure 1 Different Replication Cycles of Phages. A) Lytic cycle of phage replication B) Lysogenic cycle of phage replication C) Pseudolysogenic cycle of phage replication.

1.2 DNA Packaging into the Phage Head

A study was conducted by Black and Silverman to elucidate the phage genome packaging mechanism into heads of Bacteriophage T4.[2] The study brings attention to the fact that replicated phage T4 DNA in the cytoplasmic DNA pool can be fully glucosylated very quickly and then packaged into certain head intermediates when glucose is provided to an infected host unable to synthesize glucose. It was also noted that a reasonable hypothesis could be that glucosylation of packaged DNA within the ts49 phage head intermediate can be resisted since cytoplasmic DNA-binding proteins from the mature head are excluded. The 49T*^{P} DNA structure from the same study suggested that DNA is run into the head in a unidirectional sense from one mature end. Glucosylation experiments showed that the DNA is pumped into the head in a linear fashion. Furthermore, the reported data suggested that the DNA that enters the head also exits the head upon injection and when first deposited in the head it is packaged in the center and not on the capsid surface.[2]

Due to the energy requirement for condensation being the lowest where there is maximum curvature along the capsid wall, the first T4 DNA most likely distributes there upon entry. However, additional experimental data suggest that the process in which the DNA is actually deposited into the center of the phage head is active and it seems unlikely that DNA packaging into the head proceeds after proteolysis of core and capsid proteins. One mechanism proposed by the authors is that acidic peptides pull the DNA toward the center of the phage head. On the other hand, the process is different for bacteriophage lambda. The first DNA to condense for lambda DNA is on the inner

surface of the capsid. Though these are two different mechanisms, both may be possible for their respective phages. Data reveals that when thin sections of maturing phage T4 and P22 heads are collected, the partially packaged DNA first condenses in the center of the capsid. Given that DNA is negatively charged on its backbone and the inner surface of the capsid is also negatively charged, the repulsion experienced between the DNA and the inner capsid surface could explain a mechanism of ejection from the head during infection. Additionally this repulsive force may be generated during packaging in the head. The manner of DNA packaging in T4 phage suggests that the process could be enzymatically driven. One study found that the T4 capsid protein p20 is actively involved in driving the DNA into the head when combined with p17.[2]

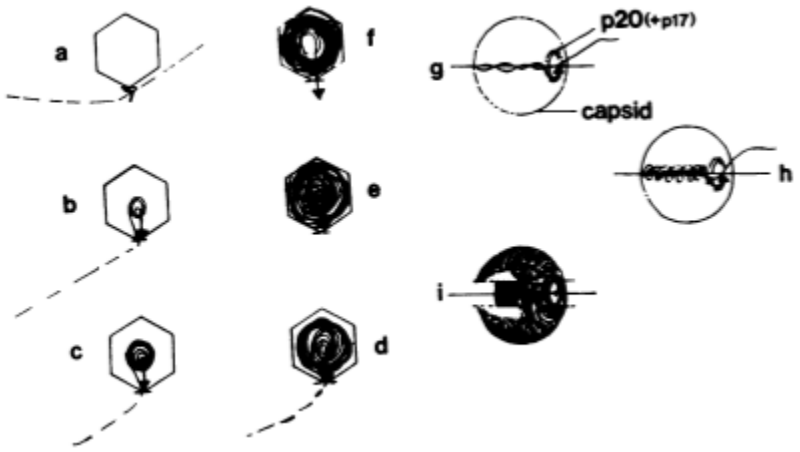

Figure 2 Schematic illustration of the suggested mechanism of in vivo DNA condensation and decondensation into the T4 capsid.

Black, LW. and Silverman, DJ. *Journal of Virology* (1978) Vol. 28, No. 2: 643-655. Copyright © 1978, American Society for Microbiology.

Figure 2. Schematic illustration of the suggested mechanism of in vivo DNA condensation and decondensation into the T4 capsid. A-F) The progression of DNA condensation inside a phage head. The DNA deposits at the center first and accumulates outward.

More recent data by Fujisawa and Morita report other mechanisms.[3] The DNA of double stranded DNA bacteriophages is highly packed in an outer shell of protein. In order to preserve the 5' end bases of the mature DNA, the mechanism to generate the DNA is important. The two important reactions for the processing of this DNA include: i) the mechanism for the end-formation of the concatemer and ii) the energetic mechanism of DNA translocation into the phage head powered by ATP-hydrolysis. A common mechanism for DNA packaging into the prohead is through the connector at the portal of the prohead vertex. A packaging enzyme known as terminase assists in this process. ATP hydrolysis or ATP as an allosteric binder affects the formation of mature DNA for translocation, ternary complexes that contain the prohead, and the packaging enzyme. This is indicative that ATP is important for different purposes in the packaging reaction. There are two subunits of the packaging enzyme: the smaller subunit, gp18, and a large subunit gp19, form separate complexes with DNA and proheads (Fig. 3).[3] Together they form a 50S complex responsible for DNA translocation into the head and requires ATP to allosterically bind. Only gp19 interacts with ATP *in vitro* and reveals multiple activities affected by interactions with ATP including specific and non-specific DNA cleavage, prohead binding, DNA translocation and ATP hydrolysis.[3]

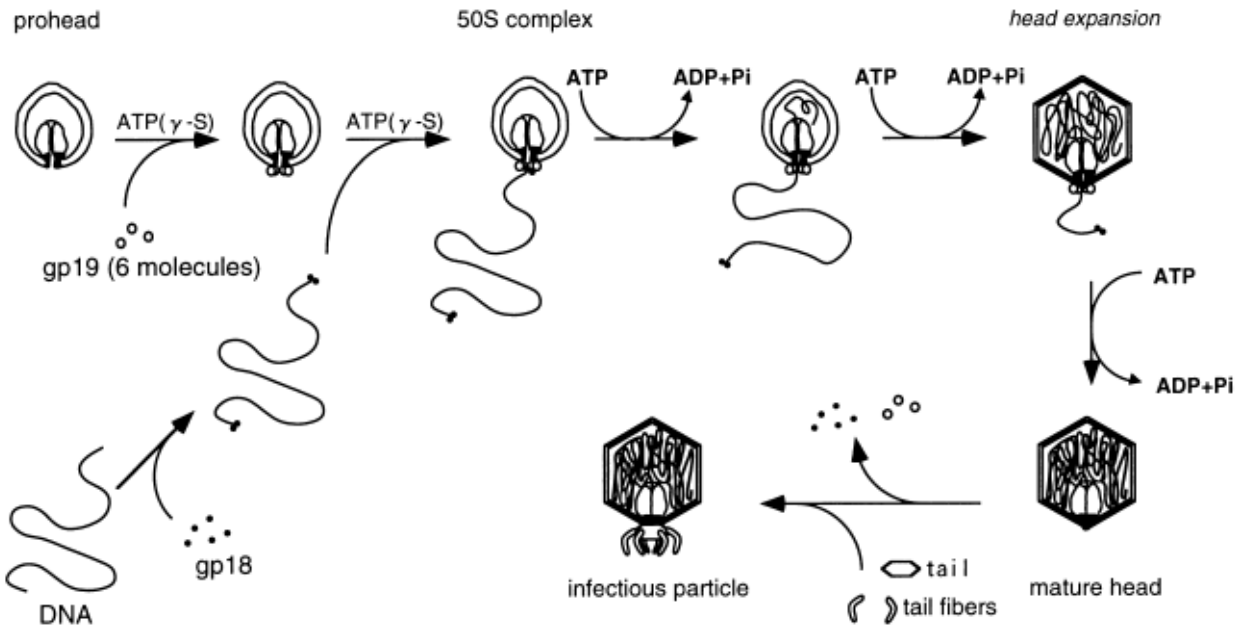

Figure 3 In vitro assembly pathway of T3 phage.

Figure 3. *In vitro* assembly pathway of T3 phage.

gp19 binds to the connector of the prohead with or without ATP. However proheads are inactivated by gp19 alone. gp19 changes conformation upon binding ATP. Therefore, the gp19-prohead complex is the main DNA packaging machine. The energy stored in the β-γ phosphoanhydride bond of ATP is coupled to the translocation of DNA into the prohead. *In vitro* packaging systems demonstrate the DNA-dependent ATPase activity, although the ATPases are stimulated by nonpackageable DNA. A G429R mutant of gp19 showed defective functionality of hydrolysis of ATP and DNA translocation, suggesting gp19 is crucial for coupling ATP hydrolysis to DNA translocation. There is transient

supercoiling in the DNA structure while another model, the rotating structure models, proposes that rotation of the connector translocates the DNA.[2]

In the assembly of phages and genome packaging, there is a requirement for a terminase enzyme, which has a small and large subunit (TerL and TerS). The terminase enzyme is necessary to cleave the genomic DNA at a site known as "cos". In HK97 phage genome, the position of the gp74 gene is adjacent to the "cos" site and the terminase gene, which suggests that gp74 plays an essential role in the DNA cleavage activity of HK97 terminase. An *in vitro* assay showed an enhanced cleavage of a cos site in an engineered DNA substrate by terminase in the presence of gp74 protein (0.2 μM gp74 and 5 μM terminase) and very small amounts of cleavage with terminase alone; however, gp74 alone did not demonstrate any cleavage of the substrate. Two very distinct bands of smaller molecular weight than the DNA substrate formed in the presence of both terminase enzyme and gp74 when analyzed by agarose gel electrophoresis; however, smaller molecular weight bands did not form with gp74 alone. Furthermore, his-tagged gp74 was mixed with untagged TerL/S and applied to a Ni-NTA purification column. After elution from the column with imidazole, analysis of the eluted sample by SDS-PAGE revealed that TerL/S was co-eluted with gp74. These data suggested that site-specific cleavage of the cos site requires an interaction between gp74 and TerL/S.[4]

