Bacteriophages and a homology model of a large terminase bacteriophage protein

Brittany Grenyer, Morgan Sperratore, Melinda Harrison*

Cabrini University, 610 King of Prussia Road, Radnor, PA 19087, USA

Received: September 24, 2018; accepted: December 13, 2018.

A bacteriophage is a virus that is able to infect and replicate inside a bacterium host. Catdawg is a *mycobacteriophage* belonging to Cluster O and the siphoviridae family. GP30, the large terminase protein of Catdawg, is responsible for the cleavage of the DNA and its translocation into the procapsid. There are two components of the large terminase protein which include the ATPase domain and the nuclease domain. The structure of GP30, the large terminase protein is unknown and the amount of structural information on phages is limited. In this study, homology modeling of GP30 was performed and the model was then compared to known crystal structures. The top result was a DNA packaging protein, GP-17 from bacteriophage T4. The predicted structure was then validated using Procheck. These results confirm that the structure of GP 30, the large terminase protein will contain a similar conformation as the known terminase crystal structures.

Keywords: Bacteriophages; homology model; large terminase protein.

Abbreviations: ATP: Adenosine triphosphate; ADP: Adenosine diphosphate; KB: Kilobases; Jelly roll: strand antiparallel β-barrel; DNA: deoxyribonucleic acid; RNA: ribonucleic acid; GP: Gene product; PDB: Protein data bank; RMSD: root-mean-square deviation; Å: Ångström.

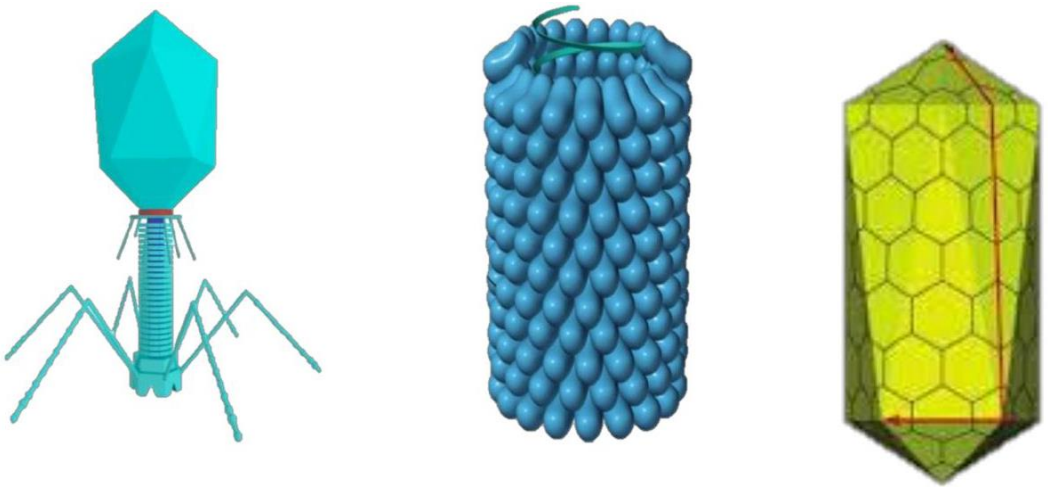
Financial support: This research project was supported by Cabrini University.

***Corresponding author:** Melinda Harrison, Cabrini University, 610 King of Prussia Road, Radnor, PA 19087, USA. E-mail: Melinda.a.harrison@cabrini.edu.

Introduction

A bacteriophage is a virus that infects and replicates inside a bacterium host. They were first discovered in 1915 by William Twort, and in 1917 Felix d'Herelle identified their ability to kill bacteria [1]. Bacteriophages are one of the most common and diverse organisms in the entire biosphere [2]. They are important in the balance of microbial life [2]. Bacteriophages are not able to carry out metabolic processes due to their inability to generate ATP which is why they are dependent on their bacterial hosts [3]. They can be found in the soil, ocean, and intestines of animals. Bacteriophage studies have provided the foundation for studies in molecular biology, including those that examine amino acid codon assignment and the identification of restriction enzymes [1].

There are numerous classes of bacteriophages which are characterized by their genomes and their morphology. The families are categorized into two known orders, Caudovirales and Ligamenvirales. The families within the Ligamenvirales include lipothrixviridae and rudiviridae, while Caudovirale includes myoviridae, podoviridae, and siphoviridae [4]. Although there are numerous families of bacteriophages, they are all similar in composition, consisting of a capsid (head), a tail, and a nucleic acid genome [5]. The phage nucleic



http ://cronodon.com/images/t4_model_v3.jp g

http s://classconnection.s3.amazonaws.com/698/flashcards/1100698/jp g/helical1330008244293.jp g

http ://p ubs.rsc.org/en/content/articlelanding/2008/jm/b805874c

Figure 1. This figure represents the various capsid morphologies of bacteriophages. The first depiction is a bacteriophage with an icosahedral capsid, the second is a helical capsid, and the third is a prolate capsid, which is an elongated icosahedral head. The bacteriophage of interest in this study, Catdawg, belongs to the siphoviridae family, which contain icosahedral capsids.

acid genome is comprised of either DNA or RNA and can be single- or double-stranded in nature.

Structurally, the head and the tail are connected by a collar protein and an adaptor protein. The head is composed of capsid proteins which help to protect the genomic material within from degradation [6]. There are currently nineteen families of phages which are recognized to infect bacteria and archaea. Of the nineteen families, two of them contain RNA genomes and five are enveloped [7]. An envelope is an outer membrane which is derived from the host that surrounds the capsid to further protect the genetic material from degradative proteins [5]. Of the seventeen families which contain DNA genomes, only two are single-stranded. Eight of the families have circular genomes while the others contain linear genomes. Nine of the families infect only bacteria, nine infect only archaea, and only one infects both [7].

