



Structural and Biophysical Investigation of the Key Hotspots on the Surface of Epstein–Barr Nuclear Antigen 1 Essential for DNA Recognition and Pathogenesis

Huma Farooque Hashmi^{1*}, Muhammad Waseem², Syed Shujait Ali³, Zahid Hussain³ and Kaoshan Chen¹

¹College of Life Sciences, Shandong University, Jinan, China, ²Faculty of Rehabilitation and Allied Health Science, Riphah International University, Islamabad, Pakistan, ³Center for Biotechnology and Microbiology, University of Swat, Swat, Pakistan

OPEN ACCESS

Edited by:

Qing-Chuan Zheng,
Jilin University, China

Reviewed by:

Xuanyu Meng,
Soochow University, China
Lin Chen,
Qiqihar University, China

*Correspondence:

Huma Farooque Hashmi
humahashmi@bs.qau.edu.pk

Specialty section:

This article was submitted to
Biological Modeling and Simulation,
a section of the journal
Frontiers in Molecular Biosciences

Received: 05 February 2021

Accepted: 14 June 2021

Published: 29 June 2021

Citation:

Hashmi HF, Waseem M, Ali SS,
Hussain Z and Chen K (2021)
Structural and Biophysical
Investigation of the Key Hotspots on
the Surface of Epstein–Barr Nuclear
Antigen 1 Essential for DNA
Recognition and Pathogenesis.
Front. Mol. Biosci. 8:664436.
doi: 10.3389/fmolb.2021.664436

Epstein-Barr Virus (EBV) is considered the most important human pathogen due to its role in infections and cellular malignancies. It has been reported that this Oncolytic virus infects 90% world's population. EBNA1 is required for DNA binding and survival of the virus and is considered an essential drug target. The biochemical and structural properties of this protein are known, but it is still unclear which residues impart a critical role in the recognition of dsDNA. Intending to disclose only the essential residues in recognition of dsDNA, this study used a computational pipeline to generate an alanine mutant of each interacting residue and determine the impact on the binding. Our analysis revealed that R469A, K514A, Y518A, R521A and R522A are the key hotspots for the recognition of dsDNA by the EBNA1. The dynamics properties, i.e. stability, flexibility, structural compactness, hydrogen bonding frequency, binding affinity, are altered by disrupting the protein-DNA contacts, thereby decreases the binding affinity. In particular, the two arginine substitution, R521A and R522A, significantly affected the total binding energy. Thus, we hypothesize that these residues impart a critical role in the dsDNA recognition and pathogenesis. This study would help to design structure-based drugs against the EBV infections.

Keywords: EBNA1, DNA recognition, MD simulation, in silico mutagenesis, free energy calculations

INTRODUCTION

Due to the prevalent infections instigated by Herpes viruses, it is considered as an important virus in human pathogens flora. This diverse pathogenic flora Epstein-Barr Virus (EBV) is regarded as the most important human pathogen due to its role in infections and cellular malignancies (Bochkarev et al., 1996; Yasuda et al., 2011) It has been reported that this Oncolytic virus infects 90% world's population. Immortalization of B lymphocytes accompanies the main EBV infection and stimulates them to replicate as lymphoblastic cell lines (Garai-Ibabe et al., 2012). Alongside the B lymphocytes infection, EBV also causes infectious mononucleosis by targeting the epithelial cells (Pope et al., 1968; Haque and Crawford, 1996). Nasopharyngeal carcinoma (NFC), muscle cell sarcoma and gastric carcinomas (GaCa), Hodgkin's lymphoma (HL), Burkitt's lymphoma (BL), extranodal lymphoma of T/NK cell origin and post-transplant lymphoproliferative disease (PT-LPD) are among the EBV associated diseases (Taylor et al.,

2015). EBV caused tumors stores the viral genome as a multi-copy episome in the nucleus of infected cells (Tikhmyanova et al., 2014). During the latent infection, progenitor virions are not reproduced, but alternatively, a set of genes essential for survival and proliferation are expressed (Matsuura et al., 2010). Epstein-Barr Nuclear Antigen 1 (EBNA1) acts to preserve the latent viral genome in proliferating cells (Okano, 1998). This protein is expressed in the malignant cell and sustain the proliferation (Leight and Sugden, 2000).

The EBV genome encodes ~100 genes, among which EBNA1 is the key nuclear antigen that works with the other five others (Hiraku et al., 2014). EBNA1 is almost detected in every kind of infection induced by the EBV in both latent and lytic infections (Gianti et al., 2016). This essential antigen is reported to be involved in mitotic segregation of episomes, replication, reactivation, viral transcription, and lytic infection of EBV (Young and Rickinson, 2004). It has also been reported that EBNA1 possess a similar structure to that of human papillomavirus (HPV) E2 protein and the Kaposi's Sarcoma Associated Herpesvirus (KSHV) LANA protein (Garber et al., 2002). In addition to structural similarity, these proteins are reported to have similar function, i.e. DNA binding and episome regulation (Wang et al., 2006). The biochemical and structural properties of this protein are known. EBNA1 works as a dimer with two functional domains (Pope et al., 1968; Haque and Crawford, 1996). The two terminals CTD (carboxy-terminal DNA-binding domain) and ATCTD (amino-terminal chromosome tethering domain) bind the 18ps DNA to initiate the plasmid and viral genome replication (Lindner and Sugden, 2007). Due to the multi-faceted role of EBNA1, it is the primary drug target for the treatment of EBV associated infections.

The crystallographic structure of ENBA1 has been solved, and reported that 459–607 residues at C-terminal are required for DNA binding (Bochkarev et al., 1998). Previous studies determining the dissociation constant (KD) for the EBNA1-DNA association reported that mutating interacting residues, R469A reduced the binding of DNA by 300-fold, Y518A by 80-fold and R522A by 1600-fold (Cruickshank et al., 2000). Additionally, other studies also reported that K514A also reduces the binding affinity significantly. At the same time, others reported that three residues R491E, R491A, and D581E significantly impair the DNA binding (Morgunova et al., 2015). These studies mutated only selected residues, while the impact of others remains a question. To understand each residue's impact and reveal only a few residues that are required explicitly for DNA recognition while others are supplementary interactions, an in-depth investigation is needed. To disclose only the essential residues, this study used a computational pipeline to generate an alanine mutant of each interacting residue and determine the impact on the binding. Highly destabilizing and affinity reducing mutations were subjected to biophysical investigation to reveal their real effect on the binding. Our

analysis would help to target the critical hotspots for future rational and structure-based drug designing to curtail the EBV associated lytic and latent infections.

MATERIALS AND METHODS

Structure Retrieval and Preparation

The RCSB protein databank (<http://www.rcsb.org/>) repository was accessed for structural retrieval. Structural deformities were detected and addressed (Rose et al., 2016). The missing hydrogens were added, and partial charges were assigned. The structure was also analyzed for structural breaks and unknown residues. The structure was minimized and prepared before *in silico* mutagenesis and molecular docking.

Epstein-Barr Nuclear Antigen 1-DNA Docking

For the docking HADDOCK (High Ambiguity Driven protein-protein Docking) (Dominguez et al., 2003) was utilized. It uses biochemical and structural data to drive the docking process. The Guru interface with approximately 500 features considered as the best to predict the docking poses. Using default parameter i.e. lowest intermolecular energies, the best structural complex was extracted. We also used NPDock (Tuszynska et al., 2015), an online web server, which uses the scoring of poses, clustering of the best-scored models, and refinement of the most promising solutions to give the best results. The best scoring complex was retrieved from NPDock and analyzed. To determine the interaction of different residues with the DNA DNAproDB (Sagendorf et al., 2020) was used to extract the interactions from DNA-nucleic acid complexes.

Interface Analysis and Mutants Library Construction

Using the machine learning protocol implemented in MOE (Vilar et al., 2008) the Alanine scanning approach was applied to compute the impact of each residue in the interaction with the DNA. The dAffinity and dStability parameters are essential considerations in the ASM (alanine scanning module) which determine the relative stability and affinity changes upon substitution. The detailed mechanism of this alanine scanning mutagenesis and residue scan approaches has been discussed previously (Junaid et al., 2019). Furthermore, we also used mCSM-NA an online webserver, for the affinity changes prediction upon the alanine substitution uses the graph-based signature model. Residues with high-affinity changes were subjected to molecular dynamics simulation investigation.

