COMPARATIVE STUDY OF ORIGIN BINDING PROTEIN RELATIVE BINDING FREE ENERGIES

An Undergraduate Research Scholars Thesis

by

SARAH FROSS

Submitted to the LAUNCH: Undergraduate Research office at
Texas A&M University
in partial fulfillment of the requirements for the designation as an

UNDERGRADUATE RESEARCH SCHOLAR

Approved by
Research Advisor: Dr. Wonmuk Hwang

May 2022

Major: Biomedical Engineering

RESEARCH COMPLIANCE CERTIFICATION

Research activities involving the use of human subjects, vertebrate animals, and/or biohazards must be reviewed and approved by the appropriate Texas A&M University regulatory research committee (i.e., IRB, IACUC, IBC) before the activity can commence. This requirement applies to activities conducted at Texas A&M and to activities conducted at non-Texas A&M facilities or institutions. In both cases, students are responsible for working with the relevant Texas A&M research compliance program to ensure and document that all Texas A&M compliance obligations are met before the study begins.

I, Sarah Fross, certify that all research compliance requirements related to this Undergraduate Research Scholars thesis have been addressed with my Research Advisor prior to the collection of any data used in this final thesis submission.

This project did not require approval from the Texas A&M University Research Compliance & Biosafety office.

TABLE OF CONTENTS

ABSTRACT

Comparative Study of Origin Binding Protein Relative Binding Free Energies

Sarah Fross
Department of Biomedical Engineering
Texas A&M University

Research Faculty Advisor: Dr. Wonmuk Hwang
Departments of Biomedical Engineering, Materials Science & Engineering, and Physics & Astronomy
Texas A&M University

The Epstein-Barr virus (EBV) and Kaposi's Sarcoma-associated herpesvirus (KSHV) are prevalent worldwide with a 95% infection rate for EBV and as high as 50% for KSHV in adults. These herpesviruses code for DNA Origin-Binding Proteins (OBPs) such as Epstein-Barr Nuclear Antigen 1 (EBNA1) in EBV and Latency-Associated Nuclear Antigen (LANA) in KSHV. Both of these OBPs are responsible for oncogenic conditions (Hodgkin's lymphoma, Kaposi's Sarcoma, and gastric cancers) and episome maintenance during latency phases, making them attractive targets for inhibitor therapy. This study focused on understanding the atomic interactions between EBNA1, LANA, and their corresponding viral DNAs (vDNA) using molecular dynamics simulations of the EBNA1-vDNA^{EBV} and LANA-vDNA^{KSHV} complexes. Our analysis focused on calculating relative binding free energies (RBFE) allowing for a comparison to be made between EBNA1 and LANA RBFE. We found that EBNA1 binds more strongly to vDNA when compared to LANA. We determined noncovalent interactions, root-mean-squared fluctuation, root-mean-squared deviation from the initial structure, solvent-accessible surface area, and calculated entropy using the maximum information spanning tree method to further characterize the bound versus free states of these proteins. We found EBNA1 to make more bonds with vDNA, have a larger binding area, and

the RBFE to be -224.56 $\pm$ 33.1 kcal/mol for EBNA1 compared to -105.16 $\pm$ 17.9 kcal/mol for LANA. Determining the RBFE and characterizing the atomic interactions from simulating these viral proteins may contribute to inhibitor development for KSHV and EBV.

DEDICATION

To my family, friends, and the Hwang lab for believing in me and helping me reach new heights.

ACKNOWLEDGMENTS

I would like to thank my faculty advisor, Dr. Hwang, and my mentoring graduate students in the Hwang lab, Sofiya Bettencourt, James Gonzales, and Jie Shi, for their guidance and support throughout the course of this research.

A majority of the molecular dynamic simulations and computations completed herein were made possible from access to Texas A&M University High Performance Research Computing.

NOMENCLATURE

EBV	Epstein-Barr virus
EBNA1	Epstein-Barr Nuclear Antigen 1
KSHV	Kaposi's sarcoma-associated herpesvirus
LANA	Latency-Associated Nuclear Antigen
MD	molecular dynamics
CHARMM	Chemistry at Harvard Macromolecular Mechanics
X3DNA	X3DNA: Dissecting the Spatial Structure of RNA
VMD	Visual Molecular Dynamics
DBD	DNA-binding domain
OMM	OpenMM
TAMU HPRC	Texas A&M University High Performance Research Computing
vDNA	viral DNA
RMSD	root-mean-square deviation
RMSF	root-mean-square fluctuation
SASA	solvent accessible surface area
MIST	maximum information spanning tree
RBFE	relative binding free energy

1. BACKGROUND

1.1 Origin Binding Proteins

The central dogma of biology states that DNA can be transcribed into RNA which can be translated into proteins. For the purposes of viral replication, the type of protein that assists in DNA replication is known as an Origin DNA-Binding Protein (OBP) [1] or Episome Maintenance Protein (EMP) [2]. OBPs bind to specific sequences of viral DNA (vDNA) [1,3], initiate unwinding of the vDNA origin (the designated region for replication initiation on the genome, also referred to as origin of replication (oriP) [1, 4]), and interact with other proteins involved in replication. Additionally, EMPs can repair damaged episomes, segregate viral genomes when the host cell undergoes mitosis, and initiate antigen production from the "lysing" of cells during the lytic phase [2, 5, 6]. EMPs typically form homodimers and also share a region known as the DNA-Binding Domain (DBD) that is responsible for the sequence specific binding to the oriP [2]. Both OBPs for the Epstein-Barr virus (EBV) and Kaposi's Sarcoma-associated herpesvirus (KSHV) maintain the viral episome and enhance infection in their respective systems [2].

1.2 Epstein-Barr virus and EBNA1

1.2.1 Epidemiology

The Epstein-Barr virus (EBV) is a variant of the human gamma herpesvirus subtype 4 (HHV-4) [7], with an infection rate of roughly 95% worldwide among adults [8], while children and young adults have infection rates as low as 25% in developed countries [9]. Those infected with EBV between the ages of 15 and 24 years old have the highest likelihood of developing acute infectious mononucleosis (IM) ranging between 9 to 48 out of 1,000 people annually, whereas other infections have a rate of incidence of 5 out of 1,000 people [10]. EBV is largely transmitted via saliva or transplanted kidneys, hearts, livers, or bone marrow [11]. Blood tests are performed before surgical procedures to recognize vDNA in an attempt to minimize transmission of EBV and consequently other post-transplant lymphoproliferative disorders [12].

1.2.2 Viral Latency and Pathogenesis

EBV exits the latency phase when physiological signals trigger viral gene expression [6,13]. Although the specific stimuli are unknown, chemical reagents such as calcium ionophore, anti-immunoglobulin, and transforming growth factor β have been confirmed *in vitro* [6]. EBV infects epithelial, Natural Killer (NK)/T cells, and B lymphocytes resulting in epithelial cancers and lymphoproliferative disorders respectively [13, 14]. Initial infection is expressed as lytic cell replication within host cells then absolute latency [15]. During these latency phases, specific Epstein-Barr Nuclear Antigens (EBNA) become activated and are responsible for inhibiting apoptosis, instigating proliferation of the host cell, and inhibiting the immune response [7]. These functions are associated with oncogenic disorders such as Burkitt's/non-Hodgkin's lymphoma, Hodgkin's lymphoma, B-lymphoproliferative disease, diffuse large B cell lymphoma, rare immunocompromised B-lymphoma, T cell or Natural Killer (NK/T) lymphoproliferations, gastric cancers, breast cancers, and nasopharyngeal cancers [7, 15].

Proliferation of infected cells can be attributed to the overstimulation of the oncogenic cellular myelocytomatosis (c-Myc) gene [15–17]. The inhibition of apoptosis arises from the mutation of host-cell tumor protein 53 (p53) gene because the normal physiological response of increased c-Myc activity results in increased p53-controlled apoptosis [15]. The usual binding to p53 is performed by a cellular ubiquitin-specific protease (USP7) that elicits apoptosis [18]. However, the first Epstein-Barr Nuclear Antigen (EBNA1) has also been found to inhibit apoptosis by out-competing USP7 for binding to p53 [7, 18]. Lastly, inhibition of the immune response can be linked to the glycine-alanine (GA) repeating sequence present on the N-terminal of the EBNA1 [19], which includes residues 90 to 324 that are not shown in **Figure 1.1** [20]. When this segment binds to the 5' end of vDNA^{EBV}, the GA repeats can halt antigen presentation to the major histocompatibility complex type 1 (MHC-1) that is responsible for antigen presentation to T cells [20, 21].

EBV is known to have four stages of latency associated with increasing amounts of viral protein activity [7, 15]; Latency 0 has no activity, Latency I with EBNA1, Latency II with EBNA1

and latent membrane protein 1 & 2 (LMP-1/2), and Latency III with EBNA1, 2, 3A, 3B, 3C, & 5. During all latency phases, the OBP EBNA1 binds to the oriP on the $vDNA^{EBV}$ to trigger DNA replication following the host cell's division, facilitates genetic segregation during replication, and allows for $vDNA^{EBV}$ expression [1].

1.2.3 EBNA1 as an Origin Binding Protein

EBNA1 specifically binds to $vDNA^{EBV}$, as seen in **Figure 2.1** over amino acids/residues 459-607, where residues 459 through 503 (the flanking domain) are responsible for DNA-binding, as seen in **Figure 1.1**, and residues 504 through 604 (the core domain) are responsible for its dimeric structure [1, 22, 23]. This dimer structure is essential for DNA-binding because both monomers constitute the DBD and alter the DNA conformation [2, 22]. During metaphase when the genome is segregated, EBNA1 uses arginine-glycine rich domains to bind to adenine-thymine rich regions on the $vDNA^{EBV}$ minor groove [2], otherwise known as the linking region that follows the GA repeating sequence [20]. This includes residues 325 to 376, and are not shown in **Figure 1.1**. EBNA1 also has the ability to "loop" the $vDNA^{EBV}$ when attached to separate DBDs to "link" those regions [3]. DNA-binding aside, EBNA1 is not enzymatically active to unwind the $vDNA^{EBV}$ and instead affects the genome structure and recruits other proteins [3, 24]. EBNA1 has a role in all activity of EBV's latency phases and is thus an appealing target for developing a treatment for EBV.