The capsid is a protein shell in which the genetic material is contained. Most viruses self-assemble within the cell, packaging the viral genome, viral proteins, and relevant host partners within a protein capsid for export from the cell. The capsid not only protects the genome, but it is also responsible for maintaining the stability of protein-protein interactions which are required for viral assembly [8]. There are five major capsid geometries or symmetries which include: icosahedral, prolate, helical, binary (phages with a head and tail), and pleomorphic (Figure 1). An icosahedral capsid contains twelve pentamers and varying numbers of hexamers both of which are the structural units of a viral capsid. A prolate capsid is an elongated icosahedral along a fivefold axis, and a helical capsid is rod shaped [5]. An icosahedral capsid contains 20 equilateral triangular facets, 12 vertices which correspond to a fivefold symmetry axis. Ten three-fold symmetry axes pass through the 20 facets and 15 two-fold axes pass through the 30 edge-to-edge contacts between each of the facets. One facet of an icosahedral capsid consists of at least three proteins since proteins cannot have intrinsic three-fold symmetry of an equilateral triangle. [8].

Evolutionary speaking, there are a few structural motifs which are favored in the formation of viral capsid proteins [8]. A common protein fold is the

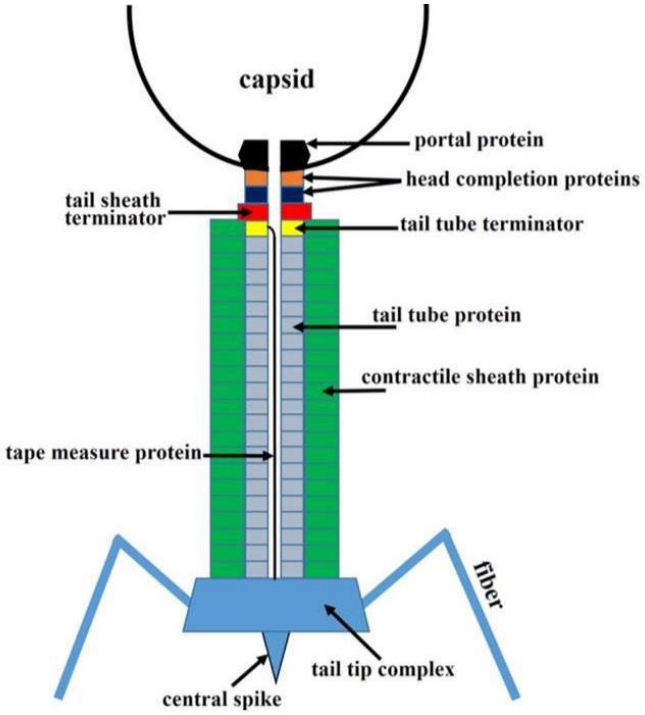


Figure 2. This figure indicates the main structural components of a myovirus phage. The protein of interested is the large terminase protein, which is located within the tail tube terminator (figure from Fokine & Rossmann, 2014).

eight-strand antiparallel β-barrel or "jelly roll". The "jelly roll" is found in numerous viruses including RNA picornaviruses, DNA parvoviruses, plant viruses, DNA polomaviruses, papillomaviruses, and even some DNA bacteriophages. The common folding patterns were first discovered in the bacteriophage HK97 which contains a unique alpha helical fold concatemer [8]. A virus must assemble with high fidelity to form a successful viral unit. The correct capsid geometry and the interaction stability is not only important for the formation of a protein shell for packaging but also for the regulation of assembly. The capsid is responsible for the activation of certain subunits to trigger assembly and packaging of the genetic material [8].

The bacteriophage must assemble a stable and protective protein shell (capsid) which is able to defend itself from a variety of chemical and physical agents which include radiation, pH fluctuations, temperature changes, and proteolytic and nucleolytic enzymes. Some bacteriophages also contain a tail structure which varies in length and size. The tail is a hollow, helical tube in which the genetic material passes through into the bacteria during infection. Connecting the capsid and the tail are adaptor proteins and collar proteins which include the portal protein, head completion proteins, tail sheath terminator (terminase), and the tail tube terminator indicated in Figure 2. The portal protein and the terminase protein work together to package the DNA into the capsid of the bacteriophage. The portal protein aids the packaging of DNA while the terminase protein cleaves the DNA when the concatemer has reached capacity within the capsid [9]. Concatemers are long connected sequences of phage DNA which are replicated through the rolling circle of replication. *Cos* sequences are located in the beginning and the end of the DNA concatemer sequences. These sequences are the sites in which the DNA is cleaved during the packaging of the DNA into the head [9]. Figure 3 shows how the DNA is packaged into the viral capsid.

