4-2019

Hinge region in DNA packaging terminase pUL15 of herpes simplex virus: A potential allosteric target for antiviral drugs

Lana Faisal Thaljeh

Hinge region in DNA packaging terminase pUL15 of herpes simplex virus: A potential allosteric target for antiviral drugs

by

Lana Faisal Thaljeh

Undergraduate Honors Thesis under the direction of

Dr. Michal Brylinski

Department of Biological Sciences, Center for Computation & Technology

Submitted to the Louisiana State University Roger Hadfield Ogden Honors College

April 2019

Louisiana State University & Agricultural and Mechanical College

Baton Rouge, Louisiana

List of Abbreviations

- Human Herpesviruses (HHV)
- Herpes simplex virus 1, 2, 6, 7, 8 (HHV-1, HHV-2, HHV-6, HHV-7, HHV-8)
- Human cytomegalovirus (CMV or HHV-5)
- Varicella-zoster virus (VZV)
- Epstein-Barr virus (EBV)
- Kaposi's sarcoma-associated herpesvirus (KSHV)
- 2,5,6-trichloro-1-β-D-ribofuranosyl benzimidazole (TCRB)
- 2-bromo-5,6-dichloro-1-(β-D-ribofuranosyl) benzimidazole (BDCRB)
- 2-bromo-4,5,6-trichloro-1-(2,3,5-tri-O-acetyl-β-d-ribofuranosyl) benzimidazole (BTCRB)
- 2,4,5,6-tetrachloro-1-(2,3,5-tri-O-acetyl-β-d-ribofuranosyl benzimidazole (Cl4RB)
- Markov Chain Monte Carlo (MCMC)
- Reference Sequence (RefSeq)
- Gaussian Network Model (GNM)
- Cross-Correlation (CC)
- Principal Component Analysis (PCA)
- Anisotropic Network Model (ANM)
- Protein Data Bank (PDB)
- Double-Stranded (ds)

Abstract

Background: Approximately 98% of adults are infected with a member of the herpesviridae family. Herpesviruses establish life-long latent infections within neurons, which may reactivate into lytic infections. There are nine human herpesviruses (HHV) posing health concerns from benign conditions to life threatening encephalitis, including cancers associated with viral infections. Several nucleoside and nucleotide analogs targeting virus DNA polymerase are current treatment options for most HHV conditions. Although these drugs are often efficacious, their common mechanism of action may lead to the development of drug resistance, particularly in immunocompromised patients. Therefore, new classes of drugs are needed to combat HHVs.

Methods: In this study, we analyze the conservation rates of all proteins in herpes simplex virus 1 (HHV-1), a representative of the HHV family. Furthermore, we generated a structure model of pUL15 based on the crystal structure of the C-terminal domain and the bacteriophage T4 ATPase domain to shed light on its structure and function. A series of computational analyses were performed on the model in order to predict some of the functionality of the protein as well as putative ATP and DNA binding sites.

Results: Our study indicates that proteins involved in HHV-1 DNA packaging and cleavage are amongst the most conserved gene products of HHVs. Particularly, the packaging protein pUL15 is the most conserved among all HHV-1 gene products. A subsequent annotation of ATP and DNA binding sites and the elastic network analysis reveal a functionally important hinge region between the two domains of pUL15.

Conclusions: The high-quality atomic information on the active and allosteric sites in the ATP- and DNA-bound model of pUL15 presented in this study can play a key role in the structure-based drug discovery of a new class of drugs for treating a wide range of HHVs.

Background

Approximately 98% of the world population are infected with a form of herpes viruses. Herpes viruses remain for the life of the host and can establish latency, a non-productive state that allows the virus to go undetected and unaffected by antiviral drugs [1, 2]. There are nine herpes viruses that infect humans and their effects can range from easily manageable to life threatening [3]. Herpes simplex viruses 1 and 2 (HHV-1 and HHV-2) are associated with blisters on the lips and genitals, respectively, whereas HHV-6 and HHV-7 cause Roseola in infants [4]. Human cytomegalovirus (CMV or HHV-5) is linked to lethal diseases involving the lungs, gastrointestinal tract, liver, retina, and central nervous system in immune compromised individuals, and is the leading cause of post-transplant infection [5, 6]. Congenital CMV results in an estimated 8,000 cases of permanent neurological disabilities a year affecting over 90% of infants who survive the initial infection [7-9]. Chicken pox and shingles, caused by varicella-zoster virus (VZV), are manageable in immunocompetent people, but can become deadly in immunocompromised persons [3, 10]. Herpes viruses are also associated with several cancers. Epstein-Barr virus (EBV) causes Burkett's lymphoma, nasopharynal carcinoma, and Hodgkin's disease, and has been detected in approximately 70% of advance breast cancer tumors [11, 12]. Kaposi's sarcoma-associated herpesvirus (KSHV) can result in body cavity-based lymphoma, and Kaposi's sarcoma, the most common malignancy present in HIV-1 patients [13, 14]. The pervasiveness and deleterious effects of herpes viruses provides incentive for new approaches to antiviral drugs.

The most common antiviral strategy against herpes viruses is based on nucleoside analogs targeting the viral DNA polymerase and thymidine kinase. While these drugs are effective, many of them are associated with significant toxicities, poor bioavailability, and resistance in immunocompromised persons [15]. The nucleoside analogues acyclovir and ganciclovir are the

standard therapy for HHV-1 and CMV, respectively [15-17]. These compounds target HHV-1 pUL30, the viral DNA polymerase, and HHV-1 pUL23, the viral thymidine kinase. The prevalence of resistance against acyclovir is 5% in immunocompromised persons and as high as 30% in allogeneic bone marrow transplant patients [18], whereas the incidence of resistance to ganciclovir is 5-10% in organ transplant recipients [19]. HHV-1 strains that are resistant to acyclovir are typically cross-resistant to thymidine kinase dependent drugs such as penciclovir and famciclovir. They may also be cross-resistant to polymerase dependent drugs, foscarnet or cidofovir [18]. Ganciclovir resistant strains also have shown cross resistance to second-line treatments foscarnet and cidofovir [20]. Targeting the DNA polymerase and thymidine kinase, while widely used, is not the only option for herpesvirus antivirals.

Due to the resistance, toxicities, and other adverse side effects of nucleoside analogs, new targets to treat herpes viruses are being perused. The viral DNA packaging motor, namely its large terminase subunit, has become a promising target for potential herpes antivirals. Mechanisms that inhibit the viral terminase or other components of the viral DNA packaging motor are expected to be more effective and have less target-related toxicity than nucleoside analogs. This is due to the molecular functions of DNA packaging motor including capsid formation, DNA cleavage, and packaging of DNA into capsids. These functions are virus-specific and not found in mammalian cells [21, 22]. Current CMV terminase inhibitors include benzimidazoles and the 3,4-dihydro-quinazoline-4-yl-acetic acid derivative, letermovir.

The first antiviral agents to inhibit the cleavage of concatenated DNA into monomeric genomes through inhibiting the terminase were 2,5,6-trichloro-1-β-D-ribofuranosyl benzimidazole (TCRB) and 2-bromo-5,6-dichloro-1-(β-D-ribofuranosyl) benzimidazole (BDCRB). These compounds are suspected to affect CMV proteins encoded by genes UL56 coding the small

terminase subunit and UL89 coding the large terminase subunit, the key components of the packaging motor. Mutations in UL56 result in resistance to TCRB, whereas mutations in UL89 cause resistance to both BDCRB and TCRB [23, 24]. The high resistance caused by mutations in UL89 were mapped to amino acids D344E and A355T [24]. There are no acidic amino acids corresponding to the position of A344 in the HHV and VZV UL89 homologs; therefore, there is likely no significant interaction of BDCB or TCRB with HHV or VZV. This accords with *in vitro* results showing that benzimidazole ribonucleosides have little to no effect against HHV, VZV, HHV-6, and HHV-8 [21]. Along with the specific nature of these compounds, they are also metabolized too rapidly *in vivo*, despite their effectiveness in cell culture [25, 26]. In order to develop more biologically stable compounds, analogs have been derived from BDCRB including acetylated, tetrahalogenated benzimidazole D-ribonucleosides, 2-bromo-4,5,6-trichloro-1-(2,3,5-tri-O-acetyl-β-d-ribofuranosyl) benzimidazole (BTCRB) and 2,4,5,6-tetrachloro-1-(2,3,5-tri-O-acetyl-β-d-ribofuranosyl benzimidazole (Cl4RB) which inhibit cleavage and packaging [27, 28]. BTCRB is suspected to inhibit the ATPase activity of pUL56 [27], while Cl4RB is believed to interfere with the interaction of pUL56 with the portal protein pUL104 [29]. These two compounds are shown to be active against VZV, rat cytomegalovirus, and human cytomegalovirus, however Cl4RB had no effect on HHV-1 and the effects of BTCRB were minimal [28].

