Biochemical Investigation of the Bacteriophage Protein HK97 gp74

by

Serisha Moodley

A thesis submitted in conformity with the requirements
for the degree of Master of Science
Graduate Department of Chemistry
University of Toronto

© Copyright by Serisha Moodley 2010
Biochemical Investigation of the Bacteriophage Protein HK97 gp74

Serisha Moodley
Master of Science
Graduate Department of Chemistry
University of Toronto
2010

Abstract

Bacteriophages are viruses that infect and propagate within bacteria by making use of the host’s biosynthetic machinery. With a global population of $10^{31}$, phages pose a significant influence on microbial populations. Studies of bacteriophage proteins can elucidate the influence that bacteriophages play on the evolution of bacteria, as well as, providing the basis for the use of phage proteins as possible therapeutics and bioengineering solutions.

This study aims to investigate the structural and functional role of the HK97 phage protein gp74. Sequence alignments indicate that gp74 is related to homing HNH endonucleases. Homing endonucleases are predominantly double-stranded DNases, suggesting that gp74 mediates integration of phage genes into the host genome or may target foreign phage DNA. DNA digestion experiments with gp74 reveals that gp74 mediates non-specific double-stranded cleavage of lambda phage DNA and single strand cleavage of plasmid DNA. Our initial work demonstrates that HK97 gp74 is an HNH endonuclease.
Acknowledgements

I owe my deepest gratitude to my family, who without their support this thesis would not have been possible. For the innumerable trips to and from campus and their saintly patience when I seemed to experience temporary insanity, I am deeply touched by the love they have shown me throughout this past year. To my mom and dad, thank you for your encouragement and for giving me the confidence to continue my studies. To my brother, thank you for always making me smile and for your continuing support. And to my Bruno for always being a source of love and silliness.

It has been an honour to study under the supervision of Prof. Voula Kanelis and whose encouragement, guidance and support from the initial to the final level enabled me to develop an understanding of biochemical research principles. I would like to thank Dr. Svetlana Tzvetkova for being a remarkable mentor and friend in my undergraduate research education. It is also a pleasure to thank my lab colleagues, Naila Ahmed, Lynn Ikeda, Leen Ghozlani, Elvin DeAraujo and Dennis Guo for their help and friendship.

I am also grateful to our collaborators, Prof. Alan Davidson and Karen Maxwell for their counsel and support. And for the use of their lab equipment and guidance, I would like to thank Prof. Scott Prosser, Prof. Peter Macdonald, Prof. Patrick Gunning, Prof. George Espie and Prof. Jumi Shin.
# Table of Contents

1. **Introduction**
   - 1.1 What are bacteriophages 1
   - 1.2 Bacteriophage Impact on Bacterial Species 8
   - 1.3 Bacteriophage HK97 11
   - 1.4 Homing Endonucleases 15
   - 1.5 Biophysical Tools
     - 1.5.1 NMR Spectroscopy 25
     - 1.5.2 Circular Dichroism Spectroscopy 27

2. **Materials & Methods**
   - 2.1 Structure Based Sequence Alignment 29
   - 2.2 Expression of HK97 gp74 30
   - 2.3 $^{15}$N labeled HK97 gp74 Protein Expression 32
   - 2.4 Imobilized Metal Affinity Chromatography Purification of HK97 gp74 33
   - 2.5 Expression & Purification of TEV Protease 34
   - 2.6 Removal of the 6xHis tag from HK97 gp74 35
   - 2.7 Size Exclusion Chromatography of HK97 gp74 36
   - 2.8 Determination of HK97 gp74 Protein Concentration 36
   - 2.9 Preparation of the NMR Sample 39
   - 2.10 Biophysical Tools
     - 2.10.1 NMR Studies: $^{15}$N-$^1$H correlation spectrum (HSQC) 39
     - 2.10.2 Structural Characterization of HK97 gp74 by Circular Dichroism 40
2.11 Substrates for DNA Cleavage

2.11.1 Plasmid DNA 41

2.11.2 Phage DNA 41

2.11.3 Purification of Single Stranded DNA 42

2.12 Tests for Endonuclease Activity

2.12.1 DNA Cleavage Assay 43

2.12.2 Reporter Methods for DNA Cleavage Assay

a) Absorbance Assay 45

b) Agarose Gel Electrophoresis 46

2.13 Reporter Methods for DNA Binding

2.13.1 UV-Vis Metal Binding Assay 47

2.13.2 Metal Binding Experiments by NMR Titration of HK97 gp74 48

2.13.3 Tryptophan Fluorescence Spectroscopy 48

2.14 Mutant Primer and QuikChange Mutagenesis 49

3. Results

3.1 BlastP Search of HK97 gp74 Protein Sequence 52

3.2 Structure Based Sequence Alignment 53

3.3 Vector Map of HK97 gp74-pET-15b 54

3.4 Expression & Purification of 6xHis-HK97 gp74

3.4.1 Ni$^{2+}$ Affinity Chromatography Purification of 6xHis-HK97 gp74 56

3.4.2 Analysis of TEV Protease Cleavage of 6xHis-HK97 gp74 59

3.5 Size Exclusion Chromatography of HK97 gp74 60
3.6 Analysis of Endonuclease Activity

3.6.1 HK97 gp74-Mediated Digestion of Plasmid DNA 62
3.6.2 HK97 gp74-Mediated Digestion of Phage DNA 68
3.6.3 pH Dependence of HK97 gp74 Activity 71
3.6.4 Metal-Ion Stoichiometry & HK97 gp74 Activity 74

3.7 Structural Characterization of HK97 gp74 by NMR Spectroscopy 77

3.8 Analysis of HK97 gp74 Metal Binding

3.8.1 Metal Binding Titration Experiments using NMR Spectroscopy 80
3.8.2 Metal Binding Studies using Tryptophan Fluorescence 81

4. Discussion and Conclusions

4.1 HK97 gp74 84
4.2 Role of HK97 gp74 HNH Endonucleases 85
4.3 Future Biochemical & Biophysical Studies 87
4.4 Future Applications of HK97 gp74 93
4.5 Conclusions 95

5. References 97

6. Appendix 1

6.1 Sample Calculations of Amino Acid Analysis 101
List of Figures

Figure 1: Schematic drawing of the lytic and lysogenic life cycles of bacteriophages 3
Figure 2: Schematic diagram of *Caudovirales* bacteriophage morphology 4
Figure 3: Schematic drawing of bacteriophage structure 6
Figure 4: Genome map showing the organization of the bacteriophage λ genome 7
Figure 5: Genome map of bacteriophage HK97 14
Figure 6: Representative diagram of the homing mechanism 16
Figure 7: Structures of homing endonucleases 19
Figure 8: Ribbon diagram showing the structure of the I-HmuI-DNA complex 20
Figure 9: Sequence and ribbon diagram of the colicin E9 HNH motif 22
Figure 10: Diagram of the DNA cleavage mechanism of I-HmuI 24
Figure 11: Structure-based sequence alignment of HK97 gp74 with colicin E9 54
Figure 12: Vector map of pET-15b-HK97 gp74 55
Figure 13: SDS-PAGE gel of 6xHis-HK97 gp74 Ni\(^{2+}\) affinity protein purification 57
Figure 14: SDS-PAGE gel of 6xHis-HK97 gp74 protein elution samples 58
Figure 15: SDS-PAGE gel of 6xHis-HK97 gp74 cleaved with TEV protease 59
Figure 16: Size exclusion chromatography of HK97 gp74 61
Figure 17: SDS-PAGE gel of size exclusion chromatography purified HK97 gp74 62
Figure 18: HK97 gp74 cleaves pUC-18 plasmid DNA 64
Figure 19: SDS-PAGE gel of HK97 gp74 in reducing and non-reducing conditions 65
Figure 20: HK97 gp74 cleaves pBluescript (SK+) plasmid DNA 67
Figure 21: HK97 gp74 cleaves λ phage DNA 69
Figure 22: HK97 gp74 cleavage of λ phage DNA with Zn\(^{2+}\) 70
Figure 23: pH Dependence of HK97 gp74 cleavage of λ phage DNA

Figure 24: CD spectra of HK97 gp74

Figure 25: Stoichiometric assay of HK97 gp74 cleavage of λ phage DNA

Figure 26: UV-Vis assay of HK97 gp74 cleavage of λ phage DNA

Figure 27: Agarose gel of UV-Vis assay of HK97 gp74 cleavage of λ phage DNA

Figure 28: $^{15}$N-$^1$H correlation spectrum of HK97 gp74

Figure 29: $^{15}$N-$^1$H correlation spectra of apo-HK97 gp74 and Zn$^{2+}$ bound HK97 gp74

Figure 30: Tryptophan fluorescence spectra of metal-bound HK97 gp74

Figure 31: Agarose gel of pET-15b-H82A-HK97 gp74 PCR reactions
List of Tables

Table 1: Summary of the data obtained from amino acid analysis 101
List of Appendices

6.1 Sample Calculations of Amino Acid Analysis 101
### List of Abbreviations

<table>
<thead>
<tr>
<th>Abbreviation</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>DNA</td>
<td>deoxyribonucleic acid</td>
</tr>
<tr>
<td>RNA</td>
<td>ribonucleic acid</td>
</tr>
<tr>
<td>HK97</td>
<td>Hong Kong 97th strain</td>
</tr>
<tr>
<td>Gp74</td>
<td>gene product 74</td>
</tr>
<tr>
<td>ROS</td>
<td>reactive oxygen species</td>
</tr>
<tr>
<td>BLAST</td>
<td>basic local alignment search tool</td>
</tr>
<tr>
<td>ORF</td>
<td>open reading frame</td>
</tr>
<tr>
<td>hnRNA</td>
<td>heteronuclear ribonucleic acid</td>
</tr>
<tr>
<td>mRNA</td>
<td>messenger ribonucleic acid</td>
</tr>
<tr>
<td>NMR</td>
<td>nuclear magnetic resonance</td>
</tr>
<tr>
<td>PDB</td>
<td>protein data bank</td>
</tr>
<tr>
<td>TEV</td>
<td>tobacco etch virus</td>
</tr>
<tr>
<td>LB</td>
<td>Luria-Bertani</td>
</tr>
<tr>
<td>OD</td>
<td>optical density</td>
</tr>
<tr>
<td>PMSF</td>
<td>phenylmethanesulfonylefluoride</td>
</tr>
<tr>
<td>IMAC</td>
<td>immobilized metal affinity chromatography</td>
</tr>
<tr>
<td>BSA</td>
<td>bovine serum albumin</td>
</tr>
<tr>
<td>DSS</td>
<td>2,2-dimethyl-2-silapentane-5-sulfonic acid</td>
</tr>
<tr>
<td>HSQC</td>
<td>heteronuclear single quantum coherence</td>
</tr>
<tr>
<td>DTT</td>
<td>dithiothreitol</td>
</tr>
<tr>
<td>HEPES</td>
<td>4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid</td>
</tr>
<tr>
<td>TE</td>
<td>tris-EDTA</td>
</tr>
<tr>
<td>Abbreviation</td>
<td>Description</td>
</tr>
<tr>
<td>--------------</td>
<td>--------------------------------------------------</td>
</tr>
<tr>
<td>EDTA</td>
<td>ethylenediaminetetraacetic acid</td>
</tr>
<tr>
<td>CD</td>
<td>circular dichroism</td>
</tr>
<tr>
<td>dNTP</td>
<td>deoxyribonucleotide triphosphate</td>
</tr>
<tr>
<td>PCR</td>
<td>polymerase chain reaction</td>
</tr>
<tr>
<td>MCS</td>
<td>multiple cloning site</td>
</tr>
<tr>
<td>SDS-PAGE</td>
<td>sodium dodecyl sulfate polyacrylamide gel electrophoresis</td>
</tr>
<tr>
<td>UV</td>
<td>ultraviolet</td>
</tr>
<tr>
<td>SEC</td>
<td>size exclusion chromatography</td>
</tr>
<tr>
<td>ppm</td>
<td>parts per million</td>
</tr>
<tr>
<td>DMD</td>
<td>Duchenne muscular dystrophy</td>
</tr>
<tr>
<td>DCA</td>
<td>deoxycholic acid</td>
</tr>
</tbody>
</table>
1. Introduction

1.1 What are Bacteriophages?

Bacteriophages are true viruses and were first observed in 1915 by the English microbiologist, F.W. Twort and Canadian microbiologist F. d’Herelle. Bacteriophages were later defined by Adams as autonomous, obligate, intracellular parasites that infect, grow and multiply within bacteria by making use of some or all of the host’s biosynthetic machinery. Bacteriophages are a diverse group of organisms that significantly influence bacterial ecology. Like other organisms, bacteria are susceptible to infection by an assortment of viruses or virus-like particles. Research in the field of microbial viruses was driven by early interest in the potential use of bacteriophages in treating bacterial diseases. Phage therapy was later abandoned after the discovery of antibiotics. However, research into bacteriophages has increased significantly since the 1940s when M. Delbruck and colleagues demonstrated that bacteriophages reproduce in a “one step” growth mechanism in contrast to the exponential growth of cellular organisms like bacteria. One step growth is described as a growth pattern where virions within a host undergo a “burst” or sudden growth period, in which viruses are released, followed by a latent period in which no viruses are reproduced or released. More recently, interest in bacteriophage research has been rekindled due to the emergence of antibiotic resistant strains of bacteria and the search for alternative means to treat human diseases.

The basic life cycle of a phage follows several common steps; adsorption, separation of nucleic acids from the protein coat, expression of genes and replication of the phage genome, virion assembly and release and transmission of phage progeny (Figure 1).
Bacteriophages undergo two different replicative cycles. In the infective or lytic cycle, the genome or nucleic acid component of the bacteriophage is injected into the host bacterial cell.\textsuperscript{5,6} Subsequently, the host cell machinery transcribes and translates phage genes, leading to the production of whole phage particles. Whole phage particles accumulate within the host cell and release virus-encoded lytic enzymes, such as lysins and holins, that cause host bacterial cell lysis.\textsuperscript{1} In the lysogenic cycle, the injected phage genome is incorporated into the bacterial cell genome.\textsuperscript{7,8} The phage genes can remain dormant until induced by a response or they can be transcribed along with bacterial genes.\textsuperscript{1} In fact, many genes identified through sequencing of bacterial genomes were originally phage genes.\textsuperscript{9} Phage genes incorporated into the bacterial genome can then be transcribed and translated to produce phage particles, which are released out of the host cell by lysis. Bacteriophages that undergo only the lytic life cycle are referred to as virulent, whereas, phages that undergo both of these life cycles or just the lysogenic life cycle are referred to as temperate.\textsuperscript{1}
Figure 1: Schematic drawing of the lytic and lysogenic life cycles of bacteriophages. Lysogenic life cycles allow the integration of bacteriophage genomes into host bacterial genomes, thus, forming a lysogen or prophage. The prophage is replicated along with the bacterium’s genome during each cell division and remains integrated until a lytic signal initiates the dissimilation of the phage genome. The independent phage genome forms a circular molecule that can be replicated and transcribed. The lytic life cycle allows for the prophage expression, replication and production of phage progeny, which are released via the lysis of the host bacterium.

Phages are much smaller in comparison to the bacteria they infect and are usually between twenty and two hundred nanometers in size. The most common bacteriophages belong to the Caudovirales order and comprise approximately 96% of the entire population of bacteriophages. Also known as tailed viruses, the Caudovirales order comprises three families of bacteriophages, including the myoviridae (bacteriophages with contractile tails), siphoviridae (bacteriophages characterized with long, non-contractile tails) and the podoviridae (or bacteriophages with short, non-contractile tails) (Figure 2).
Figure 2: Schematic diagram of *Caudovirales* bacteriophage morphology.\textsuperscript{10} Approximately 96\% of all bacteriophages belong to the double-stranded DNA (dsDNA) tailed phages, known as the *Caudovirales*.\textsuperscript{9} *Caudovirales* is an order of viruses that are characterized by dsDNA genomes and an icosahedral head.\textsuperscript{2} The *Caudovirales* order is comprised of three families of bacteriophages, including the myoviridae, siphoviridae and the podoviridae.\textsuperscript{10}

It has been observed that a difference in tail morphology between families is indicative of major differences in the viral genome.\textsuperscript{1} Tail morphology also has additional effects on the mode of infection (or mechanism of DNA injection during infection) and in virion assembly and maturation.\textsuperscript{11} For example, long tails are first completely assembled and then grafted on to a completed head, whereas, short tails are sequentially assembled onto completed heads.\textsuperscript{1} Many bacteriophages but not all phages possess tails that are attached to the phage head. *Caudovirales* bacteriophages are known for having a hollow, helix-shaped protein tail that is used to adsorb onto the host cell membrane and to inject DNA into the host cell. The size and length of tail often varies between bacteriophage species. However, a recent study illustrates the exact length of the tail of bacteriophage T7, which is confirmed to
be approximately 40-55 nm long and 8-11 nm in diameter with an axial hole 3-4 nm in diameter. In some phages, tails function to recognize and bind to the host bacterial cell and to inject the phage genome into the bacterial host cell. Some phages of this order also possess a protein base plate with tail fibers that allow for the attachment of phage onto the surface of bacteria. Another characteristic structure of Caudovirales type phages is the icosahedral-shaped head that is known as the capsid, which is composed of individual monomeric subunits of protein that form a shell (Figure 3). The capsid contains the nucleic acid molecule of the phage and proteins essential in the invasion and lysis of bacterial cells. Thus, phage genes not only encode the structural components of the phage but also encode for infectious components. For example, phages encode proteins essential for invasion, such as proteases and nucleases, and proteins for replication of the phage genome within bacterial hosts, such as unique viral polymerases, which are contained within the capsid. The viral capsid serves three primary purposes; to protect the viral genome, to detect host cells for infection and in some instances, to initiate infection by binding to the target cell’s membrane. The capsid head and tail of these bacteriophages are connected by a small structure, known as the collar (Figure 3).
Figure 3: Schematic drawing of bacteriophage structure. Bacteriophages of the *Caudovirales* order consist of a head, collar, sheath, tail fibers and base plate. These proteins constitute the structural features of bacteriophages. One such bacteriophage of the *Caudovirales* order is HK97. HK97 also belongs to the Siphoviridae class of viruses, which follow the same basic structural arrangement.

