Journal of Molecular Graphics and Modelling 114 (2022) 108196

Contents lists available at ScienceDirect

Journal of Molecular Graphics and Modelling

journal homepage: www.elsevier.com/locate/jmgm

How Epstein-Barr virus envelope glycoprotein gp350 tricks the CR2? A molecular dynamics study

Elif Naz Bingöl [a], Ilgaz Taştekil [a], Cansu Yay [b], Nursena Keskin [b], Pemra Ozbek [b,*]

[a] Department of Bioengineering, Institute of Pure and Applied Sciences, Marmara University, Istanbul, Turkey
[b] Department of Bioengineering, Faculty of Engineering, Marmara University, Istanbul, Turkey

ARTICLE INFO

Keywords:
Molecular dynamics simulations
Molecular docking
Epstein barr virus (EBV)
Envelope proteins
gp350
Type II complement Receptor (CR2)
Viral vaccines

ABSTRACT

The connection of Epstein Barr virus (EBV) with diseases such as Burkitt Lymphoma, Hodgkin disease, multiple sclerosis, systemic lupus erythematosus and various B-cell lymphomas made EBV glycoproteins one of the most popular vaccine immunogens. As a protein being encoded by EBV, the viral membrane envelope protein gp350 is studied extensively due to its abundancy on the surface and its interaction with complementary receptor, CR2. The binding of CR2 and gp350 not only leads to the entrance of the virus to the B-cells, but also prevents CR2 and C3d protein interactions that are required for immune response. Thus, understanding the inhibition of gp350 activity is crucial for vaccine development. Although, the active residues on gp350 structure were determined by several mutational studies, the exact mechanism of CR2 binding is still not clear. To this end, we have performed molecular docking followed by molecular dynamics simulations and MM-PBSA on wildtype and several mutated gp350 and CR2 structures. Apart from identifying crucial amino acids, the results of per-residue decomposition energy analysis clarified the individual energy contributions of amino acids and were also found to be accurate in differentiating the active site residues in CR2 binding. Here, we highlight the role of binding region residues (linker-1) but more interestingly, the dynamic relation between the distant sites of gp350 (linker-2 and D3 residues) and CR2. These findings can lead further vaccine development strategies by pointing to the importance of computationally found novel regions that can be potentially used to modulate gp350 activity.

1. Introduction

Epstein Barr virus (EBV) is a type of human herpes virus that is known to infect more than 90% percent of the world population. The effects of the virus are suppressed by the immune system in healthy people; whereas for immunosuppressed people, it creates health issues [1–3]. Having a close connection with many diseases such as Burkitt Lymphoma (usually Non-Hodgkin Lymphoma), Hodgkin disease, autoimmune diseases such as multiple sclerosis, systemic lupus erythematosus (SLE) and various B-cell lymphomas [4–8], many studies are focused on understanding and inhibiting the mechanism of EBV [9–16].

Several viral glycoproteins located on the surface of the virion and virus-infected cells are encoded by EBV such as glycoprotein gp350, gB, gH, gL, gp42 and BMRF2 [17]. EBV infection initiates in the B-cells with the recognition of viral epitopes by type II complement receptor (CR2) [18]. CR2, having a role in autoimmunity, is expressed on the surface of B cells. It consists of 15/16 60–70 amino acid repeats named short consensus repeats (SCRs), a transmembrane and an intracytoplasmic domain. Some SCRs are in the binding region of CR2, and are called as CR2 SCR1-2 [19,20]. The counterpart, EBV gp350, consists of 907 amino acid residues, while the N-terminus region (residues 1–470) is the receptor binding domain that was identified by X-ray crystallography [11]. Comprised of three independently folded domains (residues 4–153, 165–305, 317–426) and two linkers of 11 residues (residues 154–164 and 306–316) among them (Fig. 1), this L-shaped structure is heavily surrounded by 14 polyglycan chains, shielding the majority of the structure. However, a single glycan-free patch exists on the surface that is negatively charged and is determined to be responsible for binding to CR2 [11,21]. This region is formed by the two N-terminal domains and the linker region between them (residues 154–164). The three synthetic peptides present in the form of linear sequences (residues 16–29, 142–161 and 282–301) construct a loop on the glycan-free surface of the structure and were identified to be involved in EBV binding to CR2 [12,14]. SCR2 domain of CR2 is strongly positively charged and complements this negatively charged surface of gp350.

Furthermore, CD2/CR21 receptor of B cells can generate signals to

* Corresponding author.
E-mail address: pemra.ozbek@marmara.edu.tr (P. Ozbek).

https://doi.org/10.1016/j.jmgm.2022.108196
Received 28 July 2021; Received in revised form 9 April 2022; Accepted 11 April 2022
Available online 26 April 2022

stimulate humoral response by interacting with C3d and its deficiency is common in SLE patients [18,22–25]. Hence, attachment of CR2 and gp350 not only leads to the entrance of the virus to the B-cells, but also prevents CR2 and C3d protein interactions that are required for humoral response [26,27]. Thus, preventing EBV fusion to B cells is a key step for finding a cure for these diseases.

Most vaccine treatment studies aimed to generate monoclonal antibodies that neutralize the envelope glycoproteins having viral entry ability like gp350, gp42, glycoprotein H (gH), glycoprotein L (gL) and glycoprotein B (gB) [28,29]. Among these, viral membrane envelope protein gp350 has been studied most intensively as a vaccine immunogen due to its abundancy on the surface [9,10,12,30,31]. Many vaccine trials on non-human primates and human have been targeting EBV gp350 against the virus using various strategies [28,32]. In a vaccine treatment study performed by Bu et al., in order to determine the primary target, HeLa cells were infected with recombinant vaccine virus in which the envelope proteins (gp350, gH/gL and gp42) were expressed and followed by staining with MAbs. At the end, virus neutralization activities of glycoproteins were observed [9].

The efficiencies of monomeric vaccines [33] in preventing gp350 binding to CR2 were tested and even though the vaccination reduced the rate of infection, it did not protect B-cells from EBV infection [34]. On the other hand, multimeric vaccines were also developed to increase the immunogenicity of monomeric vaccines, and tetrameric ones with 19-fold higher performance were designed and tested on rabbits [35]. Nevertheless, none of these developments could meet the expectations [28,34,35]. On the other hand, clinical trials on mice and monkeys showed that they generate a larger amount of neutralizing antibodies using nanoparticles containing gp350-ferritin compared to the previous methods [9,28,34]. These nanoparticles have a better ability to enhance the recognition of glycan-free area of gp350 by CR2/CD21 [36]. Even though many of these vaccines were not successful, commercially available neutralizing anti-gp350 monoclonal antibody 72A1 was characterised as being able to prevent the infection of B cells by EBV by blocking the attachment of gp350 to CR2 [10,18].

Aiming to inhibit and examine the interaction mechanisms, some mutational studies were also performed on gp350 and CR2 structures. In a study by Szakonyi et al. [11], multiple-site mutations (Mut1 (P158A + Y159A + I160A), Mut2 (W162A + D163A + N164A), Mut3 (D208R + E210R) and MutC (E426A + S427A)) were conducted on the binding region of gp350. Accordingly, residues that are crucial for binding were determined [11]. In another study, a series of different single-site mutations were applied, which resulted in a decrease in the binding of both gp350 (residues E21A, Y151A, E155A, I160A, W162A, D208A, E210A and D296A) and CR2 (residues R13A, S15P, R28A, S32A, T34A, R36A, K41A, K57A, Y64A, K67A, R83A and R89A) (Fig. 2). Following this, active residues involved in the binding event were determined [15].

To date, although there were several attempts to inhibit EBV infection, neither a definite vaccine nor a specific treatment is present. To this end, we have conducted a computational study involving both molecular docking and MD simulations on gp350 and CR2 structures using their unbound forms as well as wildtype and selected mutated complex forms. Aiming to enlighten the binding mechanism, previous experimental mutation studies were described using computational simulations and the reflections of these site mutations on the overall global dynamics of the structures are studied in detail. MM-PBSA analysis were performed on the complex structures and the contribution of each residue was calculated accordingly. Further analysis is conducted on a residue-based level to identify the residues that provide connection between the binding region and linker-2, which is crucial for dynamic stability of the complex. The importance of linker-2 region is highlighted in the findings which can be used in the vaccine development strategies in the future.

2. Materials and methods

2.1. Protein structures and selected mutations

3D structures of EBV major envelope glycoprotein, gp350 (PDB ID: 2H6O [11]) and CR2 (PDB ID: ILY2, [20] were retrieved from Protein Data Bank (PDB) [37]. 4 multi-site *in silico* mutations were applied on the gp350 structure (details given in Table 1 and Fig. 2) based on a former experimental study [11].