Raltegravir, an HIV integrase inhibitor, exhibits efficiency against herpes viruses. The effect of raltegravir on herpes viruses is attributed to an inhibition of the large subunit of the viral terminase because it has the same RNase H-like fold as the HIV integrase [22]. Recently however, drug resistance to raltegravir has been traced to HHV-1 UL42 coding for the DNA polymerase accessory factor [30]. Because of the fact that pUL42 is not a part of the terminase, it was determined that raltegravir likely inhibits DNA replication through the polymerase accessory

factor, not DNA cleavage and packaging through the terminase [30]. Letermovir (AIC246 or MK-8228), is another promising new antiviral drug for the treatment of CMV. Letermovir is believed to inhibit the viral terminase complex by targeting pUL56 because L241P and R369S mutations in pUL56 correlate with resistance to letermovir [22, 31]. The inhibitory effect of letermovir is believed to be distinct due to the lack of cross-resistance of letermovir-resistant CMV strains to benzimidazoles [22, 32]. While letermovir has no target-related toxicity and has a good safety profile, it is specific for human CMV and is ineffective against all other viruses including other herpesviruses [33].

There are no drugs for HHV-1 that target proteins other than the resistance prone DNA polymerase and the viral thymidine kinase. Although numerous antivirals have been developed against HHV, many of these compounds are prone to resistance, have poor bioavailability and high toxicities, and are so specific that they cannot be utilized against other HHVs. The lack of suitable antivirals for HHV-1 and the promising nature of CMV terminase targeting drugs suggests a need for a more complete understanding of the terminase as it could potentially offer a means for discovering novel drugs against HHV-1. In this study, we first conduct a thorough review on the current state of HHV research, revealing a less than desired focus on the treatment and mechanisms of HHV. By focusing on the packaging motor, we propose pUL15 as a drug target that could also be utilized in other herpesviridae and possibly a larger set of viruses utilizing a DNA packaging motor. We also predict that the hinged region between the two domains of pUL15 may be the perfect drug target to inhibit its overall function.

Methods

Estimated rate of change

Eight previously sequenced herpesvirus strains (bold in Table1) were selected from GenBank [34] to be included in genome-wide phylogenetic analyses. These sequences were chosen to represent all major groups of alphaherpesviruses, while also focusing on those found in human, HHV-1 and HHV-2. In order to conduct phylogenetic analyses, homologous regions of the genome must be identified and aligned. We conservatively interpreted available annotations to identify homologous regions, resulting in 59 homologs that were extracted from each genome and compiled into gene-by-gene datasets. Each gene-by-gene dataset was aligned with MAFFT version 7 using the E-INS-i algorithm for the highest level of accuracy [35]. In order to select the best fitting model of amino acid substitution, we used the Bayesian Information Criterion [36] as implemented in ProtTest version 3.2 [37]. All available 120 amino acid substitution models were tested, and in each case the Jones92 model with invariant sites and a gamma distribution [38] was preferred. All individual gene datasets were then concatenated, and partitioned based on gene identity, resulting in an alignment with 41,543 amino acids. A Bayesian Markov Chain Monte Carlo (MCMC) analysis with fixed partitioning schemes was performed using MrBayes version 3.2.6 [39] using 4 independent runs of 5 million generations, with 4 Metropolis-coupled chains per run. Fixed partitioning analyses included parameter estimates across all 59 genes. For each partition, all substitution model parameters, with the exceptions of topology and branch lengths, were unlinked across data subsets, allowing each partition to independently estimate relative rates of evolution. Analyses were monitored for convergence by examining trace plots of scalar values in Tracer version 1.6 [40]. Topological convergence was assessed using the average standard deviation of

split frequencies technique. To compare the relative rates of evolution across different genes, the inferred rate multipliers from the posterior distribution were plotted using Tracer version 1.6 [40].

Phylogeny Tree

Twenty-one HHV-1 UL15 gene homologs from various herpes virus species were collected by using BLAST [41] with HHV-1 as the query. These sequences include: human alphaherpesvirus 1 (HHV-1, YP 009137089.1), chimpanzee herpesvirus strain 105640 (ChHV-2, YP_009011001.1), human alphaherpesvirus 2 (HHV-2, YP_009137166.1), equid alphaherpesvirus 4 (EHV-4, NP_045262.1), equid alphaherpesvirus 9 (EHV-9, YP_002333526.2), equid alphaherpesvirus 1 (EHV-1, YP_053090.1), human alphaherpesvirus 3 (HHV-3,VZV, NP_040165.1), bovine alphaherpesvirus 5 (BHV-5, YP_003662508.1), bovine alphaherpesvirus 1 (BHV-1, NP_045342.1), suid alphaherpesvirus 1 (SuHV-1, YP_068358.1), human betaherpesvirus 5 (HHV-5, CMV, YP_081537.1), human betaherpesvirus 6B (HHV-6B, NP_050241.1), human betaherpesvirus 7 (HHV-7, YP_073802.1), equid alphaherpesvirus 8 (EHV-8, YP_006273027.1), bovine gammaherpesvirus 4 (BHV-4, NP_076521.1), bovine gammaherpesvirus 6 (BHV-6, YP_009042007.1), human gammaherpesvirus 8 (HHV-8, YP_001129382.1), equid gammaherpesvirus 5 (EHV-5, YP_009118419.1), human gammaherpesvirus 4 (HHV-4, EBV,YP_401690.1), equid gammaherpesvirus 2 (EHV-2, NP_042630.2), and gallid alphaherpesvirus 2 (GaHV-2, AAF66813.1). All sequences had E-values $<10^{-15}$. Sequences were gathered from the Reference Sequence (RefSeq) database [42]. Multiple sequence alignment was performed with MUSCLE [43]. The ProtTest 2.4 server [44] was used to determine the best fitting substitution model. The LG+I+G+F substitution model scored by AIC to be the best suited to the alignment. A phylogenetic tree was generated for the

amino acid alignment based on the LG+I+G+F model using maximum likelihood inference preformed on RAxML v8.2.4 [45] with bootstrapping of 100 replicates. Additionally, a Bayesian MCMC estimation was conducted, and a majority-rule consensus tree was generated by MrBayes v3.2.6 [39, 46]. The trees were visualized in FigTree v1.4.3 [47] using, GaHV-2 as the outgroup to root the trees. The recovered topologies of the two trees were identical except for the reverse placement of HHV-4 and BHV-6. Both the bootstrap and posterior probabilities values are displayed on the Bayesian tree [48].

Protein structure modeling

The amino acid sequence of HHV-1 pUL15 was obtained from the RefSeq database [42] (GeneID: 2703385). Only a partial crystal structure of the C-terminus of pUL15 (PDB-ID: 4iox) [49] is available in the Protein Data Bank (PDB) [50], so in order to make a more complete model of pUL15, the sequence of pUL15 was input into HHPRED [51]. The default settings were used and DNA packaging protein gp17 from enterobacteria phage T4 (PDB-ID: 2o0j) [52] was determined to be the most reliable template to model the N-terminal domain of pUL15. To properly orient the two proteins, another structure of T4 gp17 (PDB-ID: 3cpe) [53] was chosen from the HHPRED results to have the two templates superimposed onto it due to its low E-value and high coverage. Fr-TM-align [54] was used to perform the Cα-based superposition. After the structural alignment, a sequential alignment was done using SALIGN [55] to align the sequences of 4iox and 2o0j to the sequence of pUL15. The sequential and structural alignments were utilized to create the model of pUL15 with MODELLER version 9.16 [56], excluding the first 126 amino acids because there was no template to create a valid structure of that pUL15 segment.

ATP-binding site prediction

The model was screened for potential binding pockets based on evolutionarily related proteins using *e*FindSite [57, 58]. The output from *e*FindSite was inserted into *e*SimDock [59], which determined that ATP is most likely to bind to that particular binding site, and yielded a complex structure with the ATP docked into the protein. In order to resolve any possible clashes, the ligand-protein complex was input into LigPlot+ [60]. The free energy required for the ATP binding was estimated with the distance-scale finite ideal-gas reference (DFIRE) potential [61]. The energy of association between ATP and pUL15 was compared to other proteins bound to ATP selected with PDBePISA (Proteins, Interfaces, Structures and Assemblies) [62].