Although bacteriophages are a diverse group of micro-organisms, they all include a nucleic acid or genomic component, such as DNA or RNA contained within the capsid of the bacteriophage. Different types of phages possess nucleic acid molecules in either double-stranded or single-stranded forms. For example, Caudovirales type phages contain a single molecule of linear double stranded DNA that is a arranged as a circular form (Figure 4). The nucleic acid component of phages often include modified bases, which protect the bacteriophage genome from its own nucleases that function to degrade host bacterial nucleic acids during phage infection. Phage genomes are also incredibly compact and range in size from 20,000 base pairs to 500,000 base pairs and code for approximately
3-5 average-sized gene products (approximately 150-500 amino acid residues in length) in simple phages and approximately 100 gene products in more complex phages.\textsuperscript{5,16}

![Bacteriophage λ Genome](image)

Figure 4: Genome map showing the organization of the bacteriophage λ genome.\textsuperscript{7} The lambda phage genome illustrates the characteristic structure of the circular form of double-stranded DNA found in \textit{Caudovirales} type phages.\textsuperscript{2} Genes associated with proteins that share a structural or functional relation tend to be clustered together.\textsuperscript{10} For example, the genes that encode for the capsid proteins are located in one area of the genome and similarly genes associated with DNA replication are clustered together in a distinct location of the genome.

A common feature amongst \textit{Caudovirales} genomes is the association of genes that encode for proteins that interact with each other (Figure 4).\textsuperscript{10} Consequently, the genes that encode for the proteins that constitute the structural features of the bacteriophage, such as the
capsid, neck and tail or proteins that interact with each other to accomplish specific cellular functions, such as DNA replication, tend to be clustered together. For example, the capsomere protein, which comprises the identical morphological subunits of the protein coat or capsid is located in the same region of the genome as the scaffolding protein (required for DNA packaging) and the portal protein (which forms a hole that allows the passage of DNA during packaging and forms the junction between the phage head and tail protein to allow for DNA ejection).\(^{14}\) In some instances, an entire set of structural genes are grouped and transcribed together from a single promoter.

### 1.2 Bacteriophage Impact on Bacterial Species

As the most abundant and most rapidly reproducing biologically active organism on Earth, with a global population of \(10^{31}\), bacteriophages pose a significant influence on microbial populations.\(^{17}\) These small organisms impact the species distribution, nutrient cycling, food network and population density of bacteria. Phage ecology is the study of the interaction of bacteriophages with other organisms and the environment. Phage community ecology has allowed for the observation and characterization of the interactions between phages and bacteria to determine predator-prey interactions and to understand the co-evolution of bacteriophages and bacteria.\(^{18}\) Some of the more typical relationships between bacteria and phages include mutualism (where bacteria harbor phages via lysogenic conversion and these phages may in turn kill related bacterial competitors) or parasitization (or predation of particular species) of bacteria.
Bacteriophages are highly diverse and have been found in all ecosystems colonized by bacterial populations, such as aquatic, soil, and gastrointestinal locales. Phage population density generally increases with the productivity of the ecosystem. For example, in marine environments, phage population is highest in coastal areas, due to the complexity and abundance of the natural flora and fauna, and lowest in the deep sea. Bacteriophage abundance is also higher in fresh water systems as compared to marine systems, and in shallow depths, as in lake ecosystems, as compared to deeper environments. However, this is also dependent on oxygen availability as lower depths are more likely to have anoxic conditions. Bacteriophages tolerate a vast range of environmental conditions and resources, including temperature, hydrostatic pressure, radiation, oxygen, pH and host availability. For example, temperature is an important environmental factor for phage survival and most phages have been found to function at temperatures between 15 °C and 42 °C.

Bacteriophages also affect the microenvironment of bacteria, where phages use their hosts as a source of energy and matter. Phages seize control of bacterial biosynthetic processes in order to synthesize viral macromolecules that are used for the formation of viral particles. Thus, bacterial hosts serve as the bioreactor or factory for the production of phages.

Many phages develop a symbiotic interaction with their bacterial hosts via lysogeny that involves the integration of the phage genome into the host’s replicon. These phages can exist as prophages, which are a form of the phage genome that can be inserted and replicated with the host bacterial genome. These silent infections can then be inherited by daughter cells and induced by an environmental or stress response to activate biosynthesis of phage particles and cause host lysis. Bacterial genomes have been shown to consist of 3 % - 10 % of prophage-encoded genes and often carry on average three prophages.
Baltic Sea bacterial isolates, the overall frequency of prophage particles was 28% within bacterial genomes.\textsuperscript{18} It has been suggested that these prophage inserted genes are the major contributors to genomic diversity among bacterial species. In fact many sequenced genes from bacterial strains originally thought to be bacterial are continually being identified as phage genes. For example, the bacterial strain, \textit{Escherichia coli} 0157:H7 has a genome containing 18 prophage elements, which accounts for approximately 16% of its total genome.\textsuperscript{21} The prophage element contains a gene for the production of shiga toxin, which inhibits protein synthesis within target cells by N-glycosidase activity that cleaves ribosomal RNA.\textsuperscript{22} Thus, inclusion of the prophage element bestows an extremely pathogenic effect on the bacteria. It has also been proposed that the presence of prophage genes may confer defensive advantages to bacteria by protecting against infection by other phages and increasing the pathogenicity or virulence of the host organism through the use of toxins, such as pyocins, colicins and anaredoxins.\textsuperscript{23,24} Pyocins are similar to bacteriocins, which are proteinaceous toxins that are produced by bacteria to inhibit the growth of related bacterial strains, and resemble bacteriophage tail like structures.\textsuperscript{25} Colicins are bacteriocins and act as nucleases to degrade DNA and RNA or cause the formation of pores in the cell membrane of other bacteria, thereby lysing the cell.\textsuperscript{26} Anaredoxins function as oxidoreductases to form reactive oxygen species (ROS) that are highly cytotoxic to bacteria.\textsuperscript{27} Prophage genes also supply specific fitness factors, such as proteins that allow for the uptake and use of different nutrients or nutrient biosynthetic pathways that increase the host’s selective advantage in a particular system.\textsuperscript{20} Some prophage genes are able to promote host fitness through the use of five different mechanisms. Prophages can function as transposons and lead to reorganization or replication of genes.\textsuperscript{28} Prophages can interrupt genes causing silencing of non-essential
gene functions or can offer immunity to related phage infections or destroy related phages.\textsuperscript{1} Prophages also offer bactericidal factors and can cause lysis of related nearby strains of bacteria.\textsuperscript{1} Lastly, prophage genes can introduce new fitness factors by conversion or transduction.\textsuperscript{1} Thus, bacteriophages are not just dangerous molecular machines that cause bacterial cell mortality but also function as key components to bacterial survival. Understanding the molecular details by which phages control bacterial survival may help elucidate novel therapeutics against bacterial infections.

1.3 Bacteriophage HK97

The bacteriophage HK97 was originally isolated in Hong Kong and was the ninety-seventh strain characterized amongst a series of related viruses, hence the HK97 nomenclature.\textsuperscript{13,14} HK97 belongs to the \textit{Siphoviridae} class of \textit{Caudovirales} bacteriophages (Figure 2) and is a temperate bacteriophage of \textit{Escherichia coli} bacteria, meaning that it has the ability to display both a lytic and a lysogenic life cycle after infection of \textit{E. coli} cells. A lysogenic life cycle allows bacteriophage HK97 to integrate its genome into the host \textit{E. coli} bacterium’s genome, thus, becoming a lysogen.\textsuperscript{8,9,10} The lytic life cycle permits prophage expression, replication and production of phage progeny, which are released via the lysis of the host bacterium (Figure 1).\textsuperscript{7,8}

HK97 is also a lambdoid phage, meaning that this bacteriophage belongs to a group of closely related phages that are similar in virion morphology and function to that of lambda phage. Lambdoid phages also exclusively infect \textit{Escherichia coli} and are thus, often referred
to as coliphages. Like other phages in its family, HK97 is composed of a capsid, which forms an icosahedral, concatenated chain mail-like structure using repeating units of the protein gp5, and an adaptor that connects the head to a long, non-contractile tail that functions to adsorb onto *E. coli* cell membranes and as the channel through which DNA is ejected from the head into the host bacterium.

HK97 has a genome of 39.7 kB and a total of 61 protein coding genes. Like other bacteriophages, the HK97 genome appears to cluster genes that are related by structure or function. HK97 is a relatively new phage species, and as such some of the proteins have yet to be determined. However, it is hypothesized that HK97 shares many of the same genes as the lambdoid phages. Starting at gene 1 of the HK97 genome, the first set of known genes (from gene 1 to 28) encode for the structural proteins that compose the head, tail and adaptor components of the phage (Figure 5). In the lambdoid phages, genes A-F code for phage head genes and genes J-Z code for phage tail genes. Genes encoding the proteins integrase, excisionase, and recombinase are located following the structural genes. Integrase, excisionase and recombinase are involved in lysogeny, which requires the integration of phage DNA into the host genome and excision, as well as reconstruction of phage genomes from the recombinant DNA of host genomes after induction of the lytic cycle. The next sets of genes involve the transcription of phage genes. For example, the transcription activator protein is encoded by cII, which is located near the gene cIII, which encodes a binding protein that protects the transcription activator protein. Located to the left of the transcription activator are the transcription inhibitor genes, cI and cro. Next, the genes O and P are known to be involved in DNA replication followed by the genes that have been identified as lytic cycle repressor proteins. The subsequent genes are involved in DNA
repair mechanisms. For example, the lambdoid phage RusA homolog also encodes for a resolvase that ligates nucleic acid fragments at Holliday junctions, which are formed by genetic recombination.\textsuperscript{30} The last genes of the HK97 genome are involved in lysis of host bacterial cells and it is known that the S gene encodes a holin protein that creates pores through which the R endolysin protein, which degrades the bacterial peptidoglycan wall, is released.\textsuperscript{31} The very last gene of the HK97 genome encodes the protein gp74. However, the functions of various genes of the HK97 bacterophage, including the gene gp74, have yet to be determined.
Figure 5: Genome map of bacteriophage HK97. Genes are represented by red outlined arrows. The gene size and location are represented by the length and position that the arrow spans and is indicated in kilobase pairs on a ruler below the gene. Known genes are identified as their gene number and letter designation and are named after homologous lambda phage genes. Unknown genes are only identified by their order number in the genome and their putative function or homology is identified in blue type above the genes.

The profound impact of bacteriophages on the ecology and evolution of bacterial species has resulted in incredible interest in the field of phage biology. This study aims to
investigate the structural and possible functional role of the protein gp74 of the bacteriophage HK97. At the beginning of this project there was no known function of the bacteriophage protein HK97 gp74 and as of today there has yet to be a clear identity of the role that this protein plays in the life cycle of HK97 or in the infection of host E. coli bacterial cells. This thesis describes our study into the biochemical function of HK97 gp74.

1.4 Homing Endonucleases

Early in the project we conducted a search of proteins with similar sequences to HK97 gp74 using the basic local alignment search tool (BLAST). The BLAST search indicated that HK97 gp74 is a possible homing HNH endonuclease. Homing is a transfer mechanism by which mobile genetic carriers or intervening sequences are integrated into recipient homologous alleles that lack this gene sequence. Homing processes were first described for group I introns of the budding yeast Saccharomyces cerevisiae, which was found to contain a genetic marker that encoded for an endonuclease in an open reading frame in a 1.1 kB intron. Homing endonucleases initiate transfer of intron and intein elements by generating double-stranded breaks in alleles that are homologous to the endonuclease gene and lack the intron or intein element containing the homing endonuclease gene (Figure 6). Homologous recombination, as a result of cellular repair mechanisms, at the double-stranded DNA break leads to the transfer of the homing endonuclease-containing intron or intein, thereby causing the proliferation of the homing endonuclease gene within the genome. Consequently, homing endonuclease genes are inherited in a non-Mendelian manner and tend to be the dominant allele when incorporated into the genome, resulting in efficient
proliferation of the gene and gene product in prokaryotic and viral genomes. The integration of transferred intervening sequences containing the homing endonuclease gene has been observed between DNA regions and across biological kingdoms. Homing endonucleases found in phage genomes display a greater diversity in homing mechanisms. For example, it has been observed that some endonucleases of phage origin generate single strand nicks in comparison to double-stranded digestion.

Figure 6: Representative diagram of the homing mechanism. Homing is the mechanism by which introns or inteins are integrated into recipient alleles that lack this gene sequence. Homing endonucleases cause double-stranded breaks in alleles that are homologous to the endonuclease gene and lack the intron or intein element containing the homing endonuclease gene. Homologous recombination at the double-stranded DNA break leads to the transfer of the homing endonuclease-containing intron or intein, thereby causing the proliferation of the homing endonuclease gene within the genome.
Homing endonucleases are a large family of proteins that includes several hundred members, found in bacteria, archae, fungi and algae. The majority of these proteins are encoded as open reading frames (ORF) within group I, group II or archael introns or as inteins.\textsuperscript{28} Archaeal introns are believed to be the result of splicing of hnRNAs (heterogeneous nuclear RNA that are precursors of mRNAs).\textsuperscript{2} Inteins are intervening DNA sequences that are spliced and excised as a post-translational process.\textsuperscript{28} Homing endonucleases tend to be rather small proteins, typically less than 30 kD.\textsuperscript{28} However, regardless of size, homing endonucleases recognize long DNA sequences up to approximately forty base pairs. In contrast, most restriction endonucleases recognize sequences of only a few base pairs in length. Homing endonucleases also display an extremely broad range of specificity for DNA substrates, which may be due to their recognition of target sites at lengths of 14 to 40 base pairs.\textsuperscript{35} They are also very tolerant of changes or variations in recognition sequences or target sites. Unlike restriction endonucleases, homing endonucleases do not have highly stringent or defined recognition sequences but do have specific consensus sequences. Studies have demonstrated that homing endonuclease recognition sites are extremely rare. For example, a recognition sequence of approximately 18 base pairs will occur only once per 7 x 10\textsuperscript{10} base pairs in a random DNA nucleotide sequence.\textsuperscript{36} Partial complementary binding to the DNA nucleotide sequence is sufficient for cleavage to occur and a single base change in the recognition sequence will not eliminate cleavage but may reduce the efficiency of the endonuclease reaction, possibly by reducing the binding affinity of the enzyme for its DNA substrate. Thus, homing endonucleases are known to allow some nucleotide sequence degeneracy within their recognition sequence and their observed sequence specificity is on average in the range of ten to twelve base pairs.\textsuperscript{37,38}
There are four main families of homing endonucleases, each of which are named for the conserved residues associated with their nuclease domain, including the LAGLIDADG, the His-Cys box, HNH enzymes and GIY-YIG family of enzymes (Figure 7).\textsuperscript{28} Sequence data indicates that HK97 gp74 belongs to the HNH class of homing endonucleases, which are named for two conserved histidine residues and an invariant asparagine residue. In some HNH endonucleases, such as the bacterial colicins, the first histidine and the asparagine residue are thought to be involved in DNA cleavage while the second histidine is involved in divalent metal ion binding.\textsuperscript{28}
There are four families that comprise the homing endonucleases, which include the LAGLIDADG, the His-Cys box, HNH enzymes and GIY-YIG family of enzymes. This diagram shows the representative structures of enzymes that belong to the four families of the homing endonucleases bound to DNA. The enzymes shown in the diagram are I-CreI (a LAGLIDADG enzyme), I-PpoI (a His-Cys box enzyme), I-HmuI (a HNH enzyme) and I-TevI (a GIY-YIG enzyme).