1.3 HNH Endonuclease

The close proximity of HNH and phage terminase genes in phage genomes and the activity assays suggest that HNH proteins have an important role in the endonuclease activity of the terminase enzyme, which is responsible for the formation of unit-length

virions from concatemers during phage assembly. HNH proteins are characterized by having an HNH motif, which is present in a variety of proteins exhibiting endonuclease activity, but is specific to the HNH Pfam protein family. Proteins that primarily contain an HNH-motif exhibit non-specific activity while those that also contain a DNA-recognition domain have high sequence specificity.[4] Some examples of proteins with non-specific activity include Colicins E7 and E9.[5,6] One example for proteins that demonstrate homing endonuclease behavior, on the other hand, is I-HmuI.[7] HNH endonucleases contain two highly conserved His residues and one Asn residue in an HNH motif that is about 35 amino acids. Structural data of these three enzymes show that the HNH motif is comprised of a two-stranded anti-parallel β-sheet with an α-helix on one side.[7,8] Together these secondary structures are called the ββα-metal fold.[9] The difference between I-HmuI, and Colicins E7 and E9 is that the N-terminal residue of the motif in I-HmuI is Asp whereas it is a His in Colicins E7 and E9.[7]

Researchers have aligned a wide variety of terminase-associated HNH proteins and identified a hidden Markov model (HMM) from the sequence alignment. From the 200 most significant hits encoded by genes in phages and prophages, 95% were within 10 open reading frames (ORFs) of a gene encoding a TerL subunit, the larger domain of terminase enzyme. In contrast, another alignment was obtained with the Pfam HNH HMM, and of the top 200 hits 16% of the genes were close to TerL genes. These data suggest that terminase-associated HNH proteins constitute a unique functional subfamily. *E. coli* phage HK97 gp74, a 119-residue protein, contains an HNH motif and was previously demonstrated to possess a conserved His82 residue necessary for binding the divalent metal ion.[4] gp74 was originally reported to have an endonuclease activity using lambda phage DNA cleavage assays. Due to the affinity of the His tag to

bind Ni^{2+}, the effective concentration of metal ions in the activity assays would be lower since the His tag was not cleaved from gp74 or ORF22 before the assays. Additionally, Co^{2+}, Ni^{2+}, and Zn^{2+} are three ions that can be considered as potential cofactors for gp74. In assessing these metals for use in the assays, the atomic radius and electron configuration need to be considered. In accordance with the periodicity in the properties of the elements and the Aufbau principle that was first proposed by Niels Bohr and using Madelung's rule, all three have an empty $4s$ orbital that allows lone pair electrons on other atoms to coordinate with the metal ion, but the number of electrons in the $3d$ orbitals is different.[20,21] gp74 demonstrated DNA smearing with Zn^{2+} and more efficient DNA smearing with Ni^{2+} as observed by Moodley et al.[14] The Ni^{2+} ion is missing two electrons in two of its $3d$ orbitals so the atomic radius is smaller than Zn^{2+} because the nuclear charge is pulling on fewer $3d$ orbital electrons. Since Co^{2+} is missing one more electron in a $3d$ orbital, it is smaller than Ni^{2+} and might fit more easily into the active site, therefore gp74 might demonstrate more efficient DNA smearing with Co^{2+}.

1.4 Biological Activities under Extreme Conditions

Enzyme activity and efficiency are dictated by several factors including pH, solution viscosity, solubility, and temperature. The temperature is relevant due to the fact that every chemical reaction has an activation energy barrier. Temperature is a measure of the mean kinetic energy of the environment, so higher temperatures indicate that there is more available Gibb's free energy and lower temperatures indicate that there is less available Gibb's free energy.[17] Many synthetic reactions require harsh conditions and long periods of time that do not necessarily yield large quantities of product. However organisms have adapted to various environments through evolution in order to survive

by lowering the energy of activation barrier. With respect to chemical reactions *in vivo*, organisms have evolved to maximize the available Gibb's free energy through the use of enzymes, which lower the activation energy barrier for a specific reaction.

Psychrophiles are organisms that grow at temperatures below 20 °C. Low temperatures for a typical protein decrease the mean kinetic energy for conformational movement and overall activity of a protein.[10] However, psychrophilic organisms are able to thrive under these extremely low temperatures because their proteins have adapted to optimize their activity by improving structural flexibility. When compared to proteins from mesophilic organisms that thrive under temperatures ranging from 20 to 45 °C, interactions that stabilize a protein are weakened in psychrophilic proteins. Adaptations by psychrophilic proteins include a greater number of glycine residues to allow for more movement, fewer rigid proline residues in loop regions, fewer arginine residues that form stable salt bridges and hydrogen bonds, and a reduction in size of nonpolar residues in the core in order to weaken hydrophobic interactions.[11] In the cold-adapted species *Halorubum lacusprofundi,* proteins exhibit a decrease in the number of tryptophan and glutamic acid residues, which would stabilize core hydrophobic interactions and form stabilizing hydrogen bonds, respectively, in other proteins. As the most flexible and rigid amino acids, an increase in the number of glycine residues and a decrease in proline residues, respectively, is not surprising if the evolutionary goal is to increase activity under decreased energy conditions. Therefore, the changes described above make sense when considering that lower temperatures are associated with a decrease in molecular motion. Several three-dimensional models have been created for proteins of psychrophiles *Methanogenium frigidum* and *Methanococcoides burtonii*. These models show a decrease in charged residues and an increase in glutamine and threonine residues on the protein

surface. These changes are believed to reduce the possibility of aggregation while decreasing surface charge.[12]

Although a protein structure might become more unstable with an increased population of weaker interactions, the conformational flexibility of the protein allows the enzyme activity to be increased, in contrast with mesophilic proteins that might be more rigid. The specific activity (k_{cat}) of a psychrophilic enzyme is typically ten-times greater than that of a mesophilic enzyme.[10,11,13]

1.5 Purpose of the Study

Previous studies have shown that gp74 plays an essential role in the DNA cleavage by HK97 terminase enzyme. HK97 terminase cleaves concatemeric DNA into single genome units at a site known as "cos". Amino acid sequence analysis suggests that gp74 contains a DNA recognition domain in addition to an HNH-motif.[4] Analysis by agarose gel electrophoresis of the *in vitro* lambda phage DNA cleavage assays with Ni^{2+}, which stabilizes the negative charge on the phosphate backbone of DNA, showed DNA smearing, which would indicate non-specific cleavage.[14] Non-specific cleavage is characteristic of proteins composed primarily of the HNH-motif.[4] These are two conflicting hypotheses. A nickel and cobalt resistance gene was previously identified in *E. coli*, suggesting that nickel ions are toxic to this organism.[15] Meanwhile, ORF22 is

functionally homologous to gp74 because we noticed that ORF22 from the genome of bacteriophage Psychrobacter sp. MV2 (Psymv2) that lives in the McMurdo Dry Valley soil of Antarctica has a gene location that is identical to the gp74 gene in the genome of bacteriophage HK97. We hypothesize that ORF22 is also an HNH endonuclease. Characterization of the two proteins were carried out in order to provide additional data about the functionality of gp74 and to investigate the functionality of ORF22.

1.6 Expected Results

Previous data suggested two different modes of *in vitro* phage DNA cleavage by gp74.[4,14] This protein showed non-specific cleavage of lambda phage DNA *in vitro* in the presence of Ni^{2+}, while site-specific cleavage of linear DNA was observed when terminase enzyme was present. Since Psymv2 experiences temperatures ranging from -55 °C to 25 °C and given that ORF22 and gp74 have the same gene position in their respective genomes, non-specific cleavage of lambda phage DNA *in vitro* would be expected for ORF22 in the presence of Ni^{2+} at -20 °C and would support the non-specific cleavage mechanism of gp74.

2 EXPERIMENT

2.1 Expression and Purification

The coding sequence for ORF22 in the genome of temperate bacteriophage *Psychrobacter* sp. MV2 was previously cloned by Yousuf Ali into the pET-28b (+) expression vector with

a 6x N-terminal His tag. The plasmid was transfected into competent BL21(DE3) *E. coli*. Transfected cells were grown in LB at 37 °C until the OD_{600nm} reached 0.7 and expression was induced with 0.5 mM IPTG for 18-20 h at 20 °C to maximize the yield of properly folded ORF22. Cells were harvested by centrifugation at 4,000 RPM for 25 min. Harvested cells were stored as a pellet at -80 °C in a 50 mL Falcon tube. Thawed cells were resuspended in five times volume of Lysis buffer (20 mM Tris, pH 7.9, 250 mM NaCl, 5 mM imidazole, 20 mM βME). Resuspended cells were lysed by microfluid processing. The soluble and insoluble fractions were separated from the cellular crude extract by centrifugation at 15,000 RPM for 40 min at 4 °C. The soluble fraction was loaded onto a 5-mL GE HisTrap™ FF Crude column pre-equilibrated in binding buffer (20 mM Tris, pH 7.9, 500 mM NaCl, 40 mM imidazole, 20 mM βME). Unbound protein was washed from the column with binding buffer. After the absorbance returned to baseline, bound protein was eluted from the column with a linear imidazole gradient prepared with binding buffer and Binding buffer containing 600 mM imidazole. Chromatography was analyzed with UNICORN software (Figure 2). Collected samples were analyzed by SDS-PAGE (Figure 3).