2.2. Docking of gp350 and CR2 structures

Protein-protein interactions of both wild-type and mutated models of gp350-CR2 were studied using a data driven docking program, HADDOCK 2.2 [38]. WT as well as the other four mutated structures were docked individually using the solvated docking option. Mutations on gp350, shown in Fig. 2, were applied using Pymol (Version 2.3.3 Schrödinger) [39]. Fraction for Ambiguous Interaction Restraints (AIRs) were randomly excluded by the system while the active residues were determined with respect to the previous experimental studies as ASP21, TYR151, GLU155, ILE160, TRP162, ASP208, GLU210 and ASP296 for gp350 and ARG13, SER15, ARG28, LYS41, LYS67 and ARG83 for CR2 [15,16]. Accordingly, passive residues were defined as the ones around the active residues. Docking calculations were performed using Guru interface by taking relative solvent accessibility cut-off as 0.50, number of structures for rigid body as 2000, number of structures for water refinement as 100 and solvated docking parameter as 'on'. For the clustering parameters, RMSD method was selected and adjusted to 1.5 Å. The resulting docking poses were analysed and selected based on their HADDOCK score, intermolecular energy and AIR violation as well as their orientation.

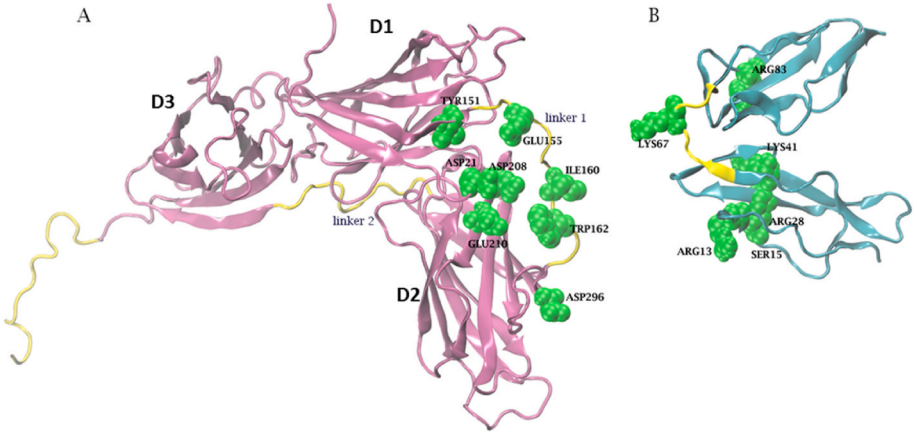

Fig. 1. 3D structure of A) gp350 (PDB ID: 2H6O) and B) CR2 (PDB ID: ILY2 (Prota et al., 2002)) are shown in a cartoon representation. Active residues are shown in green. A) The link that connects D1 (4–153) and D2 (165–305) is depicted as linker-1 and link that connects D2 and D3 (317–426) is depicted as linker-2. B) SCR1-2 domains of CR2 are shown in cyan, linker connecting these domains is shown in yellow. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

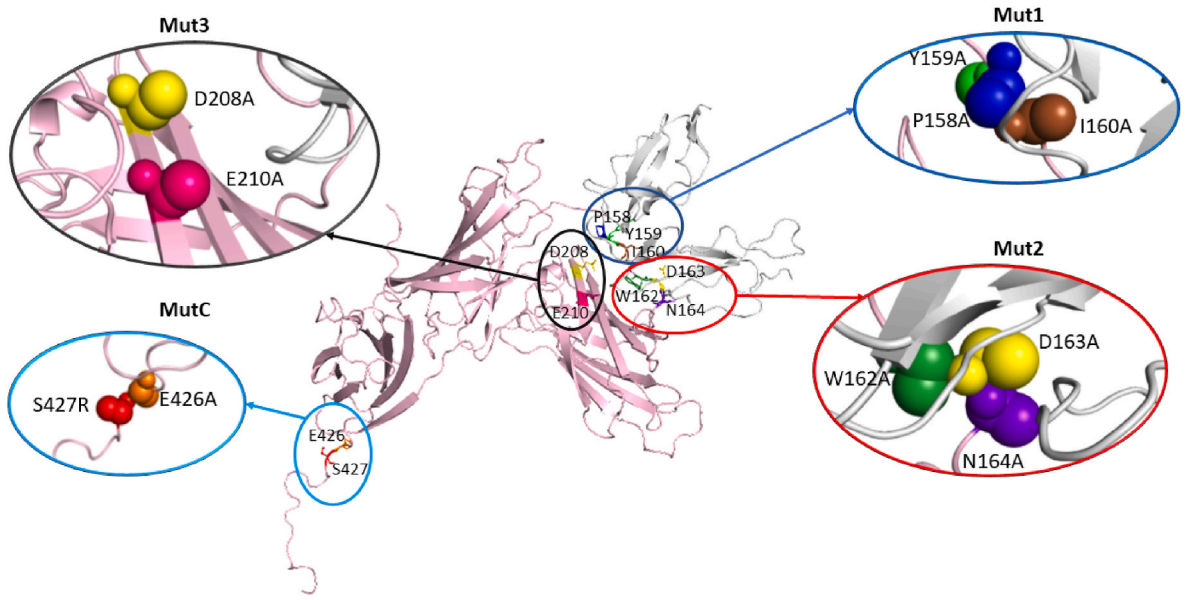

Fig. 2. Multi-site mutations are shown on the 3D structure of gp350; Mut1 (P158A + Y159A + I160A), Mut2 (W162A + D163A + N164A), Mut3 (D208A + E210A) and MutC (E426R + S427R). gp350 is shown in light pink and CR2 is shown in grey cartoon representation. Mutation sites are zoomed and shown in sphere representation. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

Table 1
Mutation Sites in gp350 (PDB ID: 2H6O [11]).

	Mutation Sites on gp350
Mut1	P158A + Y159A + I160A
Mut2	W162A + D163A + N164A
Mut3	D208A + E210A
MutC	E426R + S427R

2.3. MD simulations

The crystal structures (unbounded gp350, unbounded CR2, complex forms of WT, Mut1, Mut2, Mut3 and MutC) were solvated in a cubic water box using SPC/E water molecules. The solvated systems were neutralized by replacing water molecules with appropriate amounts of sodium/chloride ions. MD simulations were performed using GROMACS 5.0 [40,41] and OPLS-AA/L all-atom force field [42,43]. After the systems were neutralized with ions, energy minimization step was performed to avoid steric clashes or improper geometry. Then, the simulations were equilibrated in two phases: NVT (constant Number of particles, Volume, and Temperature) and NPT (constant Number of particles, Pressure and Temperature) ensembles. The simulations were then performed in the NPT ensemble with the conditions of 1 bar, 310 K and a 2 fs time step for 500ns.

2.4. Analysis of trajectories

For each system, trajectories from two parallel 500 ns simulations were concatenated after removing initial 250 ns portions of the MD simulation. The rest of the analyses were carried out on the merged trajectories of the parallel MD simulations.

The conformations generated in each trajectory were superimposed on to the average conformation to calculate Root Mean Square Deviation (RMSD) and Root Mean Square Fluctuation (RMSF) of the amino acids' alpha-carbons only.

2.4.1. Global motional correlation calculations

Aiming to characterize the motional coupling among gp350 and CR2 residues, a covariance matrix consisting alpha-carbon positional fluctuations (C) was constructed using Eq. (1),

$$C = \langle (X - \langle X \rangle)\left((X - X)^T\right) \rangle \tag{1}$$

where X is the coordinates of all alpha-carbon atoms of the protein with a dimension of 3Nx3N and $\langle X \rangle$ is the matrix of average positions of the alpha carbons along the equilibrium period of the trajectories. The Essential Dynamic Analysis (EDA) of the covariance matrix was performed [44] using matrix diagonalization according to Eq. (2),

$$\lambda = T^T C T \tag{2}$$

where λ denotes diagonal matrix of eigenvalues and T denotes eigenvectors, which are the principle components. These correspond to the modes of protein motion in characteristic directions. Eigenvalues reflect the contribution to the overall positional fluctuation of the corresponding eigenvector. A dynamic cross-correlation matrix (DCCM) is constructed using Eq. (3),

$$C_{ij} = \frac{\langle \Delta r_i \Delta r_j \rangle}{\langle \Delta r_i^2 \rangle^{1/2} \langle \Delta r_j^2 \rangle^{1/2}} \tag{3}$$

where C_{ij} denotes the cross-correlation among residues i and j; Δr_i and Δr_j denote the positional fluctuations of alpha-carbons of residues i and j, respectively in that mode. An average DCCM for each system is computed using each mode in top 20 EDA modes (with the highest contributions to the overall motion), the weighted sum of C_{ij} matrices were collected and eigenvalues were used as weights.