Molecular Dynamics Simulation

To further provide deep insight into the stability and affinity changes upon the alanine substitution, the dynamic features of

each complex was determined using the AMBER 20 simulation package. For protein ff14SB, while for DNA, the OLS3 force field was utilized (Salomon-Ferrer et al., 2013). With the TIP3P water model containing 9,784 water molecules, each complex was solvated at 10.0 Å. A total of 29 sodium ions were added to neutralize each system. Multistep energy minimization each 6,000 steps and 3,000 steps of conjugate gradient minimization were completed. Keeping the heating parameters default 300 K for 200 ps, each complex was heated. For density equilibration, using weak restraint for 2 ns at constant pressure was executed. Finally, 200 ns MD using constant pressure was achieved. Langevin thermostat with 1 atm pressure and 300 K for temperature control (Zwanzig, 1973), while Particle Mesh Ewald (PME) algorithm to evaluate long-range interactions respectively (Ryckaert et al., 1977; Roe and Cheatham, 2013) with the cutoff, distances 10 Å. For the covalent bonds involving hydrogen, the SHAKE algorithm was used (Ryckaert et al., 1977). All the simulations were GPU accelerated.

Post-Simulation Analyses

The thermodynamics state function, i.e. RMSD, residual flexibility, i.e. RMSF, structural compactness, i.e. radius of gyration (Rg) and the total number of hydrogen bonds over the simulation were computed by using CPPTRAJ and PTRAJ modules integrated with AMBER (Roe and Cheatham, 2013).

Binding Affinity Calculations

To connect the alanine mutations with the binding affinity changes, the binding free energy of each alanine substituted complex was determined. The free energy scoring function (MMGBSA) is an extensively used approach to evaluate the free energy of a protein-ligand, protein-protein and protein-nucleic acids (Khan et al., 2018; Ali et al., 2019; Khan et al., 2019; Khan et al., 2020a; Khan et al., 2020b; Khan et al., 2020c; Khan et al., 2020d; Khan et al., 2020e; Khan et al., 2020f; Hussain et al., 2020). It used the following equation to calculate the free energy.

$$\Delta G_{bind} = \Delta G_{complex} - [\Delta G_{receptor} + \Delta G_{ligand}]$$

Each term in the such as electrostatic, van der Waals interactions, polar and nonpolar were predicted using the following equation:

$$G = G_{bond} + G_{ele} + G_{vdW} + G_{pol} + G_{npol}$$

Clustering of MD trajectories using PCA and Free Energy Landscape (FEL).

An unsupervised learning approach known as Principal Component Analysis to describe the motion of MD trajectories (PCA) (Pearson, 1901; Wold et al., 1987) and gain information about the internal motion of the system using CPPTRAJ. For the eigenvector and their atomic coordinates, the spatial covariance matrix was calculated. A diagonal matrix of eigenvalues was generated using the orthogonal coordinate transformation. The Principal Components were derived based on the eigenvectors and eigenvalues. The predominant movements during the simulation were plotted using these PCs. (Balsera et al., 1996; Ernst et al., 2015).

Furthermore, Free energy landscape (FEL) was constructed to capture the different energy minima at different simulation time.

RESULTS AND DISCUSSION

Structure Retrieval and Epstein-Barr Nuclear Antigen 1-DNA Docking

Using accession number 5T7X the structure of the EBNA1 was retrieved. The structure is a dimer interface of two EBNA1 chains and 18bps DNA. The structural representation of the EBNA1-DNA complex is shown in **Figure 1A**. HADDOCK predicted the correct docking conformation with the binding energy -295.63 kcal/mol. The interactions predicted by the DNAproDB showed that G462, G463, W464, F465, R469, N475, K477, F478, R491, K514, Y518, R521 and R522 are involved in interaction with the DNA. The 3D interaction of these residues with the DNA is given in **Figure 1B**. A different number of hydrogen bonds were formed by each residues ranging from one to five at the interface. The specific hydrogen bonding interactions are shown in **Figure 1C**. These residues contributed to the total binding energy. To potentially determine the impact of each of this residue, alanine scanning revealed its impact on the binding of DNA. Among the 13 residues at the interface G462A and G463A increase the binding affinity while the remaining 11 residues decrease the binding affinity at different folds.

As tabulated in **Table 1**, it can be seen that R469A, K514A, Y518A, R521A and R522A significantly affected the binding of DNA as compared to others. In the case of the R469A, the predicted ΔG was reported to -5.784 kcal/mol, while the dAffinity was also predicted to be reduced (-1,009.21 kcal/mol). The predicted ΔG for K514A was reported to be -3.638 kcal/mol, while the dAffinity was reported to be -1,021.32 kcal/mol, respectively. For the Y518A the predicted ΔG was -3.406 kcal/mol; however, the dAffinity was reported to be -1,020.32 kcal/mol. Intriguingly the dAffinity for R521A and R522A was comparable. The predicted dAffinity for R521A and R522A was reported to be -1,009.60 kcal/mol and -1,009.37 kcal/mol, respectively. Furthermore, the predicted ΔG for R521A was -5.866 kcal/mol, while for R522A, it was -6.008 kcal/mol. In the EBNA-DNA co-crystal structure, the three targeted amino acids are oriented toward the DNA but are too far from the nearest H-bond acceptor in the bases (more than 6 Å) to form H-bonds. Hence, these results also show that R469A, K514A, Y518A, R521A and R522A are required for DNA recognition and are the key hotspots for drug discovery. Thus, these residues were selected for further evaluation and subjected to molecular dynamics simulation to understand its dynamics behaviour and reveal its binding energy differences.

Mutation Stability Correlation (Root Mean Square Fluctuation)

To demonstrate the mutation's stability correlation, the thermodynamics state function Root mean square fluctuation (RMSD) of the wild type and the mutant complexes was calculated as a function of time. A 200 ns simulation trajectory