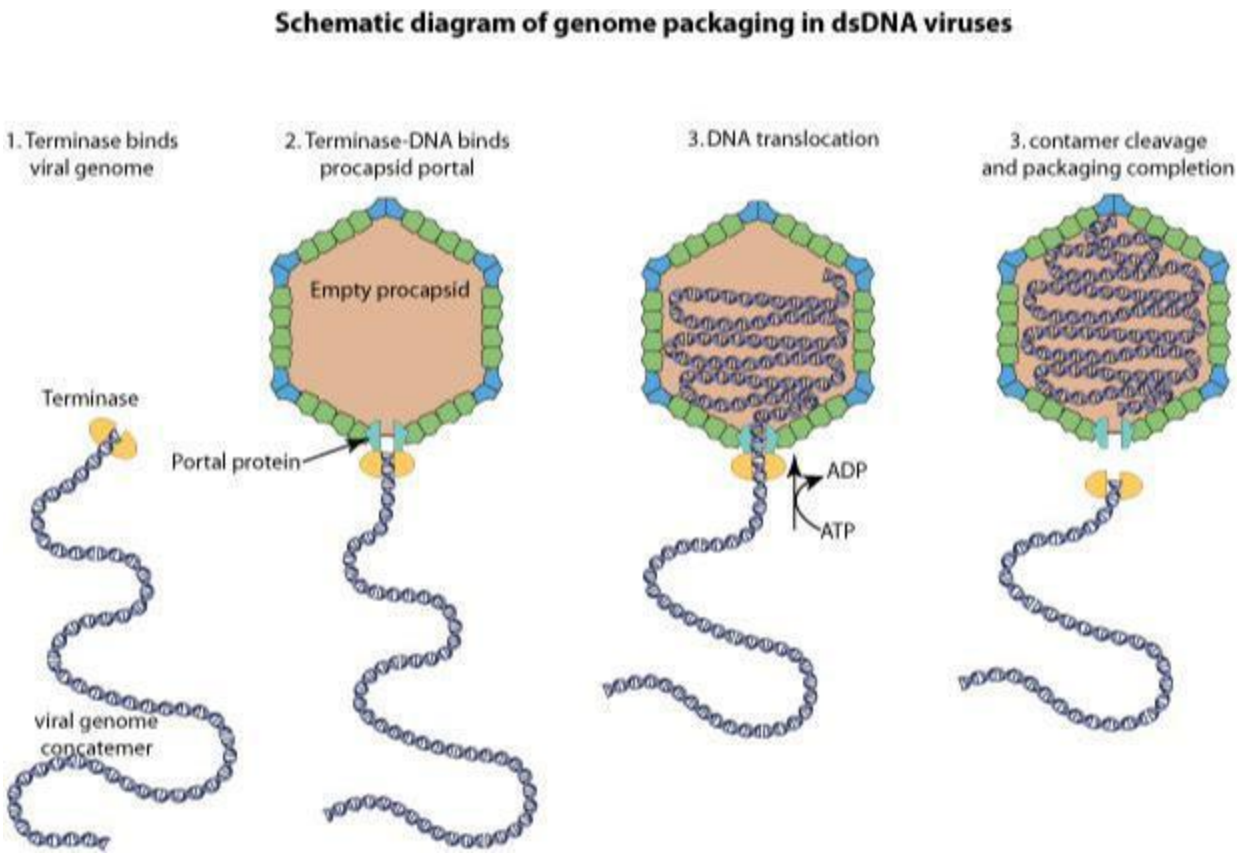


Figure 3. This figure represents the depiction of genome packaging of double-stranded DNA viruses. These viruses contain DNA machinery which aid in the translocation of the viral genome into the procapsid. The machinery consists of a portal protein, which provides the entry for DNA and an ATP-driven motor, which is the driving force behind DNA translocation. The motor consists of the large terminase protein which contains two subdomains, ATPase and nuclease. The ATPase subdomain is responsible for the movement of DNA while the nuclease subdomain cleaves the DNA.

Large terminase protein

The large terminase protein is responsible for the cleavage of the DNA and its translocation into the procapsid. There are two components of the large terminase protein, which include the ATPase domain and the nuclease domain. The small terminase protein is involved in the initiation of packaging and stimulates the ATPase domain of the large terminase protein. The packaging method is initiated with the small terminase protein, which recognizes the DNA and recruits the large terminase protein to make the initial cut. Once the DNA concatemer is cut, both the terminase proteins associate with the cleaved end of the concatemer which then bind to the portal vertex of the procapsid. The DNA packaging into the procapsid is catalyzed by the motor which consists of the large and the small terminase protein. These proteins assemble into a dodecamer of a portal protein to form a complex. The motor then packages the DNA into the empty procapsids with high efficiency and proficiency. The ATPase domain of the large terminase hydrolyzes ATP to power the translocation of the DNA into the procapsid. Once the DNA is packaged into the capsid, the large terminase protein cuts the DNA to separate the packaged DNA from the remaining concatemer [9]. The repeating concatemer typically contains up to 10 copies of the phage genome. The concatemer is produced by the rolling circle of replication and the DNA is circularized by homologous recombination.

Homologous recombination is a genetic process whereby nucleotide sequences are exchanged