The coding sequence for gp74 from the genome bacteriophage HK97 was cloned into the pETM-11 expression vector with a 6x N-terminal His tag. The plasmid was transfected into competent BL21(DE3) *E. coli*. Transfected cells were grown in LB at 37 °C until the OD_{600nm} reached 0.7 and expression was induced with 0.5 mM IPTG for 18-20 h at 20 °C. Cells were harvested by centrifugation at 4,000 RPM for 25 min at 4 °C. Harvested cells were stored at -80 °C. Thawed cells were resuspended in five times volume of lysis buffer (20 mM Tris, pH 7.9, 150 mM NaCl, 5 mM imidazole, 2 mM βME). Resuspended cells were lysed by sonication. The soluble and insoluble fractions were separated from the cellular crude extract by centrifugation at 15,000 RPM for 40 min at 4 °C. The

soluble fraction was loaded onto a 5-mL GE HisTrap™ FF Crude column pre-equilibrated in binding buffer (20 mM Tris, pH 7.9, 500 mM NaCl, 20 mM imidazole, 2 mM βME). Unbound protein was washed from the column with Binding buffer. After the absorbance returned to baseline, bound protein was eluted from the column with a linear imidazole gradient prepared with binding buffer and Binding buffer containing 600 mM imidazole. Chromatography was analyzed with UNICORN software (Figure 7). Collected samples were analyzed by SDS-PAGE (Figures 8A and 8B). Fractions 15-20 of gp74 eluted from Ni-NTA (Figure 8B) were pooled and dialyzed into 50 mM Sodium phosphate, pH 7.0, 150 mM NaCl, 5 mM βME and concentrated to 1.5 mL. The concentrated gp74 protein was purified by a Superdex 75 size exclusion column.

2.2 Enzyme Activity Assay

Purified ORF22 protein was dialyzed against buffer (20 mM HEPES, pH 8.0, 5 mM EDTA, 20 mM βME) to remove residual Ni^{2+} ions using dialysis tubes with a 6-8 kDa MWCO. The sample was dialyzed further against 20 mM HEPES pH 8.0. Lambda phage DNA (New England Biolabs) was used as a substrate and the reaction conditions were set up as follows: 20 mM HEPES, pH 8.0, 10 μg/mL ORF22, 10 μg/mL Lambda phage DNA, $NiCl_2$ (2, 5, and 10 mM). The final concentration of βME was <0.07 mM. The reaction was conducted at 23 °C and aliquots were taken from the reaction at 0, 1, 3, 6, 12, and 24 h. The reaction was stopped with 25 mM EDTA and stored at -20 °C. Reaction aliquots were analyzed using 1% agarose gel electrophoresis and DNA was visualized with ethidium bromide (Figure 4).

The same reaction was set up again with the concentration of NaCl ranging from 350 to 315 mM for 0.5 to 5 mM Ni respectively. The reaction was conducted at 23 °C and aliquots were taken of the reaction at 0, 1, 3, 6, 12, and 24 h. The reaction was stopped with 25 mM EDTA and stored at -20 °C. Reaction aliquots were analyzed using 1% agarose gel electrophoresis and DNA was visualized with ethidium bromide (Figure 5).

Other reaction conditions were set up as follows: 20 mM HEPES, pH 8.0, 368 mM NaCl, 20% glycerol, 10 μg/mL ORF22, 10 μg/mL Lambda phage DNA, $NiCl_2$ (1, 5, and 10 mM). The reaction was conducted at -20 °C for 24 h, and aliquots were taken of the reaction at 0, 1, 3, 6, 12 and 24 h. The reaction was stopped with 25 mM EDTA and stored at -20 °C. Reaction aliquots were analyzed using 1% agarose gel electrophoresis and DNA was visualized with SYBR® Safe DNA Gel Stain (Invitrogen) (Figure 6).

Purified gp74 was dialyzed against buffer (20 mM HEPES, pH 8.0, 5 mM EDTA, 20 mM βME) to remove residual Ni^{2+} ions using dialysis tubes with a 6-8 kDa MWCO. The sample was dialyzed further against 20 mM HEPES pH 8.0. Lambda phage DNA (New England Biolabs) was used as a substrate and the reaction conditions were set up as follows: 20 mM HEPES, pH 8.0, 48 μg/mL gp74, 50 μg/mL Lambda phage DNA, $NiCl_2$ (2, 5, and 10 mM). The final concentration of βME was <0.07 mM. The reaction was conducted at 23 °C and aliquots were taken of the reaction at 0, 1, 3, 6, 12, and 24 h. The reaction was stopped with 25 mM EDTA and stored at -20 °C. Reaction aliquots were analyzed using 1% agarose gel electrophoresis and DNA was visualized with ethidium bromide (Figure 7).

Other reaction conditions were set up as follows: 20 mM HEPES, pH 8.0, 10 μg/mL gp74, 10 μg/mL Lambda phage DNA, $NiCl_2$ (0.5, 1, 2, and 5 mM). The reaction was conducted at 23 °C and aliquots were taken from the reaction at 0, 1, 3, 6, 12, and 24

h. The reaction was stopped with 25 mM EDTA and stored at -20 °C. Reaction aliquots were analyzed using 1% agarose gel electrophoresis and DNA was visualized with ethidium bromide (Figure 8).

2.3 Crystallization

Purified gp74 protein (Fractions 36 and 37; Figure 10A) was concentrated to 4.9 and 9.8 mg/mL respectively. Concentrated gp74 was dispensed into 96-well plates and mixed with Classic Suite, Index, and JCSG crystallization screening kits with a Gryphon robot. Crystallization plates were incubated at 23 °C at Georgia State University.

2.4 Prediction of Intrinsically Unstructured Protein

The amino acid sequence of gp74 was analyzed by the IUPred online tool at the URL http://iupred.enzim.hu/.

3 RESULTS

ORF22 and gp74 proteins were expressed and purified

Various expression conditions were used to optimize the yield of ORF22 and gp74 proteins. Utilizing a Ni-NTA column, ORF22 protein was purified from soluble lysates and was analyzed in the chromatogram and SDS-PAGE. Utilizing a Ni-NTA affinity column and Superdex 75 size exclusion column gp74 protein was purified from soluble lysates and was analyzed in the chromatograms and SDS-PAGE.

DNA smearing was observed for gp74, but not with ORF22 protein

Purified ORF22 protein was used in activity assays. Activity assays containing ORF22 protein were analyzed by 1% agarose gel electrophoresis and visualized by UV light with Ethidium bromide or SYBR® Safe DNA Gel Stain and compared to the results of the gp74 activity assays that were used as controls. gp74 was used as a control to reproduce the DNA smearing that was observed by Moodley et al.[14] Since it was hypothesized that DNA smearing indicated DNA cleavage by gp74 with Zn^{2+} and Ni^{2+}, in separate reactions, Ni^{2+} was chosen as the control metal cofactor because it allowed more efficient cleavage. DNA smearing was observed for the gp74 protein activity control assays in the presence of 0.5, 1, 2, and 5 mM Ni^{2+} with 368 mM NaCl at 23 °C (Figure 13) and 2, 5, and 10 mM Ni^{2+} without NaCl at 23 °C (Figure 12). The ORF22 protein activity assay results showed no observable smearing of DNA in the presence of 0.5, 1, 2, and 5 mM Ni^{2+} at 23 °C (Figure 4) or 0.5 mM Ni^{2+} at -20 °C (Figure 5) as compared to the gp74 control assays.

gp74 is predicted to have an unstructured, C-terminal region

The prediction of intrinsically unstructured protein tool returned a result that showed a C-terminal region (residues 107-119) of gp74 that had disorder tendency >0.5. The disorder of a protein in this prediction refers to the protein having no single well-defined tertiary structure in its native state.[18]

Crystal hits were observed for gp74

Crystal hits were observed for gp74 (9.8 mg/mL) in JCSG+ crystal screen condition 0.2 M calcium chloride dihydrate, 0.1 M Bis-Tris pH 5.5, 45% (v/v) MPD (Figure 14) and Classic Suite crystal screen condition 0.2 M calcium chloride, 0.1 M sodium acetate pH 4.6, 20% (v/v) isopropanol (Figure 15). Crystals were also observed in the Index

crystal screen condition 0.2 M calcium chloride dihydrate, 0.1 M Bis-Tris pH 5.5, 45% (v/v) MPD with gp74 (4.9 mg/mL) (Figure 16B), however, salt crystals were observed in the control (Figure 16A) and with gp74 (9.8 mg/mL) (Figure 16C).