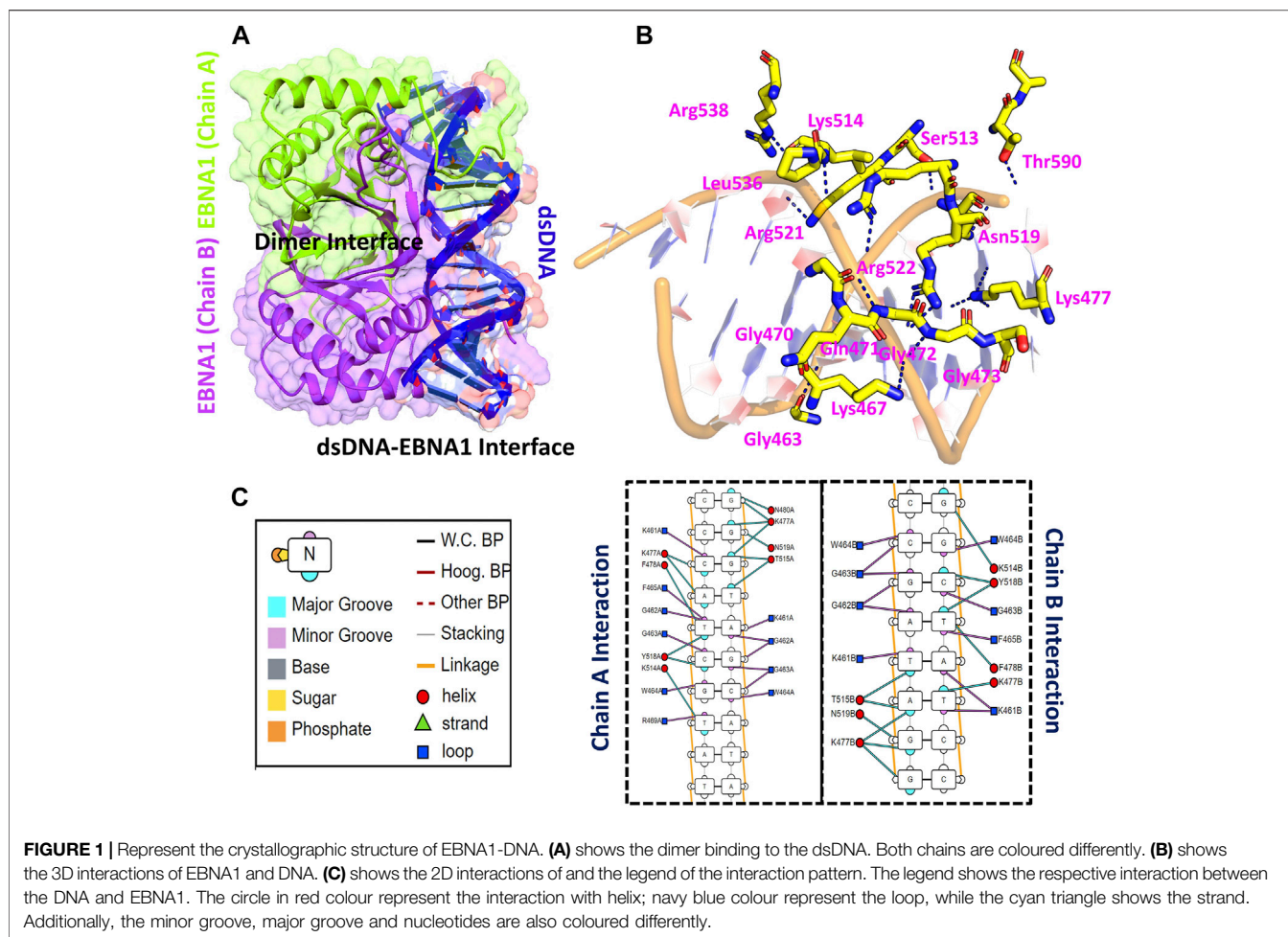


TABLE 1 | The table shows the alanine scanning results of the interacting residues. dStability, dAffinity, predicted ΔG and the outcome of each mutation upon substitution is given. Highly affinity reducing mutations are given as bold and were subjected to molecular dynamics simulation-based investigation. All the energies are given in kcal/mol.

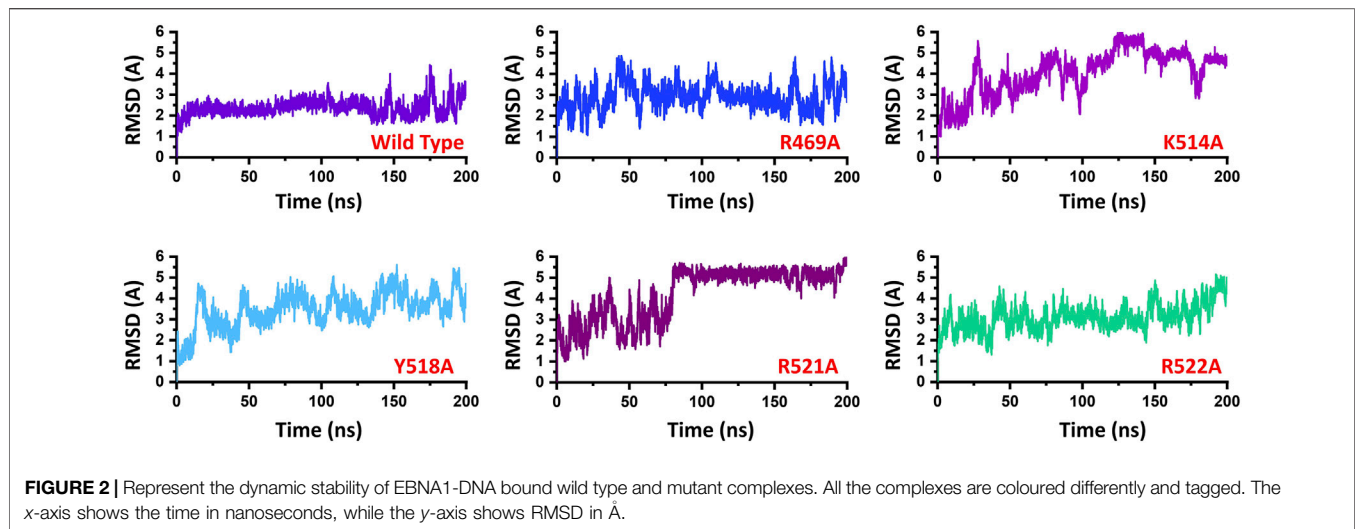
Index	Mutant residue	dStability ^a	dAffinity ^b	Predicted ΔG^c	Outcome
1	G462A	-161.94	-1,297.62	1.482	Increased affinity
2	G463A	-152.55	-1,021.36	1.488	Increased affinity
3	W464A	-153.84	-1,022.16	-1.656	Reduced affinity
4	F465A	-158.65	-1,022.34	-2.878	Reduced affinity
5	R469A	-154.65	-1,009.21	-5.784	Reduced affinity
6	N475A	-163.71	-1,016.41	-0.592	Reduced affinity
7	K477A	-160.67	-1,020.26	-0.824	Reduced affinity
8	F478A	-161.56	-1,022.69	-2.148	Reduced affinity
9	R491A	-225.26	-1,012.93	-1.986	Reduced affinity
10	K514A	-165.16	-1,021.32	-3.638	Reduced affinity
11	Y518A	-156.59	-1,020.47	-3.406	Reduced affinity
12	R521A	-158.75	-1,009.60	-5.866	Reduced affinity
13	R522A	-156.41	-1,009.37	-6.008	Reduced affinity

^adStability = it is the relative stability change upon the mutation. The more the negative the more instable the structure.

^bdAffinity = it is the relative affinity change between the wild type and mutated complex. Negative dAffinity means the mutation will increase the binding affinity while positive dAffinity mean it will decrease the binding affinity.

^cPredicted ΔG = it implies a similar formula but different algorithm to calculate the binding differences between the wild type and mutant. ΔG shows change in the binding free energy changes upon the mutation.

Bold values are Mutations selected for MD simulation.



for each complex was analyzed. Results for all the complexes are presented in **Figure 2**. In the case of the wild type, the structure gained stability at 2.0 Å. The structure remained remarkably stable during the simulation. After reaching 150 ns the structure converged, and the RMSD increased which is due to the loop opening and closing surrounded the DNA. It was observed that a loop region between 540–560 deviated from its mean structure significantly and thus the RMSD fluctuated substantially. Furthermore, the terminal of the DNA molecule packs the protein by opening and closing also causes significant structural deviation from its mean position thus causes structural instability. This can be inferred from **Figure 3A** where the loop region in all the complexes fluctuated significantly and **Figures 3B,C** shows the closer look into the loop region which is significantly deviated at different intervals. In the case of the R469A mutation, the complex experiences significant divergence from the initial structure. The equilibrium was never achieved during the 200 ns simulation time. During the first 50 ns simulation the structure owned significant convergence to the following 100 ns. Between 50 and 150 ns the RMSD remained lower and experienced only one significant convergence at 100 ns. Afterwards, the structure remained unstable until the 200 ns. The average RMSD for R469A was reported to be 3.5 Å. The K514A mutant, which is considered as an important residue for the DNA binding, caused significant perturbation upon the substitution. The complex remained significantly unstable during the 200 ns simulation time. The average RMSD for the first 25 ns remained 3.0 Å. Until the first 25 ns the RMSD remained 2.0 Å; however, a significant convergence was observed abruptly, and this trend continues until 200 ns. On the other hand, the Y518A behaviour was also comparable with the K514A. Significant convergence at different intervals was reported over the simulation and the average RMSD remained 4.0 Å. Furthermore, the two arginine replacements at position R521 and R522 significantly altered the dynamics and interaction of the EBNA1-DNA. These replacements caused significant destabilization of the EBNA1-DNA complex at different interval of the 200 ns. These residues are also reported experimentally to cause significant instability of