DNA-binding site prediction

DNA-protein interactions were predicted with DISPLAR [63] and DP-Bind [64] against the sequence of pUL15 as the input. Further, a review of the currently available literature discussing potential DNA binding residues in pUL15 was also done. Subsequently, those residues common between the literature and the sequence-based prediction as well as residues in the literature that seem to have a particular importance were noted. A small segment of HHV-1 DNA was created by putting the nucleotide sequence (GenBank ID: JQ673480.1) into 3D-DART, a DNA structure modelling server [65]. This DNA structure along with the protein model were input into HADDOCK [16] in order to generate a three-dimensional model of the DNA-pUL15 complex. In HADDOCK, pUL15 residues 517, 695, 700, 701, and 620-633 were labeled as active, whereas the first and the last base pairs of the DNA model were selected as non-active.

Cross-correlation of residue fluctuations

In order to predict residue fluctuations, an elastic network for pUL15 was built using the Gaussian Network Model (GNM) via the DynOmics online tool [66]. The default cutoff distance of 7.3 Å between GNM nodes was used. The cross-correlation (CC) map shows the extent of a connection between two nodes representing amino acid residues. DynOmics by default includes all the principal modes of structural variations on the CC map. Each mode represents the direction of the concerted motion of the entire system, which is determined upon principal component analysis (PCA) of the covariance matrix.

Results

Publications on HHV, HIV, and influenza

In order to grasp the amount of research focused on HHV compared to HIV and influenza, we assess trends in the number of papers indexed by PubMed concerning these viruses. We also look at how many of these studies utilized computational methods in their research. Figure 1 shows that the number of papers reporting computational research consistently remains fewer than the number of studies employing *in vivo* or *in vitro* methods for all three viruses. This is expected, as *in vitro* and *in vivo* techniques are traditionally the primary methods for approaching virology research. Even so, the rapid development of technology in genomics and computational science has resulted in more reliable and useful tools that can be used for making relatively accurate predictions in order to facilitate *in vitro* and *in vivo* studies. *In silico* research has provided essential information for better understanding viruses including viral structures, interactions, and evolution, which can be used to systematically develop methods to combat viruses [67].

A significant gap between the number of papers published on HIV compared to HHV and influenza is likely a result of the rapidly increasing mortality rate of HIV and significant support provided by the World Health Organization Global Special Programme on AIDS [68]. The number of influenza research papers increases corresponding to flu outbreaks this is especially true with the outbreak of the highly pathogenic H5n1 in Asia in 2003 and the Swine flu pandemic in 2009. Herpes viruses lack the dramatic increases in transmission of influenza and the wide spread mortality rate of HIV to spur research, which likely resulted in comparatively fewer papers. However, the human herpesvirus family includes the pathogens HHV-1 and HHV-2 along with VZV, EBV, CMV, HHV-6, HHV-7, and HHV-8, meaning that the human herpesvirus family is responsible for numerous diseases including roseola, post-transplant infections, neurological disabilities in infants, chicken pox, shingles, lymphoma, nasopharyngeal carcinoma, Hodgkin's disease, and infectious blisters on the lips and genitals. HHV-1 and HHV-2 alone can result in neonatal infection, keratitis meningitis, encephalitis, and HHV-2 increases the likelihood of HIV infection by three-fold [69]. The ability of HHV to establish latency gives it permanent residence in the body and has made it a difficult target for antiviral strategies [70]. Although there are some antivirals currently available to treat herpesvirus, they have been insufficient due to the increase in antiviral resistance.

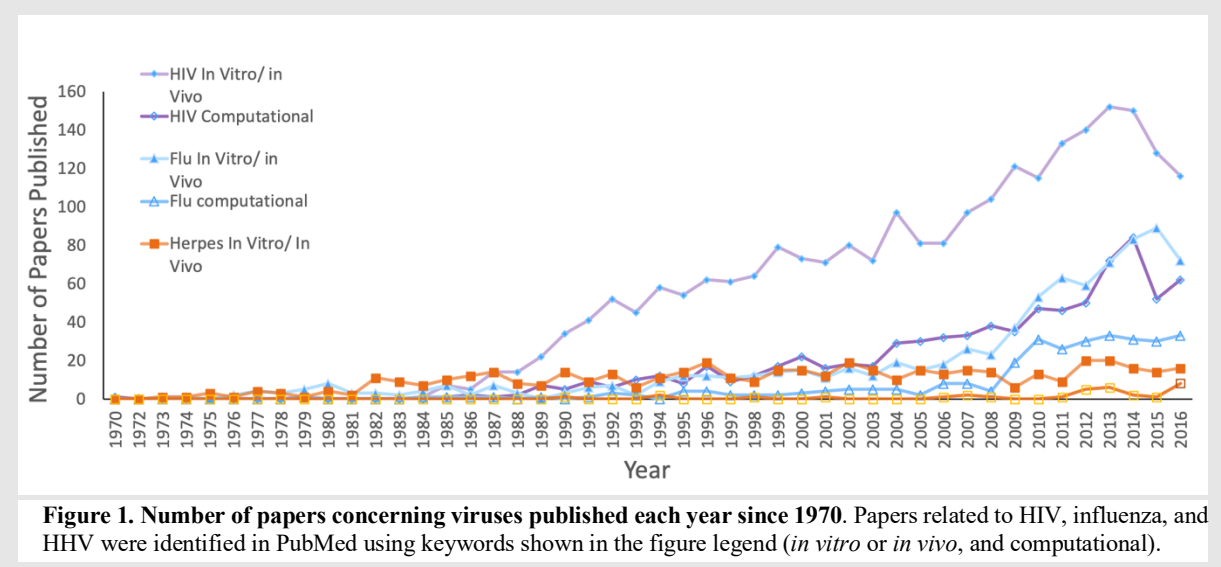

Figure 1. Number of papers concerning viruses published each year since 1970. Papers related to HIV, influenza, and HHV were identified in PubMed using keywords shown in the figure legend (*in vitro* or *in vivo*, and computational).

Protein Target

Of all proteins in the herpesviridae family, those involved in the packaging of DNA are generally the most conserved. These proteins include pUL6, pUL15, pUL17, pUL25, pUL28, pUL32, and pUL33. Mutations or removal of these proteins result in complications in the formation of the virus, typically affecting the ability to successfully cleave and package DNA into its capsid. This makes any of these proteins a potential antiviral target, because once the DNA packaging has been hindered, there will be no formation of a new virus. In order to determine the best protein target, rates of change were calculated for those protein products of HHV-1 involved in the packaging of DNA. Figure 2 shows that pUL15 is more conserved than any other packaging protein in HHV-1 as well as the current drug targets. Interestingly, pUL42 and pUL23 are the least conserved out of the proteins which were analyzed, which may explain the resistance that has emerged against raltegravir and acyclovir, respectively. With pUL15 being one of the most conserved proteins in the entire virus, utilizing it as a model for drug discovery may prove to be highly beneficial. Although letermovir has already paved the way for DNA packaging motor inhibition, it has done so by targeting the pUL28 homolog in CMV and has been observed to only affect CMV. This could be explained by the lesser degree of conservation found in the pUL28 homolog in HHV-1. Targeting a more conserved protein, such as pUL15, may allow for a more inclusive use of antivirals.

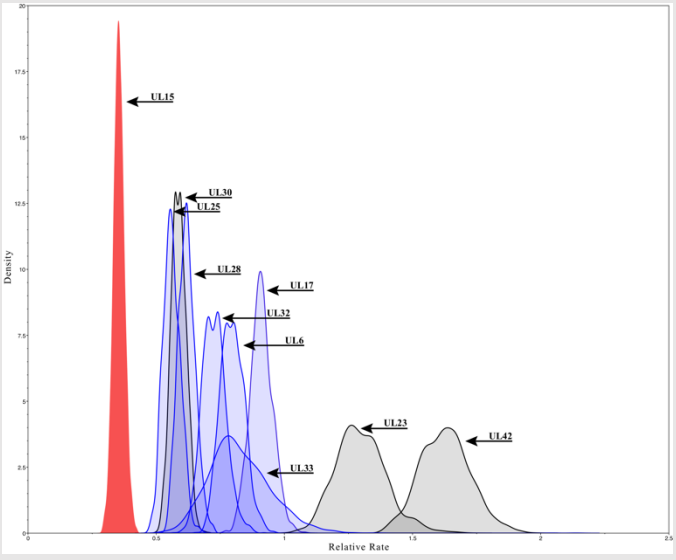

Figure 2. Posterior distributions of relative substitution rates. Protein products of HHV-1 involved in the packaging of DNA (pUL6, pUL15, pUL17, pUL25, pUL28, pUL32, and pUL33; blue) and the protein products of HHV-1 that are currently used as antiviral targets (pUL23 and pUL42; grey) are shown. pUL15 (red) has been highlighted to show an extremely conservative rate of change relative to the other packaging related genes. The current drug targets and the rest of the packaging proteins are shown to demonstrate how the packaging motor proteins are more conserved than the proteins we currently have antivirals for.