3.1 Affinity chromatography of ORF22 by Ni-NTA

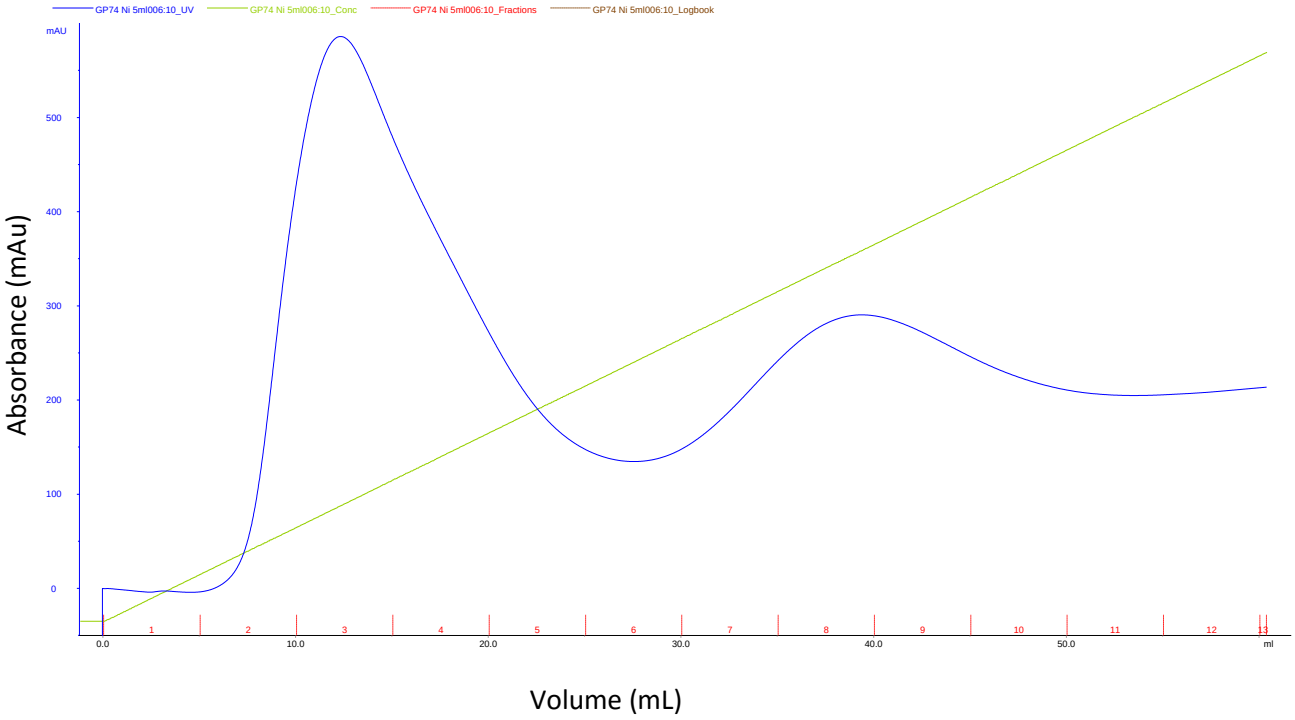

Figure 4 Affinity chromatography of ORF22 by Ni-NTA

Figure 4. Affinity chromatography of ORF22 by Ni-NTA

3.2 SDS-PAGE of ORF22 purification by Ni-NTA

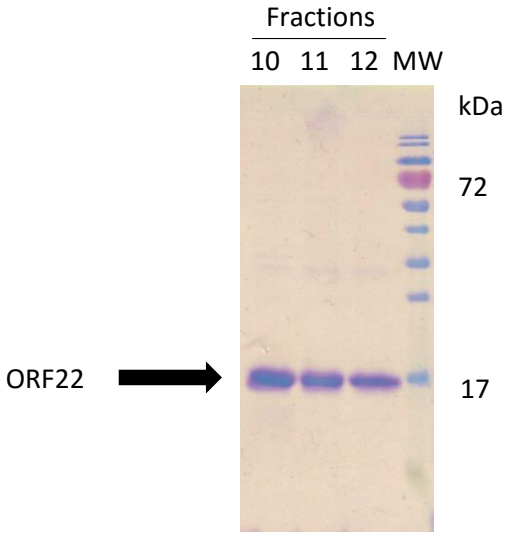

Figure 5 SDS-PAGE of ORF22 fractions 10-12 purified by Ni-NTA

Figure 5. SDS-PAGE of ORF22 fractions 10-12 purified by Ni-NTA. Molecular weight marker (MW), Fractions10-12 of Ni-NTA affinity chromatography.

3.2.1 1% Agarose Gel Electrophoresis of ORF22 Activity Assay at 23 ℃

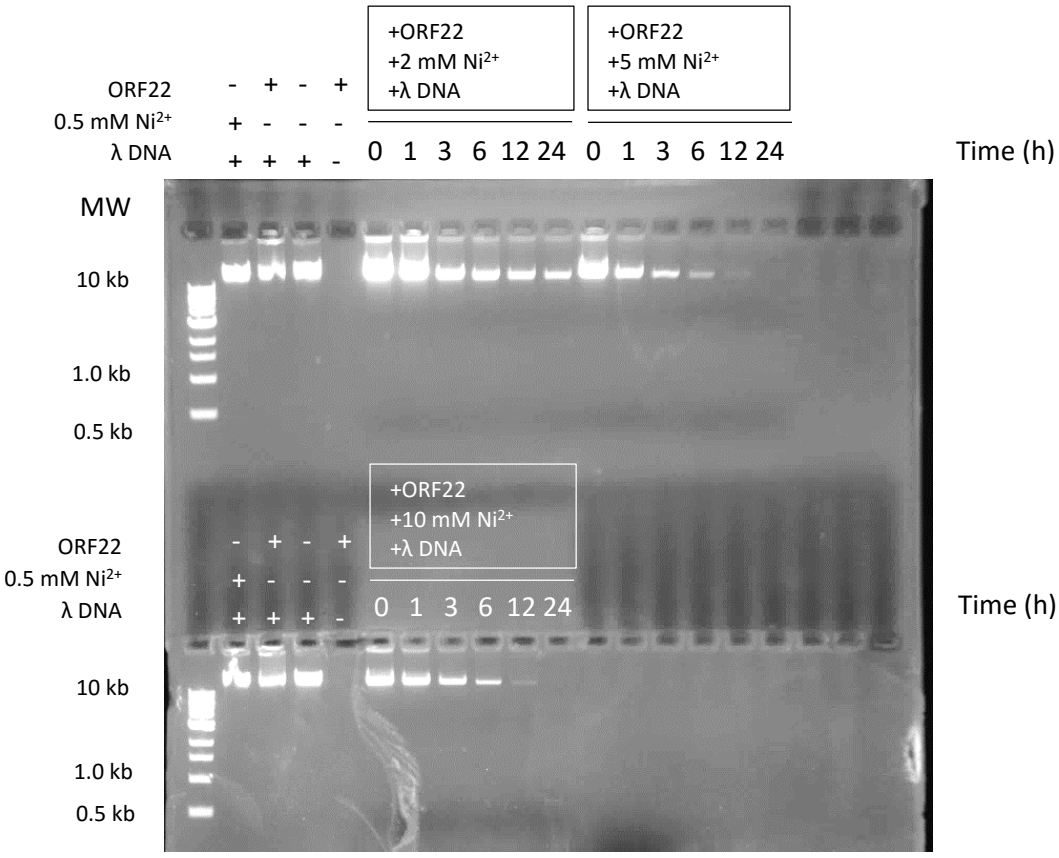

Figure 6 1% Agarose Gel Electrophoresis of ORF22 Activity Assay at 23 ℃

Figure 6. Agarose gel electrophoresis of ORF22 activity assay at 23 °C. ORF22 activity assays with 2 and 5 mM Ni^{2+} (top) and 10 mM Ni^{2+} (bottom).

3.2.2 1% Agarose Gel Electrophoresis of ORF22 Activity Assay with NaCl at 23 °C

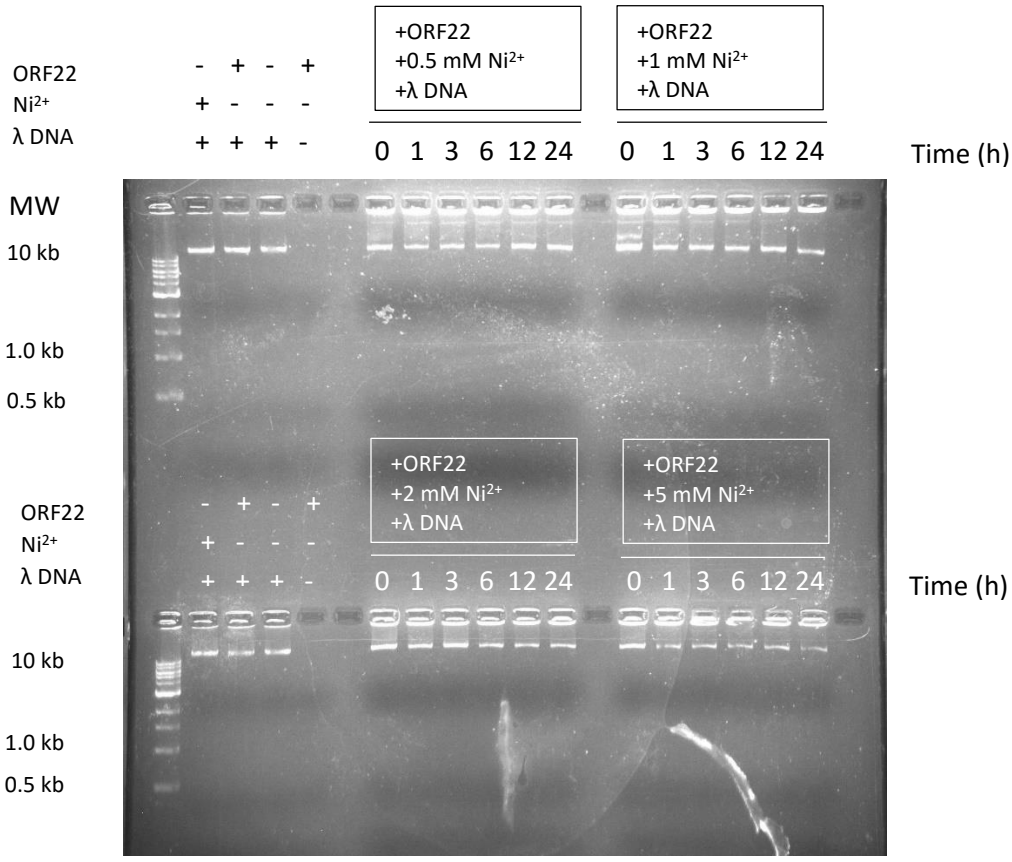

Figure 7 1% Agarose Gel Electrophoresis of ORF22 Activity Assay with NaCl at 23 °C

Figure 7. Agarose gel electrophoresis of ORF22 activity assay with NaCl at 23 °C. ORF22 activity assays with 0.5 and 1 mM Ni^{2+} (top) and 2 and 5 mM Ni^{2+} (bottom).