The representative HNH endonuclease I-HmuI contains a nuclease active site with a protein secondary structure architecture that forms a ββα-metal motif. The I-HmuI enzyme has a highly modular, extended, monomeric structure, where the nuclease domain is located near the N-terminus of the protein and is attached to structural motifs that recognize
DNA substrates for binding (Figure 8).\textsuperscript{28} By structural analysis of I-HmuI, it is believed that the HNH motif or ββα-metal active site binds to and spans approximately 25 base pairs of the minor groove of DNA and contacts the DNA phosphate and 3’ hydroxyl group of the scissile phosphate.\textsuperscript{36} Furthermore, I-HmuI exists as a monomer and binds only a single metal ion, which is coordinated by a conserved asparagine from the HNH motif and an additional aspartate residue. These residues, Asn-96 and Asp-74 and a non-bridging oxygen from the scissile phosphate group of the DNA bind the divalent metal manganese at the I-HmuI HNH endonuclease active site.\textsuperscript{28}

![Figure 8: Ribbon diagram showing the structure of the I-HmuI-DNA complex.](image)

The enzyme I-HmuI (green) has a highly modular, extended, monomeric structure.\textsuperscript{36} The HNH motif spans approximately 25 base pairs of the minor groove of DNA (blue) and contacts the DNA phosphate and 3’ hydroxyl group of the scissile phosphate via a divalent metal coordination centre (red).\textsuperscript{36}
The HNH endonuclease motif of the non-specific bacterial colicins has also been shown to be highly similar to the structure of I-HmuI. The HNH motif is comprised of 2 β-strands, containing the first histidine and asparagine residue, followed by an α-helix that contains the second histidine (Figure 9). In the colicin HNH motif, the first histidine and asparagine are involved in catalysis, and the second histidine is involved in divalent metal ion binding. In colicin E9, the first conserved His-103 interacts with the minor groove of the DNA and makes contact with the DNA backbone during catalysis.\textsuperscript{28} Investigation of colicins have revealed that binding and cleavage of DNA at these HNH domains involve a single, bound divalent cation, such as zinc or magnesium, to coordinate the phosphate and 3’ leaving group of DNA, which are susceptible to cleavage. The residues His-102 and His-127, which is a conserved HNH motif residue, of colicin E9 bind the divalent metal at the HNH endonuclease active site.\textsuperscript{35}

While it has been observed that the position and stabilization of the complexed divalent metal ion at the HNH motif of several related endonucleases is similar, the chemical identities of the metal ions, coordination and interaction at the active site of the divalent metal ions with their metal binding residues differ significantly between HNH endonucleases. The chemical identity of the divalent metal ion that binds to the HNH active site varies considerably and the range of metals include magnesium, zinc, cobalt, manganese, nickel, strontium and calcium.\textsuperscript{39} Coordination of the metal ions also varies between HNH endonucleases and studies have shown that the metal ion can be coordinated by one asparagine and one aspartic acid residue, like the I-HmuI enzyme (Figure 8) or two histidine residues, as in the bacterial E9 colicin (Figure 9).\textsuperscript{28} The differences observed in metal binding may correlate with differences in metal specificity.
HNH endonucleases are named for two conserved His residues and an invariant Asn. In the HNH endonuclease colicin E9, H103, N118, and H127 are the conserved HNH motif residues. Structurally, the HNH motif is comprised of 2 β-strands, containing the first His and Asn residue, followed by an α-helix that contains the second His.

By structural analysis it has been observed that different residues or groups of residues are responsible for the binding and cleavage of DNA. However, studies have also shown that the cleavage of DNA by various HNH endonucleases follows a predicted mechanism. The same steps are required in DNA binding to the HNH active site, formation of transition states and displacement of digested DNA substrates. For the enzyme I-HmuI,
the N-terminus of a β-strand of the HNH motif binds the major groove of the DNA substrate and an α-helix binds the minor groove and lastly, the C-terminal of the HNH motif of I-HmuI binds the opposite end of the DNA at its major groove (Figure 8).\textsuperscript{28,36} The enzyme appears to straddle the phosphate backbone of the DNA substrate twice, once at the cleavage site and again at the 3’ end of the target site.

In the HNH endonuclease cleavage mechanism, either an oxygen atom of an asparagine or backbone amide acts as a nucleophile to remove the proton of a conserved histidine residue.\textsuperscript{40} This results in resonance shift of protons around the aromatic ring of the histidine. In the case of I-HmuI, Asn-83 acts as a nucleophile to attack the conserved His-75 residue (Figure 10).\textsuperscript{28} A water molecule present in the active site is attacked by a nucleophilic nitrogen of the conserved histidine, resulting in formation of a hydroxyl anion. The activated hydroxyl anion reacts with the DNA backbone leading to the formation of a phospho-anion transition state. The DNA cleavage mechanism for HNH endonucleases is also highly dependent upon the binding of a divalent metal ion species at the active or catalytic site.\textsuperscript{40} A single asparagine or an asparagine and an aspartic acid or a group of two histidine residues binds and positions a divalent metal ion, which in turn acts to stabilize the phospho-anion transition state and the leaving group at the 3’ hydroxyl of the DNA molecule. The residues, Asn-96 and Asp-74 stabilize manganese at the I-HmuI active site with a non-bridging oxygen from the scissile phosphate group of the DNA substrate (Figure 10).\textsuperscript{36} The mechanism proposed concludes with the removal of the DNA substrates from the enzyme active site.
Asn-83 acts as a nucleophile to attack the conserved His-75. Nucleophilic attack of a water molecule present in the active site results in formation of a hydroxyl anion. The activated hydroxyl acts as a nucleophile for an in-line displacement, which causes formation of a phospho-anion transition state. Asn-96 and Asp-74 stabilize the divalent metal manganese (M+) at the I-HmuI active site with a non-bridging oxygen from the scissile phosphate group (red box) of the DNA substrate. Lastly, the cleaved DNA substrate is displaced from the enzyme active site.

This thesis encompasses our work to date in understanding the biochemical function of HK97 gp74 and our ongoing efforts to structurally characterize HK97 gp74. This thesis will present data demonstrating that gp74 is indeed a metal-specific endonuclease. Studies as a function of pH indicate that gp74 is a metal-dependent HNH homing endonuclease. DNA digestion experiments using stoichiometric concentrations of metals demonstrate that DNA digestions occur with one equivalent of metal to protein, providing evidence that only one metal is required in the binding site. Further experiments using HK97 gp74 with mutations at the putative HNH motif will confirm the biochemical function of HK97 gp74. These mutants will also help elucidate residues important for metal binding versus catalysis and
may provide a mechanism to study a ternary complex (of protein-metal-DNA) by NMR spectroscopy.

1.5 Biophysical Tools

1.5.1 NMR Spectroscopy

Nuclear magnetic resonance (NMR) spectroscopy is an analytical technique that exploits the behavior of magnetically active atomic nuclei to provide information about the structure, kinetic characteristics, and interactions of molecules.\(^{41}\) NMR spectroscopy has been used to study the structure, dynamics and interactions of proteins in solution at atom-specific resolution.\(^{42}\) NMR resonances have four basic properties: intensity, resonance frequency (chemical shift), splitting and line width.\(^{43}\) The intensity or peak height is measured as a volume and is proportional to the concentrations of nuclei.\(^{43}\) Thus, a less concentrated sample will give a weak signal and increasing the concentration of a sample increases the signal strength. The height of each peak in a frequency domain spectrum is also dependent on the molecular weight and dynamics of a biological compound.\(^{41}\) The chemical shift is the resonant frequency of a nucleus, such as \(^1\)H, \(^{15}\)N, \(^{13}\)C and \(^{31}\)P, which are biologically relevant isotopes that occur in DNA, proteins and lipids.\(^{41}\) The chemical shift is a relative scale measured in parts per million (ppm) that compares all signals in a spectrum to the signal from a calibration or reference compound. Several factors affect chemical shift and include the local electronic environment, the electronegativity of attached groups and the spatial proximity of compounds. Splitting is called spin-spin coupling, which is described by
the coupling constant J (measured in Hz) and occurs as a close group of two or more resonances that correspond to a single nucleus.\textsuperscript{43} Splitting is due to the interaction of spins through chemical bonds and is caused by the induction of magnetic fields from the interaction of the spins of nuclei and bonding electrons.\textsuperscript{43} In an HSQC experiment, coupling of $^1\text{H}$ and $^{15}\text{N}$ is eliminated during the experiment. The line width (at half height) is related to the T$_2$ relaxation time.\textsuperscript{41} T$_2$ relaxation time is dependent on molecular weight and on the motions of a molecule within an applied field.\textsuperscript{41} For example, a small molecule that tumbles quickly has a long T$_2$ and gives a narrow, sharp peak, whereas, larger molecules tumble slowly and have short T$_2$ relaxation times that give broad peaks.

One of the more common types of NMR experiments is the $^{15}\text{N}$-$^1\text{H}$ heteronuclear single quantum coherence (HSQC) correlation experiment. A HSQC experiment correlates the resonant frequency of the amide proton and the resonant frequency of the directly attached amide nitrogen.\textsuperscript{41} HSQC experiments give information about the conformation of a protein. Proteins in an unfolded conformation are comprised of residues that are at a distance from each other and are exposed to the solvent. These residues experience the same chemical and electronic environment. In a $^{15}\text{N}$-$^1\text{H}$ HSQC, this results in resonances with very similar $^1\text{H}$ chemical shifts, where the $^{15}\text{N}$ chemical shift is more dependent on amino acid type. The HSQC can also give insights into the dynamic behavior of molecules. Differential peak heights, as seen for HK97 gp74 indicate differential dynamics in different regions of the protein (see section 3.7). A future goal for this project is to elucidate the NMR solution structure of HK97 gp74, and also use NMR titration experiments to obtain information on metal binding.
1.5.2 Circular Dichroism Spectroscopy

Circular dichroism is an analytical and spectroscopic technique that rapidly determines the secondary structure, folding characteristics and binding properties of proteins. CD arises from the differential absorption of left-handed and right-handed circularly polarized light by chiral molecules. As biological molecules, DNA and protein are ideal candidates for CD spectroscopy because they exhibit both dextrorotary and levorotary components and are optically active. A far-UV CD spectrum can illustrate important secondary structure characteristics, such as the extent of α-helix, β-sheet, β-turn or random coil conformation. A near-UV CD spectrum, resulting from the absorption of aromatic residues (phenylalanines, tyrosines and tryptophans) and cysteine disulfide bridges, provides information about the tertiary structure of proteins. CD is an efficient analytical tool that requires smaller concentrations of protein, unlike NMR and X-ray crystallography, and can be used in conjunction with a large range of solvent conditions, varying temperature, pH and various cofactors and ligands.

Future applications of the techniques mentioned and the resultant data may lead to the development of phage-based approaches to combat bacterial diseases. Understanding the function of phage proteins, such as HK97 gp74, will also help elucidate the molecular basis of the variability in bacterial populations, which is critical for fighting bacterial infections and diseases. Moreover, better understanding of HK97 gp74 can enable its use as a possible homologous recombination system to allow for the integration of genes that, for example, recover deletion mutations that are involved in the development of diseases, such as
Duchenne muscular dystrophy. Homing endonucleases, thus, offer a new perspective on gene therapy for many diseases.
2. Materials and Methods

2.1 Structure Based Sequence Alignment

To gain insights into the function of HK97 gp74, the protein sequence of HK97 gp74 was subjected to the basic local alignment search tool (BLAST). The BLAST program utilizes a sequence comparison algorithm that is optimized for speed to search different databases for the optimal sequence alignments to a specific search query, such as a protein or DNA sequence. The full amino acid sequence of HK97 gp74 was used in the search. The BLAST program analyses the resultant output data for the best possible matches to the query and ranks related protein matches based on two variables, the bit score and E-value. A list of highly similar sequences are compiled by the program, along with a diagram indicating the most closely related family of proteins as determined by the conserved amino acid residues and sequences present in the protein compared to similar proteins. One of the top protein BLAST matches to HK97 gp74 is colicin E9.

Based on our BLAST search results, where one of the top search hits was the HNH endonuclease colicin E9, a structure-based sequence alignment of HK97 gp74 and colicin E9 was made. A structure-based sequence alignment of HK97 gp74 and colicin E9 was performed using ClustalW. ClustalW is a multiple sequence alignment program that is used to compare the relatedness or conservation between protein sequences. A structure-based sequence alignment uses the structural information of a protein to construct a better alignment, in which the gaps are located in loops and not in the secondary structural elements, such as α-helices and β-sheets. To align the sequences based on the secondary structural characteristics of colicin E9, the secondary structure information file was used.
from the coordinates in Protein Data Bank (PDB). Some manual alignment was necessary. The final alignment is shown in Figure 11.

2.2 Expression of HK97 gp74

A sample of pET-15b plasmid vector containing the gene coding for HK97 gp74 (pET-15b-HK97 gp74) was obtained from the laboratory of Karen Maxwell, Structural Genomics Consortium. The plasmid, pET-15b-HK97 gp74 encodes gp74 as a fusion protein with a six histidine residue tag (6xHis-tag). A recognition site specific for the tobacco etch virus (TEV) protease is located between the 6xHis-tag and HK97 gp74 sequence. The TEV site aids in removal of the 6xHis-tag. The plasmid has an ampicillin resistance marker.

*E. coli* BL21 Star (DE3) cells were transformed with the pET-15b-HK97 gp74 plasmid and grown on LB agar plates containing 100 μg/ml ampicillin, overnight at 37°C. A negative control was included for all transformations that consisted of plating cells without DNA on LB agar plates and for the liquid cultures that consisted of 5 ml of LB media, both of which contained 100 μg/ml ampicillin. The next day, a single colony was used to inoculate a 5 ml LB culture containing ampicillin (100 μg/ml) and grown at 37 °C until an OD_{600} of 0.6. Approximately 50 μl of this culture was used to inoculate a 100 ml LB culture, with ampicillin, which was incubated at a temperature of 25 °C overnight (approximately 17 hours) with very slow shaking (100 rpm for the overnight compared with 250 rpm for the day culture). The optical density at 600 nm (OD_{600}) of the overnight cultures was 0.899. The cultures were then centrifuged at 3,000 rpm (1,811xg) at 21°C for 20 minutes. The
100 ml overnight cultures were grown slowly to allow the cells to grow through log phase but to not reach stationary phase, which often leads to poor protein expression, likely because cells have lost the plasmid. Ampicillin was used as a selective agent during growth, since the plasmid vector contains an ampicillin resistant gene. Ampicillin is hydrolyzed by the enzyme β-lactamase, which is produced by the bacteria carrying the plasmids with the ampicillin resistance gene. In bacterial cells that do not contain the plasmid vector, no β-lactamase is produced and ampicillin acts as a competitive inhibitor to the bacterial protein, transpeptidase. Transpeptidase is a bacterial enzyme that forms cross links in peptidoglycan chains in order to form rigid cell walls. The antibiotic ampicillin binds to the transpeptidase enzyme by forming a stable complex and inhibits the transpeptidase enzyme activity. Inhibition of transpeptidase leads to lysis of the bacterial cell. β-lactamase produced by the transformed cells is secreted into the growth medium. Therefore, the elevated levels of β-lactamase in solution can hydrolyze most of the ampicillin in the culture medium, thereby, removing the selective pressure and resulting in the proliferation of cells that do not have the plasmid of interest. This can then also result in low levels of protein expression. Consequently, overnight cultures were centrifuged to remove secreted β-lactamase and the pellet was resuspended in fresh growth media with fresh ampicillin. The 1 L cultures were incubated at a temperature of 37 °C in a shaking incubator (250 rpm) until an OD₆₀₀ nm of 0.6 (mid-log phase) was reached. Protein synthesis was induced with 1 mM of isopropyl β-D-1-thiogalactopyranoside (IPTG). The cultures were incubated for an additional 3 hours at 37 °C with shaking at 250 rpm. The cells were collected by centrifugation at 5,000 rpm (2,800xg) at 4 °C for 15 minutes. The cell pellets were stored at -20 °C for future protein purification.
2.3 N\textsuperscript{15} labeled HK97 gp74 Protein Expression

The pET-15b-6xHis-HK97 gp74 plasmid DNA was transformed into *E. coli* BL21Star (DE3) cells and grown on LB agar plates containing 100 µg/ml ampicillin media, overnight at a temperature of 37°C. Negative controls were performed as mentioned before. The next day, a single colony was used to inoculate a 5 ml LB media containing ampicillin (100 µg/ml). The culture was incubated at 37°C with shaking at 250 rpm for three to four hours during the day. Approximately 100 µl of this culture was used to inoculate a 200 ml M9 minimal media, with ampicillin, so that the initial OD\textsubscript{600 nm} was 0.001 and the 200 ml M9 minimal media was incubated at a temperature of 30 °C overnight (approximately 17 hours) with shaking at 250 rpm. The optical density at 600 nm (OD\textsubscript{600 nm}) of the overnight cultures was approximately 1.0. The cultures were then centrifuged at 2,500 rpm (1,258xg) at 4 °C for 30 minutes. The bacterial pellet was resuspended in 15 ml of M9 minimal media containing 6 g/L Na\textsubscript{2}HPO\textsubscript{4} 7H\textsubscript{2}O, 3 g/L KH\textsubscript{2}PO\textsubscript{4} and 0.5 g/L NaCl at a pH of 7.5, 1 mM MgSO\textsubscript{4}, 500 µg/ml d-biotin, 500 µg/ml thiamine-HCl, 1 µM ZnSO\textsubscript{4}, 0.1 mM CaCl\textsubscript{2}, 4 g/L glucose, 1 g/L \textsuperscript{14}NH\textsubscript{4}Cl (or \textsuperscript{15}NH\textsubscript{4}Cl, as required) and 100 µg/ml ampicillin. The resuspended cells were used to inoculate 1 L M9 minimal media cultures containing \textsuperscript{14}N-NH\textsubscript{4}Cl (or \textsuperscript{15}N-NH\textsubscript{4}Cl, as required). The 1 L cultures were incubated at 37 °C with shaking at 250 rpm until an OD\textsubscript{600 nm} of 0.6 to 0.7 was reached and the temperature was dropped to 30 °C. The growth was monitored until an OD\textsubscript{600 nm} of 0.7 to 0.8 was reached and the temperature was dropped to 25 °C and finally when the OD\textsubscript{600 nm} reached 0.8, the temperature was dropped to 16 °C and protein synthesis was induced with 1 mM of IPTG and the cultures were incubated overnight with shaking at 250 rpm. In M9 minimal media, induction of HK97 gp74 at 16 °C, overnight results in a more soluble protein than induction
at 37 °C for 3 hours. Cell cultures were centrifuged at 6,000 rpm (3,381xg) at 4 °C for 15 minutes to pellet the cells, which are stored at -20 °C.