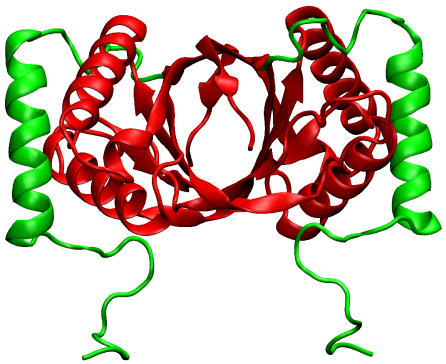

Figure 1.1: DNA Binding Domain (flanking domain) on EBNA1 in green.

1.3 Kaposi's Sarcoma-Associated Herpesvirus and LANA1

1.3.1 Epidemiology

Kaposi's Sarcoma herpesvirus (KSHV) is a human herpesvirus type 8 (HHV-8) [25–27]. KSHV can be expressed as Kaposi's Sarcoma (KS) in different forms depending on the geographical region and is mainly seen in male patients [26]. KS can be expressed as classic/indolent KS, African endemic KS, iatrogenic/post-transplant KS, or AIDS-related epidemic KS [26, 27]. For example, endemic KS is prevalent among 50% of men in Uganda compared to less than 6% in North America [26,28]. KSHV can also cause lymphoproliferative disorders such as primary effusion lymphoma (PEL), the plasmablastic variant of multicentric Castleman's disease (MCD), and KSHV inflammatory cytokine syndrome (KICS) [26, 27]. KSHV is transmitted amongst children via saliva and homosexual men via sexual contact [26]. KSHV can be detected with an immunofluorescence assay using infected cells or enzyme-linked immunoassay using capsid proteins [26]. The main latency-associated nuclear antigen (LANA) of KSHV is also used in immunohistochemistry assays to detect infection [26].

1.3.2 Viral Activity

KSHV has latency phases managed by LANA, which assists in genetic segregation [26,29], stimulates c-Myc for proliferation [26, 30], and inhibits p53 to prevent apoptosis [26]. In order to accomplish the function of evading immune response, lytic genome expression is turned off through specific signaling pathways, such as the Wingless-related integration site (Wnt) signaling, similar to EBV's LMP-1, after a few replication cycles allowing for latent infection [26,31]. KSHV becomes activated through the stress pathway of the endoplasmic reticulum where the KSHV chromosome conformation associated with latency is altered [32]. Interleukin-8 (IL-8) is a signaling chemokine that assists in the upregulation of neutrophils and T cells in the inflammatory response [31]. As part of KSHV infection, LANA increases IL-8 production to invoke infected neutrophils to transmigrate into tissue through the endothelium [33]. This is different however from EBV's action of binding to IL-8 promoters which upregulate IL-8 production and allows for

nasopharyngeal cell growth [31].

1.3.3 LANA as an Origin Binding Protein

LANA is also an OBP and as such functions similarly to EBNA1. Similar to EBNA1, LANA uses α-helices 1 and 2 on one monomer and one of the outstretched N-terminal arms on the other monomer to recognize the DBD, as seen in **Figure 1.2**. This figure denotes the α-helices 1 and 2 as residues Lys^{1030}, Pro^{1033}, Gln^{1034}, Asp^{1064}, Tyr^{1066}, Lys^{1069}, and Lys^{1070}. LANA also contains a proline-rich region for attaching to the host genome [26], and this is seen as blue 'sticks' (Licorice VMD Drawing Method) in **Figure 1.2** as Pro^{1010} and Pro^{1012}. The DBD-specific residues are listed in **Table A.1** according to [25]. For successful $vDNA^{KSHV}$ replication and episome maintenance, LANA must cooperatively bind to two neighboring DBDs [2, 34]. LANA is unique however and binds to $vDNA^{KSHV}$ asymmetrically as a result of specific binding and narrowing of the $vDNA^{KSHV}$ minor groove [25]. Although LANA binds to DBDs, it is also not enzymatically active during the latent phase and requires the recruitment of other host proteins, similar to EBNA1 [25, 26]. All of the functions of OBPs are summarized in **Figure 1.3**.

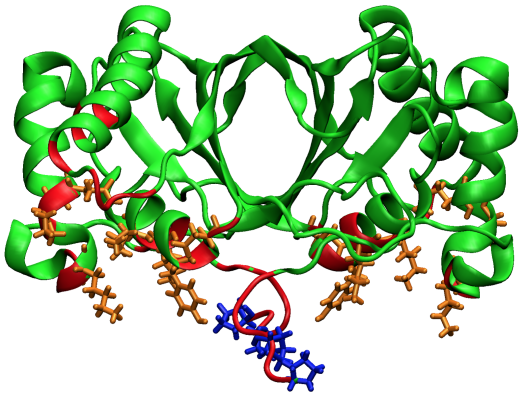

Figure 1.2: DNA Binding Domain on LANA in red, blue sticks are proline-rich regions, and orange sticks are α-helices.

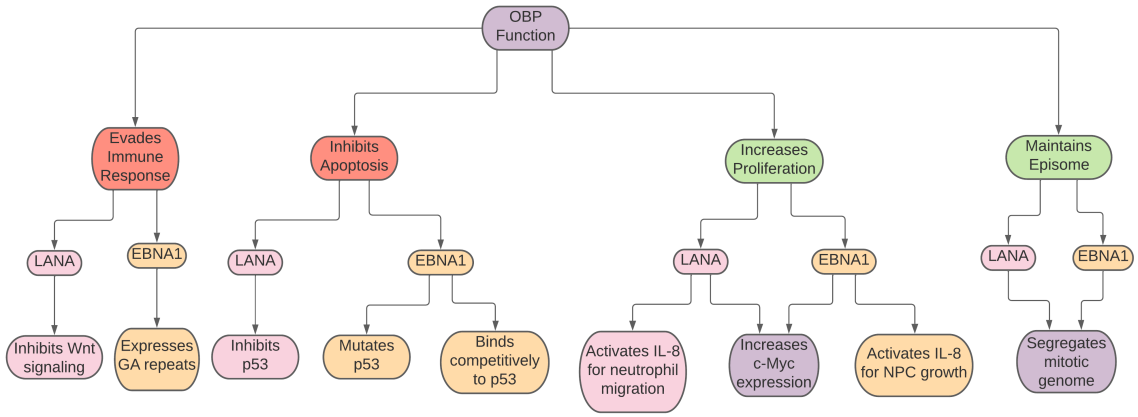

Figure 1.3: Summary of OBP functions. Green ovals represent an enhancement of normal functions whereas red ovals represent a reduction of normal functions. The unique actions of EBNA1 and LANA are shown in orange and pink, respectively and their shared actions are shown in purple.

1.4 Significance

EBNA1 has a role in all activity of EBV's latency phases, making it an appealing target for developing a treatment to combat EBV, as LANA is for KSHV. This study will compare the EBNA1 RBFE against the LANA RBFE obtained through all-atom molecular dynamics (MD) simulations. By forgoing high throughput screening of thousands of ligands [35], binding affinity studies such as these can streamline drug development [36]. With the calculation of the RBFE for EBNA1 and LANA, we hope this could be used for inhibitor development to stop OBP binding to vDNAs and provide a pre-emptive treatment for OBP-related cancers.

2. MATERIALS & METHODS

2.1 Materials

Molecular structures were obtained from the RCSB protein database (PDB) [37]. Specifically, the structures for EBNA1 and corresponding vDNA (PDB ID: 1B3T, 2.20 Å resolution) [24], and the structures for LANA and corresponding vDNA (PDB ID: 4UZB, 2.87 Å resolution) [25]. The structures of unbound vDNA were constructed using X3DNA: Dissecting the Spatial Structure of RNA (X3DNA) [38]. Simulation preparation was done using the Chemistry at Harvard Macromolecular Mechanics (CHARMM) package, version c43a2 [39]. All structures were visualized using the Visual Molecular Dynamics (VMD) software [40]. MD simulations were completed using the Texas A&M University High Performance Research Computing (TAMU HPRC) Terra cluster. Simulations were carried out using OpenMM (OMM), version 7.6.0 [41]. Analyses described below used programs developed in our lab using CHARMM, Python, and C++.

2.2 Molecular Dynamics Simulation Preparation

2.2.1 Structure Generation

The viral protein-DNA complex structures were obtained from the PDBs 1B3T (EBNA1) and 4UZB (LANA) and idealized vDNA structures for EBV and KSHV were constructed using X3DNA. Generating idealized vDNA structures minimized bias arising from deformation due to protein binding or differences in the crystallization technique used to obtain the high-resolution structures. The PDB files were converted into CHARMM protein structure files (PSF) and coordinate files (COR), a disulfide bond at residue 138 was added for EBNA1, and hydrogen atoms added to all structures. The initial structures are shown in **Figure 2.1**. After adding N- and C-termini to the ends of the proteins, the termini and side chains of the structures were energy minimized using the steepest descent (SD) CHARMM algorithm for 200 steps.