3.2.3 1% Agarose Gel Electrophoresis of ORF22 Activity Assay at -20 ℃

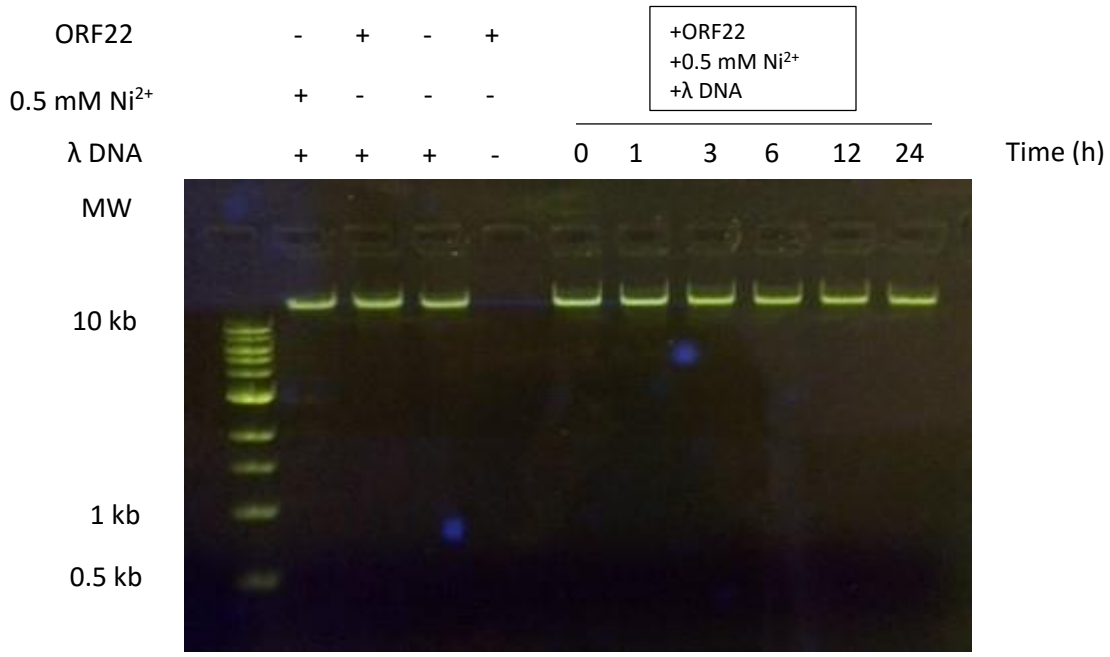

Figure 8 1% Agarose Gel Electrophoresis of ORF22 Activity Assay with NaCl at -20 ℃

Figure 8. Agarose gel electrophoresis of ORF22 activity assay with NaCl at -20 °C. ORF22 activity assay with 0.5 mM Ni^{2+}, 368 mM NaCl, and 20% glycerol.

3.3 Affinity Chromatography of gp74 by Ni-NTA

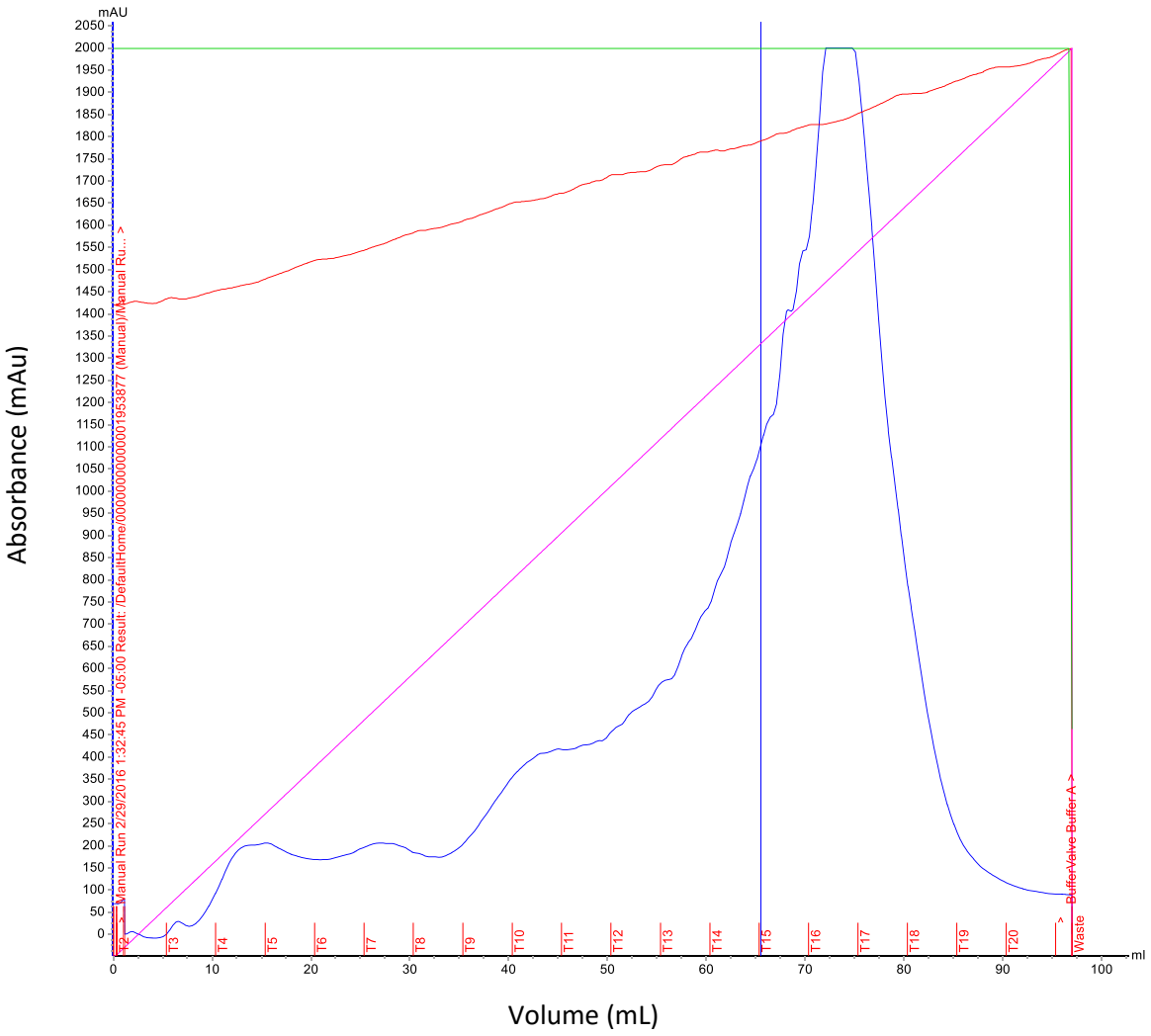

Figure 9 Affinity Chromatography of gp74 by Ni-NTA

Figure 9. Affinity chromatography of gp74 by Ni-NTA. A 20CV gradient up to 600 mM Imidazole.

3.4 SDS-PAGE of gp74 by Ni-NTA

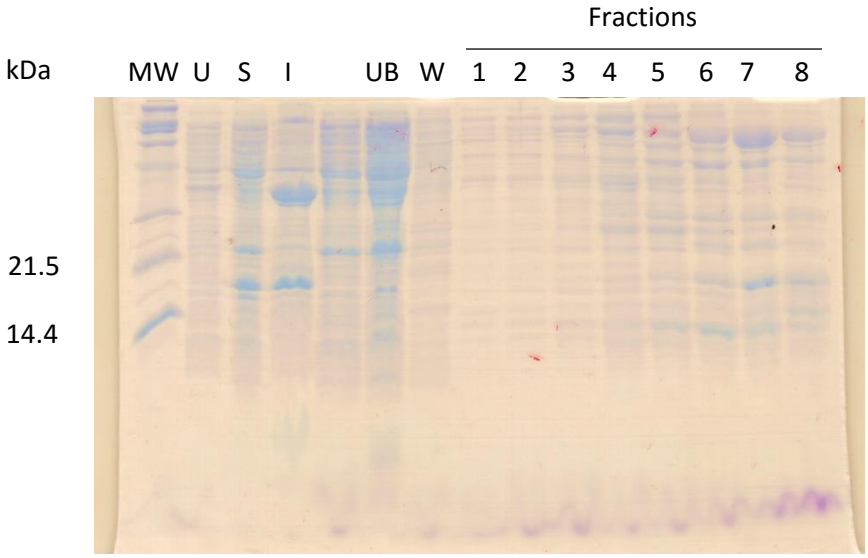

Figure 10A SDS-PAGE of gp74 by Ni-NTA. Molecular Weight Marker (MW), Uninduced expression (U), Soluble fraction after centrifugation of crude cell extract (S), Insoluble fraction after centrifugation of crude cell extract (I), Unbound protein (UB), Wash step (W), Eluted protein: Fractions 1-8.