2.4.2. Pairwise amino acid interaction energies

The non-bonded pairwise interaction energies (sum of electrostatic and van der Waals energy terms) were calculated for an ensemble of conformations obtained from MD simulations using gRINN v1.1.0 with non-bonded interaction distance cutoff of 12 Å and pre-filtering cutoff of 16 Å [45].

2.4.3. MM-PBSA calculations

The decomposition of binding free energy is calculated using MM/PBSA (Molecular Mechanics-Poisson Boltzmann Surface Area) method. The binding free energy (ΔG_{bind}) for a complex can be computed using Eq. (4),

$$\Delta G_{bind} = \langle G_C \rangle - \Delta G_A \rangle - \langle G_B \rangle \quad (4)$$

where sub-scripts C, A and B denotes complex, monomer A and monomer B, respectively. The free energy of each term can be estimated using Eq. (5),

$$G = \langle E_{MM} \rangle + \langle G_{SOL} \rangle - \langle T\Delta S \rangle \quad (5)$$

where E_{MM} is molecular mechanics of energy of the molecule, G_{SOL} denotes the solvation profile of the molecule, T is temperature and S is entropy of the system. E_{MM} can be calculated using Eq. (6),

$$E_{MM} = \langle E_{int} \rangle + \langle E_{ele} \rangle - \langle E_{vdw} \rangle \quad (6)$$

where E_{int} is total internal energy, E_{ele} is electrostatic energy and E_{vdw} is van der Waals term. The entropy calculations increase the computational cost of the MM-PBSA calculations. To avoid such a cost, it is common to calculate only a few snapshots of the trajectory which may lead to missing important conformations. Alternatively, it was emphasized in literature that the entropy term can be neglected, and significant number of published studies followed this notion [46,47]. For these reasons, we chose to neglect the entropy term. Accordingly, Eq. (4) can be rearranged as:

$$\Delta G_{bind} = \langle \Delta E_{MM} \rangle + \langle \Delta G_{SOL} \rangle \quad (7)$$

where ΔE_{MM} denotes the change in internal energy and ΔG_{SOL} denotes the change in solvation profile between monomers and final complex. The binding energies for all the systems were calculated using MM-PBSA method with gmx_MMPBSA Tool [48]. For each system, a total of 100 snapshots from equilibrium periods were used in the calculations. The ionic strength was set as 0.15 M, grid spacing and probe radius (for SASA estimation) was set as 0.5 Å and 1.4 Å, respectively. The solvent dielectric constant was set to 80, whereas the solute dielectric constant was set to 2.

3. Results and discussion

3.1. Analysis of docking results

Docking poses of gp350-CR2 complexes for WT and mutant proteins were formed using the HADDOCK 2.2 [38] program and the solvated docking mode. Since, the relative orientation between the interface residues was known from previous experimental studies [15], docking poses not reflecting this information were eliminated and the lowest score poses from the remaining ones were selected.

The complex formed between WT gp350 and CR2 was similar with a previous docking pose obtained by Young et al. [15]. Docking poses corresponding to mutant complexes, Mut1, Mut2 and Mut3 (see Fig. 2) had more positive scores (see Table 2) suggesting weaker binding, while MutC docking pose wascloser to that of WT in terms of both HADDOCK score and intermolecular energy values. The information regarding the selected poses is given in Table 2, while 3D structural figures are given in Supplementary Fig. 1.

As mutations (Mut1, Mut2 and Mut3) are introduced to the system, HADDOCK score as well as the intermolecular energies are worsened since each mutation involves an active interface residue. This effect was also observed in the experimental studies in terms of disruptions in gp350-CR2 binding, where Mut2 and Mut3 exhibited the greatest effect [11]. On the other hand, the control mutation, MutC, is score-wise similar to WT. Thus, the results from docking calculations for the WT and the mutated complexes obtained from HADDOCK supported the effect of mutations observed experimentally [11].

3.2. Analysis of MD simulations

MD simulation ensembles from unbound gp350, unbound CR2, complex forms of gp350-CR2 WT, Mut1, Mut2, Mut3 and MutC were analysed to investigate gp350-CR2 binding mechanism. The RMSD of the Cα plots suggested that the simulations reached equilibrium at 250ns. Accordingly, any measures were performed on the merged trajectories of each equilibrium period (250–500 ns) of the parallel runs totalling 500 ns simulation time for each complex structure (Supplementary Figs. 2–3).

4.2.1. Change in flexibilities upon point mutations

According to RMSF(Cα) calculations (Supplementary Fig. 4), the most mobile regions generally correspond to loops and beta sheet turns in all the structures. The dramatic changes induced by mutations can be identified by calculating the RMSF difference between the WT and the mutated models. Accordingly, residues with most notable changes are given in Fig. 3 and Supplementary Table 1.

Focusing on gp350 region, for the WT, most of the residues that display peaks have crucial roles in CR2 binding including TYR159, LYS161, GLU206, PRO250, VAL251, THR253-THR258 and PRO290-GLY295 binding [10,11,14,15,18,31,50].

For Mut1, the most significant differences are observed in domains D2 and D3. In particular, the highest peaks in ΔRMSF profile of Mut1 correspond to structurally close residue groups such as VAL221-GLY224, THR246-THR258 in D2 and VAL333-VAL346, PHE383-GLY397 in D3 domains forming beta sheets. As previously mentioned, these residues were affiliated with important roles in terms of structure-function relationship such as being resided on epitope regions [10,31].

The RMSF results of Mut2 generally have lower RMSF values compared to WT except W162A-N164A, GLN177. Peaks at these points were legitimate since they reside in close proximity to mutation sites. Moreover, similar to Mut1, D3 residues (PRO250-THR253, ASN342-PRO344 and GLY397-ALA399) of gp350 displayed higher mobility compared to WT.

One of the residues with high RMSF in Mut3, VAL157, is located on linker-1 and interacts with GLU155, THR156 and PRO158 which are active residues [11,14,22,51] in the docking process. Similar to VAL157, ASN164 also resides in linker-1 and interacts with other residues of this linker such as TRP162 and ASP163. Moreover, it also interacts with ARG13 and SER30 of CR2 and GLN300 of gp350, which have roles in gp350 binding [11,14,22,51].

From a general perspective, for MutC, the characteristics of the RMSF (Cα) plot were mainly similar with WT in the gp350 region. The only notable differences were less than 3Åwhich included the mutation sites (E426A and S427A).

For the CR2 region, the largest differences in the RMSF profiles were detected in Mut2 mostly in regions around ASP45-LYS60, LEU97-

Table 2
Docking scores, intermolecular energies and RMSD values obtained from HADDOCK 2.2 [49] for WT, MutC, Mut1, Mut2 and Mut3 complexes.

Model	HADDOCK Score	AIR Violation (Å)	Intermolecular Energies (kcal/mol)		Average Intermolecular Energies (MM-PBSA) (kcal/mol)				RMSD (Å)
			Average	*Individual*	E_{el}	E_{vdW}	G_{pol}	G_{np}	
WT	−140.90 ± 10.97	0	−575.10 ± 30.28	−559.90	−459.78	−31.95	463.07	−14.18	0.43 ± 0.28
MutC	−139.43 ± 50.59	0	−576.80 ± 47.63	−599.63	−446.60	−6.02	434.43	−11.90	1.12 ± 0.74
Mut1	−130.98 ± 31.18	0	−513.83 ± 185.96	−506.41	−422.73	−0.65	423.01	−8.53	0.50 ± 0.30
Mut2	−127.54 ± 7.90	2	−448.23 ± 44.17	−400.91	−262.86	−13.84	264.58	−10.88	1.17 ± 0.99
Mut3	−98.25 ± 10.67	1	−303.50 ± 72.52	−399.80	−358.16	−14.62	360.98	−11.65	2.49 ± 1.46

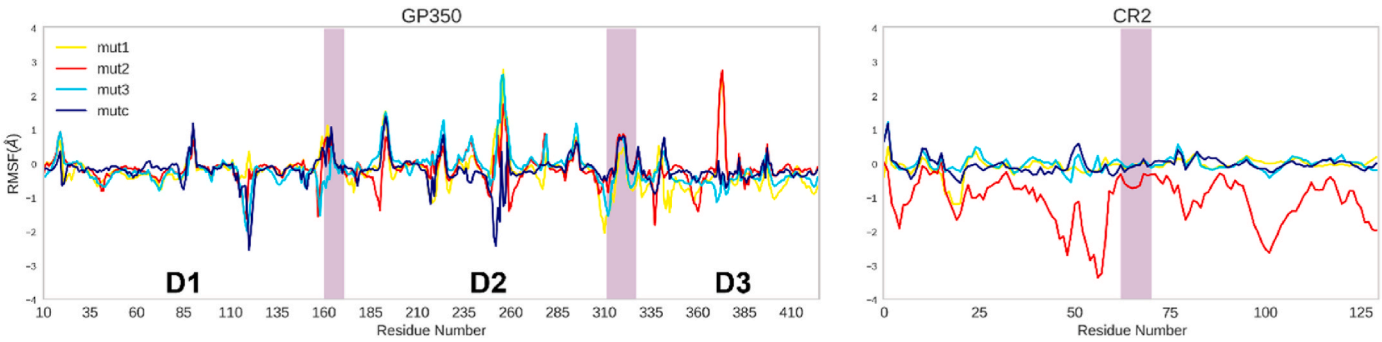

Fig. 3. ΔRMSF graphs constructed by subtracting RMSF(Cα) value of WT from the mutated structures. The domains are separated by linkers highlighted in purple. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

LYS110 and GLU120-ASP129. The ΔRMSF profile of Mut2 clearly differentiate from the rest highlighting the role of TRP162, ASP163, ASN164 in CR2 binding which also supports the experimental findings [11].