Phylogenetic Tree Analysis

The highly conserved nature of pUL15 suggests that it should have similarities to its counterparts in other human herpes viruses. The phylogenetic analysis shows the level of similarity amongst HHV-1 UL15 and its human herpes virus homologs with context through the relationships visualized in a phylogenetic tree. The phylogenetic tree in Figure 3 shows that HHV-1 UL15 is most closely related to the HHV-2 UL15 homolog. It is noteworthy that HHV-1 and HHV-2 together infect nearly 90% of the world population [71]. Our research focuses on HHV-1 as it is the more common of the two viruses affecting 65% of the United Stated population. HHV-1 and HHV-2 however are closely related, with over 80% sequence homology of their protein-coding regions, and ideally, information on one virus can elucidate the other [69, 72, 73]. HHV-1 also is closely related to the other human herpes viruses as well as those herpes viruses affecting high value domesticated animals in the agriculture industry. The close relationship between UL15 homologs of HHV-1 and the other human herpes virus proteins sequences is represented in the tree by the fact that they are on sister clades. The scale of the tree indicates that overall there is a high level of similarity between all of the viruses with only minimal changes between them. The relationships between pUL15 for HHV-1 and its homologs suggest that UL15 in HHV-1 makes it a good candidate for creating a structure model, because these models may also be applied to other closely related proteins.

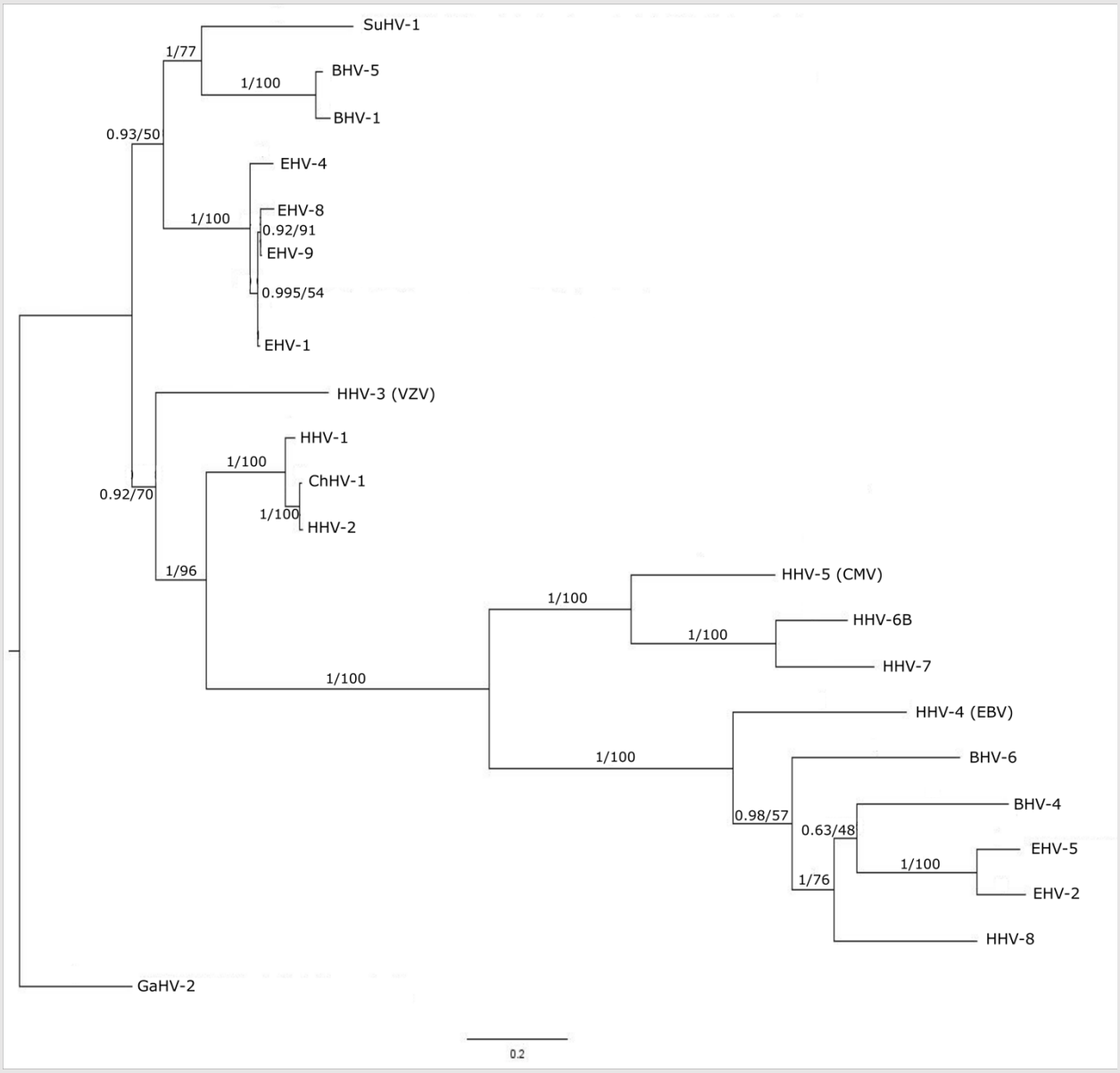

Figure 3. Phylogenetic tree of 21 homologs of HHV-1 pUL15. The tree is rooted on an out group GaHV-2. Each sequence is labeled by the abbreviation for the virus. Each node shows the posterior probability value from Bayesian analysis followed by the bootstrap support value reported as percentages from maximum likelihood analysis. The scale bar represents the number of substitutions along each branch.

Structure model for pUL15

To create this model, a partial crystal structure of the C-terminal domain of pUL15 was used, whereas the N-terminal domain was modeled based on DNA packaging protein gp17 from enterobacteria phage T4. Gp17 was chosen due to its high homology to the N-terminal domain of

pUL15. In order to create an accurate model, the two templates needed to be properly oriented. This was accomplished through the superimposition of both of the templates onto another structure of T4 gp17 (3cpe) as shown as gray ribbons in Figure 4. 3cpe was found to have both a substantially low E-value and a high coverage from the pUL15 template search. The superimposition of 2o0j (yellow ribbons in Figure 4) onto 3cpe yielded a Cα-RMSD of 1.34 Å and a TM-score of 0.62 with a 100% sequence identity. The superimposition of 4iox (purple ribbons in Figure 4) onto 3cpe yielded a Cα-RMSD of 3.21 Å and a TM-score of 0.28 with 10% sequence identity. Similarity scores obtained for 2o0j superposition onto 3cpe indicate a more successful alignment because 2o0j is actually a partial structure of 3cpe. The differences that are evident in the superimposition can most likely be attributed to the conformation that 2o0j takes when it is bound to ATP as opposed as the conformation it would take unbound to a ligand. While the scores obtained for 4iox do not show a high similarity to 3cpe, the purpose of this superimposition was simply to orient the templates to produce a plausible model. The produced model covers residues 127-735 due to the lack of suitable templates to cover the first 126 amino acids of pUL15. The model of pUL15 was subsequently subject to a series of analyses to gain insights into its molecular structure and function.

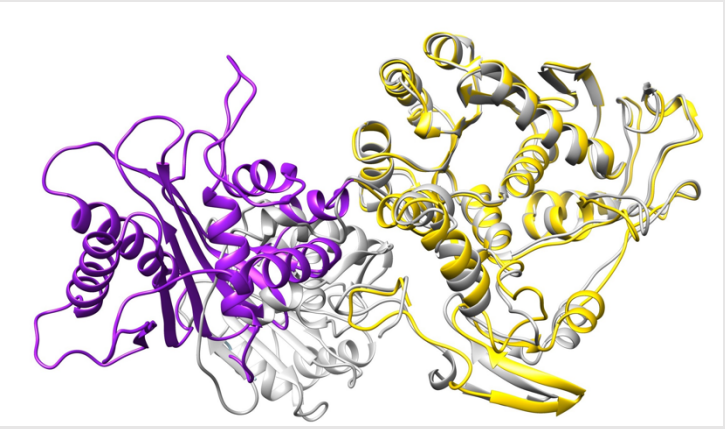

Figure 4. Superimposition of Templates. A superimposition of 4iox (purple) and 2o0j (yellow) onto 3cpe (gray) that was used to create the pUL15 model.