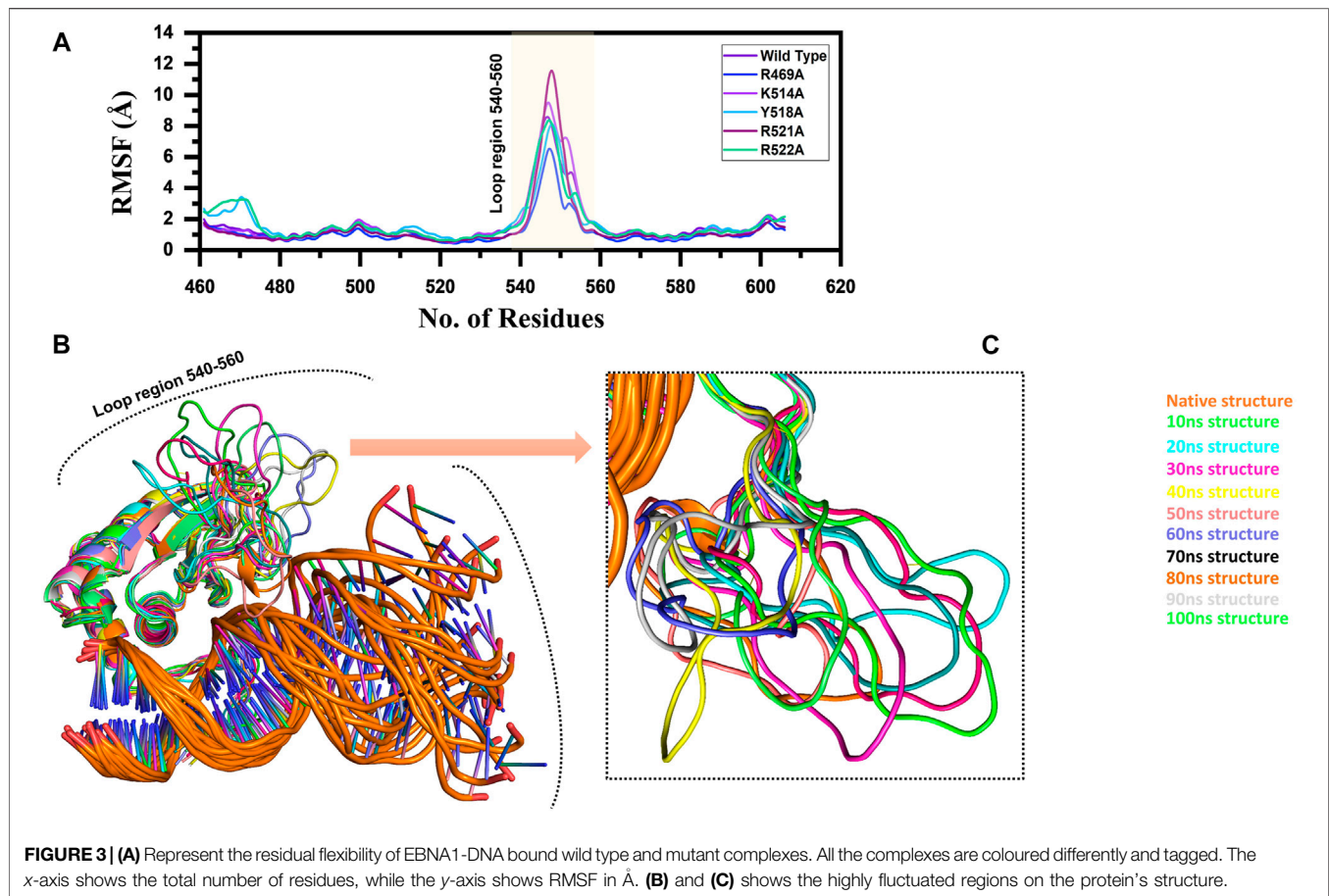
the complex. The average RMSD remained higher for R521A (5.5 Å), while the average RMSD for R522A remained lower but converged significantly. The RMSD continues to increase during the last 50 ns. Thus these results suggest that the wild type structural topology is required for stable interactions, and the mutation-induced here does not only affect the binding of the complex but also the stability. Hence further study on the impact of the substitution justified the effect of these residues on the binding of dsDNA and its druggability properties.

Residual Flexibility of the Wild Type and Mutant Complexes

Furthermore, to connect the residual flexibility with these substitutions, we estimated RMSF (root mean square fluctuation). The wild and mutant complexes owned comparable flexibility levels. It can be seen that all the structures possess a more similar pattern of flexibility. The residues 460–480, particularly in Y518A and R522A possess more flexible behaviour than the others, which is explained in **Figure 8** that it deviated more than the mean point and the mutations cause an allosteric effect on the flexibility. Significant residual flexibility for region 530–560 can be observed. These results show that complexes possess more rigid structures. The RMSF results for all the complexes are represented in **Figure 3A**. However the loop region which causes structural perturbation and flexibility is shown in **Figures 3B,C**.

Hydrogen Bonding Analysis of the Wild Type and Mutant Complexes

Furthermore, to understand the impact of these substitutions on the total number of hydrogen bonds, we calculated the total number of hydrogen bonds during the 200 ns trajectories and the bonding network between the EBNA1 and DNA. Hydrogen bonding rearrangement was observed during the simulation. Among the key bonding in the wild type R469 residue formed extra two interactions with T11 and A28. Among the others, R522

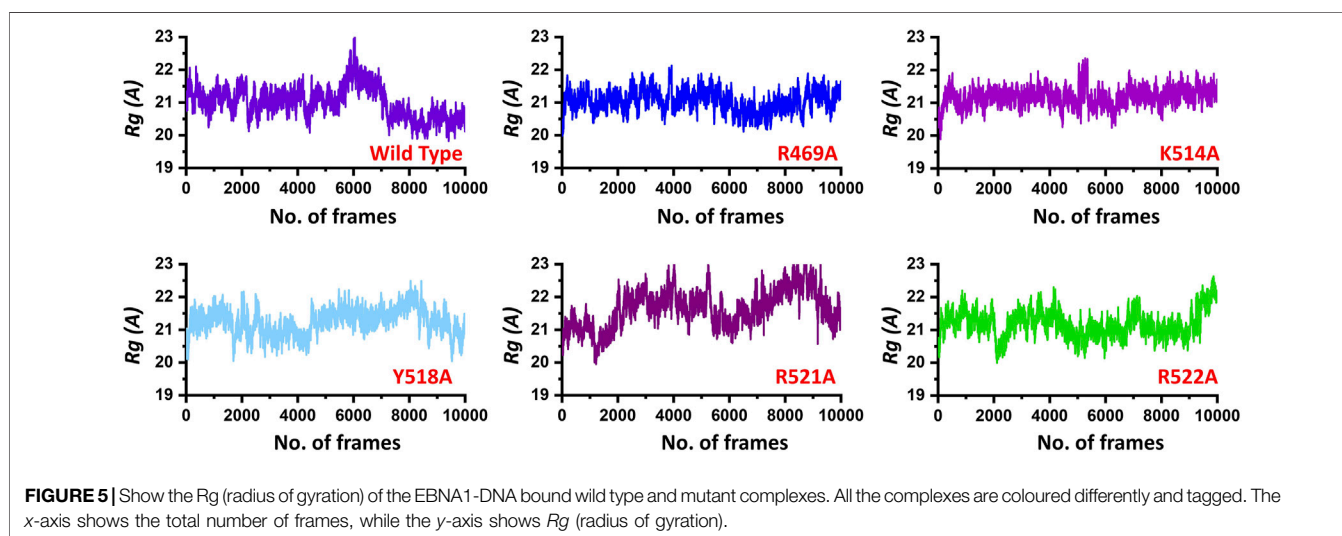
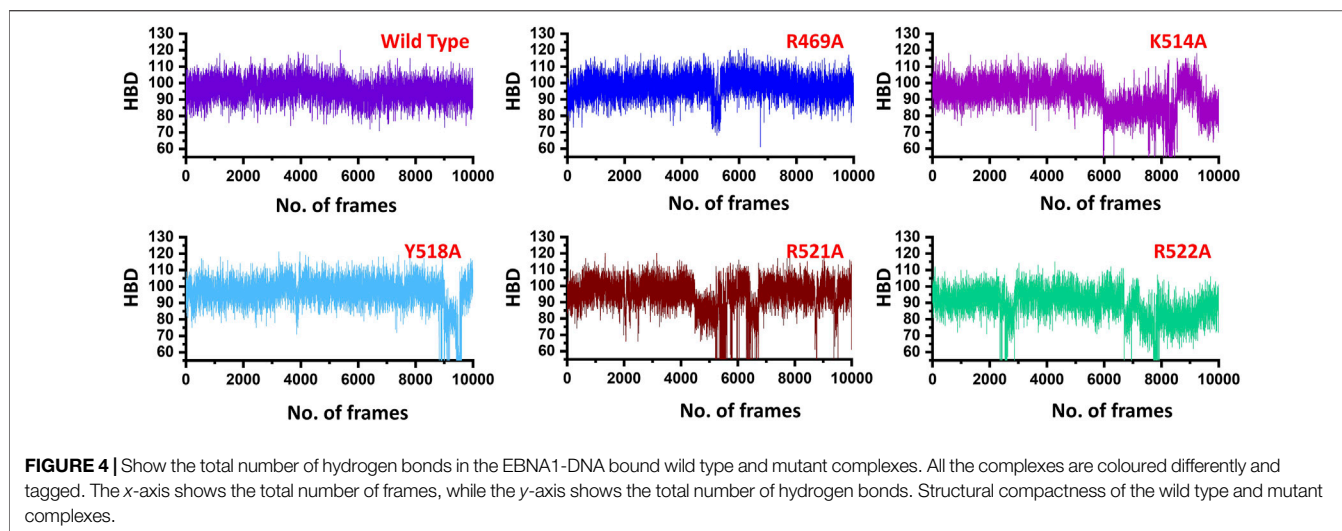