To sum up, residues around mutation sites are affected the most, as expected. The main differences are observed either in D2, D3, or both for all the mutated structures. The effects of these mutations on the global dynamics were further investigated using motional correlation analysis in the following section.

4.2.1.1. Motional correlations of WT gp350-CR2 in bound and unbounded forms. The cross-correlation of WT gp350-CR2 complex revealed that respective residues of each domain have tight positive correlations among themselves (Fig. 4-A, D). The positively correlated large residue blocks of each domain can be observed at the diagonal line of the cross-correlation map. The first two domains (D1 and D2) behave as a single domain except PHE35-TYR36 and PRO254, which display an anchor-like behavior [52], whereas D3 domain has negative correlations except the hinge-like residues (ASN345, ALA350-THR360, GLY375, THR410, ALA424). Interestingly, while D1 and D2 display negative correlations, D3 (VAL304-THR428) displays positive correlations with CR2, which is further investigated by pairwise energy calculations below.

On the other hand, cross-correlation map of the unbound gp350 displays a completely different character (Fig. 4-C). The positively correlated residue blocks of the individual domains were still definable while the high positive correlation between D1 and D2 is mostly diminished. Moreover, D3 has positive correlations on D2-CR2 interface residues and strong positive correlations were observed between SER306-VAL325 and PRO400-PRO425 and D1. The strong positive correlation between domains of CR2 (SRC1 and SRC2) is disrupted when it is non-bonded (Fig. 4-B). These domains acted as one unit in the presence of gp350, which can be explained by the effect of rigidification induced by complex formation.

4.2.2. Motional correlations of mutated systems

When cross-correlations of Mut1 are investigated (Fig. 5-A), the residue blocks present in the WT are mostly preserved; however, disruptions are observed. The strong positive correlations between D1 and D2 is decreased and the hinge-like movement of PRO254 is diminished whereas PHE35-TYR36 still retains this movement. A similar scenario can be observed in correlations among D3 and CR2. The positive correlations are decreased while preserving thehinge movement of ASN345. The results may help to explain underlying reason of the similar binding capacity of Mut1 and WT; a decrease in the binding is observed but due to these similarities, the mutation was not as effective as Mut2 and Mut3 [11].

Upon mutations introduced in Mut2 (Fig. 5-B), the correlations between residues are changed to a great extent. The strong positive inter-domain correlations detected in WT between domains D1 and D2 can no longer be observed in Mut2. Moreover, intra-domain correlations are substantially shifted from positive to negative. Additionally, strong positive correlations observed among the residues of CR2 is also disrupted and mostly transformed to negative correlations. Almost all the hinge residues are lost or cannot be detected due to the unstable relations between the residues, the correlation changes are very frequent almost at all the residues, including the mutation sites (W162A, D163A, N164A).

In Mut3, most of the inter- and intra-domain correlations are still maintained. Although positive correlations are still present among domains D1 and D2, they no longer display large positively correlated residue blocks (Fig. 5-C). A similar observation is valid for CR2 as well; the inner-domain positive correlations are disrupted and instead of maintaining the inner positive correlations, residues forged new positive correlations with D1 and D2 domains of gp350. Possible hinge residues for Mut3 are, ASP134-ALA144, PRO254, PRO344 of gp350 and ARG13, THR34, CYS62-PRO72 and ILE82-ASP92of CR2.

In MutC, the four distinct regions observed in WT are conserved with a few differences in the hinge residues (Fig. 5-D). Although the inner domain correlations are the similar to WT, the correlation among CR2 and gp350 residues are disrupted.

Introduction of Mut2 and Mut3 on the complex structures affected the cross-correlation profiles such that both gp350 and CR2 display profiles that are similar to their unbound forms (Fig. 4-B-C). The correlations within the complex structures are diminished as a result of these mutations suggesting the importance of these sites in global dynamics.

4.2.3. Change in pairwise residue interaction energies upon mutations

The motional correlations showed a direct relation between binding capacity and dynamic behavior of gp350. Although mutations induced an immense change compared to wildtype, the residues taking part in this change other than mutation sites would give a detailed explanation for the dynamical alteration. To this end, we present pairwise interaction energy results obtained from gRINN [45]. The distribution of interaction energy change of each domain (D1, D2 and D3) for each mutation with respect to WT is given in Fig. 6. The structures with most effective mutations (Mut2 and Mut3) span a wider range compared to the least effective mutation, Mut1.

A more detailed analysis is conducted on a linker basis (Supplementary Fig. 5). Since linker-1 accommodates the active site residues, interaction energies of mutated residues differ substantially from the WT structure in TYR151, GLU155, VAL157, TYR159, ILE160 and ILE303 (Supplementary Fig. 5-A-B). On the other hand, the overall interaction of linker-1 with the rest of the complex displayed a wider range of energy difference in Mut2 and Mut3 compared to Mut1(Supplementary Fig. 5B). The effect of mutations on the interaction energy profiles can be detected by focusing on the residues displaying highest interaction energy differences (Supplementary Fig. 6, Supplementary Table 2). Apart from local effects induced by mutations, distal effects are also present