### 2.4 Immobilized Metal Affinity Chromatography (IMAC) Purification of HK97 gp74

The 6xHis-HK97 gp74 was purified to homogeneity using standard immobilized nickel metal affinity chromatography procedures. The pellet from our 2 L M9 culture was resuspended in 30 ml of 20 mM tris Cl, pH 7.9, 150 mM NaCl, 2 mM β-mercaptoethanol, 5 mM imidazole, 1 mM phenylmethylsulphonyl fluoride (PMSF), 5 mM benzamidine and 5 mM n-caproic acid. Imidazole is an aromatic, heterocyclic compound that is similar in structure to the cyclic ring found in histidine and acts as a molecular mimic. Phenylmethylsulphonyl fluoride (PMSF) and benzamidine are serine protease inhibitors.\(^{49,50}\) N-caproic acid is a lysine analogue that inhibits carboxypeptidases.\(^{51}\)

To lyse the bacterial cells, 1 mg/ml lysozyme and 2 mg/ml deoxycholic acid (DCA) was added for cell lysis, along with a small amount of DNase. Lysozyme is an enzyme that catalyzes the hydrolysis of 1,4-β-linkages between N-acetylmuramic acid and N-acetyl-D-glucosamine residues in bacterial peptidoglycan cell walls.\(^6\) DCA is a mild anionic detergent that solubilises cellular and membrane components for lysis of the bacterial cell membrane. DNase was also added to degrade or cleave any contaminating DNA that would increase the viscosity of the lysate and would interfere with protein purification. The cells were placed on ice and lysed by brief sonication using 1 minute intervals, consisting of 20 seconds of sonication followed by a rest period and repeated 4-6 times. Sonication is the use of sound energy to disrupt particles and intermolecular interactions and also shears any remaining genomic DNA. Lysed cells were
centrifuged at 13,000 rpm (9,464 xg) at 4 °C for 30 minutes. Centrifugation pelleted any large macromolecular complexes, such as the lysed cell membrane and insoluble proteins. The pellets were resuspended in the lysis buffer described above and the sonication and centrifugation steps were repeated. The supernatants from these two cell lysis steps were combined and applied to a 3 ml Ni\textsuperscript{2+} column that was pre-equilibrated with 20 mM tris Cl, pH 7.9, 500 mM NaCl, 20 mM imidazole, and 2 mM β-mercaptoethanol. Non-specifically bound proteins were washed with 30 ml of the equilibration buffer. The 6xHis-HK97 gp74 was eluted with 20 mM tris Cl, pH 7.9, 500 mM NaCl, and 400 mM imidazole in 3 ml fractions. In addition, 5 mM β-mercaptoethanol was added to the elution buffer, since β-mercaptoethanol reduces disulfide bonds. There are four cysteine residues in HK97 gp74. After elution fractions of the 6xHis-HK97 gp74 protein were collected the protein expression and the efficiency of our protein purification protocol was verified with SDS polyacrylamide gel electrophoresis (SDS-PAGE).

2.5 Expression of TEV Protease

A sample of tobacco etch virus (TEV) protease DNA was obtained from the laboratory of Karen Maxwell, Structural Genomics Consortium. TEV protease was produced for removal of the six histidine residue (6xHis) tag, which is located at the amino (N) terminus of our expressed 6xHis-HK97 gp74 fusion protein. The 6xHis-TEV protease was expressed and purified by Ni\textsuperscript{2+} affinity chromatography using the same procedure as the expression and purification of 6xHis-HK97 gp74 in LB growth media.
TEV protease protein elution samples were pooled and dialyzed in 120 mM tris Cl, pH 7.0, 50 mM NaCl at 4 °C, overnight. The 10 ml sample of TEV protease was then extracted from the dialysis bag and 2 ml of 50% glycerol was added to the sample and then aliquoted into 1.5 ml eppendorf tubes for storage at -70 °C. Protein expression and the efficiency of our protein purification protocol was verified with sodium dodecyl sulphate polyacrylamide gel electrophoresis (SDS-PAGE).

2.6 Removal of the 6xHis tag from HK97 gp74

The HK97 gp74 protein elution samples were pooled and dialyzed in 50 mM phosphate, pH 7.0, 50 mM NaCl and 5 mM β-mercaptoethanol. Originally a higher salt buffer was used (consisting of 150 mM NaCl) but it was observed that high salt results in the precipitation of the HK97 gp74 protein out of solution during dialysis. TEV Protease (1 mg TEV protease/ 40 mg protein) was added to the sample to cleave the 6xHis-tag off the target protein during dialysis. Dithiothreitol (DTT) cannot be used as a reducing agent because DTT inhibits the activity of TEV protease. SDS-PAGE was used to verify cleavage of the 6xHis-tag from HK97 gp74.

The TEV-digested and dialyzed sample was concentrated using a centrifugation filter (Millipore Ultra-15 Centrifugal Filter) with a molecular weight cut off of 3 kDa by centrifugation at 2,500 rpm (1,258xg) at 4 °C in 20 minute stages. After each 20 minute spin, the retentate was mixed gently using a pipette to avoid precipitation of the protein due
to the concentration gradient that builds up in the concentrator. This procedure was repeated until a sample of 2 ml to 2.5 ml was obtained.

2.7 Size Exclusion Chromatography (SEC) Purification of HK97 gp74

The concentrated sample of the TEV protease-digested HK97 gp74 protein was applied onto a 24 ml size exclusion column (Superdex 75 Pharmacia) with a bead size of 13 μm that was pre-equilibrated with 50 mM Na₂PO₄, pH 7.0, 150 mM NaCl, 5 mM 6-aminocaproic acid, 5mM benzamidine and 1 mM PMSF. Approximately 0.2 ml of the concentrated TEV protease-digested HK97 gp74 protein sample was loaded and run through the column at a flow rate of 0.5 ml/min. 0.5 ml fractions were collected. This procedure was repeated 12 to 15 times to purify the entire sample of HK97 gp74.

2.8 Determination of HK97 gp74 Protein Concentration

Originally the concentration of HK97 gp74 was determined using the Biorad protein assay, which is based on the Bradford method. The Bradford assay is a colorimetric, analytical method that measures the concentration of proteins in solution and is based on the absorbance shift (at 595 nm) of the acidic Coomassie Brilliant Blue G250 dye used in the assay. A differential colour change of the dye in response to various concentrations of protein is observed. The unbound dye is easily identified by its reddish-brown colour and upon addition of protein, a protein: dye complex at basic and aromatic amino acid residues,
such as arginine, is formed. The Coomassie dye donates a single electron to the ionizable groups of the protein. This enables the dye to bind covalently to the basic amino acids of the protein being assayed, thus, stabilizing the blue Coomassie dye, which can be measured by spectroscopic methods at an absorbance maximum of 595 nm. The absorbance at 595 nm is directly proportional to the amount of bound dye and thus, to the amount of protein present in the sample.

The Biorad assay kit contained the Coomassie Brilliant Blue G-250 reagent, which was diluted with four parts distilled, deionized water. Ten dilutions of the bovine serum albumin (BSA) protein standard were prepared with a range from 0 mg/ml to 50 mg/ml. 100 µL of each standard was added to 1 ml of the diluted dye and gently vortexed. The samples were measured at an absorbance of 595 nm. Although BSA is commonly used as a protein standard, it is not an ideal standard since proteins have different amino acid compositions that react differently in a protein assay. Thus, the ideal condition would be to use purified samples of the protein of interest as the standard. Approximately 10 µL of HK97 gp74 concentrated protein sample was added to 1 ml of the diluted dye and vortexed. The sample was measured at an absorbance of 595 nm. Absorbance at 595 nm as a function of BSA concentration was plotted and the concentration of HK97 gp74 was interpolated from this data, where the actual HK97 gp74 concentration is ten times the interpolated concentration. However, after concentration determination by amino acid analysis and an absorbance of 280 nm, we observed that the Biorad assay was significantly different than the consensus using the other two methods. The discrepancies in concentration may be due to the use of BSA as a standard and that HK97 gp74 has more basic and aromatic residues. Alternatively, HK97 gp74, as a smaller protein, is more easily disrupted and allows for the
exposure of hydrophobic pockets for formation of the protein: dye complex, thus, increasing
the absorbance at 595 nm.

Consequently, the concentration of HK97 gp74 sample was determined by
measuring the $A_{280\,\text{nm}}$ of the protein in 8M urea.\textsuperscript{53} The concentration was calculated using
the Beer-Lambert Law, $A = \varepsilon cl$, where $\varepsilon$ describes the molar absorptivity of the protein in
urea. Because proteins are denatured in 8 M urea, and all amino acids are exposed to
solvent, the molar absorptivity can be approximated to be the sum of the individual $\varepsilon$ values
for tryptophan (Trp), tyrosine (Tyr), and cysteine (Cys) in urea. HK97 gp74 contains 4 Trp,
2 Tyr, and 4 Cys residues. Therefore, the $\varepsilon$ for HK97 gp74 in 8M urea is calculated as
follows: $\varepsilon = (4)(5500\,\text{M}^{-1}\text{cm}^{-1}) + (2)(1490\,\text{M}^{-1}\text{cm}^{-1}) + (4)(125\,\text{M}^{-1}\text{cm}^{-1})$, which was
calculated to be 25,480 L·mol$^{-1}$·cm$^{-1}$\textsuperscript{28} Rearrangement of the Beer-Lambert equation,
$c = A/\varepsilon l$, allows for the determination of the concentration of HK97 gp74.

Amino acid analysis was used as a final validation of protein concentration. A
sample of HK97 gp74 was sent to the Amino Acid Analysis Facility at the Advanced Protein
Technology Centre, The Hospital for Sick Children. A report of the amount in picomoles of
each amino acid residue in the protein was obtained and the data was compared to the
sequence or number of each amino acid residue of HK97 gp74 to determine the
concentration of protein in the sample. See calculations in Appendix 1.
2.9 Preparation of the NMR Sample

The fractions containing the purified HK97 gp74 were pooled and concentrated in a centrifugation filter (Millipore Ultra-15 Centrifugal Filter, with a molecular weight cut off of 3 kDa) by centrifugation at 2,500 rpm (1,258xg) at 4 °C in 20 minutes intervals. After each 20 minute spin, the retentate was mixed gently by pipetting. This procedure was repeated until a sample of approximately 0.5 ml was obtained. The first NMR sample consisted of the HK97 gp74 protein in a high salt buffer (containing 150 mM NaCl, 50 mM Na₂PO₄, pH 7.5, 5 mM 6-aminocaproic acid, 5 mM benzamidine and 1 mM PMSF), whereas, the other NMR sample consisted of the HK97 gp74 protein in a low salt buffer (containing 50mM NaCl, 50 mM Na₂PO₄, pH 7.5, 5 mM 6-aminocaproic acid, 5 mM benzamidine and 1 mM PMSF).

The purified and concentrated ¹⁵N-labeled HK97 gp74 was placed in an NMR tube. 50 μl of 99.9 % D₂O and 5 μl of 100 mM DSS (2,2-dimethyl-2-silapentane-5-sulfonic acid) were added to the sample. DSS is added to biological NMR samples in water as a calibration standard. DSS has an easily identifiable proton resonance that is significantly further upfield than any other resonance in proteins and nucleic acids.⁵⁴

2.10 Biophysical Analysis of HK97 gp74

2.10.1 NMR Studies: ¹⁵N-¹H correlation spectrum (HSQC)

A ¹⁵N-¹H heteronuclear single quantum coherence (HSQC) correlation spectrum was recorded for HK97 gp74 on a Varian Unity 600 at 25 °C or 15 °C, equipped with either a
triple resonance cryoprobe or a triple resonance room temperature probe. The 2D spectrum correlates the resonant frequency of the amide proton and the resonant frequency of the directly attached amide nitrogen. This experiment is most often used to determine protein conformation (i.e. folded or non-folded) before other more complex experiments are recorded, such as an experiment for resonance assignment or in structure determination. The HSQC can also be used to screen for protein interactions or conformational changes in conditions. We also used HSQC spectra to screen for ideal conditions for long term NMR studies. Different salt conditions were tested since it was observed that at higher concentrations of salt, HK97 gp74 precipitated out of solution.

2.10.2 Structural Characterization of HK97 gp74 by Circular Dichroism

In order to assess whether lower pH induces protein unfolding, which would affect activity, circular dichroism (CD) was used to assess if there were any pH-dependent structural changes. A 600 µl sample containing 2 µM HK97 gp74 in 20 mM HEPES, pH 5.0 to 8.0 was prepared, in the absence of either DTT or β-mercaptoethanol. A CD spectrum was obtained after analysis of the samples using 5 scans per sample, measuring a range of absorbance from 190 nm to 260 nm, where the absorbance was measured every 0.2 nm at 25 °C. A data file was retrieved and the absorbance was manipulated to plot intensity as ellipticity (deg cm²/dmol). The data was plotted as ellipticity as a function of absorbance.

However, it was noted that HEPES buffer appears to behave very erratically and produces a large portion of signal noise at 195 nm to 210 nm. Consequently, using 2 µM of
HK97 gp74 resulted in very poor intensity and samples of 10 µM, 20 µM and 40 µM were analyzed. Comparison of the different concentrations proved that there was a shift in the spectrum toward the right as the concentration increased. Consequently, our protein sample was dialyzed in a 20 mM phosphate, 50 mM NaCl buffer at various pHs from 5.0 to 8.0. The CD experiments were repeated using 2 µM HK97 gp74 in 20 mM phosphate, 50 mM NaCl buffer at a pH of 5.0 to 8.0 and at 25 °C. A CD spectrum was obtained after analysis of the samples using 5 scans per sample, measuring a range of absorbance from 195 nm to 260 nm, where the absorbance was measured every 0.2 nm. Controls or “blanks” were run under the same conditions containing 600 µl of 20 mM phosphate, 50 mM NaCl buffer at a pH of 5.0 to 8.0.

2.11 Substrates for DNA Cleavage

2.11.1 Plasmid DNA

The pUC-18 and pBluescript plasmid DNA was extracted from DH5α cells and purified using a chromatography step (Sigma-Aldrich GenElute HP Plasmid MiniPrep).

2.11.2 Phage DNA

Lambda phage DNA was obtained commercially from NEB.
2.11.3 Purification of Single Stranded DNA

A sample of pBluescript (SK+) was transformed into DH5α cells and was grown overnight at 37 °C on LB agar plates containing 100 µg/ml ampicillin. A 50 ml LB media containing 100 µg/ml ampicillin and $10^8$ pfu/ml R408 helper phage was inoculated with a single colony of the overnight pBluescript (SK+) transformed DH5α cells and incubated at 37 °C with vigorous aeration for 16 to 24 hours. The cell culture was centrifuged at 2,500 rpm (1,258xg) at 4 °C for 20 minutes. The supernatant was decanted to a fresh tube and 7.5 ml of a solution containing 20 % PEG-8000 and 2.5 M NaCl at pH 7.5 was added. The phage particles were allowed to precipitate on ice or at 4 °C for 15 minutes to overnight (for an increased yield). The sample was then centrifuged for 45 minutes at 14,000 rpm (23,670xg) until a pellet formed. The supernatant was removed and the sample was centrifuged again to remove all residual liquid. The pellet was resuspended by vortexing vigorously in 4 ml of 0.3 M sodium acetate (pH 6.0) and 1 mM EDTA. The sample was subjected to addition of one volume to sample of phenol-chloroform and centrifuged for 10 minutes at 4,000 rpm (3,584xg) to separate the organic phase from the aqueous phase. The aqueous phase was transferred to a fresh tube and 4 ml of ethanol was added. The sample was centrifuged for 10 minutes at 4,000 rpm (3,584xg). The ethanol was removed and the DNA pellet was dried. The pellet was dissolved in 1 ml of TE buffer containing 20 mM tris Cl⁻, pH 7.5, 10 mM EDTA. The concentration of the DNA pellet was very low and the sample was precipitated by adding 1 volume of 0.3 M sodium acetate at pH 5.2 and 2 volumes of ice cold 100 % ethanol to 1 volume of the DNA sample. The sample was mixed and stored at -20 °C for at least 1 hour to precipitate the DNA. The precipitated DNA was recovered by centrifugation at 15,000 rpm (5,040xg) for 15 minutes. The ethanol was
decanted and the pellet was washed twice with 70% ethanol. The DNA pellet was air dried and then resuspended in 100 µl of TE buffer. The resulting DNA samples ranged in concentration from 2 µg/ml to 9 µg/ml.