Cross-correlation between pairwise fluctuations

To further investigate the distinct domains in pUL15 and analyze the motion of these domains, a cross-correlation heat map was generated based on the PCA vectors using DynOmics online server [66]. The heat map presented in Figure 5 shows that there are two distinct areas where residue movements are correlated (red color). The first region corresponds to the ATP binding domain (first 350 amino acids), and the second region corresponds to the DNA binding region. Moreover, the two domains show anti-correlated movement (blue color), which can generate the movement necessary to push or screw the DNA. The heat map also shows that there is a hinge region that connects the two domains at residues 467-487. We predict that the functions of both of these domains may rely on the movement of the hinge region; therefore, we expect this hinge region to be a potential allosteric site for drug development.

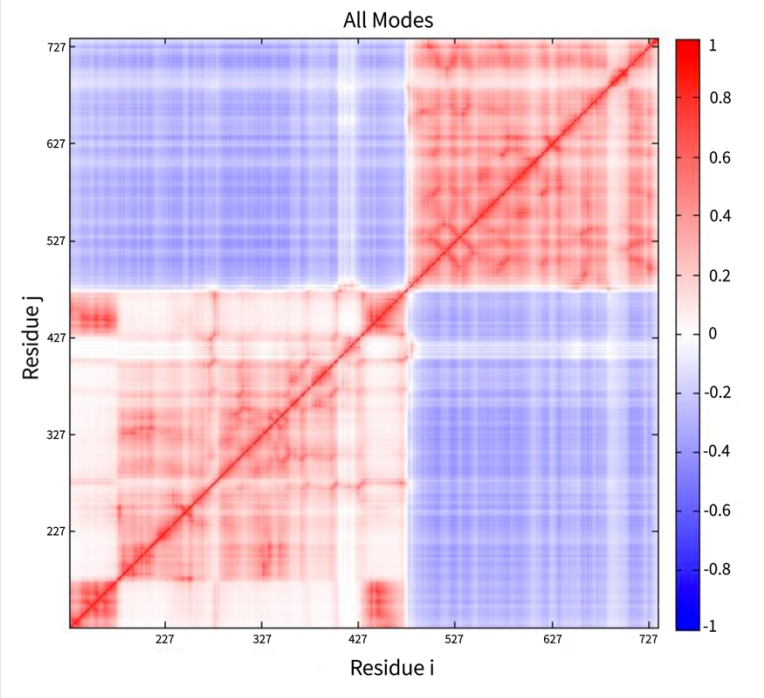

Figure 5. Orientational cross-correlations between residue fluctuations. The cross-correlation (CC) map shows the extent of connection between two nodes i, and j, where nodes represent amino acid residues. The values vary from -1 (fully anticorrelated motions; blue) to +1 (fully correlated; red). This plot shows that there are two distinct domains in this protein moving in opposite directions.

Ligand-binding site in pUL15

A putative binding site in the pUL15 model was predicted by *e*FindSite with a 97.9% confidence. This site is comprised of residues A244, A245, V246, T254, V255, L257, R260, F279, R280, G281, I282, K283, I284, G285, and G319 which are highlighted in cyan in Figure 6. All of these residues lie on the N-terminal domain of the protein, which is consistent with our previous predictions that pUL15 has two distinct domains, a ligand-binding domain and a DNA-binding domain. Also, both of the template proteins were bound to ligands, 2o0j was bound to ADP while 2o0h was bound to ATP. Furthermore, fingerprint-based virtual screening conducted with *e*FindSite suggests that this binding pocket is likely an ATP-binding site. Interestingly, the predicted binding site was also found next to two ATP binding motifs known as Walker A and Walker B boxes, as shown in Figure 6. The Walker A motif has been suggested to play a critical role in the hydrolysis and binding of ATP. It works in conjunction with the Walker B motif which is responsible for coordinating the magnesium ion of the Mg^{2+}- ATP complex. These two motifs work together to properly orient the phosphates of ATP for hydrolysis. The binding pocket for ATP have been found upstream from the Walker A motif of gp17 of the T4 bacteriophage, which was used as a template for the protein model. Walker A and Walker B boxes are found at residues 249-272 and 345-362, respectively [74]. Another study suggested somewhat narrower ranges for Walker A (residues 258-265) and Walker B (residues 352-357) motifs [75].

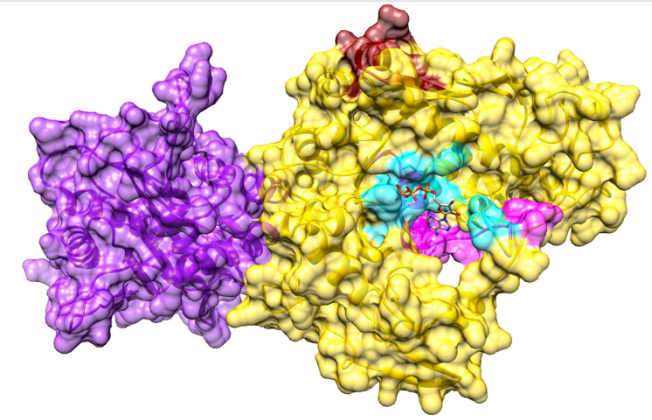

Figure 6. Model of pUL15 with docked ATP. A model of pUL15 with ATP docked. The N-terminal of the protein is highlighted in yellow and the C-terminal is highlighted in purple. The predicted ATP-binding site is in cyan, and the two walker boxes are also shown (Magenta = Walker box A; Brick Red = Walker box).

The predicted ATP binding site is consistent with several of the residues being found upstream of and overlapping with the Walker A box, further indicating that the binding site was properly predicted in the model.

Model of pUL15 complexed with ATP

*e*FindSite prediction strongly indicates that the putative binding site accommodates ATP. To further investigate this finding, an pUL15-ATP complex was modeled with *e*SimDock. Encouragingly, *e*SimDock reported a fitness score of 0.51, a binding probability of 0.84, and the estimated pK_d of 6.11. These scores corroborate the *e*FindSite prediction that ATP is the likely ligand of the predicted binding site. The ATP molecule that was docked did contain a few minor clashes between the ligand and the protein, which may account for the somewhat low fitness score. In order to remove the clashes, the model was refined by removing binding residues within 4.5Å from any ATP atom and remodeling pUL15 in the presence of ATP. The analysis of the ATP binding pose in the refined pUL15-ATP complex reveals a hydrogen bond between a phosphate moiety and K283 and numerous hydrophobic contacts with residues D243, A244, F256, F279, G281, I282, and I284 (Figure 7). In order to determine how the binding energy of the modeled pUL15-ATP complex compares to other protein-ATP complexes, a DFIRE score was calculated for the modeled complex and 1,553 other protein-ATP complexes. The DFIRE score for the modeled pUL15-ATP complex is -255.13, which corresponds to the upper range of the third quartile of the distribution of DFIRE scores across protein-ATP complexes (Figure 8). This shows that the interaction energy between the ATP molecule and pUL15 is comparable to that in many experimental complex structures involving ATP, making the model plausible.

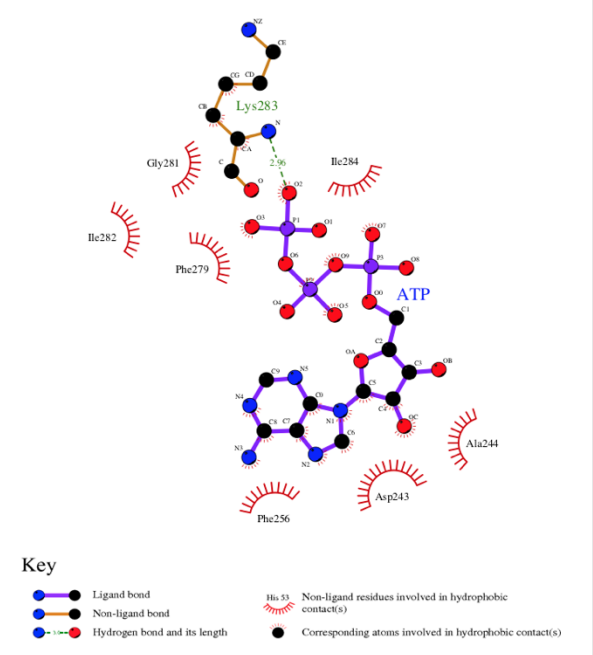

Figure 7. Ligand-Protein Contact Map. The residues in pUL15 binding pocket that interact with the docked ATP based on their proximity are displayed. The various types of bonding is also indicated.