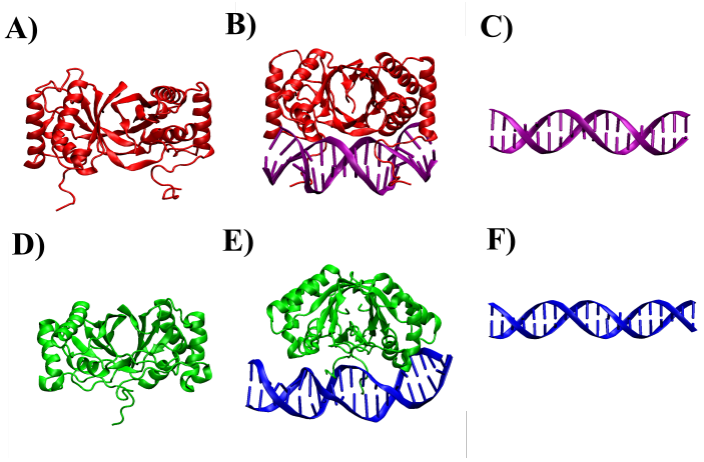

Figure 2.1: Crystal structure of viral protein, vDNA, and protein-vDNA complexes. (A) EBNA1. (B) EBNA1-vDNA^{EBV} complex. (C) vDNA^{EBV}. (D) LANA. (E) LANA-vDNA^{KSHV} complex. (F) vDNA^{KSHV}.

2.2.2 Solvation & Neutralization

In order to simulate physiological conditions, each system was solvated in a cubic TIP3 water box, which simulates the thermal fluctuations experienced by proteins. Each structure was placed in a cubic box that had faces at least 13 Å away from the protein, and the actual box dimensions are shown in **Table 2.1**. If the oxygen in a water molecule was within 2.8 Å of a structure's heavy atoms, that water molecule was removed. Sodium and chloride ions were then added to bring the net charge of the system to 0, with a total concentration of roughly 50 mM, as shown in **Table 2.1**. An example of a neutralized EBNA1 structure is shown in **Figure 2.2**.

Table 2.1: Box lengths and number of Sodium and Chloride ions added to each system.

System	EBNA1			LANA		
Structure	vDNA	Protein	Complex	vDNA	Protein	Complex
Box Lengths (Å)	89.71	83.07	83.99	96.31	86.04	96.99
Na^+ ions	60	5	41	69	12	59
Cl^- ions	24	15	15	29	29	25

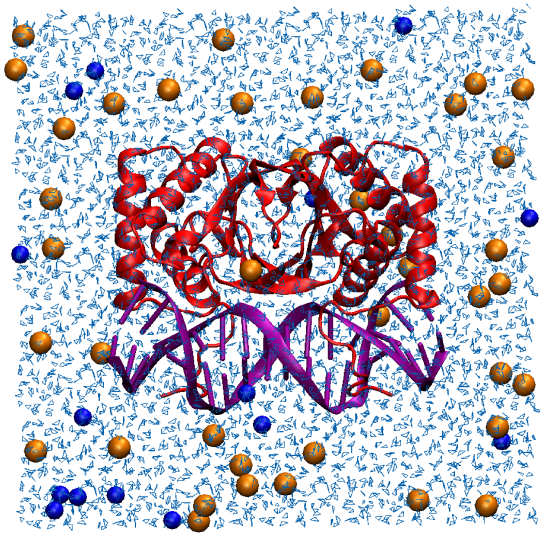

Figure 2.2: EBNA1 structure neutralized in water box. EBNA1 is shown in red and vDNA in purple, light blue lines are water molecules, and orange and blue spheres are Na^+ and Cl^- ions respectively.

2.2.3 Energy Minimization

The structures were then energy minimized according to Hwang et. al [42]. The backbone and sidechain atoms initially had harmonic constraints of 40 kcal/[mol·Å^2] and 20 kcal/[mol·Å^2], respectively. The constrained structure was then minimized using the Adopted Basis Newton-Raphson method for 200 steps. Then their constraints were reduced by 10 kcal/[mol·Å^2] for the backbone and 5 kcal/[mol·Å^2] for sidechain atoms respectively for three iterations followed by an iteration with no constraints. The structures were then heated from 30 K to 300 K with a harmonic constraint of 5 kcal/[mol·Å^2] for 100 picoseconds. The structures were allowed to reach equilibrium using the Leapfrog Verlet method of constant temperature and pressure at 300 K and 1 atm for 200 picoseconds [39].

2.2.4 Molecular Dynamics

The structures were further equilibrated with a light harmonic constraint of 0.001 kcal/[mol·Å^2] applied to the α-carbon atoms in the proteins and phosphorous atoms in the vDNA for 2 nanosec-

onds (ns). Initial simulations revealed a tendency for the DNA to denature. To combat this, harmonic distance restraints were placed on the canonical base-pairing atoms to maintain a maximum distance of 2.4 Å. Specifically, the distance restraints were applied to the oxygen, nitrogen, and hydrogen atoms involved in hydrogen-bonding on the vDNA terminal residues (H1, H21, H41, N3, O2, and O6) with a force of 1 kcal/[mol·Å^2]. The covalent bond lengths for hydrogen atoms were fixed using SHAKE, similar to [42]. All restraints were removed, except for the light harmonic constraints on the phosphorous atoms of the vDNA and the distance restraints of the terminal bases which were continued for the entirety of the simulation. The distance restraints were 'loosened' to 2.7 Å for the simulation as well. Positional coordinates were saved every 20 picoseconds resulting in 10,000 total frames for the entire 200 ns simulations.

2.3 Molecular Dynamics Analyses

2.3.1 Noncovalent interactions

The hydrogen bonds and nonpolar contacts between atoms in the protein-vDNA complexes were identified using CHARMM [39] and previous developments by Hwang et. al [43]. Nonpolar contacts were assigned a cutoff radius of 3 Å between neutral atoms with a partial charge less than $0.3e$ ($e = 1.6 \times 10^{-19}$ Coulombs). A cutoff occupancy of 40% was used to generate heatmaps of the contact trajectory throughout the simulation, as seen in **Figures A.3** and **A.4**. This occupancy was selected because both complexes had a major decrease in occupancy below roughly 40% and the increase in occupancy thereafter was relatively constant for all systems. In other words, the increase in occupancy for all bonds was gradual until roughly 30% occupancy, where the occupancy increased by a large factor followed by more gradual increases in bond occupancy.

2.3.2 Root-Mean-Square Deviation, Root-Mean-Square Fluctuation, and Solvent-Accessible Surface Area

All positional coordinate analyses involved the usage of the Coordinate Manipulation (CORMAN) module in CHARMM [39, 44]. The root-mean-square fluctuation (RMSF) was calculated to determine how mobile individual residues were throughout the simulations compared to a ref-

erence structure, in this case, the crystal structure. α-carbon atoms were analyzed for proteins and phosphorous atoms were analyzed for vDNA. All RMSF plots were generated using Python and are displayed in **Figure 3.2**. The root-mean-square deviation (RMSD) was then calculated to quantify the average fluctuation of the entire structure compared to the crystal structure over time. Protein residues with a RMSF above 1 Å and DNA residues with a RMSF above 2 Å were ignored for the RMSD calculation to avoid excessive noise from highly-mobile regions in the structure. All RMSD plots were similarly generated using Python and are shown in **Figure 3.5**. Lastly, the solvent-accessible surface area (SASA) was computed to find the area of the binding interface. For each complex, the calculation involved adding the surface area of the protein by itself with the surface area of the vDNA by itself, then subtracting that from the surface area of the complex. The equation used can be seen in **Eq. 1** where A_1 is the surface area of the protein, A_2 is the area of the DNA, and A_{12} is the shared area of A_1 and A_2 complex. The expression is halved because the numerator results in twice the complex area. By subtracting the combined area by the total areas, the resultant SASA is the change in area of the protein-vDNA complex binding interface. This calculation was repeated for each frame in the simulation and can be seen in **Figure 3.6**.

$$SASA = \frac{A_1 + A_2 - A_{12}}{2} \qquad \text{(Eq. 1)}$$

2.3.3 Conformational Entropy Calculations

The conformational entropy was found using the method from [31]. Measuring entropy relies on the information theory [45] and relating the random process, that is sidechain dynamics, to a probability distribution of how the structure is able to fluctuate in space [46,47]. The maximal information spanning tree (MIST) approximates the total entropy by finding correlations between conformational degrees of freedom (DOF). MIST also reduces the order of highly-dimensional systems in order to estimate the entropy of the entire system. The corrections are associated with correlated motion between the DOF and reduce the overall entropy [31]. The entropy was found by first calculating the dihedral angles of the backbone (BB) and sidechain (SC) atoms of both

the protein and vDNA structures. This calculation finds how much each structure rotates around a specific axis defined by three consecutive bonds and the four atoms involved. The protein BB dihedral angles included the phi (φ), psi (ψ), and omega (ω) angles, whereas the vDNA BB dihedral angles were the alpha (α), beta (β), gamma (γ), delta (δ), epsilon (ε), and zeta (ζ) angles as seen in **Figure 2.3** according to Hart et. al [48]. The SC dihedral angles for both types of structures depended on the specific residue or base. The protein SCs were defined depending on which of the 20 amino acids was present, whereas the vDNA SCs only depended on if the nucleic acid was a purine or pyrimidine. An example of the SC dihedral angles is shown in **Figure 2.4** for lysine, guanine, and cytosine. The conformational entropy results for each system are shown in **Table 3.2**.

To verify the accuracy of these values, the angles throughout the entire simulation were subsampled for five separate entropy calculations. Specifically, every fifth dihedral angle was separated resulting in five 40 ns trajectories instead of one 200 ns trajectories. From these subsampled trajectories, an average and standard deviation were obtained, as shown in **Table 3.3**. A bar graph displaying the conformational entropy and standard deviation is shown in **Figure 3.7**.

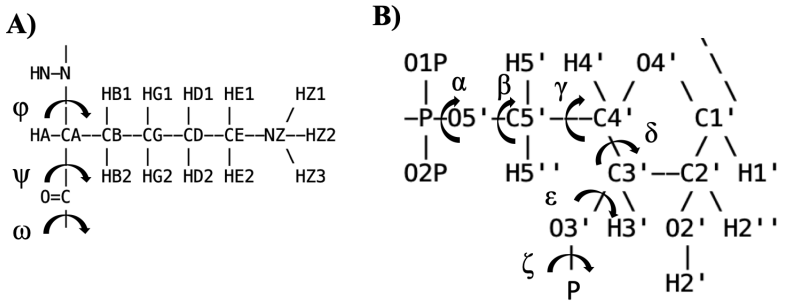

Figure 2.3: Backbone dihedral angles. A. Lysine. B. Any nucleic acid.