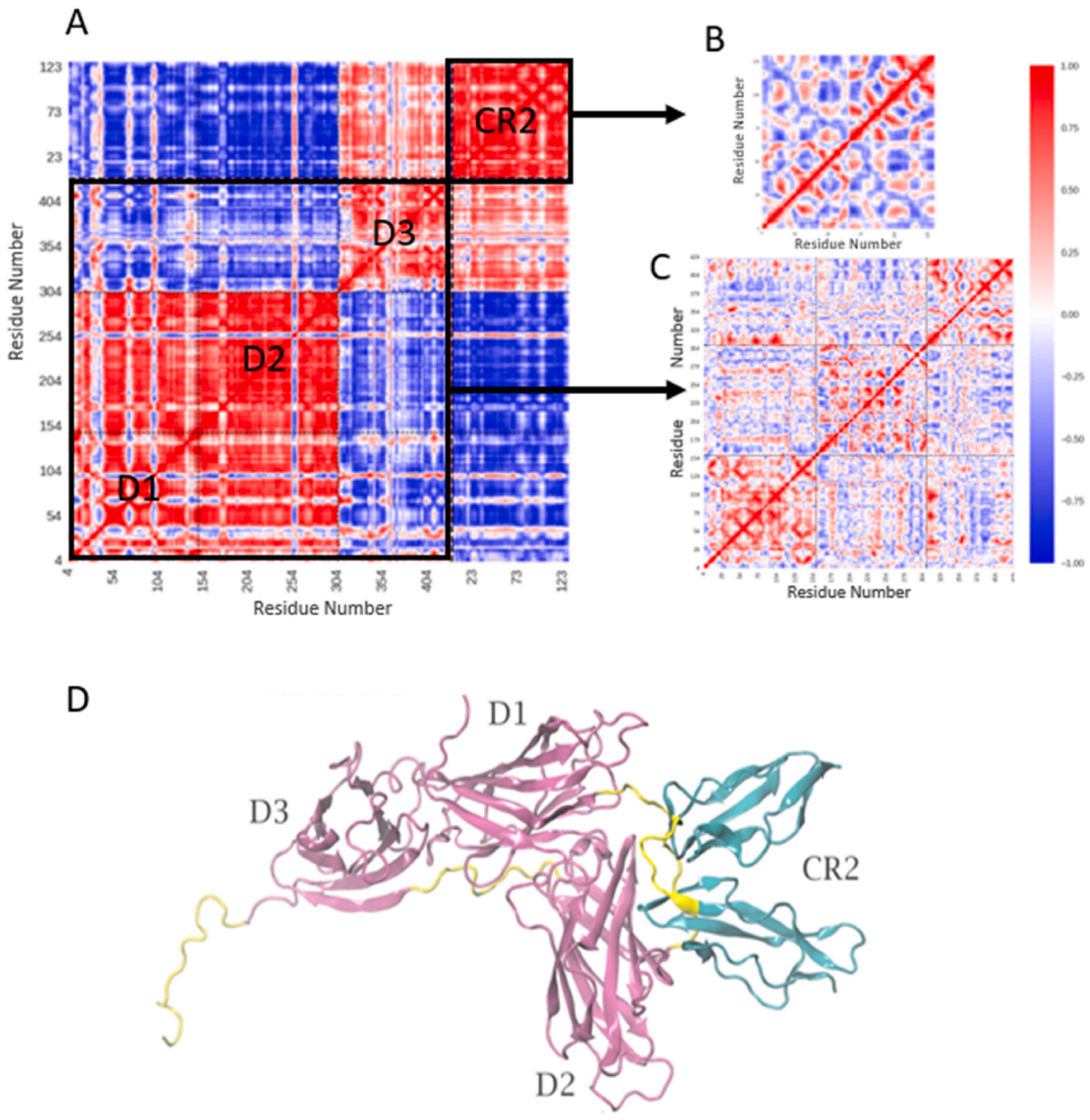

Fig. 4. Cross correlation maps of each model, where positive and negative correlations are represented with red and blue colors, respectively. A. WT gp350-CR2 complex B. Unbound CR2 structure C. Unbound gp350 structure D. 3D representation of gp350-CR2 complex domains. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

especially in D2, D3 and CR2 regions, which draws attention to the importance of these regions in particular (Fig. 6). This effect is also reflected on the global dynamics in the form of diminished correlations upon induction of mutations. Tracing the elements of this effect, we focused on the residues that potentially have a role in the establishment of domain interrelations. For this purpose, we constructed pairwise interaction energy difference maps of linker-1 and linker-2 in mutated structures compared to WT (Fig. 7). In linker-1, the residues with notable energy changes are ASN27, ILE113, GLY204-GLU214, ASN288, GLN300, VAL304 of gp350 and THR25-SER30, PHE35-GLU40, SER70, GLY84, SER85, VAL94, THR95 of CR2 (marked with green boxes in Fig. 7). In linker-2, the pairwise interaction energies of ASN40, PRO78, GLY100-ARG107, LEU179, ASN278 and PHE421 of gp350 displayed changes compared to WT (marked with orange boxes in Fig. 7). These results clearly demonstrated the dynamic relation between the distant sites of gp350 (linker-2 and D3 residues) and CR2 which is also visually displayed on the structure given in Fig. 7B.

4.2.4. *Energy decomposition of the structures*

The MM-G(P)BSA method calculates the contributions of gas phase

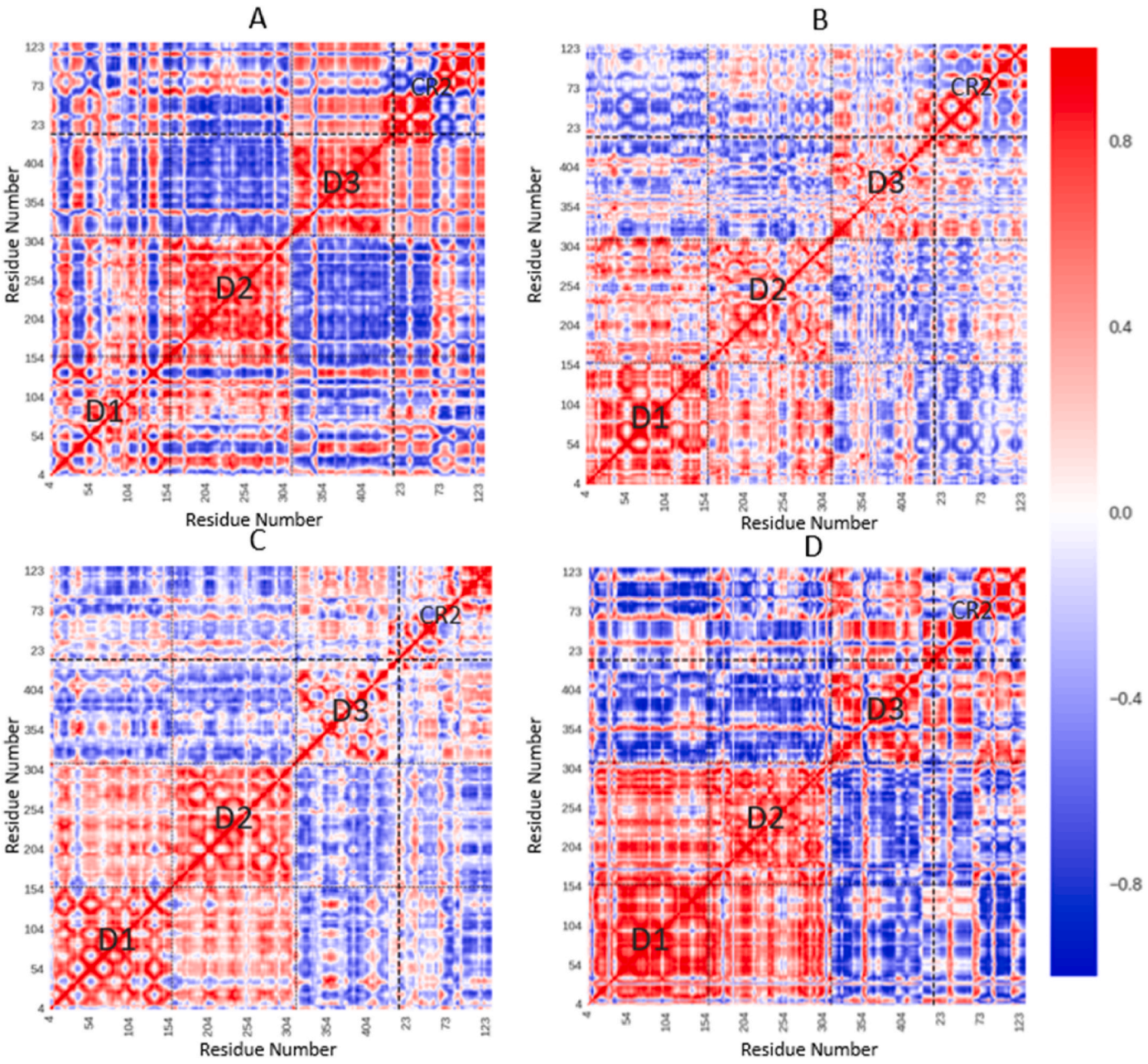

Fig. 5. Cross correlation maps of each model, where positive and negative correlations are represented with red and blue colors, respectively. Domains (D1-D4) are marked on the figures. gp350 and CR2 regions are separated by dashed lines. A. Mut1 B. Mut2 C. Mut3 D. MutC. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

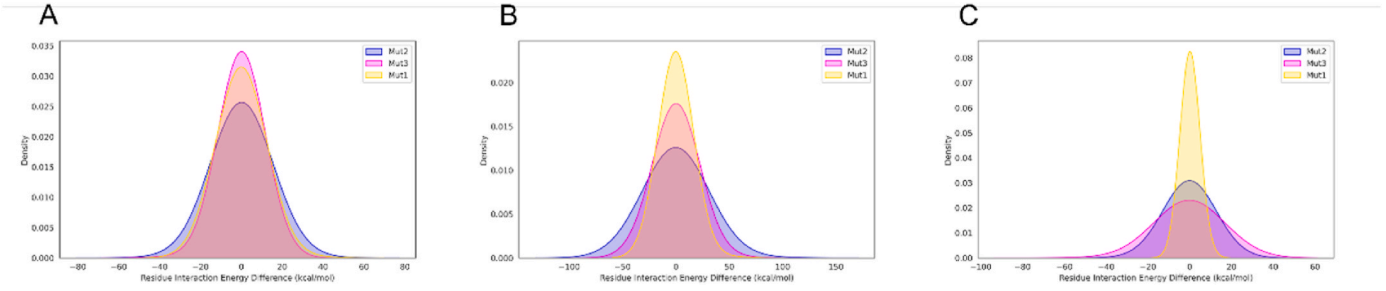

Fig. 6. Distribution of pairwise residue interaction energy differences of gp350 domains are given for each mutated structure A. D1, B. D2 and C. D3.