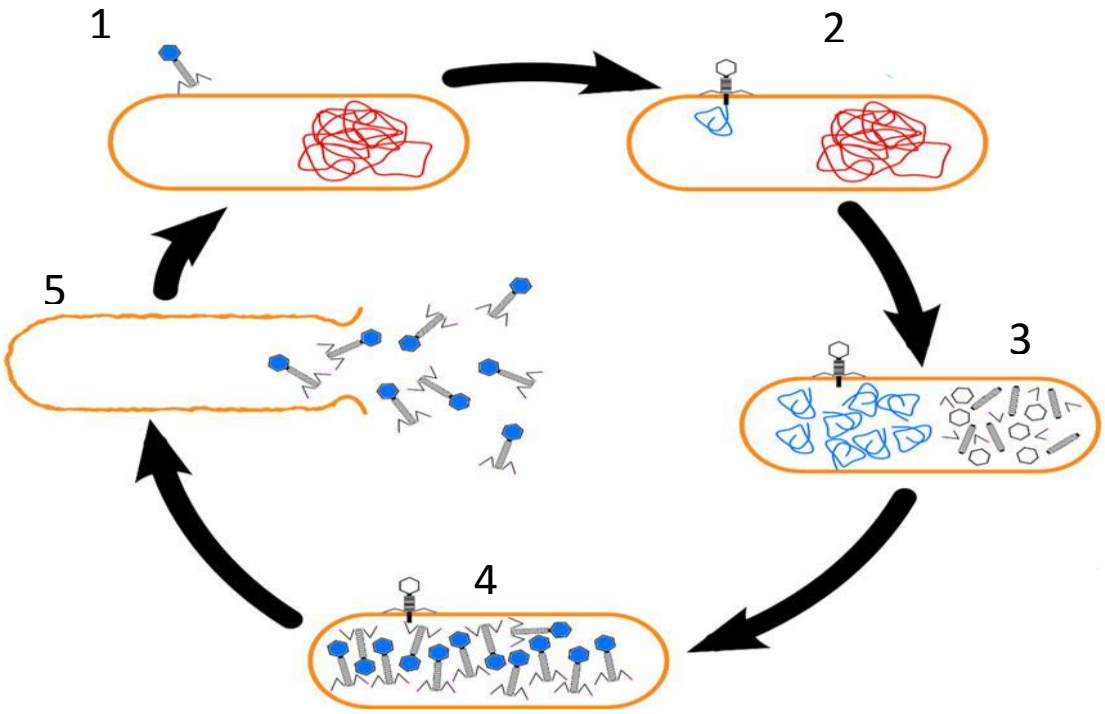


Figure 4. This figure depicts the lytic life cycle of bacteriophage. 1: First step in the lytic life cycle where the bacteriophage adheres to the surface of the bacteria. 2: Second step where the genomic material contained within the capsid is released into the bacterium. 3: Third step where the viral machinery is replicated within the host bacterium. 4: Fourth step where the viral particles are assembled. 5: Fifth step where the newly assembled phages are lysed from the bacteria. Once the phages are lysed from the bacterial host they will adsorb and infect other bacteria. (http://agrilifecdn.tamu.edu/jjgill/files/2013/ 04/phage-life-v iru lent.png)

between two similar or identical molecules of DNA. This process is commonly used to fix double-stranded breaks in a piece of DNA, which brings two ends together and ligates them. The phage DNA is packaged at the *pac* site which is located within gene 3 that encodes the small terminase protein. The nuclease domain of the large terminase protein then breaks the DNA backbone at several points which is then inserted into the procapsid. The small terminase protein recognizes the *pac* site and presents it to the motor which initiates the translocation of the DNA into the procapsid [10]. The amount of DNA packaged into the procapsid is regulated by the headful packaging mechanism. This mechanism predicts the length of the DNA which is encapsulated inside the head by the interior volume of the mature phage particle. This process is critical in the DNA packaging process [10].

Phage life cycle

There are two phage life cycles, lytic or lysogenic, which are contributors to their role in bacterial and/or archaeal biology [5]. Virulent phages undergo the lytic cycle where the phage infects and kills their bacterial host by lysing the cells and releasing viral DNA (Figure 4). The lysogenic phages, however, do not lyse the cell and are identified by either the integration into the host's genome or the formation of a circular replicon in the cytoplasm [1]. The integration of the viral DNA into the host's genome is referred as a prophage. The integration can be random or specific but is dependent on the genes, which are encoded. The formation of a prophage represses the genes and the proteins which are required for the phage to enter the lytic cycle [5]. However, if the repressor protein becomes inactivated during the cycle then the viral DNA is excised from the prophage. The DNA that was removed from the host's chromosome now acts