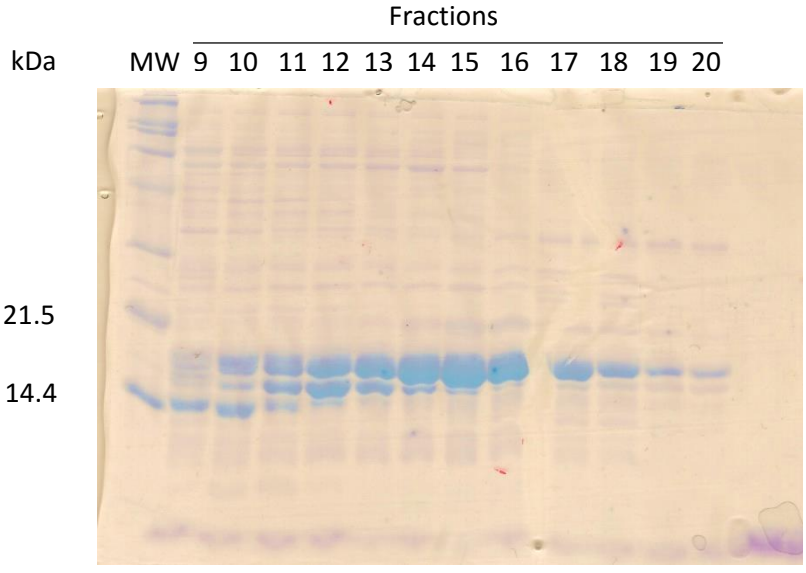

Figure 10B SDS-PAGE of gp74 Ni-NTA Extraction. Molecular Weight Marker (MW), Eluted protein: Fractions 9-20.

Figure 10 SDS-PAGE of gp74 by Ni-NTA

3.5 Size Exclusion Chromatogram of gp74

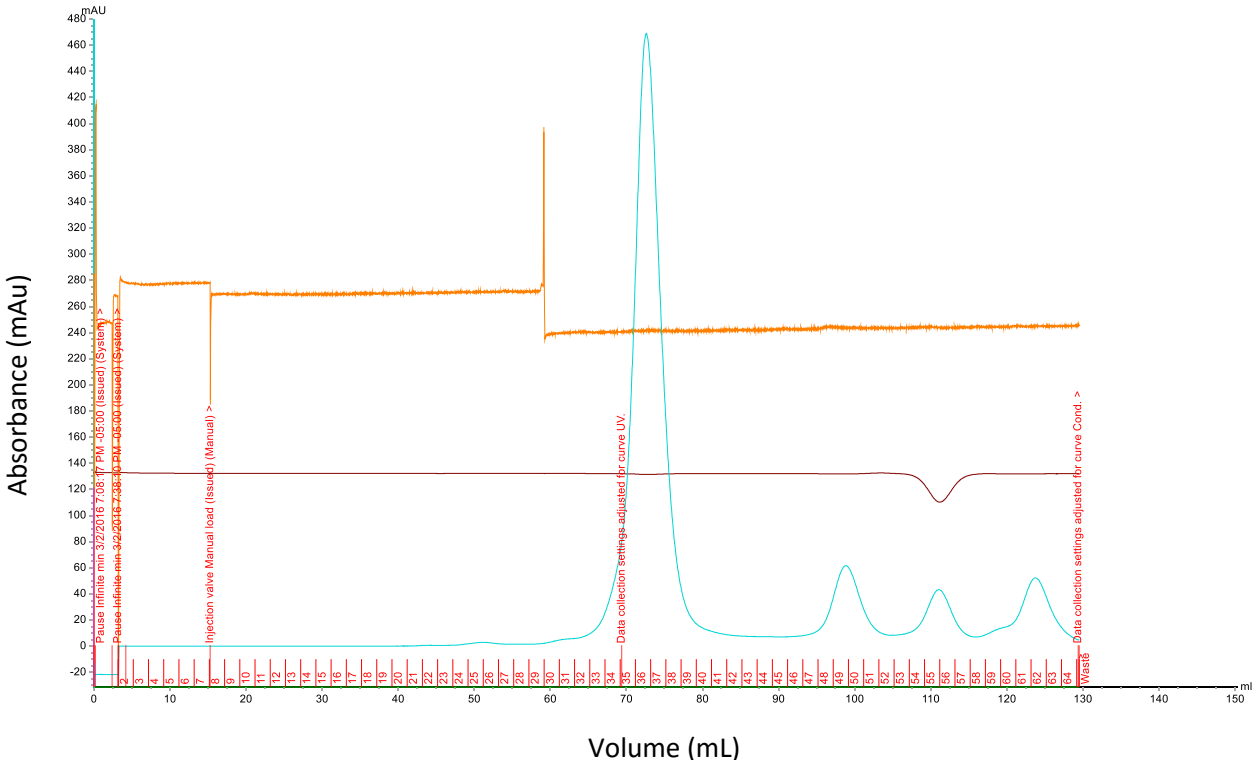

Figure 11 Size Exclusion Chromatogram of gp74

Figure 11 Size Exclusion Chromatogram of gp74. The Superdex 75 size exclusion chromatogram of gp74 from the pooled fractions 15-20 from Ni-NTA.

3.6 SDS-PAGE of gp74 Purified by Size Exclusion Chromatography

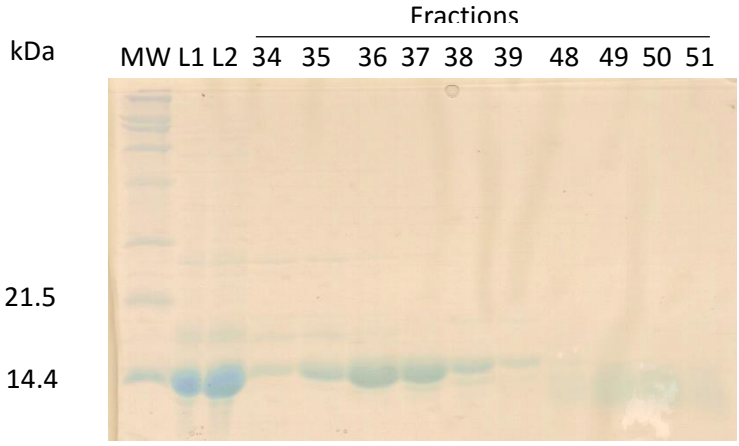

Figure 12A. SDS-PAGE of gp74 Purified by Size Exclusion Chromatography. Molecular Weight (MW), Load sample 1 µL (L1), Load sample 2 µL (L2), Fractions 34-39, 48-51.

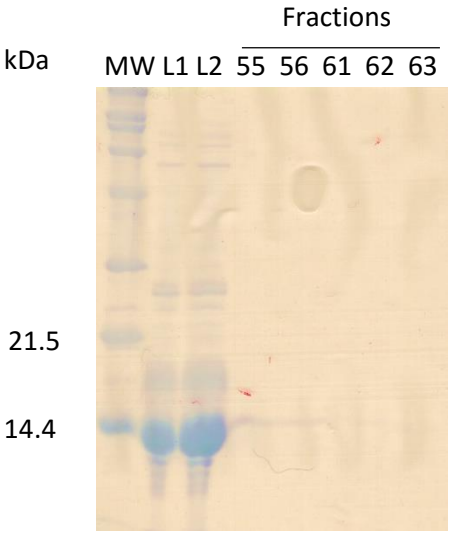

Figure 12B. SDS-PAGE of gp74 Purified by Size Exclusion Chromatography. Molecular Weight (MW), Load sample 1 µL (L1), Load sample 2 µL (L2), Fractions 55, 56, 61-63.