### 2.12 Tests for Endonuclease Activity

#### 2.12.1 DNA Cleavage Assays

A DNA cleavage assay was designed to test the digestion of pUC-18 plasmid DNA, pBluescript (SK+) plasmid DNA and lambda phage DNA. Based on work with other endonucleases, assays were performed using 10 µg/ml of DNA and 1.5 µg/ml of HK97 gp74 in 20 mM phosphate, pH 7.0, 50 mM NaCl. The assay varied the concentration of the divalent metal ion (either Ni²⁺, Mg²⁺, Ca²⁺, Zn²⁺ or Co²⁺) from 10 mM to 40 mM, since HNH endonuclease are known to require binding of a divalent metal for catalysis. Stock metal solutions were prepared at a 0.1 M concentration in a 20 mM phosphate, pH 7.0, 50 mM NaCl. Each reaction was prepared as a 0.5 ml sample and incubated at room temperature (21°C) for several hours. Controls were performed to test the cleavage of DNA alone in buffer, DNA with metal only or DNA in the presence of HK97 gp74 without metal. DNA digestion was analyzed with a 1% agarose gel stained with SybrSafe® (Invitrogen). The assay was repeated 3 to 5 times per divalent metal and substrate.

When analyzing digestion experiments with pUC-18 or pBluescript (SK+) plasmid DNA, an EcoRI digested plasmid DNA sample was included to compare the migration pattern of HK97 gp74-digested DNA to double stranded linear DNA. As mentioned above,
the assay was used to test a range of divalent metals, including magnesium, calcium, cobalt, nickel and zinc. However, in assays containing zinc, cobalt, and high concentrations of magnesium (5 mM), the formation of precipitate was observed when the metals were dissolved in 20 mM phosphate, pH 7.0, 50 mM NaCl. The solubility of the divalent metals, nickel, magnesium, zinc, cobalt and calcium were tested in a 50mM tris Cl’, pH 7.8 but a significant amount of precipitation with zinc and cobalt were observed. Zinc formed a dense, cloudy, insoluble precipitate in the buffer and cobalt formed a red, particulate precipitate. Finally, the divalent metals were tested in 50 mM 4-(2-hydroxyethyl)-1-piperazineethanesulfonic acid (HEPES), pH 7.0 and a reduction in precipitation was noted, where zinc appeared to have little to no insoluble product and cobalt formed a red, particulate fraction but appeared to be reduced in quantity. A reduction of buffer concentration, from 50 mM to 20 mM HEPES, resulted in higher divalent metal ion solubility. Therefore, subsequent DNA cleavage assays were prepared with 20 mM HEPES, pH 7.0. All experiments were repeated 3-5 times for all divalent metal ions and DNA substrates.

The concentration range of divalent metals was decreased to 1 mM, 2 mM, 5 mM and 10 mM compared to the previous assays that used 10 mM to 40 mM of divalent metal. The concentration of pUC-18 DNA used in cleavage assays was increased to a maximum of 25 µg/ml, whereas, pBluescript (SK+) DNA concentration was increased to 20 µg/ml and lambda phage DNA concentration was increased to 48 µg/ml. Only 2 µg/ml of single-stranded pBluescript was used in assays since the yield was significantly lower than expected at 2 µg/ml to 9 µg/ml of DNA. The concentration of HK97 gp74 was increased to 24 µg/ml to test if the DNA could be cleaved to completion in a given amount of time. The
temperature of incubation was also increased to 37 °C from 21 °C to increase the reaction rate of HK97 gp74 DNA cleavage.

Because there are two histidines involved in the reaction of HNH endonucleases and if our protein is an HNH endonuclease, a change in cleavage activity with lower pH was expected, thus, lambda phage DNA cleavage assays were performed at a pH range from 5 to 8. In another lambda phage DNA cleavage experiment, stoichiometric amounts of protein and metal were tested to determine the stoichiometry of binding between HK97gp74 and the divalent metal ion. In these reactions, one equivalent of protein is equal to 0.858 µM (or 11 µg/ml) and the divalent metal concentration was varied from 1 to 10 equivalents to determine at what ratio cleavage was observed.

2.12.2 Reporter Methods for DNA Cleavage Assays

a) Absorbance at 260 nm

A predominantly common technique to test for endonuclease activity is spectrophotometric assays. Such a spectrophotometric assay was first developed in 1950 by M. Kunitz to measure DNase I activity. The assay measures the increase in the absorbance of free bases of DNA at a wavelength of 260 nm (A_{260 nm}) at a pH of 5.0 and a temperature of 25 °C. Thus, an increase in the A_{260 nm} measurement indicates the degradation of a DNA substrate. Furthermore, the rate of increase of the A_{260 nm} measurement under the assay conditions should produce a linear relationship, where the rate is proportional to the concentration of active enzyme. This assay resulted in the definition of the Kunitz unit,
which is defined as the change in A$_{260\text{nm}}$ of 0.001/ minute/ml at a pH of 5.0 and a temperature of 25 °C. However, these conditions are often not optimal for many enzymes and thus, the Kunitz assay has been modified commercially and in academic literature to better study the activity of enzymes at different temperatures and pHs.

The assay was performed on pUC-18 and pBluescript (SK+) plasmid DNA in the presence of metal and HK97 gp74. The A$_{260\text{nm}}$ was measured every 10 minutes and the activity of HK97 gp74 was stopped by the addition of 1 mM EDTA to each sample.

b) Agarose Gel Electrophoresis

DNA digestion was also monitored by agarose gel electrophoresis. Separation of molecules and complexes by electrophoresis are based on size, charge and shape. DNA, which is negatively charged, will migrate from the negative diode to the positive diode. Smaller, more compact molecules, like super-coiled plasmids, are more likely to travel faster through the gel, thereby, migrating to a position farther down the gel. Larger, more loosely associated molecules, such as DNA plasmids that are cut on a single strand (nicked) and lose their supercoiled structure are more likely to travel more slowly through the gel and reach a higher position on the gel. Similarly, linear, double stranded DNA migrates to a higher position on the gel because it is in a less compact form.

A 20 µl sample was removed from the reaction tube every 30 minutes for plasmid DNA or one hour for phage DNA and enzyme activity of HK97 gp74 was stopped by the addition of 25 mM EDTA to each sample. A sample was also taken prior to addition of
HK97 gp74 and just after addition of the enzyme. To analyze the digestion of DNA after a 2 to 8 hour incubation period, a 1 % agarose gel was run showing the DNA cleavage per half hour or one hour for reactions that varied the metal concentration. Controls were run alongside the reactions and tested digestion of the DNA alone, or DNA with either divalent metal or HK97 gp74.

HK97 gp74 was analyzed by SDS-PAGE in either reducing or non-reducing conditions. Two samples of HK97 gp74 protein were prepared, one containing β-mercaptoethanol and another without any reducing agent. β-mercaptoethanol reduces disulfide bonds formed between cysteine residues. Comparison of the SDS-PAGE migration of the two samples allowed for analysis of HK97 gp74.

2.13 Reporter Methods for Metal Binding Experiments

2.13.1 UV-Vis Metal Binding Assay

Since reaction and stoichiometric conditions were determined for DNA cleavage of lambda phage DNA, an assay was performed to test digestion at an absorbance of 260 nm. In this protocol, we prepared a 500 µl reaction sample at varying equivalents of the divalent metal, nickel. The absorbance at 260 nm was measured every 30 minutes and recorded. Controls or “blanks” were used to control for any changes in absorbance due to the buffer or of the metal and protein. Consequently, the first blank contained only 500 µl of 20 mM HEPES, pH 7.0 and the second blank contained 1 µM HK97 gp74, 1 µM nickel sulfate and 20 mM HEPES, pH 7.0 buffer, which totaled a final volume of 500 µl.
2.13.2 Metal Binding Experiments by NMR Titration of HK97 gp74

Initially, metal binding experiments were done using NMR titrations. After recording a $^{15}$N-$^1$H HSQC correlation spectrum for HK97 gp74, 0.5 mM of zinc sulfate in 20 mM HEPES, pH 7.0 was added to the 0.5 mM HK97 gp74 sample. However, addition of metal at a high concentration of protein caused immediate precipitation of most of the sample. A $^{15}$N-$^1$H correlation spectrum for the metal bound HK97 gp74 was performed and recorded.

Since a highly concentrated sample of HK97 gp74 precipitated upon addition of HK97 gp74, metal titrations using unlabelled HK97 gp74 protein samples were performed. A 100 µM sample of HK97 gp74 and an equivalent (100 µM) of nickel sulfate or zinc sulfate, in 20 mM HEPES, pH 7.0, was applied to a Pall Nanosep centrifuge filtering device (with a molecular cut off of 3 kDa). The 500 µl sample was concentrated, at a speed of 2,000 rpm (112xg) at 4 ºC, down to 50 µl or a concentration of 1 mM HK97 gp74 with an equivalent concentration of divalent metal. Very little precipitate formed.

2.13.3 Tryptophan Fluorescence Spectroscopy

Metal binding was tested to correlate activity with metal dependence and to test the HNH motif mutants. Furthermore, at high concentrations of protein required for NMR, metal bound species aren’t soluble. Fluorescence spectroscopy was used to assess metal binding. Since most of the intrinsic fluorescence emissions of a folded protein are due to excitation of tryptophan residues and since tryptophan fluorescence is strongly influenced by the chemical environment, metal binding should promote a change in the fluorescence
intensity.\textsuperscript{56} HK97 gp74 contains a total of four tryptophan residues; tryptophan 111 is located directly adjacent to the metal binding histidine of the HNH motif and therefore, a change in fluorescence intensity is expected. The other tryptophan residues are distributed elsewhere in the protein. Thus, it is hypothesized that metal binding at the HNH motif may result in significant and measurable changes in fluorescence.

100 µl reactions were prepared containing 0.5 µM HK97 gp74 in the presence of the divalent metal Ni\textsuperscript{2+} at 1 µM, 3 µM and 5 µM in a 20 mM HEPES buffer, pH 7.0. A control was performed and consisted of 100 µl of 0.5 µM HK97 gp74 in 20 mM HEPES, pH 7.0. A Tecan Infinite M1000 Fluorescence Plate Reader (Gunning Group) was used to excite 60 µl to 75 µl of each sample at a wavelength of 280 nm at a gain of 255, a frequency of excitation of 400 Hz and a temperature of 24.4 °C to 24.6 °C, where the emission fluorescence intensity was measured from 300 nm to 360 nm at step sizes of 1 nm.

\textbf{2.14 Mutant Primer and QuikChange Mutagenesis}

Mutagenesis at the putative HNH motif was performed to confirm that gp74 is an HNH endonuclease and to determine which residues are essential for binding and cleavage of DNA by HK97 gp74. Mutants were generated in which each HNH motif residue is changed to alanine. Recall that the first histidine and asparagine are involved in catalysis and the second histidine of the HNH motif is involved in divalent metal ion binding. Therefore, single mutants may display differences in just metal binding or just catalysis. A triple mutant
that has all of the conserved residues of the HNH motif changed to alanine, should produce a knockout or severe reduction in cleavage activity of HK97 gp74.

HK97 gp74 mutants were made using the Stratagene Quikchange Site Directed Mutageneis Kit. Forward and reverse mutagenic oligonucleotide primers were designed that contained the desired mutation for each of the conserved HNH motif amino acid sites. Each of the primers were 45 bases in length, where the desired mutation was located in the middle of the primer with 21 bases of correct sequence on either side of the alanine mutation. The primers were also designed to have a GC content of 40% or more to terminate in a C or G. The primers were synthesized at the DNA Synthesis Facility at The Centre for Applied Genomics, The Hospital for Sick Children and were obtained as dry samples. The pellets were dissolved in 10 mM tris Cl, pH 8.5 to a concentration of 100 µM and stored at -20 °C.

A set of PCR reactions were prepared as indicated in the Stratagene QuikChange Site Directed Mutagenesis Kit Manual. A control reaction was prepared using 10x reaction buffer, pWhitescript control plasmid, oligonucleotide primer #1 and #2, dNTP mix and water. Sample reactions were prepared using 10x reaction buffer, varying the amount of pET-15b-HK97 gp74 vector DNA from 5 ng, 10 ng, 25 ng and 50 ng, forward and reverse oligonucleotide primers, dNTP mix and water. Pfu turbo DNA polymerase was added to each of the PCR reactions, including the control. A PCR protocol was run where the first phase was denaturation of DNA at 95 °C for 30 seconds, the second phase consisted of denaturation of DNA at 95 °C for 30 seconds, followed by annealing at 55 °C for 1 minute and extension at 68 °C for 12 minutes and 6 seconds (2 minutes per kilobase of vector DNA). The second phase was repeated for 16 cycles.
The PCR reactions were electrophoresed on a 1 % agarose gel. The bands containing the plasmid of interest were excised by a sterile surgical razor. The gel slice was weighed and purified by Qiagen Qiaquick Gel Extraction Kit. Following purification, 1 µl (10 U/µl) of Dpn I was added to the purified pET-15b-H82A HK97 gp74 DNA and incubated for 1 hour at 37 °C. The pET-15b-H82A HK97 gp74 plasmid was transformed into *E. coli* DH5α cells and grown on LB agar plates containing 100 µg/ml ampicillin, overnight at 37 °C. Transformation of pET-15b-H82A-HK97 gp74 plasmid in *E. coli* strain DH5α cells produced many clones, one of which was selected and grown in LB media containing 100 µg/ml of ampicillin. The plasmid DNA was purified using Sigma-Aldrich GenElute HP Plasmid MiniPrep. This DNA can now be used in future protein expression studies to produce mutant versions of the HK97 gp74 protein, where the HNH motif can be studied more extensively.
3. Results

3.1 BlastP Search of HK97 gp74 Protein Sequence

At the beginning of this study, there was no structural and functional data available for the HK97 gp74 protein. Thus, we searched for proteins with similar sequences to HK97 gp74 using the basic local alignment search tool (BLAST). This program utilizes a sequence comparison algorithm that is optimized for speed to search different databases for the optimal local alignments to a specific search query, such as a protein or DNA sequence. A protein-protein BLAST search indicated that HK97 gp74 is a possible bacterial homing endonuclease. More specifically search results demonstrate that this gene product may belong to the HNH-endonuclease family of enzymes that are conserved among bacteria and viruses. HNH endonucleases are named for two conserved histidine residues and an invariant asparagine residue. For example, the HNH endonuclease (Acidovorax avenae subsp.) was one of the top search hits for HK97 gp74 having a bit score of 128 bits and an E-value of $2 \times 10^{-28}$. The bit score is a value derived from the raw alignment score ($S$), which is calculated as the sum of the scores for aligned position and gap scores. Gap scores are calculated as the sum of the gap opening penalties and the gap extension costs. A gap penalty results from the presence of a gap or consecutive number of spaces in a sequence alignment. Gaps are caused by mutations, insertions or deletions in a sequence. The E-value is the expectation value, which is defined as the number of different alignments with scores that are equal to or better than the raw alignment score and that are expected to occur by chance in a given database search. Thus, a lower E-value corresponds to a more significant score.
3.2 Structure Based Sequence Alignment

Another top search hit was the HNH endonuclease colicin E9, which has been extensively studied and reviewed in the biochemical and biomolecular literature. Consequently, we wanted to construct a sequence alignment of HK97 gp74 with related proteins to provide us with an indication of the possible function of this protein. We made a structure based sequence alignment of HK97 gp74 and colicin E9 to determine which residues in HK97 gp74 correspond to the HNH motif. Using secondary structural information of colicin E9 in our alignment improves the alignment by adding gaps in loops, rather than α-helices in β-strands. Although we obtained a good alignment, some manual refinement was necessary. The alignment is shown in Figure 11. Our alignment indicates that HK97 gp74 contains conserved histidine and asparagine residues that likely belong to the HNH motif of HK97 gp74. In HK97 gp74, the conserved residues are H82, N101, and H110 and appear to correlate to the HNH motif found in the bacterial protein, colicin E9 (Figure 11).
Figure 11: Structure-based sequence alignment of HK97 gp74 with colicin E9. HK97 gp74 was aligned against the structure and sequence information of colicin E9, using ClustalW. The secondary structure of colicin E9 is shown above the alignment and conserved HNH residues are highlighted by a red box. Red letters represent small, hydrophobic and aromatic amino acid residues, blue letters represent acidic residues, magenta letters represent basic residues, green letters represent hydroxyl or amine residues and all other amino acids are represented by gray letters. An “*” indicates that the residues in that column are identical in all sequences in the alignment. A “::” indicates that conserved substitutions are observed and a “.” indicates that semi-conserved substitutions are observed.