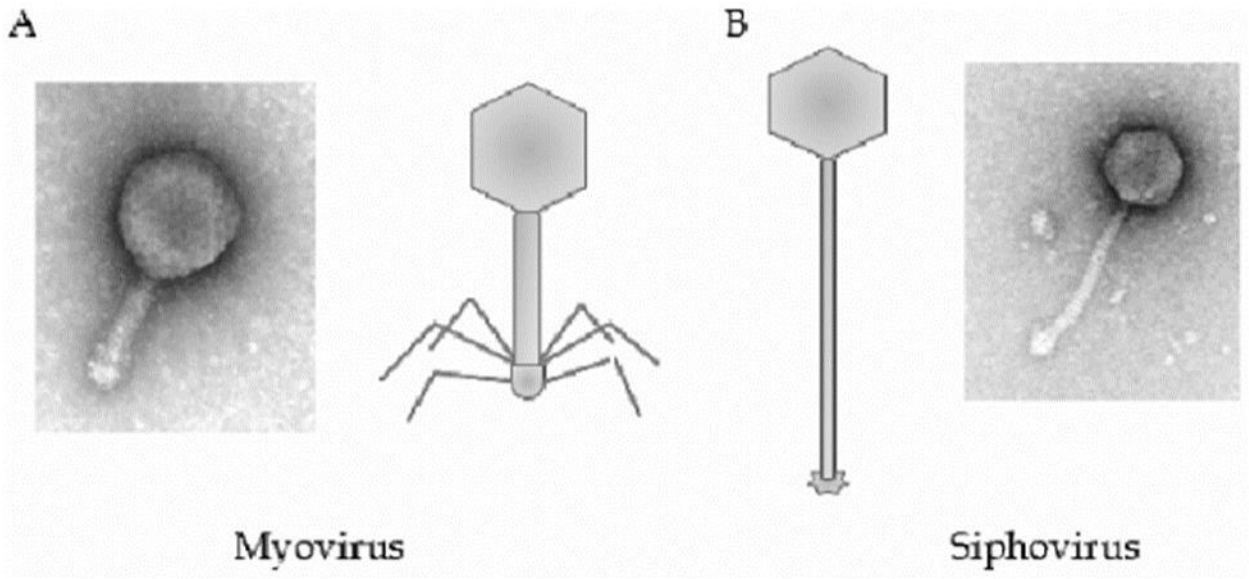


Figure 5. This figure represents the morphological differences between the myovirus and the siphovirus. The myovirus contains an icosahedral head and a contractile tail containing a base plate with numerous long tail fibers. Unlike the myovirus, the siphovirus contains an icosahedral head and a non-contractile tail with short tail fibers. The first depiction (A) is the myovirus using an electron microscope which is followed by a cartoon depiction of the composition of the myovirus. The second depiction (B) is the cartoon of the siphovirus, followed by the picture of the virus under an electron microscope. (https://www.researchgate.net/profile/Catherine_Rees2/publication/221924849/figure/fig1/AS:304780430856197@1 449676725410/Fig-1-Co mmon-Morphotypes -of-Mycobacteriophage-Myoviridae-panel-A-typically-have.png)

as a virulent phage and obtains the ability to produce new viral particles, which are released during cell lysis. Phages that are able to switch from the lysogenic cycle to the lytic cycle are referred as temperate [5].

Mycobacteriophages

Mycobacteriophages are bacteriophages which infect the host mycobacteria. These phages are commonly isolated from *Mycobacterium tuberculosis* and *Mycobacterium smegmatis* [11]. Mycobacteriophages are the most studied bacteriophages with approximately 2,413 phages isolated and more than 344 full sequenced genomes identified [12]. All the mycobacteriophages that have been identified to date have contained double stranded DNA genomes. Mycobacteriophages are characterized by either their myoviridae or siphoviridae morphology, which is shown in Figure 5 [4]. Phages with the myoviridae morphology are non-enveloped with double-stranded, linear genomic information, which is between 33-244 kilobases (kb) in length. Unlike the siphoviridae phages, long contractile tails characterize myoviridae [13].

Large terminase protein structural studies

The structure of the large terminase protein is important because the amount of structural information on phages is limited. Currently, there are five known structures for the large terminase protein in phages. There are four crystal structures of the nuclease domain and one of the ATPase domains, which are the two components of the large terminase protein. The known crystal structures of the nuclease domain are in thermophilic phage G20c, Enterobacteria P22 phage, Bacillus subtilis SPP1, and Enterobacteria RB49 phage [14]. The crystal structure of the ATPase domain is of the thermophilic phage P74-26 [14]. The ribbon structures of the known large terminase proteins can be seen in Figures 6, 7, 8, 9, and 10. The crystal structure of the nuclease domain of the thermophilic bacteriophage G20c indicates similarity to the ATPase domain of bacteriophages P74-26 and P23-45. The structure comparison shows that there is

plasticity in the Loop L1 which plays a role in the cleavage during DNA packaging. The known crystal structures for the large terminase show similarity to each other despite the low sequence identity between them. The highest sequence similarity is between the G20c (nuclease domain) and GP17 from bacteriophage T4, which has 25% identity covering 176 residues. The lowest sequence similarity is between the nuclease domain of G20c and the nuclease domain of SPP1 with 12.6% covering 143 residues. There are two main differences between the nuclease domains of thermophilic phage G20c and SPP1 which are an addition of a β strand and a terminal β hairpin [15].

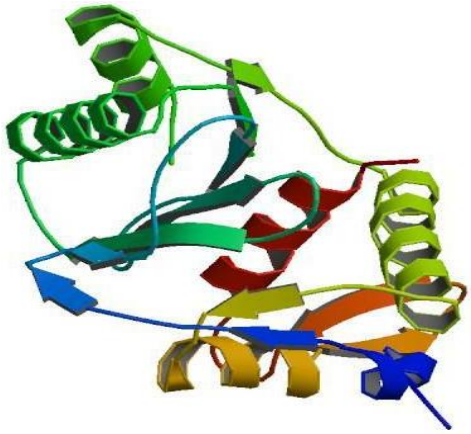


Figure 6. This figure represents the ribbon structure of the large terminase protein, the nuclease domain from thermophilic phage G20c. The expression system for this bacteriophage was *E. coli* (PDB.org)

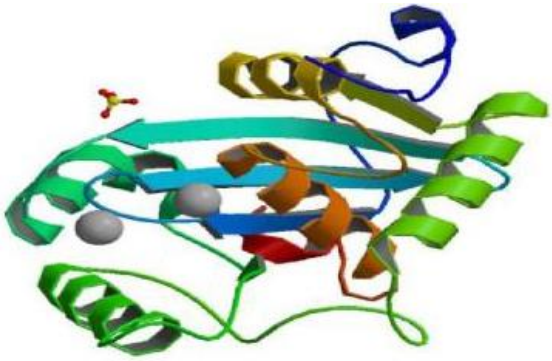


Figure 7. This figure represents the ribbon structure of the large terminase nuclease domain of P22 bacteriophage expressed in *E. coli.* (PDB.org)

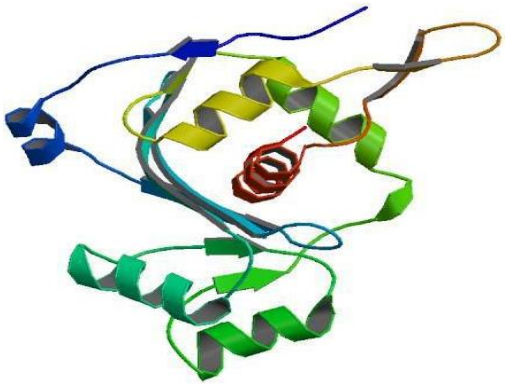


Figure 8. This figure represents the ribbon structure of the large terminase protein, the nuclease domain of SPP1 bacteriophage expressed in *E. coli* (PDB.org).