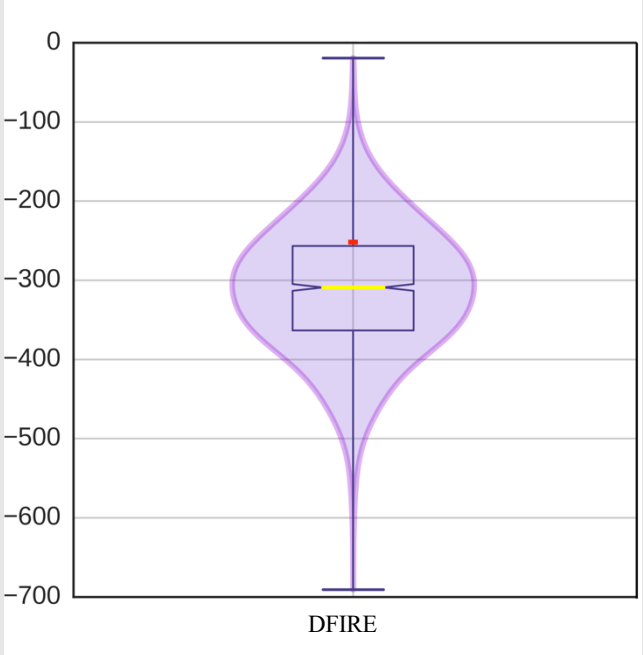

Figure 8. DFIRE Score for Protein-ATP Complexes. A violin plot of the DFIRE scores from a database of 1552 protein-ATP complexes. (Mean: -311.35, 25%: -363.31, 50%: -309.00, 75%: -256.62).

Model of pUL15 bound to DNA

A sequence-based analysis of the protein was conducted with DP-Bind and DISPLAR. Since sequence-based predictions may not be reliable, a review was done of currently available literature discussing potential DNA binding residues in pUL15. To test these predictions, we docked DNA onto the model. We computationally created a small strand of DNA made of the HHV-1 sequence using 3D-DART and docked it onto the modeled protein with HADDOCK. Of the top five results, the third cluster was used because it was most consistent with the binding residues predicted from DP-Bind, DISPLAR, and the published literature. In addition to the heat map described above, the DynOmics online server also generated a molecular motion using an Anisotropic Network Model (ANM). This ANM model provides information on how the residues in the protein are predicted

to move. Interestingly, the constructed pUL15-DNA model shown in Figure 9 reveals three loop structures flanking the DNA. Two are flanking the DNA on one side, while the third is positioned on the other side of the DNA. That, along with the ANM analysis, indicates that the protein is moving in a manner perpendicular to the DNA. This model therefore supports a hypothesis that the DNA movement likely involves either the rotation or revolution rather than a simple linear motion.

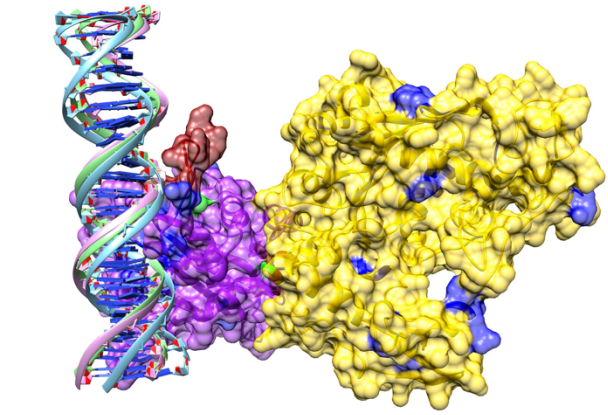

Figure 9. **Model of pUL15 with docked DNA.** A model of pUL15 with DNA docked. The N-terminal of the protein is highlighted in yellow and the C-terminal is highlighted in purple. The DNA binding residue predictions with DISPLAR (blue), DP-Bind (red) and overlap (green) are also shown.

Discussion

Many double-stranded (ds)DNA viruses utilize a motor to package their DNA into an empty capsid and therefore may be classified under type II DNA-packaging. However, the exact mechanism of the packaging of the DNA is unknown. There have been several hypotheses that have surfaced to try to answer this question. One very popular theory involves the idea that the proteins in the packaging motor push the DNA into the capsid in a linear manner. With the linear hypothesis, it is believed that the protein utilizes the hydrolysis of ATP in order to clamp down onto the DNA and push it up into the empty capsid [76]. Most dsDNA viruses contain packaging proteins that are composed of two separate domains, an ATPase domain as well as a DNA-binding/nuclease domain. These two functions are linked together, which is necessary in order for proper DNA-packaging to occur. This mechanism begins with the DNA-binding domain binding the viral DNA.

Then, once the DNA is bound, the DNA-binding domain gets the viral DNA closer to the ATPase domain of the same protein. A conformational change takes place in the ATPase domain, prepping the ATPase active site for hydrolysis. The subsequent hydrolysis of ATP causes another conformational change that moves the DNA-binding domain in such a way that it pushes the viral DNA into the capsid. When ADP and Pi are released, the DNA-binding domain releases the DNA and returns to its original conformation. When the DNA is released from the DNA-binding domain, it ends up being bound to the DNA-binding domain of an adjacent subunit [77].

A second hypothesis, the rotation theory, states that the DNA is packaged into the capsid with a screw-like motion as the packaging proteins rotate to facilitate this motion. More specifically, it refers to the possibility of a connecter protein in the packaging motor rotating around the viral DNA, and therefore, using this movement to screw the DNA into the capsid. More recent studies have suggested that the rotational method does not happen for ds-DNA viruses. Additionally, it was reported that rotational motors have relatively small channels that are typically smaller than the width of dsDNA, so when the DNA goes through that channel, it would require the dsDNA to split into ssDNA [76]. A more plausible theory states that the packaging motors utilize a revolving mechanism instead. This means that instead of the proteins rotating in order push the DNA through, the proteins are relatively stationary, and the DNA revolves around the channel due to the backbone interactions with the walls of the channel. Revolution motors have been found to have diameters that are substantially larger than the width of dsDNA, allowing the DNA to go through the revolving mechanism [76].

Another hypothesis considers DNA to be what provides the energy for its own translocation rather than the protein doing all of the work [78]. This is referred to as the "scrunchworm hypothesis." It is based on how during the process of packaging, DNA is found to compress or

"scrunch." This is the result of the proteins removing the DNA from solvent by binding to it, causing dehydration and compression of the DNA. The energy stored in the DNA at this compressed state is then responsible for the ultimate movement of the DNA into the empty capsid.

The GNM of our modeled pUL15 indicated that it contains two domains which move in opposite directions of one another and are connected by a hinge in the protein. When dsDNA was docked onto our model, we found that the DNA was surrounded by three loop structures which flanked it on the right and left. Taking a look at the ANM model, it showed that the movement of the protein upon ATP hydrolysis would most likely be perpendicular to that of the DNA, contrary to the more popular linear theory. Our data indicates that packaging of DNA would follow more closely with either the rotation or revolution theory. Due to the limitations of the rotation theory, particularly the small size of the channel for the dsDNA to pass through, it is more likely that the DNA is packaged according to the revolution theory in HHV-1 due to its abundant space for the dsDNA to pass through the channel.

The proteins that are directly involved with the propagation of the concatemeric viral DNA into the capsid have the most potential as drug targets due to their essential role in the formation of new viruses. Out of all of the packaging proteins, the terminase proteins maintain the crucial functions of utilizing ATP hydrolysis to push the concatemeric DNA into the capsid and cleave it once packaging is complete. This makes it a promising drug target because there are multiple functions that may be disturbed to prevent the formation of new viruses. With this information, we predict that an effective allosteric drug target would be the hinge region of pUL15, connecting the DNA and ATP binding domains. By blocking this hinge region, it would be possible to prevent the conformational change in pUL15 that results in the movement of the dsDNA into the capsid. It is also a favorable site for drug targeting both because of its uniqueness to HHV, which could

decrease drug toxicity, and also due to the significant conservation of this protein, which could allow the drug to be used across multiple HHVs.

Conclusions

Through evolutionary analysis, we determined that focusing on HHV-1 would be most effective not just because it is the most abundant form of HHV, but also because it is closely related to the other HHVs. We determined that the most ideal target would be the packaging motor due to its crucial role in the lifecycle of the virus and how it is highly conserved throughout all herpesviridae. The rate of change for all of the packaging motor proteins was experimentally found, and the large terminase pUL15 was the most conserved, not just in the packaging motor, but in all of the viral proteins. Currently, there is only a partial model of this protein (PDB: 4IOX), so we constructed a more complete protein in order to analyze its functions. From this model, we were able to predict putative ATP and DNA binding sites (with high confidence). In addition to this, we used Gaussian network model (GNM) to predict the correlation between residue fluctuations, which not only supported the presence of two distinct domains in the protein but also revealed a potential hinge region between the ATP-binding and DNA-binding domains. An Anisotropic Network Model (ANM) was used to observe the probable movements of the protein and indicated that the HHV-1 packaging motor likely utilizes a revolutionary mechanism to package DNA. If the hinge region that was predicted can be confirmed in subsequent studies, then it will be a promising target for antiviral drug development against HHVs.