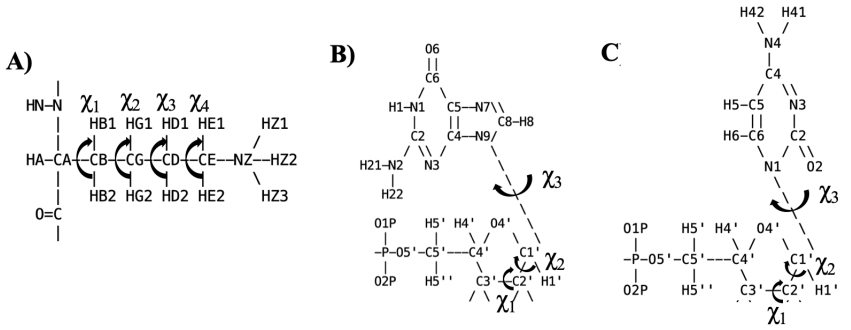

Figure 2.4: Sidechain dihedral angles. A. Lysine. B. Guanine. C. Cytosine

2.3.4 Relative Binding Free Energy Calculations

The energy of each structure as well as the solvation energy was calculated to determine the relative binding free energy. Negative RBFE values correspond to complexes favoring a bound state [49]. This was done according to the methods in [49] where the binding free energy is the sum of the energy contributions from the van der Waals and electrostatic energies (ΔG^0_{bind}), the difference between the solvation energy of the complex and the sum of the two components (ΔG_{desolv}), and the difference of the entropy (-TΔS). The equation of the differences in solvation energy can be seen in **Eq. 2**, while the binding free energy equation can be seen in **Eq. 3**. The solvation energy used the generalized Born model for molecular volumes (GBMV) [50] and generalized Born model with a simple switching function (GBSW) [51] using the GBMV and GBSW modules in CHARMM [39] as seen in **Tables 3.4** and **A.3**. The solvation energy terms for the complex, ΔG^{D}, and the monomers, ΔG^{M}, use the calculated GBMV or GBSW for the complex and protein and vDNA from the complex trajectory, respectively. Dissimilar to **Eq. 2**, the difference in entropy is calculated using the conformational entropy values from the complex, sole protein, and sole vDNA simulations.

$$\Delta G_{desolv} = \Delta G^{D}_{solv} - (\Delta G^{M1}_{solv} + \Delta G^{M2}_{solv}) \qquad \text{(Eq. 2)}$$

$$\Delta G_{bind} = \langle \Delta G^{0}_{bind} \rangle + \langle \Delta G_{solv} \rangle - \langle T\Delta S \rangle \qquad \text{(Eq. 3)}$$

3. RESULTS

3.1 Noncovalent interactions Analyses

The noncovalent interaction analysis concluded EBNA1 created a higher number of contacts to vDNA compared to LANA. Specifically, EBNA1 had 31 different hydrogen bonds and 32 nonpolar contacts while LANA had 14 hydrogen bonds and 17 nonpolar contacts. However, some of these bonds did not last for the entirety of the simulation, and are shown in **Figures A.3** and **A.4**. The occupancy is indicative that a bond was fully formed at a certain time. Both the hydrogen and nonpolar contacts are shown in **Figure 3.1** in contrasting-colored 'sticks' (Licorice VMD Drawing Method). Over the simulation, if a specific complex had a higher total number of bonds compared to another complex, then it can be assumed that complex has a higher RBFE.

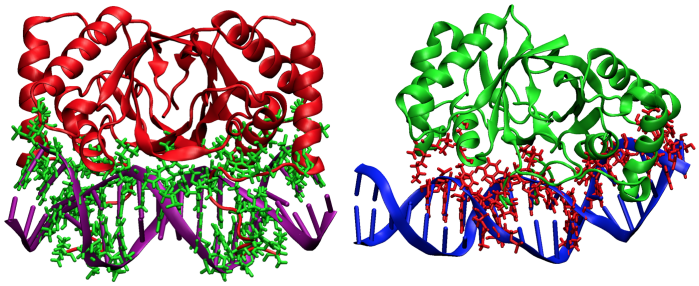

Figure 3.1: Contacts with 40% occupancy or higher. A) EBNA1 complex. B) LANA complex.

3.2 Root-Mean-Square Deviation, Root-Mean-Square Fluctuation, and Solvent-Accessible Surface Area

The RMSF calculation was performed first to determine which amino and nucleic acids were the most mobile relative to the reference structure. The combined RMSF of the proteins and vDNA from the two complexes can be seen in **Figure 3.2**. The fluctuations within each structure directly influence the following RMSD calculation. An example of this can be seen in **Figure 3.4** where the only restrained residues were the stable Thr^{575}:Thr^{576} (Residue 115 on the graph) amino acids to be able to perform the calculation.

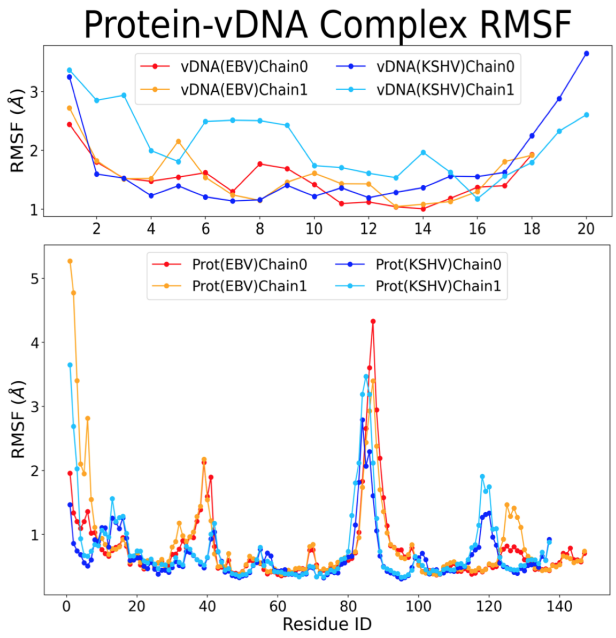

Figure 3.2: RMSF of parts of vDNA and protein in EBNA1 and LANA complexes. Residue IDs associated with EBNA1 residue ID 461:607 and LANA residue ID 1010:1146.

As shown in **Figure 3.2**, there are peaks at ranges of residues so those specific residues are excluded in the RMSD calculation. This can be attributed to whether that residue forms contacts with vDNA where interacting residues have a lower RMSF and free residues have a higher RMSF, as shown in **Figure 3.1**. This can be seen in **Figure 3.4** where the highly-mobile regions are included for RMSD calculations. These highly-mobile regions are visually shown in **Figure 3.3** as blue sticks. It is apparent from **Figure 3.3** that the blue regions would contribute to increasing the RMSD because they are on the outside of the tertiary structure and their position is thus more likely to be altered from thermal fluctuations. For structures like the vDNA$^{\text{KSHV}}$ Chain 1 where the baseline is greater than the other structures, only the ends were excluded to retain an average baseline for that structure. The residue selections are shown in **Table A.2** where vDNA residues specifically select the phosphorous atoms and protein residues select the α-carbon atoms in the backbones.

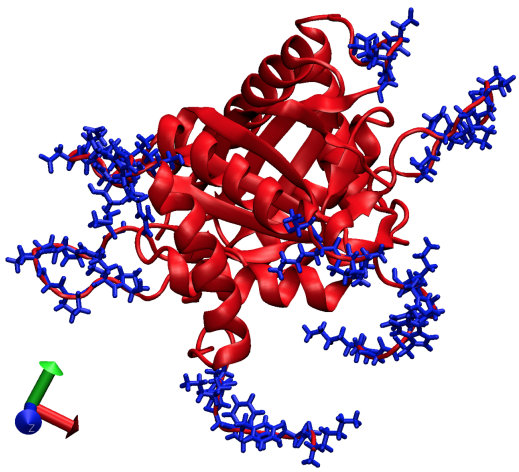

Figure 3.3: Excluded highly-mobile regions within EBNA1 as blue sticks.

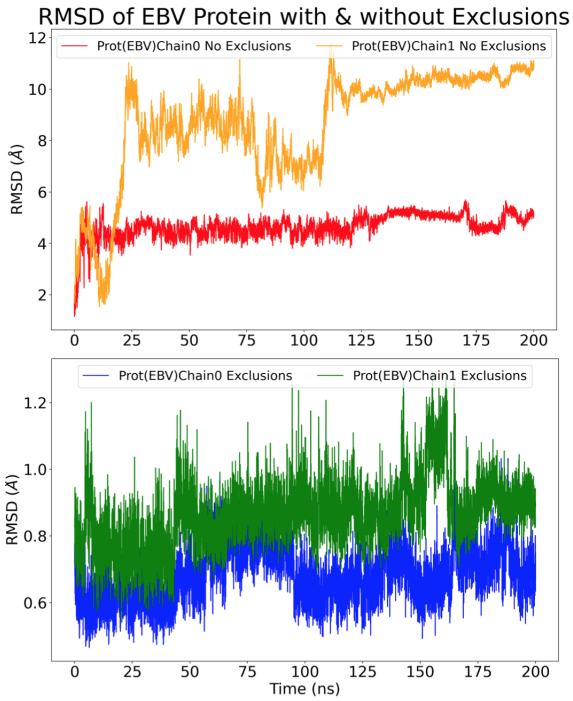

Figure 3.4: Comparison of EBNA1 RMSD with and without excluding highly-mobile residues.