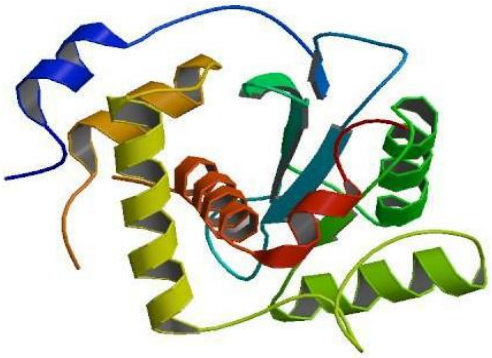


Figure 9. This figure represents the ribbon structure for the nuclease domain of the large terminase protein in RB49 bacteriophage expressed in *E. coli* (PDB.org).

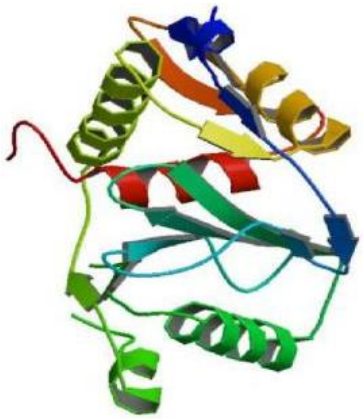


Figure 10. This figure represents the ribbon structure for Thermus phage P74-26 the large terminse protein, the ATPase domain. This protein was expressed in *E. coli.*

The nuclease domain of Enterobacteria RB49 is a mutant of the T4 bacteriophage with 72% sequence identity. The last 10 amino acid

residues are disordered, indicating that the sequence identity could increase if the protein was cleaved at the amino acid residue 567. The nuclease domain of RB49 was compared to the RNAse H/resolvase family due to functional similarities, even though there was no sequence identity between them the sequences were able to be superimposed on top of each other indicating structural similarities [16]. The nuclease domain of bacteriophage P22 was also compared to the known crystal structure for GP17 for bacteriophage T4. The nuclease domain of P22 and T4 contain the Walker A motif, which is a domain that utilizes ADP or ATP for nucleotide binding; this motif is characteristic of DNA packaging proteins. As mentioned previously, there are two subdomains of the large terminase protein, the ATPase domain and the nuclease domain. The tertiary structure of the nuclease domains from the known crystal structures of SPP1, T4, RB49, and HHV-5 contain structural similarity again despite their low sequence identities (16%, 14%, 12%, and 12%). The four proteins were superimposed on top of each other indicating the presence of a conserved amino acid residue at Asp324 [17]. Structural analysis indicated that when the nuclease domains of T4 and SPP1 are isolated they maintain a monomeric conformation; however, when they are in active packaging mode, they maintain a pentameric conformation. The analysis showed that there is a conserved central β sheet and a C-terminal β hairpin in the large terminases [18].

As mentioned previously, there is little structural information about bacteriophages; however, they have evolved to be extremely efficient so that very little genetic information is required to assemble a robust and functional protein. Due to this, the morphology of phages has remained relatively conserved throughout their evolution. The small genomes of bacteriophages have made them targets for a variety of clinical applications including the area of antibiotic resistant bacterial infections. The lytic phage named Catdawg, isolated from *Mycobacterium smegmatis mc²155* (*M. smeg*) at Cabrini College

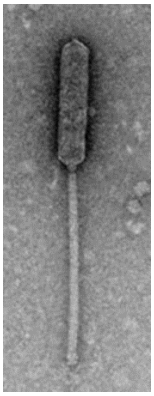


Figure 11. Electromicrograph picture was taken of the mycobacteriophage Catdawg. Catdawg has a prolate head and a long non-contractile tail [7].

in 2011 was selected for this study due to its unusual prolate capsid and long flexible tail (Figure 11) [7]. Catdawg is an O cluster siphovirus phage along with phages Dylan, Corndog, Firecracker, Smooch, YungJamal and Zakhe101 [7, 13]. Determining the structural information on Catdawg's large terminase protein could be beneficial in elucidating the capsid formation, DNA packaging, and ultimately, resistance in other *Mycobacterium* species, including *M. tuberculosis*. Within this study, we have generated a homology model of the large terminase protein in phage Catdawg and have validated the model using Procheck.

Methods

Homology model

The program PHYRE2 was used to generate a predicted homology model of GP30, the large terminase protein [19]. The FASTA sequence of the large terminase protein was obtained from phagesDB.org and uploaded into the program. The program searched the known crystal structures that contained similarity to the uploaded sequence. After the generation of a predicted structure, the PDB 3D model was uploaded to Procheck, a validation program. Procheck is a program used to determine how "good" the predicted model is compared to other models [20].