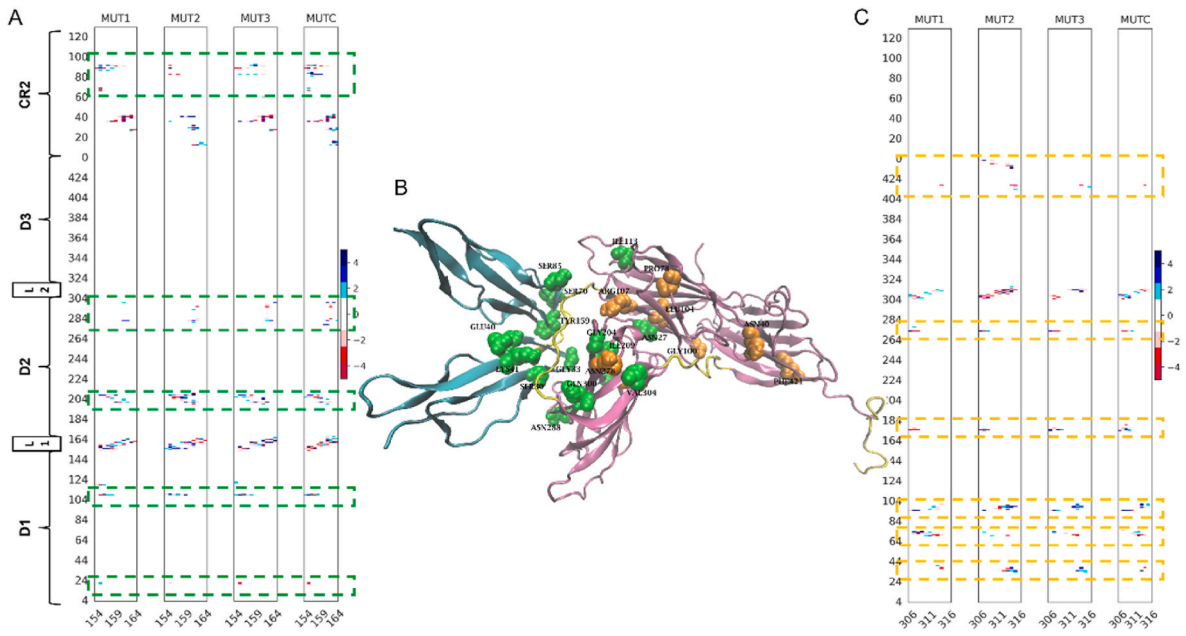

Fig. 7. The pairwise interaction energy difference map of A. linker-1 and C. linker-2. The link that connects D1 (4–153) and D2 (165–305) is depicted as linker-1 and the link that connects D2 and D3 (317–426) is depicted as linker-2. The residues with notable differences are marked with green and orange dashed boxes for linker-1 and linker-2, respectively. B. The residues with notable differences are demonstrated as spheres on the 3D structure of gp350-CR2 complex (green (linker-1) and orange (linker-2)). (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

energies, solvent energies and solute entropies using snapshots taken from the trajectories of free molecules as well as the complex molecule. The binding free energy is then determined by taking the difference of the complex and free molecules. To quantify the energy contribution of the amino acids highlighted in the pairwise interaction energy results, MM-PBSA calculations were carried out for each system (Table 2). The energy contributions of the amino acids upon 7 Å radius of the binding interface are given in Fig. 8. Apart from identifying crucial amino acids, the results of per-residue decomposition energy analysis were helpful in understanding the individual energy contributions of amino acids and were also found to be accurate in differentiating the active site residues in CR2 binding. The active site residues that were previously identified experimentally can also be recognized in the WT energy decomposition map, where residues with low energy values are shown in blue (Fig. 8A)

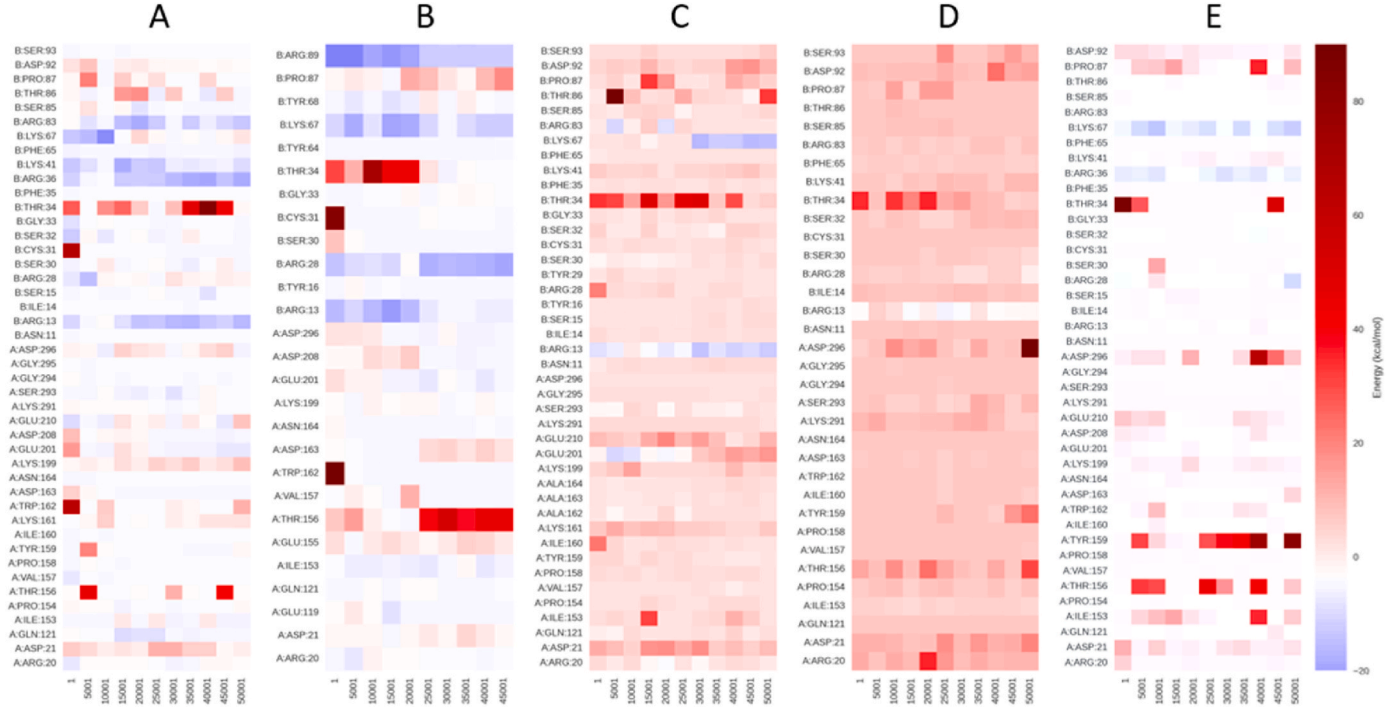

Fig. 8. The heatmaps of energy decomposition data obtained from MM-PBSA. The snapshots were taken with 1000 ps intervals. The colorbar indicates energy values from lower (blue) to higher (red) energy. A. WT B. Mut1 C. Mut2 D. Mut3 E. MutC. (For interpretation of the references to color in this figure legend, the reader is referred to the Web version of this article.)

[15]. The comparison of decomposition maps of the mutated structures aid in differentiating the effects induced by the amino acid substitutions at the active sites. The profiles of Mut1 and MutC are similar to WT, while Mut2 and Mut3 result in an overall higher energy map indicating lower affinity in CR2 binding. Both pairwise interaction energy analysis of gRINN and energy decomposition analysis of MM-PBSA emphasized mostly common residues. Among these are ARG20, ILE153-PRO158, ASN164, LYS199, GLU210, LYS291 and ASP296 for gp350 and ARG13, ARG28, GLY33-ARG36, LYS67 and ARG83-SER93 for CR2.

5. Conclusion

Immunoinformatics plays an important role in the development of antibody and immunodiagnostic agents and in the design of vaccines. The early phases of the development of vaccines and other therapeutic agents were focused only on immunological studies. These early developmental approaches were often repetitive and expensive. The use of computational approaches significantly reduces time and expense of the production of such agents by gathering data faster than experimental techniques. Here, we compared WT and mutation-containing structures using docking and MD simulations to investigate the binding mechanism of the EBV glycoprotein-350 (gp350) and complementary receptor 2 (CR2). As a result, along with the compatibility of the findings with the literature, novel information was also obtained about the structural dynamics of EBV. To the extent of our knowledge, this is the most extensive MD simulation study on gp350 and CR2 system.

To understand the functional mechanism of a protein and to obtain further knowledge of its dynamics, it is crucial to identify the structural fluctuations induced by mutations. In the light of the RMSF results, more in-depth knowledge was obtained on the role of residues that form linker-1 in binding. The terminal residues of linker-1 play a less active role in binding than those close to the domains. Moreover, anchoring residues of linker-2 to the D2 and D3 and decreased RMSF values for Mut1, Mut2 and Mut3 on the linker-2 highlights the active role of D2 and D3 in binding.