formed one additional hydrogen bonds with 1.83 Å. Initially, a total of 15 bonds were observed, while after simulation with these three extra interactions formed and the total bonds were reported to be 18 in total in wild type. On the other hand, the R469A lost multiple interactions during the simulation, particularly those formed with R469 residue, consequently remaining 13 hydrogen bonds between EBNA1 and DNA. Among these, Lys461, Arg521 and Arg522 created multiple interactions while the other residues were involved in single interaction only. This shows that in the wild type complex, R469 formed three interactions while those are lost here signifies its role in recognition.

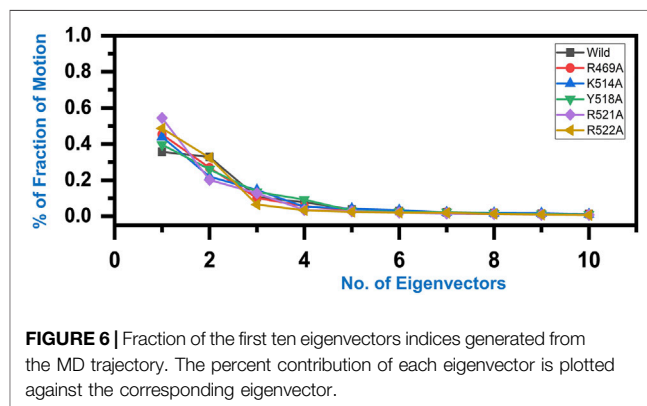
Similarly, only 11 hydrogen bonds were observed in K514A complex. Among the hydrogen bonding interaction, K514 lost its interaction while R521 and R522 also lost three interactions during the simulation. This shows that the mutation has allosterically affected the other residues and destabilized the interaction with the DNA molecule. Moreover, with 12 hydrogen bonds between EBNA1 (Y518A) and DNA complex R469, R521 and R522 lost their multiple interactions which were reported to be sustained in wild type complex. Furthermore, the two essential residues R521A and R522A which significantly contributed to the total binding energy reported in substantial hydrogen bonds reduction between EBNA1 and DNA. In the case of R521A only 10 hydrogen bonds were reported, while only nine bonds were reported between R522A and DNA. In R522A five

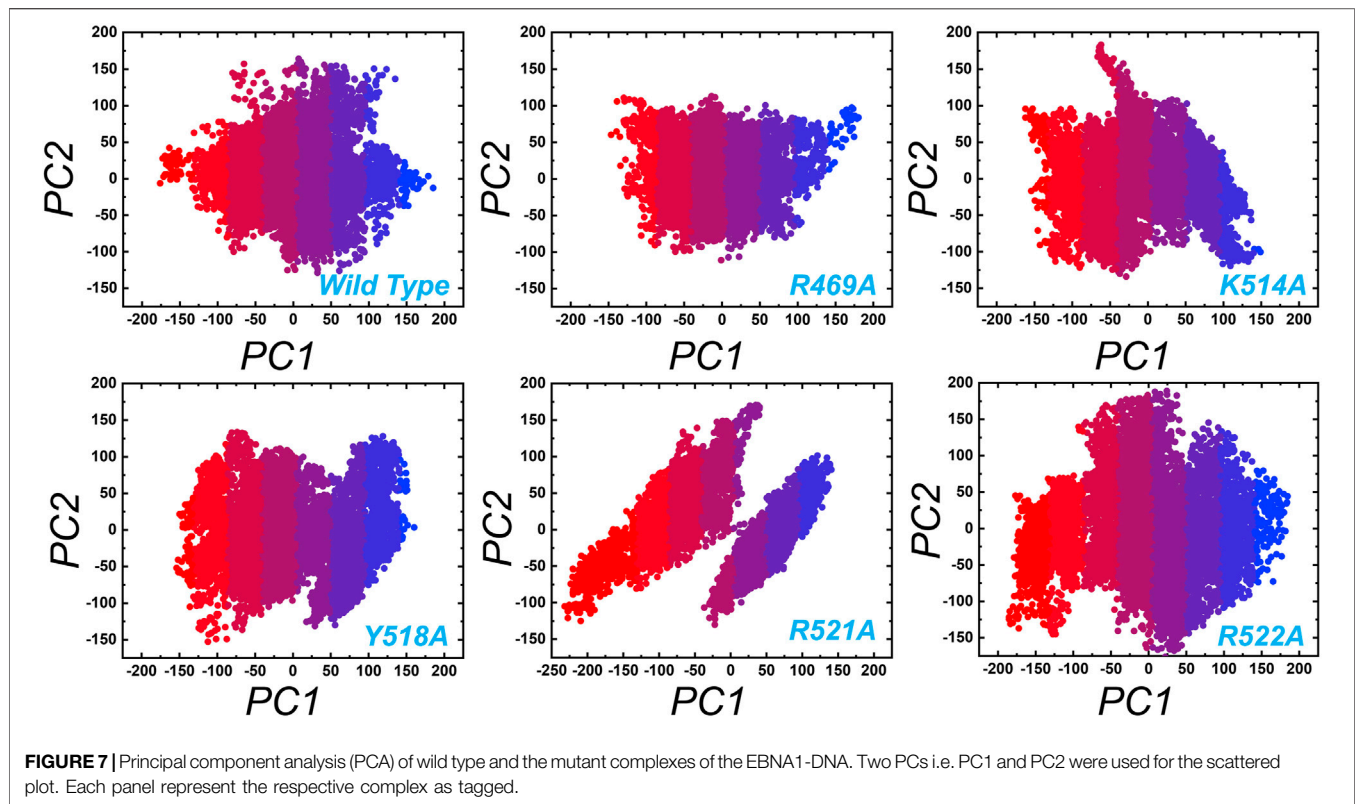
bonds formed by R521 and R522 are lost. This shows that the two arginine moieties play a significant role in recognition of DNA. After all, the average number of hydrogen bonds were calculated for each complex. As given in **Figure 4**, a significant drop can be observed in the mutant complexes, particularly in the K514A, R521A and R522A complexes. The average number of hydrogen bonds in the wild type was reported to be 98, while in the mutant complexes (R469A), the H-bonds were reported to be 94, (K514A) H-bonds were observed to be 90, Y518A reported 94, while the significant drop was observed in R521A (86) and R522A the total H-bonds were 88. Thus it can be seen that upon substitution, H-bonds count was decreased and thus, these residues potentially act as druggable hotspots. The hydrogen bonding results for all the complexes are represented in **Figure 4**.