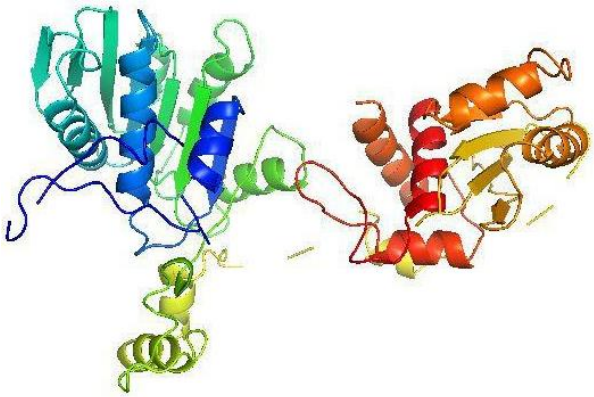


Figure 12. This figure represents the predicted 3D structure of GP 30, the large terminase protein from Catdawg. The PDB file was uploaded into PHYRE which generated the predicted structure.

```
DFHPLNEAGKILIKHPSLAERKDEDGIHWIKSQWDGKWYPEKFSDYLRLHKIVKIPNNSDKPELFQTYKDKNNKRSRYMGL
PNLKRANIKTQWTREMVEEWKKCRDDIVYFAETYCAITHIDYGVIKVQLRDYQRDMLKIMSSKRMTVCNLSRQLGKTTVVA
IFLAHFVCFNKDKAVGILAHKGSMSAEVLDRTKQAIELLPDFLQPGIVEWNKGSIELDNGSSIGAYASSPDAVRGNSFAMI
YIDECAFIPNFHDSWLAIQPVISSGRRSKIIITTTPNGLNHFYDIWTAAVEGKSGFEPYTAIWNSVKERLYNDEDIFDDGN
QWSIQTINGSSLAQFRQEHTAAFEGTSGTLISGMKLAVMDFIEVTPDDHGFHQFKKPEPDRKYIATLDCSEGRGQDYHALH
IIDVTDDVWEQVGVLHSNTISHLILPDIVMRYLVEYNECPVYIELNSTGVSVAKSLYMDLEYEGVICDSYTDLGMKQTKRI
KAVGCSTLKDLIEKDKLIIHHRATIQEFRTFSEKGVSWAAEEGYHDDLVMSLVIFGWLSTQSKFIDY
```

Sequence length: **553** residues.

Your sequence search returned the following **132** hits:

	PDB code	Model	Length	%-tage identity	a.a. overlap	z-score	Ligands	Protein name
1.	3ezk(A)		553	100.0%	553	4676.7		Bacteriophage t4 gp17 motor assembly based on crystal structures and cryo-em reconstructions
2.	3cpe(A)	X-ray 2.80Å	553	99.6%	553	4664.3	PO4.	Crystal structure of t4 gp17
3.	2o0j(A)	X-ray 1.80Å	360	99.4%	351	2976.0	ADP.	T4 gp17 atpase domain mutant complexed with adp
4.	2o0h(A)	X-ray 1.88Å	357	99.4%	348	2953.6	ATP.	T4 gp17 atpase domain mutant complexed with atp
5.	2o0k(A)	X-ray 2.50Å	357	99.4%	348	2953.6		T4 gp17 atpase domain mutant
6.	3c6a(A)	X-ray 1.16Å	198	69.3%	202	1167.9		Crystal structure of the rb49 gp17 nuclease domain
7.	3c6h(A)	X-ray 2.80Å	182	67.0%	194	922.6		Crystal structure of the rb49 gp17 nuclease domain
8.	4znk(A)	X-ray 1.93Å	258	25.4%	236	210.1	SO4.	Thermus phage p74-26 large terminase atpase domain from (p 3 space group)
9.	4znl(A)	X-ray 2.10Å	251	25.8%	233	200.3	SO4.	Thermus phage p74-26 large terminase atpase domain (i 2 3 sp
10.	4znl(A)	X-ray 2.07Å	253	25.8%	233	200.2	BEF-ADP, SO4.	Thermus phage p74-26 large terminase atpase domain bound to beryllium fluoride
11.	4znj(A)	X-ray 2.53Å	250	25.3%	233	189.1	SO4.	Thermus phage p74-26 large terminase atpase domain mutant r1 space group)
12.	5m1p(A)	X-ray 1.10Å	187	27.3%	165	153.6		Crystal structure of the large terminase nuclease from therm phage g20c with bound calcium
13.	5m1o(A)	X-ray 1.60Å	188	27.3%	165	153.6		Crystal structure of the large terminase nuclease from therm phage g20c with bound cobalt
14.	5m1k(A)	X-ray 1.20Å	189	27.3%	165	153.5	SO4, BTB.	Crystal structure of the large terminase nuclease from therm phage g20c with bound magnesium
15.	5m1n(A)	X-ray 1.20Å	189	27.3%	165	153.5	SO4, BTB.	Crystal structure of the large terminase nuclease from therm phage g20c with bound manganese
16.	5m1f(A)	X-ray 2.15Å	186	26.5%	166	141.2		Crystal structure of the large terminase nuclease from therm phage g20c
17.	1uly(A)	X-ray 2.50Å	190	51.4%	37	139.8		Crystal structure analysis of the arsr homologue DNA-binding from p. Horikoshii ot3
18.	2cwe(A)	X-ray 2.70Å	191	51.4%	37	139.7		Crystal structure of hypothetical transcriptional regulator ph1932 from pyrococcus horikoshii ot3

Figure 13. This figure depicts the results obtained from Procheck after submitting the predicted PDB structure. The top results report 100% identity for bacteriophage T4 the motor assembly protein, followed by two results with identities 99.6 and 99.4% identity indicating ATPase domains from bacteriophage T4. The protein of interest the large terminase protein from Catdawg is predicted to be a large terminase protein. These results confirm the structure of GP 30; the large terminase protein will contain a similar conformation as the known terminase crystal structures.

Results and conclusions

PHYRE2, a bioinformatics program was used to predict a structure of GP-30, the large terminase protein depicted in Figure 12 [19]. The program searched the known crystal structures that contained similarity to the uploaded FASTA sequence. Numerous results were generated that showed the protein template, alignment coverage, a depiction of the 3D model, the confidence, the % identity, and more information regarding the generated protein template. The top result was a DNA packaging protein, GP-17 from bacteriophage T4. This template accepted 408 residues and modelled 81% of the amino acid sequence with 100% confidence. After the generation of a predicted structure, the PDB 3D model was uploaded to Procheck a validation program [20]. Procheck is a program used to determine how "good" the predicted model is compared to other models. The top result reported 100% identity which was the DNA packaging protein GP 17 from bacteriophage T4 which can be seen in Figure 13. These results confirm that the structure of GP 30, the large terminase protein will contain a similar conformation as the known terminase crystal structures.