References

1. Fishman JA, Emery V, Freeman R, Pascual M, Rostaing L, Schlitt HJ, Sgarabotto D, Torre-Cisneros J, Uknis ME: **Cytomegalovirus in transplantation - challenging the status quo.** *Clin Transplant* 2007, **21:**149-158.
2. Penkert RR, Kalejta RF: **Tegument protein control of latent herpesvirus establishment and animation.** *Herpesviridae* 2011, **2:**3.
3. Chisholm C, Lopez L: **Cutaneous infections caused by Herpesviridae: a review.** *Arch Pathol Lab Med* 2011, **135:**1357-1362.
4. Grinde B: **Herpesviruses: latency and reactivation - viral strategies and host response.** *J Oral Microbiol* 2013, **5**.
5. Kotton CN: **Management of cytomegalovirus infection in solid organ transplantation.** *Nat Rev Nephrol* 2010, **6:**711-721.
6. Whitley RJ: **Herpesviruses.** In *Medical Microbiology.* Edited by th, Baron S. Galveston (TX); 1996
7. Boppana SB, Fowler KB, Britt WJ, Stagno S, Pass RF: **Symptomatic congenital cytomegalovirus infection in infants born to mothers with preexisting immunity to cytomegalovirus.** *Pediatrics* 1999, **104:**55-60.
8. Boppana SB, Pass RF, Britt WJ, Stagno S, Alford CA: **Symptomatic congenital cytomegalovirus infection: neonatal morbidity and mortality.** *Pediatr Infect Dis J* 1992, **11:**93-99.
9. Coll O, Benoist G, Ville Y, Weisman LE, Botet F, Anceschi MM, Greenough A, Gibbs RS, Carbonell-Estrany X, Group WPIW: **Guidelines on CMV congenital infection.** *J Perinat Med* 2009, **37:**433-445.

10. Peritz DC, Duncan C, Kurek K, Perez-Atayde AR, Lehmann LE: **Visceral varicella zoster virus (VZV) after allogeneic hematopoietic stem cell transplant (HSCT) in pediatric patients with chronic graft-versus-host disease (cGVHD).** *J Pediatr Hematol Oncol* 2008, **30:**931-934.

11. Bonnet M, Guinebretiere JM, Kremmer E, Grunewald V, Benhamou E, Contesso G, Joab I: **Detection of Epstein-Barr virus in invasive breast cancers.** *J Natl Cancer Inst* 1999, **91:**1376-1381.

12. Wong M, Pagano JS, Schiller JT, Tevethia SS, Raab-Traub N, Gruber J: **New associations of human papillomavirus, Simian virus 40, and Epstein-Barr virus with human cancer.** *J Natl Cancer Inst* 2002, **94:**1832-1836.

13. Schulz TF: **The pleiotropic effects of Kaposi's sarcoma herpesvirus.** *J Pathol* 2006, **208:**187-198.

14. Sunil M, Reid E, Lechowicz MJ: **Update on HHV-8-Associated Malignancies.** *Curr Infect Dis Rep* 2010, **12:**147-154.

15. Razonable RR: **Antiviral drugs for viruses other than human immunodeficiency virus.** *Mayo Clin Proc* 2011, **86:**1009-1026.

16. Griffiths PD, Boeck M: **Antiviral therapy for human cytomegalovirus.** In *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Edited by Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K. Cambridge: Cambridge University Press; 2007

17. Kimberlin DW, Whitley RJ: **Antiviral therapy of HSV-1 and -2.** In *Human Herpesviruses: Biology, Therapy, and Immunoprophylaxis.* Edited by Arvin A,

Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K. Cambridge: Cambridge University Press; 2007

18. Morfin F, Thouvenot D: **Herpes simplex virus resistance to antiviral drugs.** *J Clin Virol* 2003, **26:**29-37.

19. Lurain NS, Chou S: **Antiviral drug resistance of human cytomegalovirus.** *Clin Microbiol Rev* 2010, **23:**689-712.

20. Schreiber A, Harter G, Schubert A, Bunjes D, Mertens T, Michel D: **Antiviral treatment of cytomegalovirus infection and resistant strains.** *Expert Opin Pharmacother* 2009, **10:**191-209.

21. Dittmer A, Woskobojnik I, Adfeldt R, Drach JC, Townsend LB, Voigt S, Bogner E: **Tetrahalogenated benzimidazole D-ribonucleosides are active against rat cytomegalovirus.** *Antiviral Res* 2017, **137:**102-107.

22. Goldner T, Hewlett G, Ettischer N, Ruebsamen-Schaeff H, Zimmermann H, Lischka P: **The novel anticytomegalovirus compound AIC246 (Letermovir) inhibits human cytomegalovirus replication through a specific antiviral mechanism that involves the viral terminase.** *J Virol* 2011, **85:**10884-10893.

23. Krosky PM, Underwood MR, Turk SR, Feng KW, Jain RK, Ptak RG, Westerman AC, Biron KK, Townsend LB, Drach JC: **Resistance of human cytomegalovirus to benzimidazole ribonucleosides maps to two open reading frames: UL89 and UL56.** *J Virol* 1998, **72:**4721-4728.

24. Underwood MR, Harvey RJ, Stanat SC, Hemphill ML, Miller T, Drach JC, Townsend LB, Biron KK: **Inhibition of human cytomegalovirus DNA maturation by a benzimidazole ribonucleoside is mediated through the UL89 gene product.** *J Virol* 1998, **72:**717-725.

25. Biron KK: **Antiviral drugs for cytomegalovirus diseases.** *Antiviral Res* 2006, **71:**154-163.

26. Good SS, Owens BS, Townsend LB, Drach JC: **The disposition in rats and monkey of 2-bromo-5,6-dichloro-1-(ß-D-ribofuranosyl)-benzimidazole (BDCRB) and its 2,5,6-trichloro conger (TCRB).** *Antiviral Res* 1994, **23:**103.

27. Hwang JS, Kregler O, Schilf R, Bannert N, Drach JC, Townsend LB, Bogner E: **Identification of acetylated, tetrahalogenated benzimidazole D-ribonucleosides with enhanced activity against human cytomegalovirus.** *J Virol* 2007, **81:**11604-11611.

28. Hwang JS, Schilf R, Drach JC, Townsend LB, Bogner E: **Susceptibilities of human cytomegalovirus clinical isolates and other herpesviruses to new acetylated, tetrahalogenated benzimidazole D-ribonucleosides.** *Antimicrob Agents Chemother* 2009, **53:**5095-5101.

29. Dittmer A, Drach JC, Townsend LB, Fischer A, Bogner E: **Interaction of the putative human cytomegalovirus portal protein pUL104 with the large terminase subunit pUL56 and its inhibition by benzimidazole-D-ribonucleosides.** *J Virol* 2005, **79:**14660-14667.

30. Zhou B, Yang K, Wills E, Tang L, Baines JD: **A mutation in the DNA polymerase accessory factor of herpes simplex virus 1 restores viral DNA replication in the presence of raltegravir.** *J Virol* 2014, **88:**11121-11129.

31. Goldner T, Zimmermann H, Lischka P: **Phenotypic characterization of two naturally occurring human Cytomegalovirus sequence polymorphisms located in a distinct region of ORF UL56 known to be involved in in vitro resistance to letermovir.** *Antiviral Res* 2015, **116:**48-50.

32. Melendez DP, Razonable RR: **Letermovir and inhibitors of the terminase complex: a promising new class of investigational antiviral drugs against human cytomegalovirus.** *Infect Drug Resist* 2015, **8:**269-277.

33. Marschall M, Stamminger T, Urban A, Wildum S, Ruebsamen-Schaeff H, Zimmermann H, Lischka P: **In vitro evaluation of the activities of the novel anticytomegalovirus compound AIC246 (letermovir) against herpesviruses and other human pathogenic viruses.** *Antimicrob Agents Chemother* 2012, **56:**1135-1137.

34. Clark K, Karsch-Mizrachi I, Lipman DJ, Ostell J, Sayers EW: **GenBank.** *Nucleic Acids Res* 2016, **44:**D67-72.

35. Katoh K, Standley DM: **MAFFT multiple sequence alignment software version 7: improvements in performance and usability.** *Mol Biol Evol* 2013, **30:**772-780.