3.3 Vector Map of HK97 gp74-pET-15b

The pET-15b-HK97 gp74 plasmid was obtained from our collaborators (Karen Maxwell, Structural Genomics Consortium). The sequence of the insert and multiple cloning site was confirmed (ACGT). A vector map or diagram and linear DNA sequence of the plasmid illustrating the significant attributes of the plasmid, such as the multiple cloning site (MCS), origin of replication (ori), antibiotic resistance marker and complete inserted gene is
shown in Figure 12. The plasmid vector, pET-15b contains an origin of replication, an ampicillin resistance gene that acts as a selective agent, an abbreviated lac operon to control protein expression by addition of the lactose analogue, IPTG, and a multiple cloning site, which has various restriction enzyme recognition sites for insertion of a gene of interest. The pET-15b plasmid vector used for protein biosynthesis includes the gene encoding a 6xHis tagged HK97 gp74 bacteriophage protein with a tobacco etch virus (TEV) protease cleavage site between the tag and HK97 gp74 protein (Figure 12).

Figure 12: Vector map of pET-15b-HK97 gp74.
3.4 Expression & Purification of 6xHis-HK97 gp74

3.4.1 Ni\textsuperscript{2+} Affinity Chromatography Purification of 6xHis-HK97 gp74

An SDS-PAGE gel was run for all purifications to verify the presence of our target protein HK97 gp74, the size of the protein product, and the point of elution of the protein. SDS-PAGE is a molecular technique used to separate and analyze proteins according to their electrophoretic mobility, which is based on a protein’s size. The 6xHis-HK97 gp74 was expressed in the soluble fraction in *E. coli* BL21 Star (DE3) cells and purified to homogeneity using immobilized Ni\textsuperscript{2+} affinity chromatography. The basic principle of affinity chromatography relies on the fact that the target protein to be purified must have particular structural or chemical characteristics to allow for interaction of that molecule with the resin and the addition of a structural or molecular mimic is required for the elution of the target protein.\textsuperscript{6} Consequently, the HK97 gp74 protein was expressed as a fusion with an N-terminal 6xHis tag, allowing the HK97 gp74 protein to be purified with Ni\textsuperscript{2+} affinity chromatography. The protein was washed with 20 mM imidazole and eluted with 400 mM imidazole. Imidazole is a molecular mimic that is an aromatic, heterocyclic compound similar in structure to the cyclic ring found in histidine. The buffers contained 150 mM NaCl to prevent the non-specific binding or interactions of proteins other than the 6xHis-HK97 gp74 protein with the Ni\textsuperscript{2+} resin agarose beads. Several other components were included in the buffers. β-mercaptoethanol (2 mM) was included because it reduces cysteine disulphide bonds. Non-native disulfide bonds may result in protein precipitation or the loss of activity. Protease inhibitors were also included, such as phenylmethanesulfonylfluoride (150 μM) and benzamidine (5 mM), which are serine protease inhibitors and n-caproic acid,
which is a lysine analogue that inhibits carboxypeptidases.\textsuperscript{49-51} 17% SDS polyacrylamide gels were loaded with the following samples: the pellet and supernatant after lysis and centrifugation, the cell lysate that was loaded onto the Ni\textsuperscript{2+} column, the flow-through from the column, and the column wash to analyse the purification (Figure 13).

\begin{figure}[h]
\centering
\includegraphics[width=\textwidth]{figure13.png}
\caption{SDS-PAGE gel of 6xHis-HK97 gp74 Ni\textsuperscript{2+} affinity protein purification. 6xHis-HK97 gp74 was expressed in \textit{E. coli} BL21 Star (DE3) cells grown in M9 Minimal Media. Each protein fraction from the immobilized Ni\textsuperscript{2+} affinity chromatography protein purification procedure was electrophoresed to verify protein expression. The cell sample was lysed and centrifuged, forming a “pellet” that includes cell membrane, organelles and waste by-products and a “supernatant” comprised of soluble proteins, including the expressed 6xHis-HK97 gp74, and genomic material. The lysate that was passed over the Ni\textsuperscript{2+} affinity chromatography column comprised the “load” and the resulting sample that exited the column comprised the “flow-through”. Non-specifically bound proteins were washed with 30 ml of the 20 mM imidazole-containing buffer producing the “wash” sample.}
\end{figure}
Identity of 6xHis-HK97 gp74 was verified at a position between 6 kD and 16 kD, where the actual size of the 6xHis-HK97 gp74 protein is approximately 13.6 kDa (Figure 14). The immobilized Ni\textsuperscript{2+} affinity purified eluant contained several different proteins. Although the 6xHis-HK97 gp74 protein appears as a very intense band between 6 kD and 16 kD on the gels, other protein bands are also visible. Thus, the sample did not consist of a pure 6xHis-HK97 gp74 protein and was not at a purity required for NMR structural analyses.

Figure 14: SDS-PAGE gel of 6xHis-HK97 gp74 protein elution samples. 1/300\textsuperscript{th} of the elution samples obtained from Ni\textsuperscript{2+} purification was loaded onto the SDS polyacrylamide gels. The gel was stained with Coomassie Brilliant Blue R-250 dye.
3.4.2 Analysis of TEV Protease Cleavage of 6xHis-HK97 gp74

The 6xHis-HK97 gp74 protein contains a recognition site for TEV protease in order to remove the 6xHis tag. Because the 6xHis tag is only 0.9 kDa, analysis of the efficiency of TEV protease digestion is difficult by simply running pre-digested and post-digested samples on an SDS-PAGE gel. An undigested 6xHis-HK97 gp74 protein sample was electrophoresed beside individually TEV protease digested samples of 6xHis-HK97 gp74. The 6xHis-HK97 gp74 samples incubated with TEV protease appear to have only a slight difference in mobility shift as compared to the uncleaved 6xHis-HK97 gp74 (Figure 15).

![Figure 15: SDS-PAGE gel of 6xHis-HK97 gp74 cleaved with TEV protease. The uncleaved 6xHis-HK97 gp74 protein has a molecular weight of 13.6 kD, whereas, the TEV protease cleaved 6xHis-HK97 gp74 protein samples have a molecular weight of 12.8 kD. 1/200th of the elution samples obtained from Ni\textsuperscript{2+} affinity purification was loaded onto the SDS polyacrylamide gels. The gel was stained with Coomassie Brilliant Blue R-250 dye.](image-url)
3.5 Size Exclusion Chromatography of HK97 gp74

After dialysis and concomitant incubation of the 6xHis-HK97 gp74 protein with TEV protease to remove the 6xHis tag, the protein sample was further purified using size exclusion chromatography (SEC). In size exclusion chromatography, size separation of molecules occurs via the principle that molecules of different sizes will move or elute through the stationary phase (resin) at different rates. Each column is defined by a permeation rate, which is the lowest molecular weight and smallest molecule that can penetrate into the pores of the stationary phase completely. Thus, molecules larger than the pores of the stationary phase will elute first, since they are not hindered by these porous channels within the resin. Purification of HK97 gp74 from the 6xHis tag and the TEV protease was accomplished with a 24 ml Superdex 75 column (GE Healthcare) composed of cross-linked agarose and dextran with an average particle size of 13 µm and an optimum separation range for proteins of 3,000 Da to 70,000 Da. This is an optimal separation range for Ni²⁺ purified fractions of HK97 gp74, which contain proteins of high molecular weight in addition to HK97 gp74. The size exclusion purification was monitored using UV spectroscopy at an absorbance of 280 nm. Our $A_{280}$ nm trace shows a large peak at 13.5 ml, which corresponds to a protein with a monomeric molecular weight (Figure 16). SDS-PAGE analysis verifies this peak migrates to a position between 6 kD and 16 kD, which is indicative of HK97 gp74 with a molecular weight of 12.8 kDa (Figure 17). The size exclusion purified HK97 gp74 sample was pure and contained no other proteins.
Figure 16: Size exclusion chromatography of HK97 gp74. The $A_{280\,\text{nm}}$ absorbance is measured and plotted against elution volume.
Figure 17: SDS-PAGE gel of size exclusion chromatography purified HK97 gp74. 1/375<sup>th</sup> of the elution fraction was loaded onto the SDS polyacrylamide gel. The gel was stained with Coomassie Brilliant Blue R-250 dye.

3.6 Analysis of Endonuclease Activity

3.6.1 HK97 gp74-Mediated Digestion of Plasmid DNA

Preliminary protein sequence comparisons and sequence alignment results suggest that HK97 gp74 is a possible homing HNH endonuclease. Thus, we attempted to study digestion
of DNA by HK97 gp74 using the absorbance at 260 nm. The $A_{260\ nm}$ measures the absorbance of free DNA bases and an increase in the $A_{260\ nm}$ measurement indicates the degradation of a DNA substrate. However, problems were encountered, such as precipitation of the metal solutions in 20 mM phosphate, pH 7.0, 50 mM NaCl and in 50 mM tris Cl, pH 7.8. No significant change in the $A_{260\ nm}$ measurement of pUC-18 or pBluescript (SK+) DNA digestion was observed. $A_{260\ nm}$ measurements ranged from an OD of 0.006 to 0.026 and appeared to fluctuate between this range.

Consequently, investigation of the DNA cleavage activity of HK97 gp74 was analyzed by agarose gel electrophoresis. These experiments were performed using pUC-18 and pBluescript (SK+) plasmid DNA in either 50 mM tris Cl, pH 7.8 or 20 mM HEPES, pH 7.0. In our experimental reactions we varied the concentration of divalent metal from 1 mM to 10 mM. The assay also tested a range of divalent metals, including nickel, magnesium, calcium, cobalt and zinc. DTT is a potent disulfide reductant but also chelates metal ions. Thus, DTT was not added to the HK97 gp74 protein sample. As a substitute, β-mercaptoethanol (5 mM) was used as a disulfide reductant. There was no DNA digestion in control reactions with DNA alone, DNA with metal, or DNA with HK97 gp74 but without metal. In contrast, incubation of DNA with HK97 gp74 and metal results in DNA digestion and indicates that our protein is a metal-specific endonuclease. A comparison with EcoR1 digested DNA, which is linear dsDNA, indicates that in the presence of divalent metal ions, HK97 gp74 mediates single-stranded DNA digestion of plasmid DNA. We observed that bands of pUC-18 DNA migrated to a position between 3 kB and 4 kB, which is much larger than the supercoiled plasmid (which migrates to a position just above 2 kB) and the linear DNA (at 3 kB) (Figure 18).
Figure 18: HK97 gp74 cleaves pUC-18 plasmid DNA (2688 bp). The pUC-18 plasmid DNA (25 μg/ml) was incubated with HK97 gp74 (3 μg/ml) with varying concentrations of metal in 50 mM tris Cl-, pH 7.8. Controls show no DNA digestion. Reactions show HK97 gp74 cleavage of pUC-18 DNA that results in bands between 3.0 kB and 4.0 kB. This data provides evidence of ssDNA cleavage of plasmid DNA by HK97 gp74.
HK97 gp74 analysis by SDS-PAGE in either reducing or non-reducing conditions was performed after each digestion assay. The SDS polyacrylamide gel shows the presence of HK97 gp74 at a position that corresponds to its molecular weight of 12.8 kD in both reducing and non-reducing conditions (Figure 19).

![Figure 19: SDS-PAGE gel of HK97 gp74 in reducing and non-reducing conditions. A 67 uM sample of HK97 gp74 was analyzed. β-mercaptoethanol was used as a disulfide reductant. The gel was stained in Coomassie Brilliant Blue R-250 dye.](image)

For pBluescript (SK+) DNA, we observed that bands of HK97 gp74 digested DNA migrate at a position between 4 kB and 5 kB, which is much larger than the supercoiled
phagemid (which migrates to a position just above 2 kB) and the linear DNA (at 3 kB) (Figure 20). Thus, in the presence of divalent metal ions, HK97 gp74 also mediates single strand cleavage of the pBluescript (SK+) phagemid DNA, similar to results observed for pUC-18 plasmid DNA. Our results to date do not conclude whether single-stranded digestion of plasmid DNA occurs at a specific site or multiple sites.
Figure 20: HK97 gp74 cleaves pBluescript (SK+) plasmid DNA (2961 bp). The DNA (20 μg/ml) was incubated with HK97 gp74 (24 μg/ml) with varying concentrations of metal in 20 mM HEPES, pH 7.0. Controls show no DNA digestion. A sample of pBluescript was also cleaved with EcoRI to show a representative single strand cleavage reaction and results in a band at 3.0 kb. Reactions show HK97 gp74 cleavage of pBluescript (SK+) that results in bands between 2.0 kb and 3.0 kb and between 4.0 kb and 5.0 kb. This data provides evidence of ssDNA cleavage of plasmid DNA by HK97 gp74.
Thus, we attempted to generate ssDNA using pBluescript (SK+). However, we could never view single-stranded DNA on an agarose gel, possibly due to low DNA concentrations.

3.6.2 HK97 gp74-Mediated Digestion of Phage DNA

Since gp74 is a possible HNH endonuclease encoded by the bacteriophage HK97, we tested the digestion of phage DNA by HK97 gp74. Upon incubation of lambda phage DNA with HK97 gp74 and divalent metals, we observed a smear of DNA. This suggests that in the presence of divalent metal ions (such as Ni$^{2+}$) HK97 gp74 mediates non-specific double-stranded digestion of λ phage DNA (Figure 21). Control reactions indicate that both HK97 gp74 and metal are required for DNA digestion, consistent with our previous observations.
Figure 21: HK97 gp74 cleaves λ phage DNA (48,502 bp). The DNA (25 μg/ml) was incubated with HK97 gp74 (24 μg/ml) with varying concentrations of metal in 20 mM HEPES, pH 7.0. Controls show no DNA digestion. Reactions show HK97 gp74 cleavage of lambda phage DNA that results in many bands along the length of the gel. This data provides evidence of double stranded phage DNA cleavage by HK97 gp74.

Lambda phage digestion assays were performed with various divalent metals, including Co²⁺, Ni²⁺, Mg²⁺, Ca²⁺ and Zn²⁺. Comparison and analysis of digestion assays performed with different metals reveals that some divalent metals do not cleave λ phage DNA as effectively, such as Zn²⁺ (as compared to cleavage reactions in the presence of Ni²⁺) (Figure 22). We observed that under the same conditions but with different metals at an
equivalent time interval, it appears that DNA is not digested to an equal extent. For example, with 0.5 mM Zn\(^{2+}\) it takes 4 hours to completely digest lambda phage DNA, whereas, with 0.5 mM Ni\(^{2+}\) all lambda phage DNA is digested after 2 hours.

Figure 22: HK97 gp74 cleavage of λ phage DNA (48 502 bp) with Zn\(^{2+}\). The DNA (25 µg/ml) was incubated with HK97 gp74 (24 µg/ml) with varying concentrations of metal in 20 mM HEPES, pH7.0. Controls show no DNA digestion. Reactions show HK97 gp74 cleavage of lambda phage DNA that results in many bands along the length of the gel. Digestion assays performed with Zn\(^{2+}\) reveals that some metals do not allow HK97 gp74 to cleave lambda phage DNA as effectively, as compared to results observed in the presence of Ni\(^{2+}\). We also see that at higher concentrations of divalent metal, HK97 gp74 cleavage of lambda phage DNA is inhibited.
Also, we observed that at higher concentrations of divalent metal cleavage of DNA by HK97 gp74 was inhibited (Figures 21 and 22). At concentrations of 5 mM divalent metal and higher, digestion of lambda phage DNA was inhibited by Zn\(^{2+}\) (Figure 22) and reduced with Ni\(^{2+}\) (Figure 21).