Figure 3.4 pictured above shows the effects of not excluding the selected highly-mobile amino acids in EBNA1 as listed in **Table A.2** and as shown in **Figure 3.3**. The average RMSDs with exclusions equates to 0.68 $\pm$ 0.082 Å and 0.86 $\pm$ 0.10 Å, whereas the average RMSDs

without exclusions equates to 4.65 $\pm$ 0.47 Å and 8.66 $\pm$ 2.09 Å. The fluctuation in chain 1 without exclusions can be attributed to the tails that were originally free, but were ultimately bound to other amino acids on the same structure, as shown in **Figure A.2**. This shows that the exclusions of specific ranges of residues is worthwhile in finding the accurate deviation of the structure as a whole over time. The resultant RMSD plot is shown in **Figure 3.5** and shows the deviation of the simulated structure from the crystal structure position over time. Specifically, **Figure 3.5** is comparing the RMSD of the protein and vDNA chains from the EBNA1 and LANA complexes.

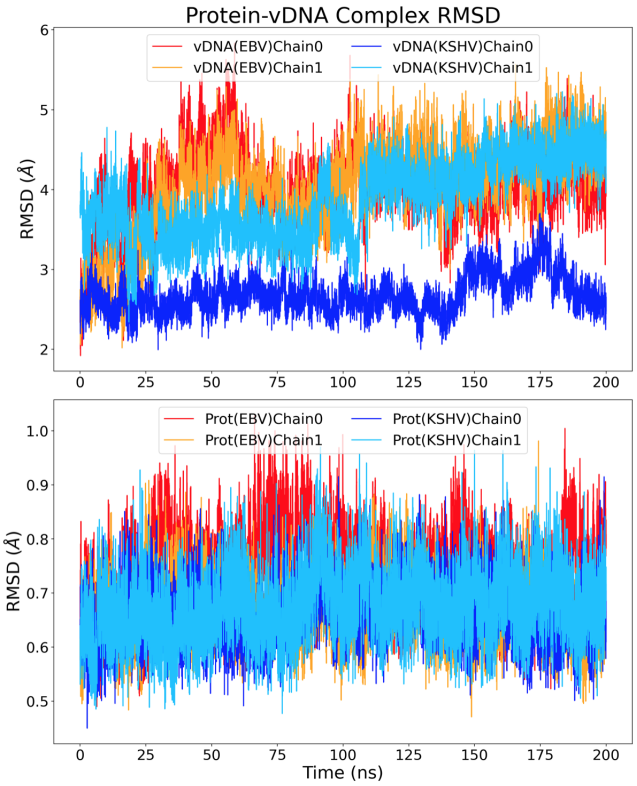

Figure 3.5: RMSD of whole EBNA1 and LANA structures in space over 200ns.

The average RMSD values and standard deviations for the vDNA and protein chains are shown in **Table 3.1**. On average, the DNA chains deviated 3.66 $\pm$ 0.43 Å whereas the protein chains deviated 0.68 $\pm$ 0.06 Å. This difference in movement can be attributed to the amount of binding interactions present. The binding of whether the DNA is strongly-held by the protein can determine if the DNA has the ability to deviate in position. Although there are more nonco-

valent bonds for EBNA1 than LANA, the average RMSD of vDNA^{EBV} was higher than that of vDNA^{KSHV}. Additionally, the average RMSD of chain 0 for vDNA^{KSHV} was roughly 1.5 Å less than that of the other vDNA chains and this can be attributed to how its backbone was relatively stable throughout the simulation.

Table 3.1: Average and standard deviation of RMSD from protein-vDNA complexes.

Molecule Chain	Average (Å)	Standard Deviation (Å)
vDNA^{EBV}0	4.044	0.474
vDNA^{EBV}1	4.040	0.567
vDNA^{KSHV}0	2.683	0.232
vDNA^{KSHV}1	3.876	0.484
Protein^{EBV}0	0.744	0.072
Protein^{EBV}1	0.668	0.058
Protein^{KSHV}0	0.663	0.055
Protein^{KSHV}1	0.676	0.067

Finally, the calculated SASA or buried surface area of both complexes can be seen in **Figure 3.6**. This plot shows the difference in surface area on each structure that is used for binding. A direct correlation can be made between an increased SASA and increased binding energy. Zoette et al. [49] describe SASA has a linear relation to nonpolar solvation free energy, and thus an increased SASA is proportional to an increased RBFE. The average and standard deviation SASA for the EBNA1 complex for the most stable last 100 ns is 2616.90 ± 107.73 Å^2, whereas the LANA SASA is 1734.74 ± 91.39 Å^2 for the last 100 ns. According to these values, it can be assumed that the EBNA1 complex has a stronger binding free energy compared to that of the LANA complex.

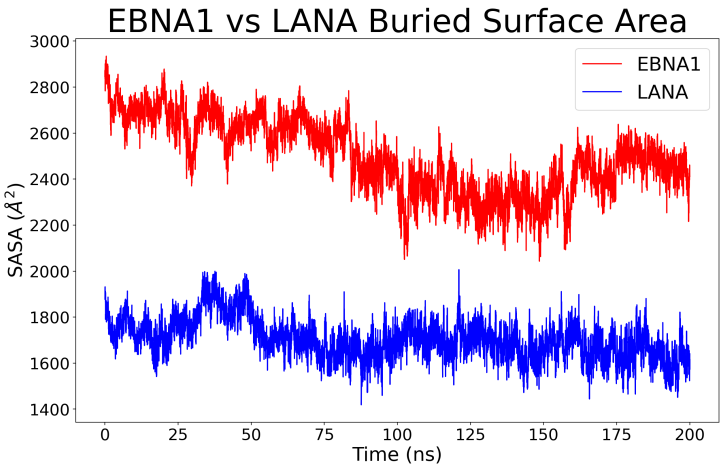

Figure 3.6: Buried surface area of EBNA1 and LANA protein-vDNA complexes over 200 ns.

3.3 Conformational Entropy

The calculated conformational entropy values can be seen in **Table 3.2** and show the contribution of the DOF the dihedral angles exhibited, or how much they rotated about their axes. The MIST2 column designates the calculated correction according to [52]. The proportional increase of values according to structure size is valid from a proportional increase in number of dihedral angles and conformational entropy.

Table 3.2: Entropy values across the systems.

System	EBNA1			LANA		
Structure	vDNA	Protein	Complex	vDNA	Protein	Complex
BB (kcal/mol)	528.587	1969.931	2468.639	602.882	1803.030	2382.108
SC (kcal/mol)	241.370	1360.428	1589.724	266.681	1333.274	1580.989
BB + SC (kcal/mol)	769.957	3330.360	4058.364	869.564	3136.305	3963.098
MIST2 (kcal/mol)	-54.938	-146.131	-196.567	-71.059	-128.840	-188.766
MIST1+2 (kcal/mol)	715.019	3184.229	3861.797	798.504	3007.464	3774.332

If there is a higher order of DOF, it can be assumed there will be more similarities between the movement of the structure as a whole, or an increased correlation. The subsampled average and standard deviation entropy values are shown in **Table 3.3**. Compared to **Table 3.2**, these values are relatively similar to the computed entropy over the entire simulation. This result shows there were

small dihedral angle fluctuations throughout the simulation. The averages and standard deviations from **Table 3.3** are shown as error bars on **Figure 3.7**, but appear as lines because of the small deviation.

Table 3.3: Sampled simulation average and standard deviation of MIST1+2 from all structures.

System	Molecule	Average (kcal/mol)	Standard Deviation (kcal/mol)
EBNA1	vDNA	708.025	0.132
	Protein	3157.546	0.337
	Complex	3831.652	0.660
LANA	vDNA	790.909	0.232
	Protein	2985.442	0.353
	Complex	3560.935	0.571

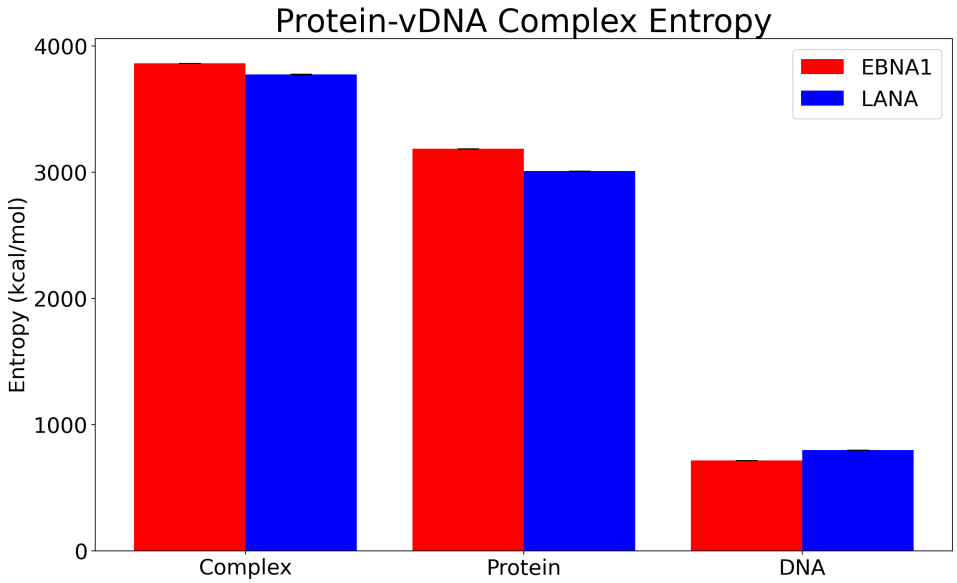

Figure 3.7: Conformational entropy of protein-vDNA complexes, sole proteins, and sole vDNAs.

Figure 3.7 accurately depicts the conformational changes that can occur from the studied structures because of the difference in structure size. For example, the EBNA1 protein has 147 amino acids compared to 137 for LANA, and this matches the results in **Table 3.2**. Similarly, the $vDNA^{KSHV}$ has 20 nucleic acids and an entropy of 798.504 kcal/mol, while $vDNA^{EBV}$ only has 18

nucleic acids and an entropy of 715.019 kcal/mol. With relation to binding energy, an increase in dihedral rotation around bonds is related to an increase in number of bonds. This can be seen in **Figure 3.1** since the EBNA1 complex creates more bonds between the protein and vDNA.