36. Schwarz G: **Estimating the dimension of a model.** *Ann Stat* 1978, **6:**461-464.

37. Darriba D, Taboada GL, Doallo R, Posada D: **ProtTest 3: fast selection of best-fit models of protein evolution.** *Bioinformatics* 2011, **27:**1164-1165.

38. Jones DT, Taylor WR, Thornton JM: **The rapid generation of mutation data matrices from protein sequences.** *Comput Appl Biosci* 1992, **8:**275-282.

39. Ronquist F, Huelsenbeck JP: **MrBayes 3: Bayesian phylogenetic inference under mixed models.** *Bioinformatics* 2003, **19:**1572-1574.

40. Rambaut A, Drummond AJ, Xie D, Baele G, Suchard MA: **Posterior summarization in bayesian phylogenetics using Tracer 1.7.** *Syst Biol* 2018, **67:**901-904.

41. Altschul SF, Gish W, Miller W, Myers EW, Lipman DJ: **Basic local alignment search tool.** *J Mol Biol* 1990, **215:**403-410.

42. O'Leary NA, Wright MW, Brister JR, Ciufo S, Haddad D, McVeigh R, Rajput B, Robbertse B, Smith-White B, Ako-Adjei D, et al: **Reference sequence (RefSeq) database at NCBI: current status, taxonomic expansion, and functional annotation.** *Nucleic Acids Res* 2016, **44:**D733-745.

43. Edgar RC: **MUSCLE: multiple sequence alignment with high accuracy and high throughput.** *Nucleic Acids Res* 2004, **32:**1792-1797.

44. Abascal F, Zardoya R, Posada D: **ProtTest: selection of best-fit models of protein evolution.** *Bioinformatics* 2005, **21:**2104-2105.

45. Stamatakis A: **RAxML version 8: a tool for phylogenetic analysis and post-analysis of large phylogenies.** *Bioinformatics* 2014, **30:**1312-1313.

46. Huelsenbeck JP, Ronquist F, Nielsen R, Bollback JP: **Bayesian inference of phylogeny and its impact on evolutionary biology.** *Science* 2001, **294:**2310-2314.

47. Rambaut A: **http://tree.bio.ed.ac.uk/software/figtree/.**

48. Harrington B: **https://inkscape.org/.**

49. Selvarajan Sigamani S, Zhao H, Kamau YN, Baines JD, Tang L: **The structure of the herpes simplex virus DNA-packaging terminase pUL15 nuclease domain suggests an evolutionary lineage among eukaryotic and prokaryotic viruses.** *J Virol* 2013, **87:**7140-7148.

50. Berman HM, Battistuz T, Bhat TN, Bluhm WF, Bourne PE, Burkhardt K, Feng Z, Gilliland GL, Iype L, Jain S, et al: **The Protein Data Bank.** *Acta Crystallogr D Biol Crystallogr* 2002, **58:**899-907.

51. Zimmermann L, Stephens A, Nam SZ, Rau D, Kubler J, Lozajic M, Gabler F, Soding J, Lupas AN, Alva V: **A completely reimplemented MPI Bioinformatics Toolkit with a new HHpred server at its core.** *J Mol Biol* 2018, **430:**2237-2243.

52. Sun S, Kondabagil K, Gentz PM, Rossmann MG, Rao VB: **The structure of the ATPase that powers DNA packaging into bacteriophage T4 procapsids.** *Mol Cell* 2007, **25:**943-949.

53. Sun S, Kondabagil K, Draper B, Alam TI, Bowman VD, Zhang Z, Hegde S, Fokine A, Rossmann MG, Rao VB: **The structure of the phage T4 DNA packaging motor suggests a mechanism dependent on electrostatic forces.** *Cell* 2008, **135:**1251-1262.

54. Pandit SB, Skolnick J: **Fr-TM-align: a new protein structural alignment method based on fragment alignments and the TM-score.** *BMC Bioinformatics* 2008, **9:**531.

55. Braberg H, Webb BM, Tjioe E, Pieper U, Sali A, Madhusudhan MS: **SALIGN: a web server for alignment of multiple protein sequences and structures.** *Bioinformatics* 2012, **28:**2072-2073.

56. Webb B, Sali A: **Comparative protein structure modeling using MODELLER.** *Curr Protoc Protein Sci* 2016, **86:**2 9 1-2 9 37.

57. Brylinski M, Feinstein WP: **eFindSite: improved prediction of ligand binding sites in protein models using meta-threading, machine learning and auxiliary ligands.** *J Comput Aided Mol Des* 2013, **27:**551-567.

58. Feinstein WP, Brylinski M: **eFindSite: Enhanced fingerprint-based virtual screening against predicted ligand binding sites in protein models.** *Mol Inform* 2014, **33:**135-150.

59. Brylinski M: **Nonlinear scoring functions for similarity-based ligand docking and binding affinity prediction.** *J Chem Inf Model* 2013, **53:**3097-3112.

60. Laskowski RA, Swindells MB: **LigPlot+: multiple ligand-protein interaction diagrams for drug discovery.** *J Chem Inf Model* 2011, **51:**2778-2786.

61. Zhang C, Liu S, Zhu Q, Zhou Y: **A knowledge-based energy function for protein-ligand, protein-protein, and protein-DNA complexes.** *J Med Chem* 2005, **48:**2325-2335.

62. Krissinel E, Henrick K: **Inference of macromolecular assemblies from crystalline state.** *J Mol Biol* 2007, **372:**774-797.

63. Tjong H, Zhou HX: **DISPLAR: an accurate method for predicting DNA-binding sites on protein surfaces.** *Nucleic Acids Res* 2007, **35:**1465-1477.

64. Hwang S, Gou Z, Kuznetsov IB: **DP-Bind: a web server for sequence-based prediction of DNA-binding residues in DNA-binding proteins.** *Bioinformatics* 2007, **23:**634-636.

65. van Dijk M, Bonvin AM: **3D-DART: a DNA structure modelling server.** *Nucleic Acids Res* 2009, **37:**W235-239.

66. Li H, Chang YY, Lee JY, Bahar I, Yang LW: **DynOmics: dynamics of structural proteome and beyond.** *Nucleic Acids Res* 2017, **45:**W374-W380.

67. Sato H, Yokoyama M, Toh H: **Genomics and computational science for virus research.** *Front Microbiol* 2013, **4:**42.

68. Mann JM: **The World Health Organization's global strategy for the prevention and control of AIDS.** *West J Med* 1987, **147:**732-734.

69. Johnston C, Gottlieb SL, Wald A: **Status of vaccine research and development of vaccines for herpes simplex virus.** *Vaccine* 2016, **34:**2948-2952.

70. Deshpande SP, Kumaraguru U, Rouse BT: **Why do we lack an effective vaccine against herpes simplex virus infections?** *Microbes Infect* 2000, **2:**973-978.

71. Wald A, Corey L: **Persistence in the population: epidemiology, transmission.** In *Human herpesviruses: Biology, therapy, and immunoprophylaxis.* Edited by Arvin A, Campadelli-Fiume G, Mocarski E, Moore PS, Roizman B, Whitley R, Yamanishi K. Cambridge: Cambridge University Press; 2007

72. Dolan A, Jamieson FE, Cunningham C, Barnett BC, McGeoch DJ: **The genome sequence of herpes simplex virus type 2.** *J Virol* 1998, **72:**2010-2021.

73. Xu F, Schillinger JA, Sternberg MR, Johnson RE, Lee FK, Nahmias AJ, Markowitz LE: **Seroprevalence and coinfection with herpes simplex virus type 1 and type 2 in the United States, 1988-1994.** *J Infect Dis* 2002, **185:**1019-1024.

74. Przech AJ, Yu D, Weller SK: **Point mutations in exon I of the herpes simplex virus putative terminase subunit, UL15, indicate that the most conserved residues are essential for cleavage and packaging.** *J Virol* 2003, **77:**9613-9621.

75. Rao VB, Mitchell MS: **The N-terminal ATPase site in the large terminase protein gp17 is critically required for DNA packaging in bacteriophage T4.** *J Mol Biol* 2001, **314:**401-411.

76. Pi F, Zhao Z, Chelikani V, Yoder K, Kvaratskhelia M, Guo P: **Development of Potent Antiviral Drugs Inspired by Viral Hexameric DNA-Packaging Motors with Revolving Mechanism.** *J Virol* 2016, **90:**8036-8046.

77. Sun S, Rao VB, Rossmann MG: **Genome packaging in viruses.** *Curr Opin Struct Biol* 2010, **20:**114-120.

78. Waters JT, Kim HD, Gumbart JC, Lu XJ, Harvey SC: **DNA Scrunching in the Packaging of Viral Genomes.** *J Phys Chem B* 2016, **120:**6200-6207.