Next, to connect the protein conformation changes with the compactness of each complex, we calculated Rg (radius of gyration) as a function of total frames in a trajectory. In the case of the wild, the complex remained more compact than the others. The average Rg for the wild type was reported to be 21.0 Å. In the case of R469A, the same pattern was observed. The results of R469A and wild type was comparable. On the other hand, the K514A the Rg remained higher during the simulation time period. During the first 100 ns. the Rg was observed to be higher, which significantly increased between 100 and 125 ns. Afterwards, the Rg remained uniform. In the case of Y518A, the



structural compactness also remained haphazard. Initially, it remained higher but then decreased between 40 and 100 ns while then increased and decreased continuously until the 200 ns. In the case of R521A and R522A, the structural compactness is disrupted significantly. This shows that the loss of structural compactness is due to the binding and unbinding events that occurred during the simulation, and this can be clearly concluded from **Figure 4** as the total number of hydrogen bonds are vary in numbers. Besides the packing of EBNA1 by the DNA terminal and the opening and closing of the loop region 530–560 also demonstrates the compactness variations. The calculated R_g (radius of gyration) results for all the complexes are represented in **Figure 5**.





Principal Motions of the Wild Type and Decoy Designed Peptides

Variations in the proteins' trajectories motions were exhibited by each system was captured through PCA. PCA would help to comprehend conformational changes induced variations in the proteins' motion of the wild type and mutant complexes. The internal motion was shown by the first three eigenvectors, while localized fluctuations in the remaining eigenvectors in each complex were observed (Figure 6). In the case of wild type peptide complex, the first three eigenvectors contributed 35% variance to the total observed motion, while in R49A, 45%, K514 43%, Y518A 39%, R521A 54% while R522A accounted for 48% variance in motion. This shows the increased motion in the mutated systems and may explain the structural rearrangement due to the mutations in the binding site.

In order to obtain plausibly attributed movements, the first two eigenvectors were projected against one another. The continuous representation of the red to blue colour indicates the transition from one conformation to another over the simulation period. The dots, starting with red and ending in blue, represent each frame. In each complex periodic jumps and continuous overlapping can be observed (Figure 7). consequently, all these annotations infer that mutations expressively affected the structure and variations in the internal dynamics of the complexes.

Furthermore, a free energy landscape (FEL) was constructed to relate the structural features and thermodynamics properties. To obtain the energy minima based on probability of given data

points of MD trajectories and to map the minimum energy conformation of the all the complexes during the explored time scale, and finally to connect the structural changeovers between these minima. Figure 8 represent the FEL of all the complexes i.e. wild type, R469A, K514A, Y518A, R521A and R522A. The wild type, K514A, Y518A and R522A shows only one energy minima while R469A and R521A exhibit two lowest energy minima separated by a small subspace, thus explaining global conformational variations adjusted by the mutant complexes in response to mutations. The major variations in these conformations were the loop deviation and nearby beta-sheets conversion. All the variations are highlighted in the Figure 8.

Binding Free Energy Calculations

To further connect the protein conformation changes with the binding affinity, we calculated the total binding energy using MM-GBSA approach. This method is considered as the best tool for calculating the real time-binding energy of the biological macromolecules complex. Herein to estimate the impact of the alanine substitutions at a specific position, we estimated the binding free energy. As given in Table 2, significant differences in the binding energy can be observed. The electrostatic contribution is significant in each complex. In the wild complex, the total binding free energy was estimated to be -145.18 ± 0.269 kcal/mol. However, in the R469A, this total binding energy was calculated to be -110.75 ± 0.230 kcal/mol. This reduction in the binding energy is due to the disruption of the hydrogen bonding network caused by the alanine

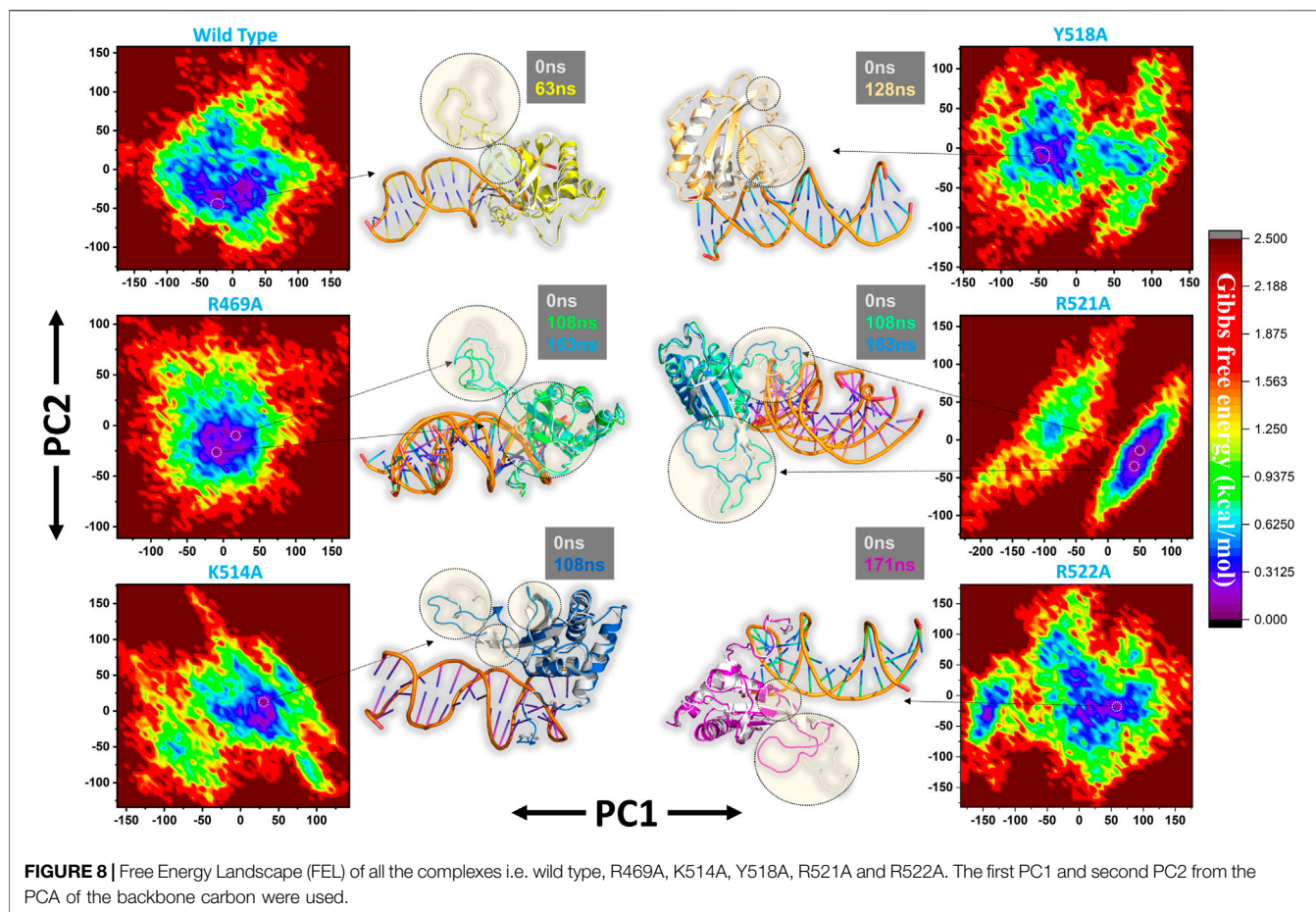


TABLE 2 | Display the total binding energy of the wild type and mutant complexes. Van Der Waal forces, electrostatic energy, generalized born, surface area and the total binding energy values for each complex is given. All the energies are calculated in kcal/mol.