Figure 12 SDS-PAGE of gp74 Purified by Size Exclusion Chromatography

3.6.1 1% Agarose Gel Electrophoresis of gp74 Activity Assay at 23 °C

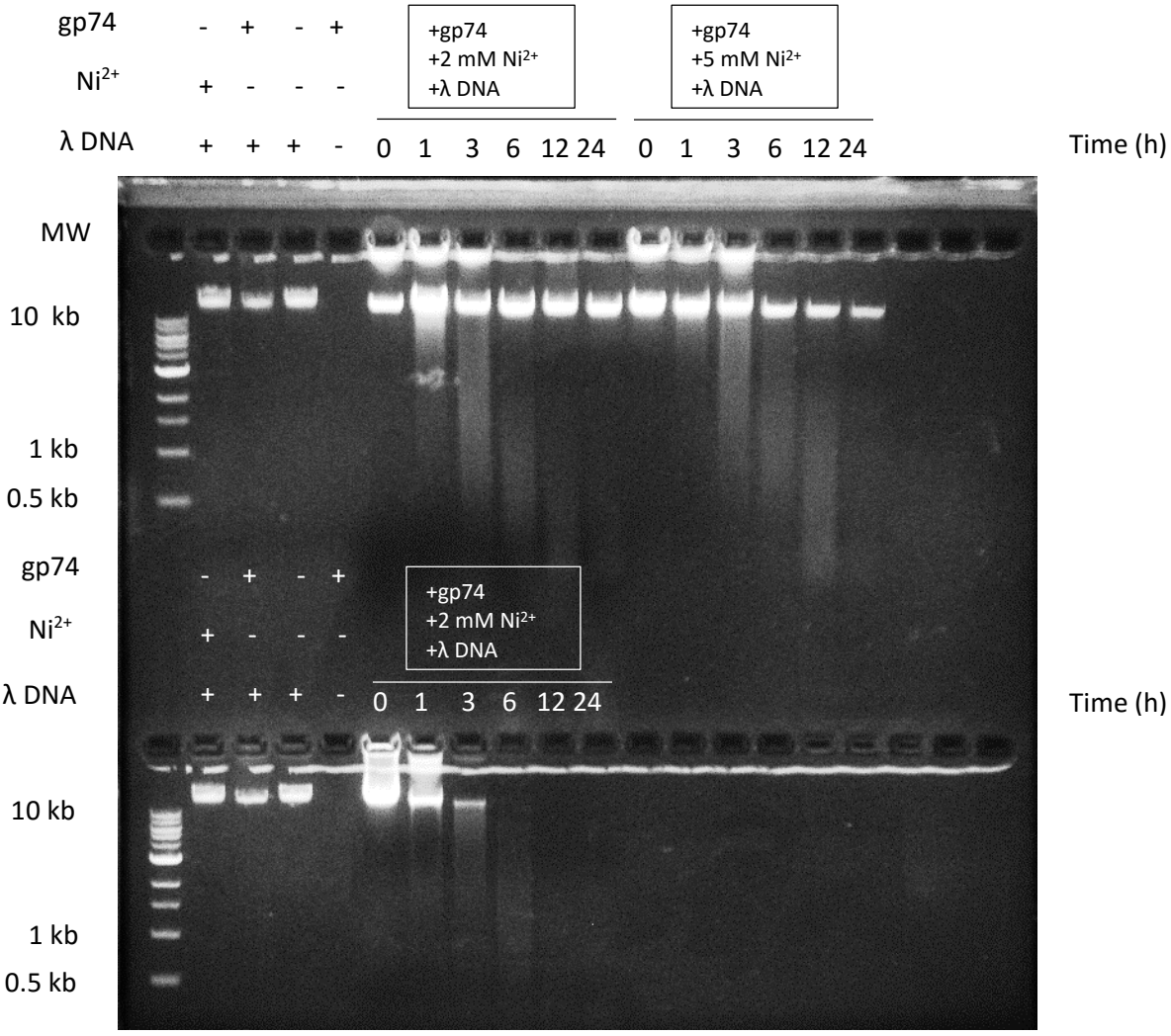

Figure 13 1% Agarose Gel Electrophoresis of gp74 Activity Assay at 23 °C

Figure 13. Agarose Gel Electrophoresis of the gp74 activity assay conducted at 23 °C. gp74 activity assays with 2 and 5 mM Ni^{2+} (top) and 10 mM Ni^{2+} (bottom).

3.6.2 1% Agarose Gel Electrophoresis of gp74 Activity Assay with NaCl at 23 °C

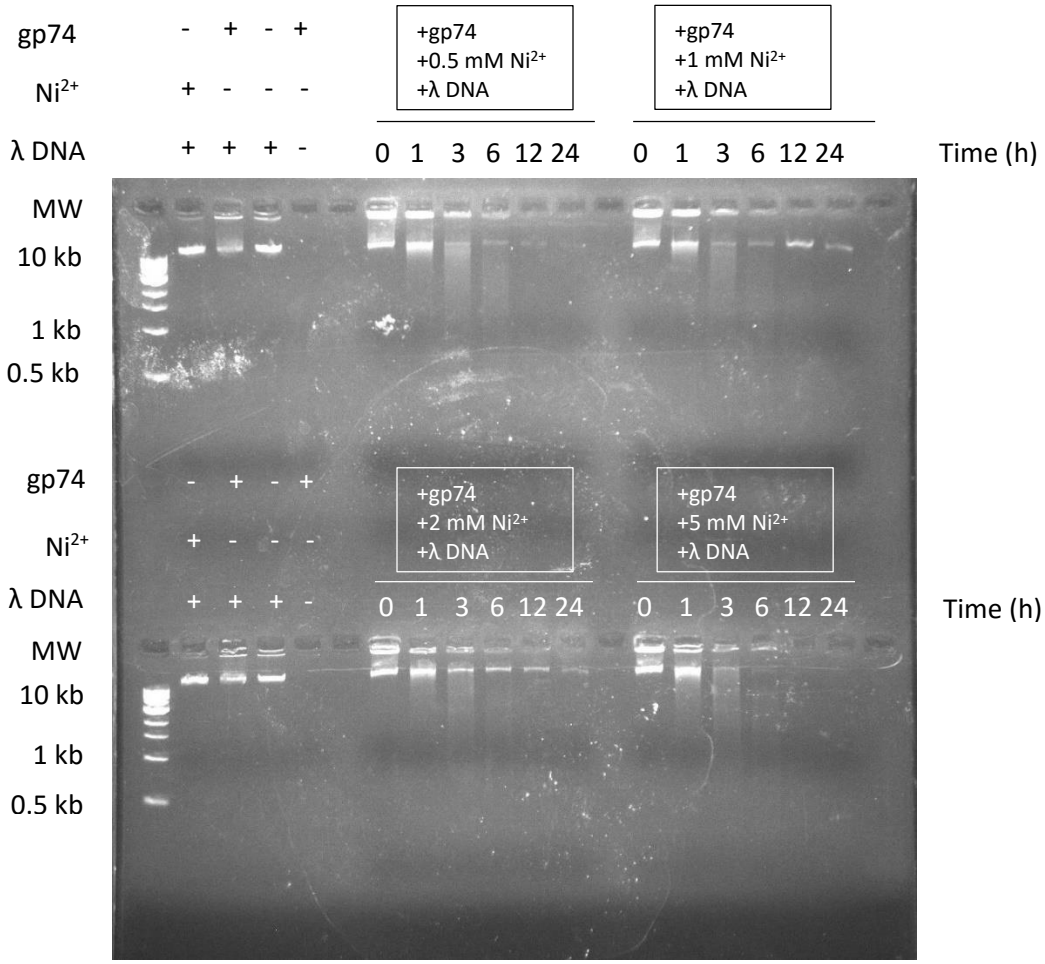

Figure 14 1% Agarose Gel Electrophoresis of gp74 Activity Assay with NaCl at 23 °C

Figure 14. Agarose Gel Electrophoresis of gp74 Activity Assay Conducted at 23 °C with 368 mM NaCl. gp74 activity assays with 0.5 and 1 mM Ni^{2+} (top) and 2 and 5 mM Ni^{2+} (bottom).

3.7 Prediction of Intrinsically Unstructured Protein

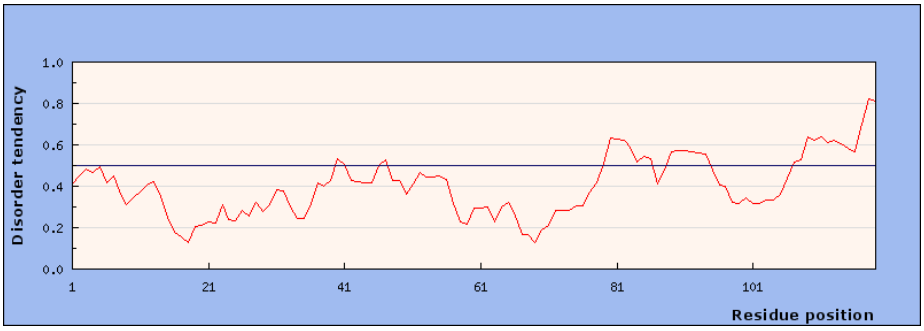

Figure 15 Prediction of Intrinsically Unstructured Protein.

Figure 15. Values greater than 0.5 of disorder tendency indicate unstructured residues.

3.7.1 gp74 Crystal Screen Hit-JCSG+

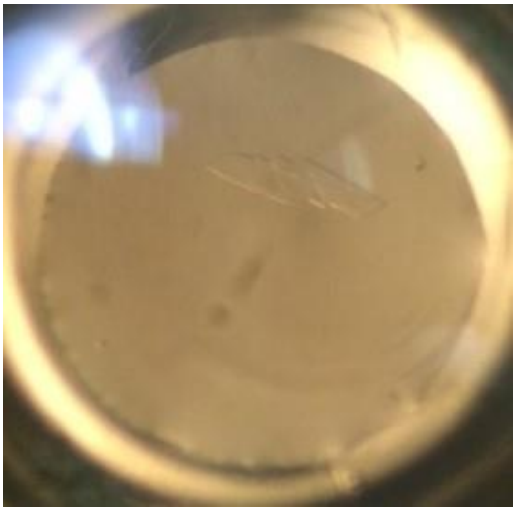

Figure 16 gp74 Crystal Screen Hit-JCSG+

Figure 16. gp74 Crystal Screen Hit-JCSG+. 0.2 M calcium chloride dihydrate, 0.1 M Bis-Tris pH 5.5, 45% v/v MPD, gp74 (9.8 mg/mL)

3.7.2 gp74 Crystal Screen Hit-Classic Suite

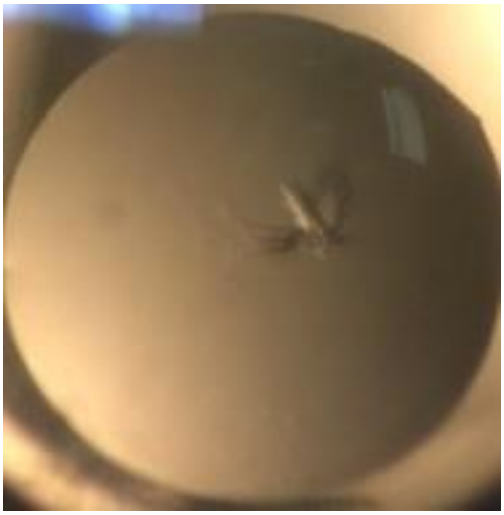

Figure 17 gp74 Crystal Screen Hit-Classic Suite

Figure 17. gp74 Crystal Screen Hit-Classic Suite. 0.2 M calcium chloride dihydrate, 0.1 M Sodium acetate pH 4.6, 20% (v/v) isopropanol, gp74 (9.8 mg/mL).