3.4 Relative Binding Free Energy Calculations

The final GBSW RBFE value is tabulated in **Table 3.4** where a more negative value for ΔG_{bind} can be associated with stronger binding free energy. As shown below, the EBNA1 protein-vDNA complex elicited a lower RBFE compared to LANA.

Table 3.4: GBSW RBFE values for EBNA1 and LANA complexes.

Energy	EBNA1 GBSW (kcal/mol)	LANA GBSW (kcal/mol)
ΔG^{0}_{bind}	-9928.58 $\pm$ 225.5	-5788.49 $\pm$ 136.0
ΔG^{M1}_{solv}	-6143.89 $\pm$ 122.1	-4175.28 $\pm$ 104.0
ΔG^{M2}_{solv}	-12278.78 $\pm$ 87.3	-14299.80 $\pm$ 138.5
ΔG^{D}_{solv}	-8756.11 $\pm$ 140.0	-12823.39 $\pm$ 156.0
-Tds	37.45 $\pm$ 0	31.64 $\pm$ 0
ΔG_{bind}	-224.56 $\pm$ 33.1	-105.16 $\pm$ 17.9

Overall, the increased number of bonds as seen in **Figures A.3** and **A.4**, the larger buried surface area as seen in **Figure 3.6**, and the increase in conformational entropy as seen in **Figure 3.7** are all validated with the comparison of EBNA1's RBFE relative to LANA's. Moreover, EBNA1 has a dissociation constant of 0.5 nM from fluorescence anisotropy assays [53] while LANA has a dissociation constant of 1.51 $\pm$ 0.16 nM from electrophoretic mobility shift assays [54]. These values correlate with the RBFE calculations and verify EBNA1 has stronger binding to vDNA. We hope the RBFE values of -224.56 $\pm$ 33.1 kcal/mol for EBNA1 and -105.16 $\pm$ 17.9 kcal/mol for LANA can be used to further inhibitor development for EBV and KSHV similar to [35]. In this fashion, potential inhibitors exhibiting similar RBFE values to those reported here via *in silico* analyses can be assumed to have a similar binding affinity to the OBPs. Furthermore, Gianti et. al [36] similarly uses the RBFE values of ligands to find potential "druggability pockets" in the EBNA1-vDNA structure.

3.5 Limitations and Sources of Error

As seen in **Figure A.1**, the vDNA had a tendency to break the hydrogen bonds during simulations as a result of the adenine-thymine (AT) instability [55]. This led to the attempted implementations of additional distance restraints on these base pairs, but were inevitably only utilized for the terminal base pairs. Additional restraints on AT bases could be ignored in order to keep the similarity to physiological conditions. Similarly, physiological conditions for the terminal base pairs of the vDNA were maintained by adding distance restraints to prevent DNA "fraying" [55–57]. Finally, small errors in input values for computing the RBFE produce large errors in resultant RBFE values [49], which leads to the need for simulations longer than 200ns. This is likely the reasoning for the positive GBMV RBFE values as seen in **Table A.3**.

4. CONCLUSION

The EBNA1 and LANA OBPs are responsible for prolonging the longevity of infected host cells and stimulating cancer development. If the binding of these OBPs to vDNA was stopped, the virus would not be able to replicate vDNA and inhibit the immune response and apoptosis. We found that the EBNA1 protein binds more strongly to $vDNA^{EBV}$ than LANA to $vDNA^{KSHV}$. This result is validated through the noncovalent interactions study, since EBNA1 formed more noncovalent bonds and had more high-occupancy contacts. The SASA calculation showed EBNA1 had a larger area for binding, and therefore a more robust binding interface. This was all confirmed with the RBFE calculation that gave a direct comparison of the increase in relative energy for EBNA1 DNA-binding compared to LANA. According to [58], EBNA1 attaches to vDNA with specific "linking domains" that are separate from the DBD. This contrasts how LANA uses the same C-terminal region as a DBD and chromosome tethering [58, 59]. With relation to RBFE, this can explain the lower RBFE of LANA because its binding regions are expended for tethering as well as sequence-specific vDNA-binding. On the other hand, EBNA1 can allocate one region for sequence-specific vDNA-binding and maintain high-affinity binding. Overall, we hope these calculations could be used for inhibitor development to disrupt the binding of these OBPs to vDNA. For future studies, we could perform a docking simulation to visualize the sequence-specific binding. Additionally, both EBNA1 and LANA are known to form oligomers, hexamer and decamer respectively [2]. With this in mind, a comparison of entropy and RBFE between the dimer and higher-order oligomers can be made to find the more stable oligomer for DNA-binding.

REFERENCES

[1] A. M. Edwards, A. Bochkarev, and L. Frappier, "Origin dna-binding proteins," *Current opinion in structural biology*, vol. 8, no. 1, pp. 49–53, 1998.

[2] A. De Leo, A. Calderon, and P. M. Lieberman, "Control of viral latency by episome maintenance proteins," *Trends in microbiology*, vol. 28, no. 2, pp. 150–162, 2020.

[3] B. Sugden and E. Leight, "Ebv's plasmid replicon: an enigma in cis and trans," *Epstein-Barr Virus and Human Cancer*, pp. 3–11, 2001.

[4] M. L. Mott and J. M. Berger, "Dna replication initiation: mechanisms and regulation in bacteria," *Nature Reviews Microbiology*, vol. 5, no. 5, pp. 343–354, 2007.

[5] Q. Rosemarie and B. Sugden, "Epstein–barr virus: How its lytic phase contributes to oncogenesis," *Microorganisms*, vol. 8, no. 11, p. 1824, 2020.

[6] T. Murata and T. Tsurumi, "Switching of ebv cycles between latent and lytic states," *Reviews in medical virology*, vol. 24, no. 3, pp. 142–153, 2014.

[7] R. Ayee, M. E. O. Ofori, E. Wright, and O. Quaye, "Epstein barr virus associated lymphomas and epithelia cancers in humans," *Journal of Cancer*, vol. 11, no. 7, p. 1737, 2020.

[8] K. Hoover and K. Higginbotham, "Epstein barr virus (ebv)," *StatPearls [Internet]*, 2020.

[9] V. Gares, L. Panico, R. Castagne, C. Delpierre, and M. Kelly-Irving, "The role of the early social environment on epstein barr virus infection: a prospective observational design using the millennium cohort study," *Epidemiology & Infection*, vol. 145, no. 16, pp. 3405–3412, 2017.

[10] J. P. Womack and M. Jimenez, "Common questions about infectious mononucleosis," *American family physician*, vol. 91, no. 6, pp. 372–376, 2015.

[11] J. Slots, I. Saygun, M. Sabeti, and A. Kubar, "Epstein–barr virus in oral diseases," *Journal of periodontal research*, vol. 41, no. 4, pp. 235–244, 2006.

[12] S. Gottschalk, C. M. Rooney, and H. E. Heslop, "Post-transplant lymphoproliferative disorders," *Annu. Rev. Med.*, vol. 56, pp. 29–44, 2005.

[13] S. E. Straus, J. I. Cohen, G. Tosato, and J. Meier, "Epstein-barr virus infections: biology, pathogenesis, and management," *Annals of internal medicine*, vol. 118, no. 1, pp. 45–58, 1993.

[14] S. H. Speck and D. Ganem, "Viral latency and its regulation: lessons from the γ-herpesviruses," *Cell host & microbe*, vol. 8, no. 1, pp. 100–115, 2010.

[15] C. Shannon-Lowe, A. B. Rickinson, and A. I. Bell, "Epstein–barr virus-associated lymphomas," *Philosophical Transactions of the Royal Society B: Biological Sciences*, vol. 372, no. 1732, p. 20160271, 2017.

[16] J. M. Phang, W. Liu, C. N. Hancock, and J. W. Fischer, "Proline metabolism and cancer: emerging links to glutamine and collagen," *Current opinion in clinical nutrition and metabolic care*, vol. 18, no. 1, p. 71, 2015.

[17] C. V. Dang, L. M. Resar, E. Emison, S. Kim, Q. Li, J. E. Prescott, D. Wonsey, and K. Zeller, "Function of the c-myc oncogenic transcription factor," *Experimental cell research*, vol. 253, no. 1, pp. 63–77, 1999.

[18] L. Frappier, "Contributions of epstein–barr nuclear antigen 1 (ebna1) to cell immortalization and survival," *Viruses*, vol. 4, no. 9, pp. 1537–1547, 2012.

[19] S. Apcher, C. Daskalogianni, B. Manoury, and R. Fåhraeus, "Epstein barr virus-encoded ebna1 interference with mhc class i antigen presentation reveals a close correlation between mrna translation initiation and antigen presentation," *PLoS pathogens*, vol. 6, no. 10, p. e1001151, 2010.

[20] J. B. Wilson, E. Manet, H. Gruffat, P. Busson, M. Blondel, and R. Fahraeus, "Ebna1: oncogenic activity, immune evasion and biochemical functions provide targets for novel therapeutic strategies against epstein-barr virus-associated cancers," *Cancers*, vol. 10, no. 4, p. 109, 2018.

[21] M. Wieczorek, E. T. Abualrous, J. Sticht, M. Álvaro Benito, S. Stolzenberg, F. Noé, and C. Freund, "Major histocompatibility complex (mhc) class i and mhc class ii proteins: conformational plasticity in antigen presentation," *Frontiers in immunology*, vol. 8, p. 292, 2017.

[22] A. Bochkarev, J. A. Barwell, R. A. Pfuetzner, E. Bochkareva, L. Frappier, and A. M. Edwards, "Crystal structure of the dna-binding domain of the epstein–barr virus origin-binding protein, ebna1, bound to dna," *Cell*, vol. 84, no. 5, pp. 791–800, 1996.