### 3.6.3 pH Dependence of HK97 gp74 Activity

We performed the lambda DNA cleavage assays at a pH range from 5 to 8 to test the cleavage activity of HK97 gp74 at different conditions. Because there are two histidines involved in the reaction of HNH endonucleases and if our protein is an HNH endonuclease, we expect a change in cleavage activity with lower pH. Agarose gel electrophoresis reveals that at pH 5 and 6, DNA digestion is inhibited (Figure 23). However, digestion of DNA is equally efficient at pH 7 and 8 and these results are consistent with the presence of histidine residues in the metal binding and catalytic site. Recall that the pKa of the imidazole group of histidine is approximately 6.7. Therefore, we expect the histidine involved in metal binding to be deprotenated. From the mechanism (shown on page 24), we expect histidine involved in catalysis to also be deprotenated.
Figure 23: pH Dependence of HK97 gp74 cleavage of λ phage DNA (48 502 bp). The DNA (25 µg/ml) was incubated with HK97 gp74 (24 µg/ml) with 0.5 mM divalent metal (Ni²⁺) in 20 mM HEPES, pH 5.0 - 8.0. Controls show no DNA digestion. At pH 5 and 6, DNA digestion is inhibited. Digestion of DNA is equally efficient at pH 7 and 8.
In order to assess whether lower pH induces protein unfolding, which would affect activity, circular dichroism (CD) was used to assess pH-dependent structural changes. The spectra show negative bands at 222 nm and 208 nm, which is indicative of α-helical structure. This is consistent and comparable with the predominantly α-helical structure of colicin E9. The CD spectra are similar from pH 7 to pH 8 and slightly different at pH 6. There are significant differences in CD spectra at pH 5, due to protein unfolding as indicated by more positive ellipticity (Figure 24). Therefore, lower activity at pH 6 is likely due to histidine ionization but at pH 5, we cannot rule out gross structural changes.
Figure 24: CD spectra of HK97 gp74 in 20 mM NaH$_2$PO$_4$, 50 mM NaCl at pH 5-8. A 600 µl sample containing 2 µM HK97 gp74 in 20 mM HEPES at a pH of 5.0 to 8.0 was prepared, in the absence of either DTT or β-mercaptoethanol. A CD spectrum was obtained after analysis of the samples using 5 scans per sample, measuring a range of absorbance from 190 nm to 260 nm, where the absorbance was measured every 0.2 nm at 25 °C. A control or “blank” was run under the same conditions containing 600 µl of 20 mM HEPES at a pH of 5, 6, 7 or 8. The spectra show negative bands at 222 nm and 208 nm, which is indicative of α-helical structure. The CD spectra are similar from pH 6-8 but significantly different at pH 5.

3.6.4 Metal Ion Stoichiometry & HK97 gp74 Activity

Our initial experiments used an excess of divalent metals. Thus, we wanted to assay HK97 gp74-mediated digestion of DNA with stoichiometric amounts of HK97 gp74 and metals to obtain insights into the mode of metal-protein binding. In these stoichiometric assays, 1 equivalent of protein was equal to 0.858 µM (or 11 µg/ml) and we varied the
divalent metal concentration from 1 to 10 equivalents to determine at what ratio we observe cleavage. We observed that cleavage occurs with stoichiometric amounts of metal and protein, suggesting that one divalent metal ion binds to a single HK97 gp74 protein unit for cleavage to occur (Figure 25).

Figure 25: Stoichiometric assay of HK97 gp74 cleavage of λ phage DNA (48,502 bp) with Ni$^{2+}$. In these reactions, 1 equivalent of protein is equal to 0.858 μM (or 11 μg/ml) and thus, we varied the divalent metal concentration from 1-10 equivalents. The DNA (25 μg/ml) was incubated with HK97 gp74 (11 μg/ml) with varying concentrations of metal in 20 mM HEPES, pH 7.0. Controls show no DNA digestion. At stoichiometric concentrations of divalent metal (Ni$^{2+}$) and HK97 gp74, cleavage occurs at 1 equivalent of metal and protein.
The UV-vis metal binding assay was used to measure the increase in the absorbance of free bases of DNA at a wavelength of 260 nm (A$_{260}$ nm) due to DNA digestion by HK97 gp74 in the presence of stoichiometric amounts of metal. Like the A$_{260}$ nm digestion assays, we observed no significant change in A$_{260}$ nm measurements over time. Furthermore, the measurements appeared to decrease in absorbance in the absence of divalent metal ions and fluctuated erratically as time passed with a difference in A$_{260}$nm (OD) of approximately 0.05 from the lowest measurement recorded to the highest measurement for reactions that included the presence of divalent metal (Figure 26).

Figure 26: UV-Vis assay of HK97 gp74 cleavage of λ phage DNA (48 502 bp) with Ni$^{2+}$. In these reactions, 1 equivalent of protein is equal to 0.858 µM (or 11 µg/ml) and we varied the divalent metal concentration from 0, 1, 5 and 10 equivalents. The DNA (25 µg/ml) was incubated with HK97 gp74 (11 µg/ml) with varying concentrations of metal in 20 mM HEPES, pH7.0. Controls were used to reduce any signal resulting from the buffer, protein or metal. Digestion of DNA is marked by the increase in absorbance at a wavelength of 260 nm due to the presence of free nucleotides. No significant changes in absorbance were observed and the absorbance appears to rise and fall slightly as time progresses.
Analysis of the UV-vis metal binding assay by agarose gel electrophoresis showed double-stranded DNA digestion and revealed that DNA digestion is visibly apparent after 1 hour of incubation at 37 °C in the presence of one equivalent of Ni\(^{2+}\) and HK97 gp74 (Figure 27).

Figure 27: Agarose gel of UV-Vis assay of HK97 gp74 cleavage of λ phage DNA (48 502 bp) with Ni\(^{2+}\). In these reactions, 1 equivalent of protein is equal to 0.858 µM (11 µg/ml) and we varied the divalent metal concentration from 0, 1, 5 and 10 equivalents. The DNA (25 µg/ml) was incubated with HK97 gp74 (11 µg/ml) with varying concentrations of metal in 20 mM HEPES, pH7.0. Reactions show HK97 gp74 cleavage of lambda phage DNA in the presence of divalent metal that results in many bands along the length of the gel.

3.7 Structural Characterization of HK97 gp74 by NMR Spectroscopy

A \(^{15}\text{N}^{-1}\text{H}\) correlation spectrum was recorded for HK97 gp74. The 2D spectrum correlates the resonant frequency of the amide proton and the resonant frequency of the directly attached amide nitrogen. In these spectra we see one peak per backbone NH and one
peak per side chain NH. We observed approximately 107 peaks in the spectrum, which correlate to our protein of 113 amino acid residues (Figure 28). However, this spectrum does not show proline residues, which accounts for the loss of peaks in our spectrum. The group of four peaks, three peaks that are resolved and one that is overlapped, at approximately 10 ppm in the $^{1}\text{H}$ dimension and at approximately 127-129 ppm in the $^{15}\text{N}$ dimension are from the indole NH in the four tryptophan residues in HK97 gp74 (Figure 28, indicated within a red box). There are also a set of “double” peaks at approximately 6.6 ppm to 7.5 ppm in the $^{1}\text{H}$ dimension and 109-114 ppm in the $^{15}\text{N}$ dimension that arise from the side chains of the four asparagine and six glutamine residues in our protein (Figure 28, indicated with blue dashed lines). The high degree of dispersion of the peaks in the spectrum suggests that the protein is in a folded conformation. However, the spectrum shows some sharp peaks that are centered at approximately 8.2 ppm in the $^{1}\text{H}$ axis, which is indicative of an unfolded region (Figure 28, indicated with green arrows). This unfolded region may be restricted to a single unstructured motif, such as a loop, within the otherwise stably folded protein. The structure of the HNH motif of colicin E9 shows a loop that in the metal-bound state forms hydrogen bonds (Figure 9). Thus, metal binding may stabilize some disordered structure in HK97 gp74.
Figure 28: $^{15}$N-$^1$H correlation spectrum of HK97 gp74 in 50 mM NaCl, 50 mM NaH$_2$PO$_4$, pH 7 at 25 °C. The 2D spectrum correlates the resonant frequency of the amide proton and the resonant frequency of the directly attached amide nitrogen. 2D $^1$H-$^{15}$N correlation spectra of HK97 gp74, indicates that the protein is folded. Approximately 107 of 113 backbone resonances are observed. NMR spectra show side chain correlations from 4 Asn and 6 Gln residues (-----) and 4 Trp residues (□□□□). The spectra also indicate that there are regions of disorder in the protein (←), such as the Asn-containing loop.
3.8 Analysis of HK97 gp74 Metal Binding

3.8.1 Metal Binding Titration Experiments using NMR Spectroscopy

Our results indicate that HK97 gp74 is a possible HNH homing endonuclease, which binds and cleaves DNA in the presence of divalent metal cations. The structural or conformational changes that occur as a result of divalent metal binding at the HNH motif or in the global structure of HK97 gp74 were investigated. Addition of metal at a high concentration of protein caused immediate precipitation of most of the sample. $^{15}$N-$^1$H correlation spectra for the apo-HK97 gp74 and metal bound HK97 gp74 samples were performed and recorded. However, comparison of the apo-HK97 gp74 $^{15}$N-$^1$H correlation spectrum and the metal-bound species of HK97 gp74 $^{15}$N-$^1$H correlation spectrum showed no change or shift of peaks in the spectra, likely because of precipitation of the metal-bound species with only the apo-HK97 gp74 protein left in solution (Figure 29). Nonetheless, high quality spectra of apo-HK97 gp74 indicate that our proposed structural studies are feasible.
Figure 29: $^{15}$N-$^1$H correlation spectra of apo-HK97 gp74 and Zn$^{2+}$ bound HK97 gp74 in 20 mM HEPES, pH 7 at 15°C. The apo-HK97 gp74 peaks appear in black and the Zn$^{2+}$ bound HK97 gp74 peaks appear in blue. 0.5 mM of Zn$^{2+}$ in 20 mM HEPES, pH of 7.0 was added to the 0.5 mM HK97 gp74 sample. The 2D spectrum correlates the resonant frequency of the amide proton and the resonant frequency of the directly attached amide nitrogen. Approximately 107 of 113 backbone resonances are observed in both spectra. NMR spectra show side chain correlations from 4 Asn and 6 Gln residues (-----) and 4 Trp residues (□□□). The spectra also show regions of disorder in the protein (►►►), such as the Asn-containing loop. There are few, if any, changes in spectra of HK97 gp74 with Zn$^{2+}$. Much of the sample precipitated upon addition of Zn$^{2+}$, indicating that perhaps the metal-bound species is insoluble and only apo-HK97 gp74 is in solution.

3.8.2 Metal Binding Studies using Tryptophan Fluorescence Spectroscopy

At high concentrations of protein required for NMR, metal bound species of HK97 gp74 are not soluble. Consequently, tryptophan fluorescence spectroscopy was used to test metal binding in order to correlate activity with metal dependence. HK97 gp74 contains a
total of four tryptophan residues, where the residue, Trp-111, is located directly adjacent to the metal binding histidine of the HNH motif. However, no conclusive evidence was obtained to show that metal binding at the HNH motif promotes conformational changes. Changes in tryptophan fluorescence of metal-bound HK97 gp74 were not significantly different than changes in tryptophan fluorescence of the control, which contained the protein in the absence of divalent metal (Figure 30). Furthermore, it is unclear why the tryptophan fluorescence intensity of HK97 gp74 in the presence of 3 µM Ni\(^{2+}\) is lower than the fluorescence intensity of HK97 gp74 in the presence of 1 µM Ni\(^{2+}\). Changes to the protocol (such as the concentration of protein or metal, temperature, buffer, controls, amount of sample scanned, etc.) or equipment (such as the sample plate, scanning attributes, etc.) may be required to obtain better tryptophan fluorescence intensity data in order to construct a complete conclusion about the metal binding interactions of HK97 gp74.
Figure 30: Tryptophan fluorescence spectra of metal-bound HK97 gp74. Reactions were prepared containing 0.5 µM HK97 gp74 in the presence of the divalent metal Ni$^{2+}$ at 1 µM, 3 µM and 5 µM in 20 mM HEPES, pH 7.0. A control consisted of 0.5 µM HK97 gp74 in 20 mM HEPES, pH 7.0. A Tecan Infinite M1000 Fluorescence Plate Reader (Gunning Group) was used to excite 60 µL of each sample at a wavelength of 280 nm at 24 °C and the emission fluorescence intensity was measured from 300 nm to 360 nm.
4. Discussion and Conclusions:

4.1 HK97 gp74

At the beginning of this project, the HK97 bacteriophage protein gp74 was uncharacterized and previously had no known function or known structural characteristics. The purpose of this study was to investigate the structural and possible functional role of the HK97 bacteriophage protein gp74. Our preliminary search of proteins with similar sequences to HK97 gp74 using the basic local alignment search tool (BLAST) indicated that HK97 gp74 is a possible homing HNH endonuclease. Homing endonucleases bind double-stranded DNA and cause double-stranded breaks in alleles that are homologous to the endonuclease gene but lack the intron or intein element, which encodes the gene for the homing endonuclease.\(^\text{33}\)

There are four main families of homing endonucleases, each of which are named for the conserved residues associated with their nuclease domain, which include the LAGLIDADG, the His-Cys box, HNH enzymes and GIY-YIG family of enzymes.\(^\text{28}\) A top BLAST search hit was the HNH endonuclease colicin E9, which has been extensively studied and set the foundation of our functional investigations of HK97 gp74. Subsequent structure-based sequence alignment and analysis provided us with evidence that HK97 gp74 contains conserved histidine and asparagine residues that likely compose the HNH motif of HK97 gp74. In HK97 gp74, the conserved residues are H82, N101, and H110 and appear to correlate to the HNH motif found in the bacterial protein, colicin E9 (Figure 11). Thus, basic sequence alignment results obtained in this study suggest that HK97 gp74 protein belongs to the HNH-endonuclease family of enzymes, which are conserved among bacteria and viruses.
The HNH class of homing endonucleases are named for two conserved histidine residues and an invariant asparagine residue that comprise the HNH motif. This family is known to include the pyocins, colicins and anaredoxins, which are bacterial proteins that inhibit the growth of similar or related bacteria. For example, colicins are non-specific bacterial HNH endonucleases and comparison to these family members may suggest that HK97 gp74 may also play a role as a bacteriocin.

### 4.2 Role of HK97 gp74 HNH Endonucleases

Homing endonucleases are highly conserved in bacterial populations and have been suggested to play a significant role in the homologous recombination of newly integrated genes from one bacteria to another. However, very little is known as to the mechanism of integration of the endonuclease gene among bacterial species and why this type of protein is conserved among bacterial species. However, we hypothesize that the existence of an HNH homing endonuclease in bacteriophages would provide a mechanism by which phage genes are incorporated into the bacterial genome. It is known that phage invasion and genomic integration through the lysogenic life cycle leads to the formation of prophage elements, and this process may be a major contributor to genomic diversity amongst bacterial species. The HK97 gp74 protein may play a possible role in the phage lysogenic cycle. From the gene map of the HK97 genome sequence, neighbouring genes to gp74 are involved in DNA repair mechanisms and lysis of bacterial cells (Figure 5). The existence of genes involved in lysis of bacterial cells and degradation of peptidoglycan, such as the S gene that encodes for a holin protein and the R gene that encodes for an endolysin, further support the theory that
HK97 gp74 is involved in phage infection processes.\textsuperscript{31} Furthermore, gp69 is hypothesized to be closely related to the RusA gene of lambda phage, which encodes for a resolvase that ligates nucleic acid fragments at Holliday junctions.\textsuperscript{30} This offers evidence to support the theory that HK97 gp74 is involved in the lysogenic cycle and more specifically that HNH homing endonuclease activity of the HK97 gp74 protein would allow for the incorporation and recombination of phage genes into the bacterial genome. Consequently, we suggest that HK97 gp74 creates double stranded DNA breaks, thereby, initiating bacterial host cell repair machinery that allows for homologous recombination at the cleaved DNA site to occur and resulting in integration of the phage genome.

Furthermore, many phage endonucleases have been observed to target and cleave other phage DNA. This strategy may be a result of competition between bacteriophage species or mutualism with bacterial species to protect against infection and cell death. A consequence of our study has lead us to suggest that HK97 gp74 may be responsible for the digestion of foreign phage DNA that pose a competitive disadvantage to the integration and proliferation of HK97 bacteriophage particles in a bacterial host and to protect the host cell from invasion of foreign phage that may be virulent or bactericidal. Thus, HK97 gp74 is an important target for study to understand the role of bacteriophages and their influence and manipulation of bacterial populations. Future studies in biological function of HK97 gp74 may discriminate between these and other possibilities.
4.3 Future Biochemical & Biophysical Studies

The structural part of the HK97 bacteriophage protein gp74 project encompasses our work to date in protein biochemistry to generate suitable samples and screen solution conditions for our biochemical and NMR studies. A $^{15}$N-$^{1}$H correlation spectrum was recorded for HK97 gp74 in the absence of metal and provided evidence that HK97 gp74 is in a folded conformation. Future NMR studies involving the divalent metal binding of HK97 gp74 using Zn$^{2+}$ or Ni$^{2+}$ ions at stoichiometric concentrations will help to elucidate the conformational changes associated with divalent metal binding at the HNH motif of HK97 gp74. Because samples of gp74 at high concentrations with metal precipitate, preparation of metal-bound samples need to be investigated and may involve changing the buffer or adding metal in dilute solutions and concentrating the metal-bound species to NMR concentrations. We would also like to solve the solution structure of HK97 gp74 in order to gain insights into the mechanism of DNA digestion. Resonance assignments will also be useful to study metal-binding and DNA binding (provided that catalysis deficient mutants are amenable to NMR studies).