3.7.3 gp74 Crystal Screen Hit-Index

Figure 18A gp74 Crystal Screen Hit-Index. 0.2M calcium chloride dihydrate, 0.1 M Bis-Tris pH 5.5, 45% (v/v) MPD (Buffer only).

Figure 18B gp74 Crystal Screen Hit-Index. 0.2M calcium chloride dihydrate, 0.1 M Bis-Tris pH 5.5, 45% (v/v) MPD, gp74 (4.9 mg/mL).

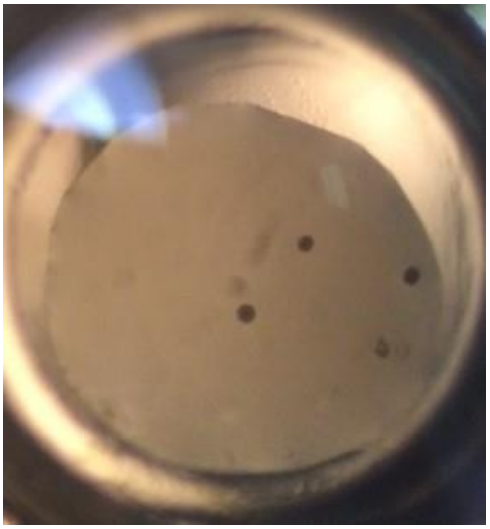

Figure 18C gp74 Crystal Screen Hit-Index. 0.2M calcium chloride dihydrate, 0.1 M Bis-Tris pH 5.5, 45% (v/v) MPD, gp74 (9.8 mg/mL).

Figure 18 gp74 Crystal Screen Hit-Index

Table 1 ORF22 and gp74 Expression Conditions

Protein	Medium	Growth Temperature (°C)	OD_{600nm} at Induction	Expression Temperature (°C)	Induction Time (h)	[IPTG] (mM)	yield per liter of culture (mg)	Purity (%)
ORF22	LB	37	0.7	20	18	0.5	4.6	99
gp74	LB	37	0.8	20	18	0.5	0.6	>95

Table 2 ORF22 and gp74 Activity Assay Conditions

Protein	Reaction Volume (uL)	[Protein] (ug/uL)	[Lambda DNA] (ug/uL)	$[Ni^{2+}]$ (mM)	[NaCl] (mM)	% Glycerol	Temperature (° C)
ORF22	150	10	10	0.5	368	0	23
gp74	150	10	10	0.5	368	0	23
ORF22	150	10	10	0.5	368	20	-20

Table 3 gp74 Crystallization Screening Conditions

Salt	Buffer	Precipitant
0.2 M calcium chloride dihydrate	0.1 M bis-tris pH 5.5	45% (v/v) MPD
0.2 M calcium chloride dihydrate	0.1 M sodium acetate pH 4.6	20 % (v/v) isopropanol

4 DISCUSSION

Previous data from Kala et al. suggested that gp74 is a cofactor in the cleavage activity of terminase enzyme during DNA packaging into the phage head.[4] The increased intensity of the DNA bands as analyzed by agarose gel electrophoresis suggested that gp74 enhances the activity of the terminase enzyme during DNA packaging of the phage head. Since the DNA bands were distinct and no additional DNA bands were observed with gp74 alone, this suggested that DNA cleavage was site specific by the terminase

enzyme and that gp74 did not cleave the DNA at the cos site. Data from Moodley et al. showed that gp74 cleaved lambda phage DNA in the presence of Ni^{2+} ions as observed by DNA smearing during analysis by agarose gel electrophoresis.[14] This suggested that cleavage activity is non-specific. In order to address these two opposing hypotheses a protein encoded by the open reading frame 22 (ORF22) in the Psymv2 genome with predicted HNH endonuclease DNA cleavage activity was used to compare to data observed by gp74 DNA cleavage activity.

The ORF22 protein was expressed with 0.5 mM IPTG at 20 °C for 18 h in LB and purified from BL21(DE3) soluble lysate by Ni-NTA affinity chromatography. Purified ORF22 protein was dialyzed into 20 mM HEPES, pH 8.0 and used for activity assays. The ORF22 activity assays resulted in no smearing of DNA at either temperature or salt concentration over a twenty-four hour period when analyzed by agarose gel electrophoresis. Lambda DNA cleavage assays were conducted with gp74 as a control at 23 °C. The results of the gp74 cleavage assays were consistent with the data collected by Moodley et al.[14] Smearing of DNA by gp74 suggested that gp74 was cleaving Lambda DNA non-specifically.[4] One might expect that Lambda DNA would smear in the ORF22 activity assays given the similar genome positions of the ORF22 protein and gp74, however, this was not observed.

The disorder tendency of >0.5 for residues 107-119 of gp74 as returned from the intrinsically unstructured protein online prediction tool suggests that this C-terminal region is unstable and could be causing degradation of gp74 as observed as lower molecular weight bands by SDS-PAGE analysis.

The crystal hits observed for 0.2 M Calcium chloride, 0.1 M Bis-Tris pH 6.5, and 45% MPD were from protein crystals because the under a polarized lens the crystals had a colorful shine, which protein crystals have, but some salt crystals have it as well.

5 CONCLUSIONS

Originally the goal was to understand the structural aspects of ORF22 to allow it to function under extremely cold conditions and to implement those structural aspects into gp74, which functions at 23 °C, in order to allow it to function under colder temperatures and could demonstrate how this technique can be used for other proteins to be used in an industrial setting.

The data and observations changed the set goals to understanding the function of gp74. The expression of ORF22 was challenging in that it was consistently produced in low yields despite optimizing the media, temperature of induction, induction time, induction concentration of IPTG and cell line. Another challenge with ORF22 was that the conditions for the DNA activity assays needed to be optimized in order for clear visualization by agarose gel electrophoretic analysis when subjected to UV light and to prevent protein precipitation during the assay. Due to the identification of a nickel and cobalt resistance gene in *E. coli*, the data from the ORF22 activity assays, which shows no smearing of DNA by agarose gel electrophoretic analysis, and the identical gene location of gp74 and ORF22 in their respective genomes, we hypothesize that the reported *in vitro* activity of gp74 in the presence of Ni^{2+} might not be essential for DNA packaging. Since the DNA did not smear in the ORF22 assay, but the assay still contained Ni^{2+}, then it seems possible that a conformational change occurred for gp74

that could explain why DNA smearing was observed. The required endonuclease activity for specific DNA cleavage during packaging is provided by the terminase, but gp74 is a cofactor associated with the DNA packaging machinery. In one possible experiment to further analyze how gp74 contributes to the terminase activity, we have initiated structural studies of gp74 by screening crystallization conditions. The structural analysis will provide insights on the mechanism of gp74 regulation of terminase activity. Understanding that gp74 enzyme activity might not be essential for DNA packaging, but that the protein still be important would allow us to develop inhibitors for the terminase-gp74 complex and could be used to prevent phage infection of *E. coli* during protein expression.

REFERENCES

1. Feiner, R. et al. *Nature Reviews* (2015) Vol. 13: 641-650.
2. Black, LW. and Silverman, DJ. *Journal of Virology* (1978) Vol. 28, No. 2: 643-655.
3. Fujisawa, H. and Morita, M. *Genes to Cells* (1997) Vol. 2, 537-545.
4. Kala et al. *PNAS* (2014) Vol. 111, No. 16: 6022-6027.
5. Gyurcsik, B. et al. *Protein Expression and Purification*. (2013) 89: 210-218.
6. Pommer AJ. et al. *JMB* (1998) 334 (Pt 2):387–392.
7. Shen, BW. et al. *JMB* (2004) 342:43-56.
8. Cheng, YS. et al. *JMB* (2002) 324: 227-236.
9. Keeble, AH. et al. *HNH Endonucleases. Nucleic Acids and Molecular Biology.* (2005) Vol. 16: 49-65.
10. Reed, CJ. et al. *Archaea.* Vol. 2013: 1-14.
11. Feller, G. *JPCM*. (2010) 22: 1-17.
12. Saunders, NFW. et al. *Genome Research.* (2003) 13:1580-1588.
13. Georlette, D. et al. *JBC.* (2003) Vol. 278, No. 39: 37015-37023.
14. Moodley et al. *Protein Science.* (2012) Vol.21:809-818.
15. Rodrigue, A. et al. *J. Bacteriol.*(2005): 2912-2916.
16. Campbell and Reece (2005). *Biology*. San Francisco: Pearson. pp. 338–339.
17. J.W. Gibbs. *Transactions of the Connecticut Academy of Arts and Sciences.* (1873), pp. 382-404.
18. Dosztáyani, Z. et al. *Bioinformatics* (2005) 21: 3433-3434.
19. Stoddard, B.L. *Q Rev Biophys.* (2005) 38(1):49-95. Epub 2005 Dec 9.
20. Bohr, N. *Zeitschrift für Physik.* (1923) 13(1): 117-165.
21. Scerri, E.R. *Journal of Chemical Education.* (1998) 75(11): 1384-1385.