[23] K. A. Malecka, J. Dheekollu, J. S. Deakyne, A. Wiedmer, U. D. Ramirez, P. M. Lieberman, and T. E. Messick, "Structural basis for cooperative binding of ebna1 to the epstein-barr virus dyad symmetry minimal origin of replication," *Journal of virology*, vol. 93, no. 20, pp. e00487–19, 2019.

[24] A. Bochkarev, E. Bochkareva, L. Frappier, and A. M. Edwards, "The 2.2 Å structure of a permanganate-sensitive dna site bound by the epstein-barr virus origin binding protein, ebna1," *Journal of molecular biology*, vol. 284, no. 5, pp. 1273–1278, 1998.

[25] J. Hellert, M. Weidner-Glunde, J. Krausze, H. Lünsdorf, C. Ritter, T. F. Schulz, and T. Lührs, "The 3d structure of kaposi sarcoma herpesvirus lana c-terminal domain bound to dna," *Proceedings of the National Academy of Sciences*, vol. 112, no. 21, pp. 6694–6699, 2015.

[26] T. Uppal, S. Banerjee, Z. Sun, S. C. Verma, and E. S. Robertson, "Kshv lana—the master regulator of kshv latency," *Viruses*, vol. 6, no. 12, pp. 4961–4998, 2014.

[27] E. A. Mesri, E. Cesarman, and C. Boshoff, "Kaposi's sarcoma and its associated herpesvirus," *Nature Reviews Cancer*, vol. 10, no. 10, pp. 707–719, 2010.

[28] S. Li, L. Bai, J. Dong, R. Sun, and K. Lan, "Kaposi's sarcoma-associated herpesvirus: epidemiology and molecular biology," *Infectious Agents Associated Cancers: Epidemiology and Molecular Biology*, pp. 91–127, 2017.

[29] L. Frappier, "The epstein-barr virus ebna1 protein," *Scientifica*, vol. 2012, 2012.

[30] P. Purushothaman, T. Uppal, R. Sarkar, and S. C. Verma, "Kshv-mediated angiogenesis in tumor progression," *Viruses*, vol. 8, no. 7, p. 198, 2016.

[31] J. Qin and C. Lu, "Infection of kshv and interaction with hiv: the bad romance," *Infectious Agents Associated Cancers: Epidemiology and Molecular Biology*, pp. 237–251, 2017.

[32] A. De Leo, H.-S. Chen, C.-C. A. Hu, and P. M. Lieberman, "Deregulation of kshv latency conformation by er-stress and caspase-dependent rad21-cleavage," *PLoS Pathogens*, vol. 13, no. 8, p. e1006596, 2017.

[33] X. Li, D. Liang, X. Lin, E. S. Robertson, and K. Lan, "Kaposi's sarcoma-associated herpesvirus-encoded latency-associated nuclear antigen reduces interleukin-8 expression in endothelial cells and impairs neutrophil chemotaxis by degrading nuclear p65," *Journal of virology*, vol. 85, no. 17, pp. 8606–8615, 2011.

[34] J. F. Domsic, H.-S. Chen, F. Lu, R. Marmorstein, and P. M. Lieberman, "Molecular basis for oligomeric-dna binding and episome maintenance by kshv lana," *PLoS pathogens*, vol. 9, no. 10, p. e1003672, 2013.

[35] I. Ancy, M. Sivanandam, and P. Kumaradhas, "Possibility of hiv-1 protease inhibitors-clinical trial drugs as repurposed drugs for sars-cov-2 main protease: a molecular docking, molecular dynamics and binding free energy simulation study," *Journal of Biomolecular Structure and Dynamics*, vol. 39, no. 15, pp. 5368–5375, 2021.

[36] E. Gianti, T. E. Messick, P. M. Lieberman, and R. J. Zauhar, "Computational analysis of ebna1 "druggability" suggests novel insights for epstein-barr virus inhibitor design," *Journal of computer-aided molecular design*, vol. 30, no. 4, pp. 285–303, 2016.

[37] H. M. Berman, J. Westbrook, Z. Feng, G. Gilliland, T. N. Bhat, H. Weissig, I. N. Shindyalov, and P. E. Bourne, "The protein data bank," *Nucleic acids research*, vol. 28, no. 1, pp. 235–242, 2000.

[38] X.-J. Lu and W. K. Olson, "3dna: a versatile, integrated software system for the analysis, rebuilding and visualization of three-dimensional nucleic-acid structures," *Nature protocols*, vol. 3, no. 7, pp. 1213–1227, 2008.

[39] B. R. Brooks, C. L. Brooks III, A. D. Mackerell Jr, L. Nilsson, R. J. Petrella, B. Roux, Y. Won, G. Archontis, C. Bartels, and S. Boresch, "Charmm: the biomolecular simulation program," *Journal of computational chemistry*, vol. 30, no. 10, pp. 1545–1614, 2009.

[40] W. Humphrey, A. Dalke, and K. Schulten, "Vmd: visual molecular dynamics," *Journal of molecular graphics*, vol. 14, no. 1, pp. 33–38, 1996.

[41] P. Eastman, J. Swails, J. D. Chodera, R. T. McGibbon, Y. Zhao, K. A. Beauchamp, L.-P. Wang, A. C. Simmonett, M. P. Harrigan, and C. D. Stern, "Openmm 7: Rapid development of high performance algorithms for molecular dynamics," *PLoS computational biology*, vol. 13, no. 7, p. e1005659, 2017.

[42] W. Hwang, R. J. Mallis, M. J. Lang, and E. L. Reinherz, "The αβtcr mechanosensor exploits dynamic ectodomain allostery to optimize its ligand recognition site," *Proceedings of the National Academy of Sciences*, vol. 117, no. 35, pp. 21336–21345, 2020.

[43] W. Hwang, M. J. Lang, and M. Karplus, "Kinesin motility is driven by subdomain dynamics," *Elife*, vol. 6, p. e28948, 2017.

[44] A. D. MacKerell Jr, B. Brooks, C. L. Brooks III, L. Nilsson, B. Roux, Y. Won, and M. Karplus, "Charmm: the energy function and its parameterization," *Encyclopedia of computational chemistry*, vol. 1, 2002.

[45] R. M. Gray, *Entropy and information theory*. Springer Science & Business Media, 2011.

[46] B. M. King and B. Tidor, "Mist: Maximum information spanning trees for dimension reduction of biological data sets," *Bioinformatics*, vol. 25, no. 9, pp. 1165–1172, 2009.

[47] B. M. King, N. W. Silver, and B. Tidor, "Efficient calculation of molecular configurational entropies using an information theoretic approximation," *The Journal of Physical Chemistry B*, vol. 116, no. 9, pp. 2891–2904, 2012.

[48] K. Hart, N. Foloppe, C. M. Baker, E. J. Denning, L. Nilsson, and A. D. MacKerell Jr, "Optimization of the charmm additive force field for dna: Improved treatment of the bi/bii conformational equilibrium," *Journal of chemical theory and computation*, vol. 8, no. 1, pp. 348–362, 2012.

[49] V. Zoete, M. Meuwly, and M. Karplus, "Study of the insulin dimerization: binding free energy calculations and per-residue free energy decomposition," *Proteins: Structure, Function, and Bioinformatics*, vol. 61, no. 1, pp. 79–93, 2005.

[50] M. S. Lee, M. Feig, F. R. Salsbury Jr, and C. L. Brooks III, "New analytic approximation to the standard molecular volume definition and its application to generalized born calculations," *Journal of computational chemistry*, vol. 24, no. 11, pp. 1348–1356, 2003.

[51] W. Im, M. S. Lee, and C. L. Brooks III, "Generalized born model with a simple smoothing function," *Journal of computational chemistry*, vol. 24, no. 14, pp. 1691–1702, 2003.

[52] J. Shi, Q. Shen, J.-H. Cho, and W. Hwang, "Entropy hotspots for the binding of intrinsically disordered ligands to a receptor domain," *Biophysical journal*, vol. 118, no. 10, pp. 2502–2512, 2020.

[53] C. Oddo, E. Freire, L. Frappier, and G. de Prat-Gay, "Mechanism of dna recognition at a viral replication origin," *Journal of Biological Chemistry*, vol. 281, no. 37, pp. 26893–26903, 2006.

[54] A. C. Garber, J. Hu, and R. Renne, "Latency-associated nuclear antigen (lana) cooperatively binds to two sites within the terminal repeat, and both sites contribute to the ability of lana to suppress transcription and to facilitate dna replication," *Journal of Biological Chemistry*, vol. 277, no. 30, pp. 27401–27411, 2002.

[55] M. Zgarbová, M. Otyepka, J. Sponer, F. Lankas, and P. Jurecka, "Base pair fraying in molecular dynamics simulations of dna and rna," *Journal of chemical theory and computation*, vol. 10, no. 8, pp. 3177–3189, 2014.

[56] F. Landuzzi, P. L. Palla, and F. Cleri, "Stability of radiation-damaged dna after multiple strand breaks," *Physical Chemistry Chemical Physics*, vol. 19, no. 22, pp. 14641–14651, 2017.

[57] R. Galindo-Murillo, J. C. Robertson, M. Zgarbová, J. Sponer, M. Otyepka, P. Jurecka, and T. E. Cheatham III, "Assessing the current state of amber force field modifications for dna," *Journal of chemical theory and computation*, vol. 12, no. 8, pp. 4114–4127, 2016.

[58] J. Hellert, M. Weidner-Glunde, J. Krausze, U. Richter, H. Adler, R. Fedorov, M. Pietrek, J. Rückert, C. Ritter, and T. F. Schulz, "A structural basis for brd2/4-mediated host chromatin interaction and oligomer assembly of kaposi sarcoma-associated herpesvirus and murine gammaherpesvirus lana proteins," *PLoS pathogens*, vol. 9, no. 10, p. e1003640, 2013.

[59] T. L. Hodin, T. Najrana, and J. L. Yates, "Efficient replication of epstein-barr virus-derived plasmids requires tethering by ebna1 to host chromosomes," *Journal of virology*, vol. 87, no. 23, pp. 13020–13028, 2013.