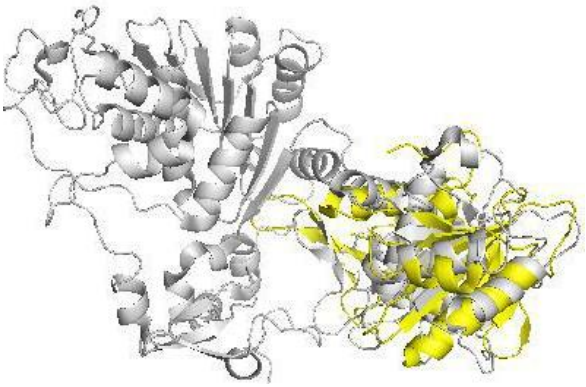


Figure 14. This figure represents the alignment for the large terminase protein in Catdawg to the ATPase domain of P74-26. The RMSD value generated between the models was 8.125Å. The yellow figure represents the nuclease domain of the large terminase protein from SPP1 and the gray figure represents the large terminase protein generated from PHYRE2, T4 bacteriophage.

Figure 14 shows the results for the comparison between the ATPase domain of bacteriophage P72-26 compared to the large terminase protein from Catdawg with an RMSD value of 8.125 Ångström (Å). The generated RMSD values for each of the protein structures were higher than the favorable RMSD value of 2 (Å) or less indicating that the generated homology model of the large terminase protein from Catdawg may not be as structurally similar as we had hypothesized. Although the RMSD values were high indicating differences in their structures, the bioinformatics programs used in this study, PHYRE2 and Procheck, both indicate similarity to T4 DNA packaging protein which demonstrates that some of the structure is conserved.

Acknowledgements

The authors would like to acknowledge Dr. Anne Coleman, Dr. Krystle Mclaughlin, and Morgan Sperratore for their support and guidance while researching and writing this paper.

Reference

1. Clokie MR, Millard AD, Letarov AV, Heaphy S. 2011. Phages in nature. Bacteriophage. 1(1):31-45.
2. Pedulla ML, Ford ME, Houtz JM, Karthikeyan T, Wadsworth C, Lewis JA, et al. 2003. Origins of highly mosaic mycobacteriophage genomes. Cell. 113(2):171-182.
3. Raoult D, Forterre P. 2008. Redefining viruses: lessons from Mimivirus. Nature Reviews Microbiology. 6:315–319.
4. Hatfull GF. 2014. Molecular genetics of mycobacteriophages. Microbiology spectrum. 2(2):1.
5. Shors T. 2013. Understanding viruses. Burlington, MA: Jones & Bartlett Learning.
6. Campbell A. 2003. The future of bacteriophages biology. Nature Reviews Genetics. 4:471– 477.
7. Cresawn SG, Pope WH, Jacobs-Sera D, Bowman CA, Russell DA, Dedrick RM, et al. 2015. Comparative genomics of cluster O mycobacteriophages. Plos ONE. 10(3):1-27. doi:10.1371/journal.pone.0118725
8. Katen S, Zlotnick A. 2009. The thermodynamics of virus capsid assembly. Methods in enzymology. 455:395-417.
9. Fokine A, Rossmann MG. 2014. Molecular architecture of tailed double-stranded DNA phages. Bacteriophage. 4(2):e28281.

10. Rao VB, Feiss M. 2015. Mechanisms of DNA packaging by large double-stranded DNA viruses. Annual review of virology. 2:351-378.
11. Mehla J, Dedrick RM, Caufield JH, Siefring R, Mair M, Johnson A, et al. 2015. The protein interactome of mycobacteriophage Giles predicts functions for unknown proteins. Journal of bacteriology. 197(15):2508-2516.
12. Smith KC, Castro-Nallar E, Fisher JN, Breakwell DP, Grose JH, Burnett SH. 2013. Phage cluster relationships identified through single gene analysis. BMC genomics. 14(1):1.
13. The Actinobacteriophage Database | Home. (n.d.). Retrieved September 04, 2018, available from http://phagesdb.org/
14. The worldwide Protein Data Bank| Retrieved September 18, 2018, available from https://www.wwpdb.org/
15. Xu R, Jenkins HT, Chechik M, Blagova E, Lopatina A, Klimuk E, et al. 2017. Viral genome packaging terminase cleaves DNA using the canonical RuvC-like two-metal catalysis mechanism. Nucleic Acids Res. 45(6):3580-3590.
16. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, et al. 2008. The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces. Cell. 135(7):1251-1262.
17. Roy A, Cingolani G. 2012. Structure of p22 headful packaging nuclease. Journal of Biological Chemistry. 287(33):28196-28205.
18. Smits C, Chechik M, Kovalevskiy OV, Shevtsov MB, Foster AW, Alonso JC, Antson AA. 2009. Structural basis for the nuclease activity of a bacteriophage large terminase. EMBO reports. 10(6):592-598.
19. Kelley LA, Mezulis S, Yates CM, Wass MN, Sternberg MJE. 2015. The Phyre2 web portal for protein modeling, prediction and analysis. Nature Protocols. 10(6):845-858.
20. Laskowski RA, MacArthur MW, Moss DS, Thornton JM. 1993. PROCHECK - a program to check the stereochemical quality of protein structures. J. App. Cryst. 26:283-291.