Complex name	vdW	Electrostatic	EGB	ESURF	Total ΔG
Wild type	-132.12 \pm 0.142	-4,900.21 \pm 2.54	4,904.28 \pm 2.43	-17.12 \pm 0.017	-145.18 \pm 0.269
R469A	-124.36 \pm 0.165	-4,210.70 \pm 2.300	4,240.52 \pm 2.204	-16.21 \pm 0.015	-110.75 \pm 0.230
K514A	-131.48 \pm 0.127	-4,331.61 \pm 2.0239	4,375.12 \pm 1.967	-17.00 \pm 0.010	-104.97 \pm 0.175
Y518A	-127.19 \pm 0.137	-5,030.86 \pm 2.253	5,072.45 \pm 2.198	-16.56 \pm 0.013	-102.17 \pm 0.190
R521A	-132.93 \pm 0.1584	-4,343.05 \pm 2.353	4,395.21 \pm 2.299	-17.16 \pm 0.016	-97.93 \pm 0.226
R522A	-132.66 \pm 0.187	-4,420.46 \pm 2.280	4,469.95 \pm 2.241	-16.87 \pm 0.019	-100.04 \pm 0.215

Bold values are the total binding energy.

substitution. A particular interaction formed by the minor groove of DNA with the loop residue R469 causes a significant decline in the total binding energy. The total binding energy results of K514A and Y518A mutant complexes are comparable. For the K514A complex, the total binding energy was reported to be -104.97 ± 0.175 kcal/mol while the total binding energy for Y518A was -102.17 ± 0.190 kcal/mol. This is due to the loss of three hydrogen bonds formed by helix residues with the major groove are diminished upon the substitution. In case of K514A only one hydrogen bond while in case of Y518A, two important hydrogen bonds are lost. Significant drop out was observed in the total binding energy of the R521A mutant complex. The TBE was for R517A was reported to be -97.93 ± 0.226 kcal/mol. On the

other hand, the estimated binding energy for R522A was -100.04 ± 0.215 kcal/mol. Overall, these results show that the mutations induced significant energy drop out but the R521A and R522A reduced the binding energy by many folds. Hence these residues play a vital role in recognition of dsDNA and contribute to the infection initiation and progression.

CONCLUSION

In conclusion, herein, we systematically investigated the mechanism of dsDNA recognition by the EBNA1 protein. Our analysis revealed that R469A, K514A, Y518A, R521A and R522A

are the key hotspots for drug discovery against the various tumors caused by EBV. In particular, the two arginine substitution R521A and R522A, significantly affected the total binding energy. Thus we hypothesize that these residues impart a critical role in the dsDNA recognition and pathogenesis. This study would help to design structure-based drugs against EBV infections.

DATA AVAILABILITY STATEMENT

Publicly available datasets were analyzed in this study. This data can be found here: <http://www.rcsb.org/>, the accession number(s) can be found in the article/Supplementary Material.

REFERENCES

- Ali, A., Khan, A., Kaushik, A. C., Wang, Y., Ali, S. S., Junaid, M., et al. (2019). Immunoinformatic and Systems Biology Approaches to Predict and Validate Peptide Vaccines against Epstein-Barr Virus (EBV). *Scientific Rep.* 9 (1), 1–12. doi:10.1038/s41598-018-37070-z
- Balsera, M. A., Wriggers, W., Oono, Y., and Schulten, K. (1996). Principal Component Analysis and Long Time Protein Dynamics. *J. Phys. Chem.* 100 (7), 2567–2572. doi:10.1021/jp9536920
- Bochkarev, A., Barwell, J. A., Pfueter, R. A., Bochkareva, E., Frappier, L., and Edwards, A. M. (1996). Crystal Structure of the DNA-Binding Domain of the Epstein-Barr Virus Origin-Binding Protein, EBNA1, Bound to DNA. *Cell* 84 (5), 791–800. doi:10.1016/s0092-8674(00)81056-9
- Bochkarev, A., Bochkareva, E., Frappier, L., and Edwards, A. M. (1998). The 2.2 Å Structure of a Permanganate-Sensitive DNA Site Bound by the Epstein-Barr Virus Origin Binding Protein, EBNA1 1 Edited by T. Richmond. *J. Mol. Biol.* 284 (5), 1273–1278. doi:10.1006/jmbi.1998.2247
- Cruickshank, J., Shire, K., Davidson, A. R., Edwards, A. M., and Frappier, L. (2000). Two Domains of the Epstein-Barr Virus Origin DNA-Binding Protein, EBNA1, Orchestrate Sequence-specific DNA Binding. *J. Biol. Chem.* 275 (29), 22273–22277. doi:10.1074/jbc.m001414200
- Dominguez, C., Boelens, R., and Bonvin, A. M. J. J. (2003). HADDOCK: A Protein-Protein Docking Approach Based on Biochemical or Biophysical Information. *J. Am. Chem. Soc.* 125 (7), 1731–1737. doi:10.1021/ja026939x
- Ernst, M., Sittel, F., and Stock, G. (2015). Contact-and Distance-Based Principal Component Analysis of Protein Dynamics. *J. Chem. Phys.* 143 (24), 244114. doi:10.1063/1.4938249
- Garai-Ibabe, G., Grinyte, R., Canaan, A., and Pavlov, V. (2012). Homogeneous Assay for Detection of Active Epstein-Barr Nuclear Antigen 1 by Thrombin Activity Modulation. *Anal. Chem.* 84 (14), 5834–5837. doi:10.1021/ac301250f
- Garber, A. C., Hu, J., and Renne, R. (2002). 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. *J. Biol. Chem.* 277 (30), 27401–27411. doi:10.1074/jbc.m203489200
- Gianti, E., Messick, T. E., Lieberman, P. M., and Zauhar, R. J. (2016). Computational Analysis of EBNA1 "druggability" Suggests Novel Insights for Epstein-Barr Virus Inhibitor Design. *J. Comput. Aided Mol. Des.* 30 (4), 285–303. doi:10.1007/s10822-016-9899-y
- Haque, T. J., and Crawford, D. H. (1996). Transmission of Epstein - Barr Virus during Transplantation. *Rev. Med. Virol.* 6 (2), 77–84. doi:10.1002/(sici)1099-1654(199606)6:2<77::aid-rmv166>3.0.co;2-c
- Hiraku, Y., Kawanishi, S., and Ohshima, H. (2014). *Cancer and Inflammation Mechanisms: Chemical, Biological, and Clinical Aspects*. John Wiley & Sons.
- Hussain, I., Pervaiz, N., Khan, A., Saleem, S., Shireen, H., Wei, D.-Q., et al. (2020). Evolutionary and Structural Analysis of SARS-CoV-2 Specific Evasion of Host Immunity. *Genes Immun.*, 1–11.

AUTHOR CONTRIBUTIONS

HH, MW, and SA conceptualized the methodology. HH, SA, MW, and ZH did the analysis. HH, SA, and KC revised the methodology, worked on graphics and supervised the study. HH, MW, and KC wrote the manuscript. ZH did the revised work and improved the manuscript.

FUNDING

The work is supported by the National Key Research and Development Program of China under the grant number 2018YFA0902000.