APPENDIX

Table A.1: LANA DBD residues.

Chain A	Chain B	Both Chains
1016	1085	1011:1015
1019		1028
1033:1034		1030
1064		1066
1119		1069:1070
1122		1128
1126		
1128:1130		

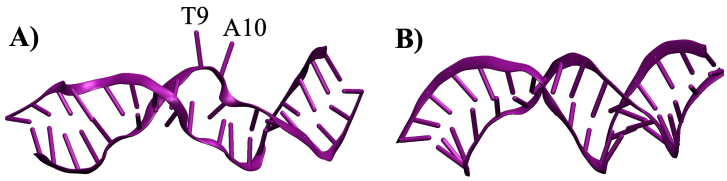

Figure A.1: Change in EBNA1 vDNA conformation as a result of simulations. A) Frame at roughly 20ns. B) Frame at 140ns.

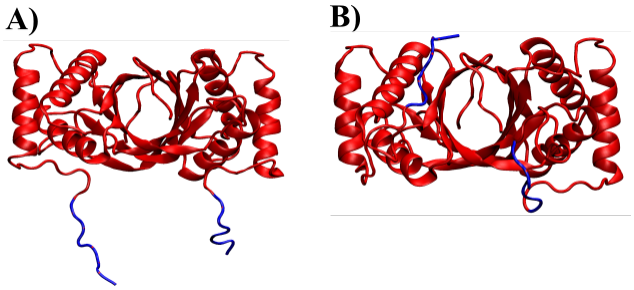

Figure A.2: Change in EBNA1 conformation from simulations. A) Frame at 0ns. B) Frame at 200ns.

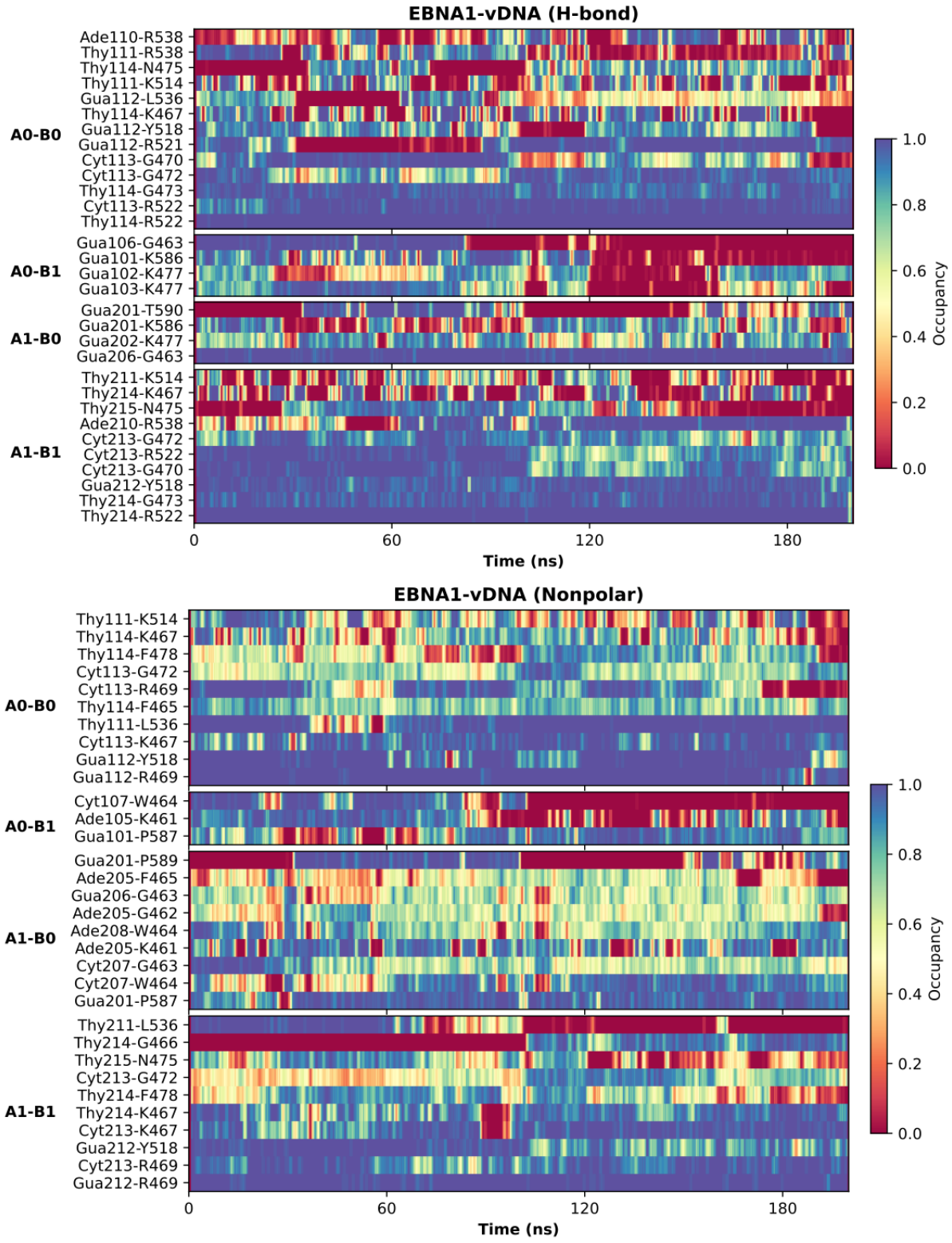

Figure A.3: Occupancy heatmap of EBNA1 bond retention.

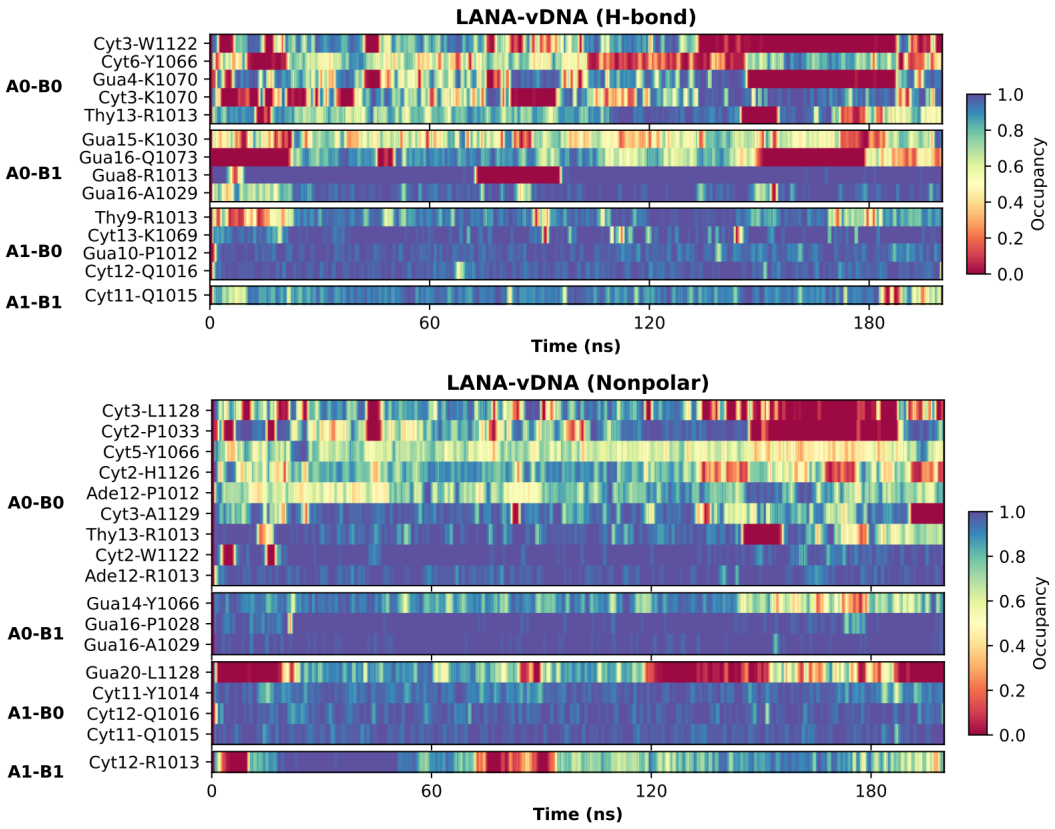

Figure A.4: Occupancy heatmap of LANA bond retention.

Table A.2: Selection of highly-mobile residues to exclude for RMSD.

Structure	Residue Selection
vDNA^{EBV}0	101:102
vDNA^{EBV}1	201:202 and 205:206
vDNA^{KSHV}0	1:2 and 18:20
vDNA^{KSHV}1	1:3, 6:9, and 19:20
Protein^{EBV}0	461:468, 497:501, and 543:551
Protein^{EBV}1	461:468, 492:501, 544:551, and 585:589
Protein^{KSHV}0	1010:1011, 1019:1025, 1091:1097, and 1127:1129
Protein^{KSHV}1	1010:1012, 1019:1026, 1090:1097, and 1126:1131

Table A.3: GBMV RBFE values for EBNA1 and LANA complexes.

Energy	EBNA 1 GBMV (kcal/mol)	LANA GBMV (kcal/mol)
ΔG^{0}_{bind}	-9928.58 $\pm$ 225.5	-5788.49 $\pm$ 136.0
ΔG^{M1}_{solv}	-5727.51 $\pm$ 115.1	-3704.74 $\pm$ 105.3
ΔG^{M2}_{solv}	-12370.78 $\pm$ 86.0	-14394.09 $\pm$ 135.1
ΔG^{D}_{solv}	-8169.30 $\pm$ 152.9	-12220.37 $\pm$ 153.6
-Tds	37.45 $\pm$ 0	31.64 $\pm$ 0
ΔG_{bind}	37.86 $\pm$ 32.1	121.61 $\pm$ 26.0