In our findings based on the DCCM, it was observed that D1 and D2 behave as a single rigid domain in the wildtype structure. All the mutations affected the cross-correlation results to varying degrees and these effects are in line with the binding properties of these structures. Among the factors that cause Mut2 to be the most disrupted structure, a mutation on the binding residue lies. In addition, it can also be concluded that it affects not only binding but also the dynamic character of the structure. MutC and WT have similar cross correlation profiles since the mutations introduced in MutC reside far from the binding region, hence have limited effects on the global dynamics of the structure. The dynamically correlated residues with the whole complex are referred as hinge residues which inherits a rotational movement around itself and resides mostly in loop or terminal regions [53]. The hinge residues observed in the cross-correlation maps remained constant in most of the mutated structures. This highlights the critical roles of these residues in global movement and binding.

As can be seen in RMSF and DCCM, mutations applied on linker-1 had a considerable effect on the dynamics of the structure. Furthermore, when the pairwise energies were examined, it was determined that these mutations did not only change the binding site, but also the residues that provide connection between the binding region and linker-2. Hence, linker-2 is also considered to be highly effective in the binding event and plays a role in establishing the positive dynamic correlation of D3 and CR2, which is crucial for the stability of the complex. Moreover, MM-PBSA results further emphasized the hotspot residues and the effect of mutations in energy contribution of residues.

With this study, we were able to highlight the role of binding region residues (linker-1) but more interestingly, the dynamic relation between the distant sites of gp350 (linker-2 and D3 residues) and CR2, pointing out new target elements for further vaccine development strategies.

Funding

ENB acknowledges YOK 100/2000, PO acknowledges TUSEB project number 3454.

Availability of data and material

Data will made available on reasonable request.

Declaration of competing interest

The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.jmgm.2022.108196.

References

[1] A. Rickinson, E. Kieff, No title, in: D. Knipe, P.M. Howley (Eds.), Fields' Virology, fifth ed., Lippincott Williams & Wilkins, Philadelphia, 2007, pp. 2655–2700.
[2] E. Kieff, A. Rickinson, Epstein-Barr virus and its replication, in: D. Knipe, P. M. Howley (Eds.), Fields' Virology, fifth ed., Lippincott Williams & Wilkins, Philadelphia, 2007, pp. 2603–2654.
[3] P. Pellet, B. Roizman, The family: herpesviridae, a brief introduction, in: D. Knipe, P.M. Howley (Eds.), Fields' Virology, fifth ed., Lippincott Williams & Wilkins, Philadelphia, 2007, pp. 2479–2500.
[4] P. Boddu, A.S. Mohammed, C. Annem, W. Sequeira, SLE and non-hodgkin's lymphoma: a case series and review of the literature, Case Rep. Rheumatol. (2017) 1–7, https://doi.org/10.1155/2017/1658473, 2017.
[5] A. Lossius, J.N. Johansen, Ø. Torkildsen, F. Vartdal, T. Holmoy, Epstein-barr virus in systemic lupus erythematosus, rheumatoid arthritis and multiple sclerosis-association and causation, Viruses 4 (12) (2012) 3701–3730, https://doi.org/10.3390/v4123701.
[6] R.F. Miller, E.L. Jones, M.J. Duddy, M. Shahmanesh, Progressive intrathoracic lymphadenopathy: EBV associated non-Hodgkin's lymphoma, Sex. Transm. Infect. 78 (1) (2002) 13–17, https://doi.org/10.1136/sti.78.1.13.
[7] U.Y. Moon, et al., Patients with systemic lupus erythematosus have abnormally elevated Epstein-Barr virus load in blood, Arthritis Res. Ther. 6 (4) (2004) 295–302, https://doi.org/10.1186/ar1181.
[8] L.H. Wang, W.M. Wang, S.H. Lin, C.C. Shieh, Bidirectional relationship between systemic lupus erythematosus and non-Hodgkin's lymphoma: a nationwide population-based study, Rheumatol. (United Kingdom) 58 (7) (2019) 1245–1249, https://doi.org/10.1093/rheumatology/kez011.
[9] W. Bu, et al., Immunization with components of the viral fusion apparatus elicits antibodies that neutralize epstein-barr virus in B cells and epithelial cells, Immunity 50 (5) (May 2019) 1305–1316, https://doi.org/10.1016/j.immuni.2019.03.010, e6.
[10] L.Z. Mutsvunguma, et al., Identification of a novel neutralizing and two non-neutralizing epitopes on epstein-barr Virus gp350 Protein 1 2, bioRxiv (2018), https://doi.org/10.1101/302844.
[11] G. Szakonyi, et al., Structure of the Epstein-Barr virus major envelope glycoprotein, Nat. Struct. Mol. Biol. 13 (11) (2006) 996–1001, https://doi.org/10.1038/nsmb1161. Nov.
[12] G.R. Nemerow, C. Mold, V.K. Schwend, V. Tollefson, N.R. Cooper, Identification of gp350 as the viral glycoprotein mediating attachment of Epstein-Barr virus (EBV) to the EBV/C3d receptor of B cells: sequence homology of gp350 and C3 complement fragment C3d, J. Virol. 61 (5) (1987) 1416–1420, https://doi.org/10.1128/jvi.61.5.1416-1420.1987.
[13] J. Tanner, ', Y. Whang, J. Sample, A. Sears, E. Kieff12, Soluble Gp350/220 and Deletion Mutant Glycoproteins Block Epstein-Barr Virus Adsorption to Lymphocytes, 1988.
[14] M. Urquiza, R. Lopez, H. Patiño, J.E. Rosas, M.E. Patarroyo, Identification of three gp350/220 regions involved in epstein-barr virus invasion of host cells, J. Biol. Chem. 280 (42) (Oct. 2005) 35598–35605, https://doi.org/10.1074/jbc.M504544200.
[15] K. A. Young, A. P. Herbert, P. N. Barlow, V. M. Holers, and J. P. Hannan, "Molecular basis of the interaction between complement receptor type 2 (CR2/CD21) and epstein-barr virus glycoprotein gp350," J. Virol., vol. 82, no. 22, pp. 11217–11227, Nov. 2008, doi: 10.1128/jvi.01673-08.
[16] K.A. Young, X.S. Chen, V.M. Holers, J.P. Hannan, IsolatingtheEpstein-BarrVirusgp350/220BindingSiteon ComplementReceptorType2(CR2/CD21), J. Biol. Chem. 282 (50) (2007) 36614–36625.
[17] R. Longnecker, E. Kieff, J. Cohen, No title, in: D. Knipe, P.M. Howley, J.I. Cohen, D. E. Griffin, R.A. Lamb, M.A. Martin, V.R. Racaniello, B. Roizman (Eds.), Fields'