- Junaid, M., Shah, M., Khan, A., Li, C.-D., Khan, M. T., Kaushik, A. C., et al. (2019). Structural-dynamic Insights into the *H. pylori* Cytotoxin-Associated Gene A (CagA) and its Abrogation to Interact with the Tumor Suppressor Protein ASPP2 Using Decoy Peptides. *J. Biomol. Struct. Dyn.* 37 (15), 4035–4050. doi:10.1080/07391102.2018.1537895
- Khan, A., Ali, S. S., Khan, M. T., Saleem, S., Ali, A., Suleman, M., et al. (2020). Combined Drug Repurposing and Virtual Screening Strategies with Molecular Dynamics Simulation Identified Potent Inhibitors for SARS-CoV-2 Main Protease (3CLpro). *J. Biomol. Struct. Dyn.*, 1–12. doi:10.1080/07391102.2020.1779128
- Khan, A., Junaid, M., Kaushik, A. C., Ali, A., Ali, S. S., Mehmood, A., et al. (2018). Computational Identification, Characterization and Validation of Potential Antigenic Peptide Vaccines from hrHPVs E6 Proteins Using Immunoinformatics and Computational Systems Biology Approaches. *PLoS one* 13 (5). doi:10.1371/journal.pone.0196484
- Khan, A., Junaid, M., Li, C.-D., Saleem, S., Humayun, F., Shamas, S., et al. (2020). Dynamics Insights into the Gain of Flexibility by Helix-12 in ESR1 as a Mechanism of Resistance to Drugs in Breast Cancer Cell Lines. *Front. Mol. Biosciences* 6, 159. doi:10.3389/fmolb.2019.00159
- Khan, A., Khan, M., Saleem, S., Babar, Z., Ali, A., Khan, A. A., et al. (2020). Decoding the Structure of RNA-dependent RNA-Polymerase (RdRp), Understanding the Ancestral Relationship and Dispersion Pattern of 2019 Wuhan Coronavirus.
- Khan, A., Khan, M., Saleem, S., Babar, Z., Ali, A., Khan, A. A., et al. (2020). Phylogenetic Analysis and Structural Perspectives of RNA-dependent RNA-Polymerase Inhibition from SARs-CoV-2 with Natural Products. *Interdiscip. Sci. Comput. Life Sci.* 12 (3), 335–348. doi:10.1007/s12539-020-00381-9
- Khan, A., Khan, M. T., Saleem, S., Junaid, M., Ali, A., Ali, S. S., et al. (2020). Structural Insights into the Mechanism of RNA Recognition by the N-Terminal RNA-Binding Domain of the SARS-CoV-2 Nucleocapsid Phosphoprotein. *Comput. Struct. Biotechnol. J.*
- Khan, A., Rehman, Z., Hashmi, H. F., Khan, A. A., Junaid, M., Sayaf, A. M., et al. (2020). An Integrated Systems Biology and Network-Based Approaches to Identify Novel Biomarkers in Breast Cancer Cell Lines Using Gene Expression Data. *Interdiscip. Sci. Comput. Life Sci.*, 1–14.
- Khan, M. T., Khan, A., Rehman, A. U., Wang, Y., Akhtar, K., Malik, S. I., et al. (2019). Structural and Free Energy Landscape of Novel Mutations in Ribosomal Protein S1 (rpsA) Associated with Pyrazinamide Resistance. *Scientific Rep.* 9 (1), 1–12. doi:10.1038/s41598-019-44013-9
- Leight, E. R., and Sugden, B. (2000). EBNA-1: a Protein Pivotal to Latent Infection by Epstein-Barr Virus. *Rev. Med. Virol.* 10 (2), 83–100. doi:10.1002/(sici)1099-1654(200003/04)10:2<83::aid-rmv262>3.0.co;2-t
- Lindner, S. E., and Sugden, B. (2007). The Plasmid Replicon of Epstein-Barr Virus: Mechanistic Insights into Efficient, Licensed, Extrachromosomal Replication in Human Cells. *Plasmid* 58 (1), 1–12. doi:10.1016/j.plasmid.2007.01.003
- Matsuura, H., Kirschner, A. N., Longnecker, R., and Jardtzyk, T. S. (2010). Crystal Structure of the Epstein-Barr Virus (EBV) Glycoprotein H/glycoprotein L (gH/gL) Complex. *Proc. Natl. Acad. Sci.* 107 (52), 22641–22646. doi:10.1073/pnas.1011806108

- Morgunova, E., Yin, Y., Jolma, A., Dave, K., Schmierer, B., Popov, A., et al. (2015). Structural Insights into the DNA-Binding Specificity of E2F Family Transcription Factors. *Nat. Commun.* 6 (1), 1–8. doi:10.1038/ncomms10050
- Okano, M. (1998). Epstein-Barr Virus Infection and its Role in the Expanding Spectrum of Human Diseases. *Acta Paediatr.* 87 (1), 11–18.
- Pearson, K. (1901). LIII. On Lines and Planes of Closest Fit to Systems of Points in Space. *Lond. Edinb. Dublin Philosophical Mag. J. Sci.* 2 (11), 559–572. doi:10.1080/14786440109462720
- Pope, J. H., Horne, M. K., and Scott, W. (1968). Transformation of Foetal Human Leukocytes *In Vitro* by Filtrates of a Human Leukaemic Cell Line Containing Herpes-like Virus. *Int. J. Cancer* 3 (6), 857–866. doi:10.1002/ijc.2910030619
- Roe, D. R., and Cheatham, T. E., III (2013). PTRAJ and CPPTRAJ: Software for Processing and Analysis of Molecular Dynamics Trajectory Data. *J. Chem. Theor. Comput.* 9 (7), 3084–3095. doi:10.1021/ct400341p
- Rose, P. W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., et al. (2016). The RCSB Protein Data Bank: Integrative View of Protein, Gene and 3D Structural Information. *Nucleic acids research*, gkw1000.
- Ryckaert, J.-P., Ciccotti, G., and Berendsen, H. J. C. (1977). Numerical Integration of the Cartesian Equations of Motion of a System with Constraints: Molecular Dynamics of N-Alkanes. *J. Comput. Phys.* 23 (3), 327–341. doi:10.1016/0021-9991(77)90098-5
- Salomon-Ferrer, R., Case, D. A., and Walker, R. C. (2013). An Overview of the Amber Biomolecular Simulation Package. *Wiley Interdiscip. Rev. Comput. Mol. Sci.* 3 (2), 198–210.
- Sagendorf, J. M., Markarian, N., Berman, H. M., and Rohs, R. (2020). DNAProDB: an Expanded Database and Web-Based Tool for Structural Analysis of DNA-Protein Complexes. *Nucleic Acids Res.* 48 (D1), D277–D287. doi:10.1093/nar/gkz889
- Taylor, G. S., Long, H. M., Brooks, J. M., Rickinson, A. B., and Hislop, A. D. (2015). The Immunology of Epstein-Barr Virus-Induced Disease. *Annu. Rev. Immunol.* 33, 787–821. doi:10.1146/annurev-immunol-032414-112326
- Tikhmyanova, N., Schultz, D. C., Lee, T., Salvino, J. M., and Lieberman, P. M. (2014). Identification of a New Class of Small Molecules that Efficiently Reactivate Latent Epstein-Barr Virus. *ACS Chem. Biol.* 9 (3), 785–795. doi:10.1021/cb4006326
- Tuszynska, I., Magnus, M., Jonak, K., Dawson, W., and Bujnicki, J. M. (2015). NPDock: a Web Server for Protein-Nucleic Acid Docking. *Nucleic Acids Res.* 43 (W1), W425–W430. doi:10.1093/nar/gkv493
- Vilar, S., Cozza, G., and Moro, S. (2008). Medicinal Chemistry and the Molecular Operating Environment (MOE): Application of QSAR and Molecular Docking to Drug Discovery. *Ctmc* 8 (18), 1555–1572. doi:10.2174/156802608786786624
- Wang, J., Lindner, S. E., Leight, E. R., and Sugden, B. (2006). Essential Elements of a Licensed, Mammalian Plasmid Origin of DNA Synthesis. *Mol. Cell Biol* 26 (3), 1124–1134. doi:10.1128/mcb.26.3.1124-1134.2006
- Wold, S., Esbensen, K., and Geladi, P. (1987). Principal Component Analysis. *Chemometrics Intell. Lab. Syst. 2* (1-3), 37–52. doi:10.1016/0169-7439(87)80084-9
- Yasuda, A., Noguchi, K., Minoshima, M., Kashiwazaki, G., Kanda, T., Katayama, K., et al. (2011). DNA Ligand Designed to Antagonize EBNA1 Represses Epstein-Barr Virus-Induced Immortalization. *Cancer Sci.* 102 (12), 2221–2230. doi:10.1111/j.1349-7006.2011.02098.x
- Young, L. S., and Rickinson, A. B. (2004). Epstein-Barr Virus: 40 Years on. *Nat. Rev. Cancer* 4 (10), 757–768. doi:10.1038/nrc1452
- Zwanzig, R. (1973). Nonlinear Generalized Langevin Equations. *J. Stat. Phys.* 9 (3), 215–220. doi:10.1007/bf01008729

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

Copyright © 2021 Hashmi, Waseem, Ali, Hussain and Chen. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.