Our biochemical and biophysical studies indicate that HK97 gp74 is a possible bacterial homing HNH endonuclease. One of the future goals in this project is to identify residues involved in DNA binding and those involved in DNA digestion. Our digestion assays at different pH values suggest that protonation of one or both His residues decrease DNA digestion. Protonation of the His residue involved in metal binding would likely decrease the affinity for the metal, whereas protonation of the catalytic His would decrease the catalytic efficiency of gp74. One of the first steps in the HNH endonuclease DNA
digestion mechanism is the abstraction of a proton from a water molecule by the catalytic His (Figure 10 in Introduction 1.4 Homing Endonucleases), with the resulting hydroxide ion attacking the phosphodiester DNA backbone.\textsuperscript{28} A protonated His residue would be less efficient in this step.

In order to determine the protonation state of each His residue in the HNH motif at physiological pH (pH 7), we need to know their specific pKas. NMR spectroscopy is an ideal tool for measuring pKa values. Because we do not currently have resonance assignments for HK97 gp74, we will need to specifically label our protein with \textsuperscript{15}N-labeled His residues. \textsuperscript{15}N-\textsuperscript{1}H HSQC spectra of \textsuperscript{15}N-His gp74 would consist of only 3 peaks. Titration with acid would result in protonation of the His residues and concomitant chemical shift changes, which can be used to determine the pKa values. A separate pH titration on uniformly \textsuperscript{15}N-labeled gp74 would need to be done to ensure that the protein is folded at different pH values, and hence that the pKa values measured reflect changes in the protonation state of the His residues and not global structural changes.

Changes in backbone \textsuperscript{15}N-\textsuperscript{1}H chemical shifts with His side chain protonation state may not be large enough due to the distance of the His side chain from the backbone. Further, backbone chemical shift changes may be influenced by ionization state of other groups as the chemical shift dependence on through-space effects of titratable groups has a 1/r\textsuperscript{2} dependence and can be long range.\textsuperscript{57,58} Thus, it would be advantageous to have side chain probes for measuring pKa values. We could also do a similar experiment with gp74 that is uniformly labeled with \textsuperscript{15}N that also contains \textsuperscript{19}F-labeled His residues at the N\textsubscript{\textbeta} position. In this way, we could assess structural changes and changes in pKa values with one
sample. 2D $^{15}\text{N}^{1}\text{H}$ HSQC experiments would assess the folded state of the protein and 1D $^{19}\text{F}$ spectra would probe protonation state. Perturbations of $^{19}\text{F}$ to protein structure are minimal and $^{19}\text{F}$ probes have a wide range of chemical shifts associated with different local electronic environments in folded proteins. However, $^{19}\text{F}$ probes in His residues can shift the pKa values by more than 2 units, which may affect the biochemical characteristics of the protein. Thus, use of $^{19}\text{F}$ probes must be carefully analyzed.

We could also measure pKa values in His by using $^{1}\text{H}^{15}\text{N}$ heteronuclear multiple bond correlation (HMBC) NMR experiments. HMBC experiments provide information about weak proton-carbon or proton-nitrogen interactions. Thus, carbon-bound protons that are two, three or four bonds away from the heteroatom of interest are observed. Thus, these experiments require $^{15}\text{N}$ and $^{13}\text{C}$ labeled protein. In a $^{1}\text{H}^{15}\text{N}$ HMBC spectra of His residues, the chemical shift of the $^{15}\text{N}$ resonance is correlated to the protonation state of the imidazole nitrogen. Protonated nitrogens are observed at ~170 ppm, whereas, deprotenated nitrogens are observed at ~250 ppm. In addition, we can also differentiate between two neutral His tautomers using these experiments. Different spectral patterns are produced in the $^{15}\text{N}^{1}\text{H}$ HMBC experiment due to different two-bond and three-bond $^{1}\text{H}^{15}\text{N}$ coupling constants in each state.

Thus, there are a number of different approaches that we can use to obtain pKa values for His residues in gp74. In absence of backbone and side chain resonance assignments, identification of specific resonances will require making His point mutants and recording spectra to obtain the assignments. Nonetheless, NMR spectroscopy provides us with many options to determine pKa values of the histidine imidazole groups involved in catalysis and
metal binding. These experiments are planned and will yield additional information regarding the biochemical function and mechanism of gp74.

As stated above, future work will also include the protein biosynthesis of mutant HK97 gp74 proteins. Mutant HK97 gp74 proteins will be designed to have alanine residues in the place of the conserved residues of the putative HNH motif, to confirm that HK97 gp74 is an HNH endonuclease and to determine which residues are essential for binding and cleavage of DNA by HK97 gp74. Mutagenesis studies are currently underway and the mutant H82A-HK97 gp74 was designed and synthesized by PCR thermal cycling. Agarose gel electrophoresis shows the appearance of the control product, in addition to, mutant HK97 gp74-pET-15b plasmid DNA that appears at 6 kB (Figure 31). The expected mutant pET-15b-H82A-HK97 gp74 plasmid has a size of 6.1 kB. The mutant DNA will be used in future protein expression studies to produce mutant versions of HK97 gp74, in which the conserved residues of the HNH motif have been changed to alanine in order to study the decrease or lack of activity of DNA digestion, as well as, divalent metal binding at the active site. Furthermore, a triple mutant will be used to study the total inhibition or severe reduction in cleavage activity. We hope to perform our DNA digestion assays to compare the wild type activity to the mutant enzymes.
Figure 31: Agarose gel of pET-15b-H82A-HK97 gp74 PCR reactions. A control reaction was prepared using 2 µL (10 ng) of pWhitescript control plasmid, 1.25 µL (125 ng) each of oligonucleotide primer #1 and #2 and 1 µL of dNTP mix. Sample reactions were prepared using varying amounts of pET-15b-HK97 gp74 vector DNA (from 5 ng, 10 ng, 25 ng and 50 ng), 1.25 µL of each primer stock at a concentration of 100 ng/µL, and 1 µL 20 mM dNTP mix. Pfu turbo DNA polymerase was added to each of the PCR reactions. A PCR protocol consisting of DNA denaturation at 95 °C for 30 seconds, followed by annealing at 55 °C for 1 minute and extension at 68 °C for 12 minutes and 6 seconds. This was repeated for 16 cycles and ended at 4 °C. Agarose gel electrophoresis showed the appearance of the control product and mutant pET-15b-HK97 gp74 plasmid DNA.
In order to determine the efficiency of cleavage by HK97 gp74, we would like to compare the activity of gp74 cleavage against the activity of a control endonuclease. Colicin E9 is an extensively studied HNH endonuclease that is closely related to HK97 gp74 (see Figure 11 in Results 3.2 Structure Based Sequence Alignment). A BLAST search and our structure-based sequence alignment highlight the similarities of HK97 gp74 and colicin E9, particularly in the $\beta\beta\alpha$ structural motif (Figure 11). Consequently, as an HNH endonuclease that is structurally similar to HK97 gp74, the bacterial protein colicin E9 would be the best choice as a control endonuclease to compare against the DNA cleavage activity of HK97 gp74. Colicin E9, like HK97 gp74, also shows a broad range of metal dependent DNA digestion but catalytic activity is greatest in the presence of Ni$^{2+}$. Colicin E9 also displays Michaelis-Menten kinetics and in the presence of Ni$^{2+}$ and calf thymus DNA, the $K_m$ was determined to be 0.93 M. Colicin E9 DNase is known to cleave DNA preferentially after thymine residues of randomly designed DNA primers. The kinetics of DNA digestion by HK97 gp74 and DNA specificity is currently not known, but these studies comprise future experiments. In further tests, colicin E9 was also shown to mediate double stranded digestion of supercoiled pUC-18 DNA, whereas our data indicate the HK97 gp74 mediates only single-stranded digestion of supercoiled pUC-18 DNA (see Results 3.6.1 HK97 gp74-Mediated Digestion of Plasmid DNA). Thus, a comparison of the enzymatic properties of colicin E9 and HK97 gp74 will help determine features common to these HNH endonucleases, but will also highlight differences which may determine the \textit{in vivo} specificity of these enzymes.

We would also like to study the affinity of HK97 gp74 for different divalent metals to determine if different metal-binding affinity correlates with activity. We would like to use
tryptophan fluorescence for these experiments since HK97 gp74 has a Trp residue located directly adjacent to the metal binding site of the HNH motif. Therefore, it is likely that the tryptophan fluorescence spectra of free and bound protein differs.

Our future work will yield insights into the function of this protein and its role in the HK97 phage. We would like to elucidate the significance of each amino acid residue in the HNH motif and to determine which residue binds metal and which catalyzes activity. In addition, we would like to determine the host or DNA specificity of HK97 gp74 using different DNA sources.

4.4 Future Application of HK97 gp74

Understanding the HNH endonuclease function of phage proteins will also help elucidate the molecular variability in bacterial populations, which is critical for fighting bacterial infections and diseases. Future applications of the data obtained during investigation of HK97 gp74 should demonstrate and lead to the development of phage-based approaches to combat bacterial diseases. For example, like antibiotics, phage proteins can target bacterial cells and through the introduction of bactericidal factors can cause bacterial cell death. As an endonuclease, of which we do not know the host DNA substrate specificity and activity, HK97 gp74 may in fact cleave bacterial genomes at many sites or can be bioengineered to include recognition sequences from other nucleases in order to digest bacterial genomes. This strategy may help to fight off bacterial infections or the spread of bacterial diseases, such as pneumonia, which is the result of infection by pathogenic bacterial
species, like *Streptococcus* and *Pseudomonas* bacteria. Phage therapy may provide the solutions required to combat some of the most challenging modern medical problems, such as multidrug resistant bacteria.

Homing endonucleases may also offer a new perspective on gene therapy using phage protein activity. Moreover, better understanding of HK97 gp74 can enable its use as a possible homologous recombination system to allow for the integration of genes. For example, a bioengineered vector could be constructed to recover deletion mutations that are involved in the development of diseases, such as Duchenne muscular dystrophy. Duchenne muscular dystrophy is a form of muscular dystrophy that is caused by a mutation in the DMD gene, which encodes the protein dystrophin, a structural component of muscle tissue.\(^{61,62}\) HK97 gp74 as a homing endonuclease offers a mechanism for the integration of genes. Bioengineering of the HK97 gp74 gene that incorporates recognition sequences and functional alleles could act as a method to treat diseases that are caused by deleterious mutant alleles. Release of the vector into a somatic cell would allow for expression of the HK97 gp74 fusion protein. Recognition sequences would direct the protein to the correct site in the genome. Endonuclease activity of HK97 gp74 at a target location would lead to double-stranded DNA breaks, which would initiate DNA repair mechanisms. This mechanism would allow for the integration of the vector into the cell genome by homologous recombination, whereby the mutant allele would be replaced by a functional allele of the same gene.
4.5 Conclusions

A protein-protein BLAST search and structure based sequence alignment provided the first indication of protein function. We determined that the bacteriophage protein HK97 gp74 is related to the homing HNH endonuclease family of proteins. Functional studies pertaining to the nuclease activity of HK97 gp74 with divalent metals provided us with the first visible evidence to conclude that HK97 gp74 digests DNA. In the presence of divalent metal ions, such as Ni$^{2+}$, HK97 gp74 mediates single-stranded cleavage of plasmid and phagemid DNA. At this time, we hypothesize that gp74 may mediate integration of phage genes into the host genome. Furthermore, cleavage assays involving phage DNA, in the presence of divalent metal ions (such as Ni$^{2+}$), revealed that HK97 gp74 mediates non-specific double-stranded digestion of λ phage DNA. Comparison of digestion assays performed with different metals reveals that some divalent metals do not cleave λ phage DNA as effectively, such as Zn$^{2+}$ (as compared to cleavage reactions in the presence of Ni$^{2+}$). Also, we observed that cleavage of DNA by HK97 gp74 is inhibited at higher concentrations of divalent metal, which is comparable to results obtained for other small metal binding proteins. Recent investigation of the stoichiometry of divalent metal binding to HK97 gp74 suggests that one divalent metal ion binds to a single HK97 gp74 protein unit for cleavage to occur. Moreover, the pH dependent analysis of DNA digestion reveals that at pH 5 and 6, DNA digestion is inhibited, whereas, digestion of DNA is equally efficient at pH 7 and 8. Circular dichroism (CD) was used to assess pH-dependent structural changes, where the CD spectra is similar from pH 7 to pH 8 and slightly different at pH 6, which may be due to histidine ionization, and significantly different at pH 5, possibly due to protein unfolding.
Through the use of protein biosynthesis and analysis, we have shown that high levels of HK97 gp74 can be solubly expressed in *E. coli* BL21 Star (DE3) cells in minimal M9 media, which is required for isotopic enrichment of proteins with $^{15}$N nuclei for NMR structural experiments. A $^{15}$N-$^1$H correlation spectrum was recorded for HK97 gp74 and provided evidence that HK97 gp74 is in a folded conformation. The $^{15}$N-$^1$H correlation spectrum shows some sharp peaks that are centered at approximately 8.2 ppm in the $^1$H axis, which is indicative of an unfolded region, which may be a single unstructured motif, such as a loop. The high quality of our spectra will enable further NMR experiments to characterize the structure and function of HK97 gp74.

Thus, we have obtained preliminary results that serve as the basis to understanding the function and determining the solution structure of HK97 gp74. Investigation of the protein HK97 gp74 has lead to the identification of a new endonuclease of bacteriophage origin. Moreover, we have determined that HK97 gp74 requires a divalent metal cofactor for digestion of DNA substrates. We have determined that the HK97 bacteriophage gp74 protein cleaves phage DNA non-specifically, which may suggest that HK97 gp74 targets foreign phage DNA as a defensive mechanism to protect against competition by other phages in the surrounding environment.
5. References


BLAST (2009).


Edward, J. (Process NMR Associates).


6. Appendix 1

6.1 Sample Calculation of Amino Acid Analysis

Amino analysis results of the amount of each amino acid residue (in pmoles) were obtained from the Amino Acid Analysis Facility at the Advanced Protein Technology Centre, The Hospital for Sick Children. The data was summarized in a table along with the number of known residues and concentration of each amino acid.

Table 1: Summary of the data obtained from amino acid analysis. The table shows the amino acids and their corresponding amount (in pmoles), the number of each residue in the HK97 gp74 protein sequence and the concentration of each amino acid.

<table>
<thead>
<tr>
<th>Amino Acid</th>
<th>pmol analyzed</th>
<th>Known Residues</th>
<th>Concentration</th>
</tr>
</thead>
<tbody>
<tr>
<td>Asp/Asn</td>
<td>6455.47</td>
<td>10</td>
<td>645.55</td>
</tr>
<tr>
<td>Glu/Gln</td>
<td>7381.36</td>
<td>11</td>
<td>671.03</td>
</tr>
<tr>
<td>Ser</td>
<td>4381.57</td>
<td>8</td>
<td>644.35</td>
</tr>
<tr>
<td>Gly</td>
<td>4600.39</td>
<td>7</td>
<td>657.20</td>
</tr>
<tr>
<td>His</td>
<td>4648.55</td>
<td>7</td>
<td>664.08</td>
</tr>
<tr>
<td>Arg</td>
<td>4564.42</td>
<td>6</td>
<td>760.74</td>
</tr>
<tr>
<td>Thr</td>
<td>2446.38</td>
<td>3</td>
<td>959.37</td>
</tr>
<tr>
<td>Ala</td>
<td>5948.50</td>
<td>9</td>
<td>660.94</td>
</tr>
<tr>
<td>Pro</td>
<td>550.35</td>
<td>8</td>
<td>68.79</td>
</tr>
<tr>
<td>Tyr</td>
<td>1423.04</td>
<td>2</td>
<td>711.52</td>
</tr>
<tr>
<td>Val</td>
<td>3923.15</td>
<td>6</td>
<td>653.86</td>
</tr>
<tr>
<td>Met</td>
<td>2104.90</td>
<td>3</td>
<td>701.63</td>
</tr>
<tr>
<td>Cys</td>
<td>390.79</td>
<td>4</td>
<td>97.70</td>
</tr>
<tr>
<td>Ile</td>
<td>2426.00</td>
<td>4</td>
<td>606.50</td>
</tr>
<tr>
<td>Leu</td>
<td>6414.75</td>
<td>10</td>
<td>641.48</td>
</tr>
<tr>
<td>Phe</td>
<td>1269.55</td>
<td>2</td>
<td>634.78</td>
</tr>
<tr>
<td>Lys</td>
<td>4989.73</td>
<td>10</td>
<td>498.97</td>
</tr>
</tbody>
</table>

The concentration of each individual amino acid was determined by dividing the amount of each amino acid (in pmoles) by the number of each corresponding amino acid in
the HK97 gp74 protein sequence. The concentration of the sample was determined by first finding the sum of the individual amino acid residue concentrations. Outliers, such as proline and cysteine, which would skew the data, were removed from the summation. The corrected sum was determined to be 674.13 pmol. The corrected sum of amino acid residue concentrations was then divided by the volume of sample used (10 µl) in amino acid analysis.

Sample Calculation for the concentration of HK97 gp74 protein sample:

\[
\text{Average concentration} = \frac{\text{Σ amino acid residue concentrations}}{\text{volume of sample}}
\]

Average concentration = 674.13 pmol / 10 µl

Average concentration = 67 pmol/µl (or 67 µmol/L)

Therefore, the concentration of the HK97 gp74 sample was 67 µM.