Virology, sixth ed., Lippincott Williams & Wilkins, Philadelphia, 2013, pp. 1898–1959.
[18] J. Tanner, J. Weis, D. Fearon, Y. Whang, E. Kieff, Epstein-Barr virus gp350/220 binding to the B lymphocyte C3d receptor mediates adsorption, capping, and endocytosis, Cell 50 (2) (1987) 203–213, https://doi.org/10.1016/0092-8674(87)90216-9.
[19] H.E. Gilbert, J.T. Eaton, J.P. Hannan, V.M. Holers, S.J. Perkins, Solution structure of the complex between CR2 SCR 1-2 and C3d of human complement: an X-ray scattering and sedimentation modelling study, J. Mol. Biol. 346 (3) (2005) 859–873, https://doi.org/10.1016/j.jmb.2004.12.006.
[20] A.E. Prota, D.R. Sage, T. Stehle, J.D. Fingeroth, The crystal structure of human CD21: implications for Epstein–Barr virus and C3d binding, Med. Sci. 99 (16) (2002) 10641–10646. www.pnas.org_cgi_doi_10.1073_pnas.162360499.
[21] J. P. Hannan et al., "Mutational analysis of the complement receptor type 2 (CR2/CD21)-C3d interaction reveals a putative charged SCR1 binding site for C3d," J. Mol. Biol., vol. 346, no. 3, pp. 845–858, Feb. 2005, doi: 10.1016/j.jmb.2004.12.007.
[22] J.M. Guthridge, et al., Epitope mapping using the X-ray crystallographic structure of complement receptor type 2 (CR2)/CD21: identification of a highly inhibitory monoclonal antibody that directly recognizes the CR2-C3d interface, J. Immunol. 167 (10) (2001) 5758–5766, https://doi.org/10.4049/jimmunol.167.10.5758.
[23] V. Sandhu, M. Quan, SLE and serum complement: causative, concomitant or coincidental? Open Rheumatol. J. 11 (1) (2017) 113–122, https://doi.org/10.2174/1874312901711010113.
[24] M.R. Sarrias, et al., Kinetic analysis of the interactions of complement receptor 2 (CR2, CD21) with its ligands C3d, iC3b, and the EBV glycoprotein gp350/220, J. Immunol. 167 (3) (2001) 1490–1499, https://doi.org/10.4049/jimmunol.167.3.1490.
[25] G. Szakonyi, Structure of complement receptor 2 in complex with its C3d ligand, Science (80-.) 292 (5522) (2001) 1725–1728, https://doi.org/10.1126/science.1059118. Jun.
[26] B. Y. C. A. Lowell, L. B. Klickstein, R. H. Carter, J. A. Mitchell, D. T. Fearon, and J. M. Ahearn, "C3dg BINDING SITES TO A COMMON DOMAIN ON COMPLEMENT RECEPTOR TYPE 2 MAPPING OF THE EPSTEIN-BARR VIRUS AND Epstein-Bars Virus (EBV), a herpesvirus that infects most humans , causes infectious mononucleosis and has been implicated in the pathogenesis of," vol. 170, no. December, 1989..
[27] B.D.R. Martin, A. Yuryev, K.R. Kalli, D.T. Fearon, J.M. Ahearn, in: From the "Graduate Program in Immunology and *D~'vision of Molecular and Clinical Rheumatology, vol. 174, Department of Medicine, The Johns Hopkins University School of Medicine, Baltimore, Maryland, December, 1991, 21205.
[28] J.I. Cohen, Vaccine development for Epstein-Barr virus, in: Advances in Experimental Medicine and Biology, vol. 1045, Springer New York LLC, 2018, pp. 477–493.
[29] J. Snijder, et al., An antibody targeting the fusion machinery neutralizes dual-tropic infection and defines a site of vulnerability on epstein-barr virus, Immunity 48 (4) (2018) 799–811, https://doi.org/10.1016/j.immuni.2018.03.026, e9.
[30] N. Widodo, , Veronica Margarecaesha Anyndita, N. Dluha, M. Rifa'i, K. Himmah, M.D. Wahyuningsih, Designing and overproducing a tandem epitope of gp350/220 that shows a potential to become an EBV vaccine, Heliyon 4 (3) (2018), e00564, https://doi.org/10.1016/j.heliyon.2018.e00564.
[31] A. Ali, et al., Immunoinformatic and systems biology approaches to predict and validate peptide vaccines against Epstein–Barr virus (EBV), Sci. Rep. 9 (1) (Dec. 2019), https://doi.org/10.1038/s41598-018-37070-z.
[32] J.I. Cohen, Epstein–barr Virus Vaccines," *Clinical and Translational Immunology*, vol. 4, John Wiley and Sons Inc., Jan, 2015, p. 1, https://doi.org/10.1038/cti.2014.27.
[33] W.T. Jackman, K.A. Mann, H.J. Hoffmann, R.R. Spaete, Expression of Epstein-Barr virus gp350 as a single chain glycoprotein for an EBV subunit vaccine, Vaccine 17 (7–8) (1999) 660–668, https://doi.org/10.1016/S0264-410X(98)00248-5.
[34] D.G. Van Zyl, J. Mautner, H.J. Delecluse, Progress in EBV Vaccines," *Frontiers in Oncology*, vol. 9, FEB. Frontiers Media S.A., 2019, https://doi.org/10.3389/fonc.2019.00104.
[35] X. Cui, et al., A novel tetrameric gp3501-470 as a potential Epstein-Barr virus vaccine, Vaccine 31 (30) (2013) 3039–3045, https://doi.org/10.1016/j.vaccine.2013.04.071.
[36] M. Kanekiyo, et al., Rational design of an epstein-barr virus vaccine targeting the receptor-binding site, Cell 162 (5) (Aug. 2015) 1090–1100, https://doi.org/10.1016/j.cell.2015.07.043.
[37] H.M. Berman, et al., The protein data bank, Acta Crystallogr. Sect. D Biol. Crystallogr. 58 (6 I) (2002) 899–907, https://doi.org/10.1107/S0907444902003451.
[38] G.C.P. Van Zundert, et al., The HADDOCK2.2 web server: user-friendly integrative modeling of biomolecular complexes, J. Mol. Biol. 428 (4) (2016) 720–725, https://doi.org/10.1016/j.jmb.2015.09.014.
[39] "The PyMol molecular graphics system, Version #, Schrodinger, LLC.".
[40] S. Páll, M.J. Abraham, C. Kutzner, B. Hess, E. Lindahl, Tackling exascale software challenges in molecular dynamics simulations with, GROMACS (2015), https://doi.org/10.1007/978-3-319-15976-8.
[41] M.J. Abraham, et al., GROMACS: high performance molecular simulations through multi-level parallelism from laptops to supercomputers, Software 1 (2) (Sep. 2015) 19–25, https://doi.org/10.1016/J.SOFTX.2015.06.001.
[42] W.L. Jorgensen, D.S. Maxwell, J. Tirado-Rives, Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids, J. Am. Chem. Soc. 118 (45) (Nov. 1996) 11225–11236, https://doi.org/10.1021/ja9621760.
[43] G.A. Kaminski, R.A. Friesner, J. Tirado-Rives, W.L. Jorgensen, Evaluation and reparametrization of the OPLS-AA force field for proteins via comparison with accurate quantum chemical calculations on peptides †, J. Phys. Chem. B 105 (28) (Jul. 2001) 6474–6487, https://doi.org/10.1021/jp003919d.
[44] A. Amadei, A.B.M. Linssen, H.J.C. Berendsen, Essential dynamics of proteins, Proteins Struct. Funct. Genet. 17 (4) (Dec. 1993) 412–425, https://doi.org/10.1002/prot.340170408.
[45] O. Serçinoğlu, P. Ozbek, gRINN: a tool for calculation of residue interaction energies and protein energy network analysis of molecular dynamics simulations, Nucleic Acids Res. 46 (W1) (Jul. 2018) W554–W562, https://doi.org/10.1093/nar/gky381.
[46] A. Kumar, G. Srivastava, A.S. Negi, A. Sharma, Docking, molecular dynamics, binding energy-MM-PBSA studies of naphthofuran derivatives to identify potential dual inhibitors against BACE-1 and GSK-3β, J. Biomolecular Struct. Dyn. 37 (2) (Jan. 2018) 275–290, https://doi.org/10.1080/07391102.2018.1426043.
[47] S. Genheden, U. Ryde, The MM/PBSA and MM/GBSA methods to estimate ligand-binding affinities, Expet Opin. Drug Discov. 10 (5) (May 2015) 449, https://doi.org/10.1517/17460441.2015.1032936.
[48] M.S. Valdés-Tresanco, M.E. Valdés-Tresanco, P.A. Valiente, E. Moreno, Gmx_MMPBSA: a new tool to perform end-state free energy calculations with GROMACS, in: J. Chem. Theory Comput., vol. 17, PDF, Oct. 2021, pp. 6281–6291, https://doi.org/10.1021/ACS.JCTC.1C00645/SUPPL_FILE/CT1C00645_SI_001, 10.
[49] C. Dominguez, R. Boelens, A.M.J.J. Bonvin, HADDOCK, A protein-protein docking approach based on biochemical or biophysical information, J. Am. Chem. Soc. 125 (7) (2003) 1731–1737, https://doi.org/10.1021/ja026939x.
[50] J.E. Tanner, et al., Peptides designed to spatially depict the epstein-barr virus major virion glycoprotein gp350 neutralization epitope elicit antibodies that block virus-neutralizing antibody 72A1 interaction with the native gp350 molecul, J. Virol. 89 (9) (2015) 4932–4941, https://doi.org/10.1128/JVI.03269-14.
[51] L.S. Sitompul, N. Widodo, M.S. Djati, D.H. Utomo, Epitope mapping of gp350/220 conserved domain of epstein barr virus to develop nasopharyngeal carcinoma (npc) vaccine, Bioinformation 8 (10) (2012) 479–482, https://doi.org/10.6026/97320630008479.
[52] S. Soner, P. Ozbek, J.I. Garzon, N. Ben-Tal, T. Haliloglu, DynaFace: discrimination between obligatory and non-obligatory protein-protein interactions based on the complex's dynamics, PLoS Comput. Biol. 11 (10) (2015) 1–19, https://doi.org/10.1371/journal.pcbi.1004461.
[53] U. Emekli, D. Schneidman-Duhovny, H.J. Wolfson, R. Nussinov, T. Haliloglu, HingeProt: automated prediction of hinges in protein structures, Proteins Struct. Funct. Bioinforma. 70 (4) (Sep. 2007) 1219–1227, https://doi.org/10.1002/prot.21613.