DYNAMIC NATURE OF H1N1 INFLUENZA VIRUS HEMAGGLUTININ

by

Zeynep Erge Akbaş

B.S., Chemical Engineering, Middle East Technical University, 2011

Submitted to the Institute for Graduate Studies in
Science and Engineering in partial fulfillment of
the requirements for the degree of
Master of Science

Graduate Program in Chemical Engineering
Boğaziçi University

2014

ACKNOWLEDGEMENTS

I would like to express my sincere thanks to my thesis supervisor Prof. Türkan Haliloğlu; without her motivation, support and understanding, I would never achieve such success in my studies or in my life. It was a great pleasure and honor for me to work with her. I am very grateful to my thesis committee members; Prof. Pemra Doruker Turgut and Assist. Prof. Pemra Özbek for spending time on my thesis and making valuable recommendations.

I would like to thank specially to Fidan Sümbül for mentoring me through my studies and answering even the silliest questions that I keep asking during the simulations and Arzu Uyar for helping me in every problem that I confronted with and also for being like a sister to me since the beginning.

Special thanks to İrem Şen, Begüm Alaybeyoğlu and Cihan Kaya, the life with you was so much fun in our "Dunkin meetings" and beyond. I would like to thank to the members of Polymer Research Center, Burçin Acar, Şefik Kerem Ovalı, Melis Yıldırım, Serdar Özsezen, Burcu Aykaç Fas, Zeynep Kürkçüoğlu Soner, Seren Soner, Gökçe Ezeroğlu, Nurver Tezcan, Canan Dedeoğlu and Yüksel Gülten. I would like to thank everyone in Chemical Engineering Department from staff to student and faculty members for their support and help. I would like to thank especially to Hilal Fidan Acar, Sedef Özçelik, Emre Can Bilir, Serhat Erşahin, Özgür Yaşar Çağlar, Can Ekici and Emre Demirel.

I would like to express my very special gratitude to Ezgi Ezdar and Ayşegül Seyrek for being like a sister to me for so many years, Engin Buz and Özgün Gözmen for being so supportive and motivating with their lovely Bisou, and to Gizem Akbaş for her motivation and smile every day in our house.

And lastly I cannot thank enough to show my love and gratitude to my family. This thesis is dedicated to my family, without them nothing would be possible or would have meaning.

ABSTRACT

DYNAMIC NATURE OF H1N1 INFLUENZA VIRUS HEMAGGLUTININ

Dynamics of antigenic and host receptor binding behavior of a virus binding protein can be altered via mutations that lead to escape from immune recognition. The main focus of this study is to investigate the dynamic behavior of Influenza A virus Hemagglutinin (HA) of the 2009 human pandemic H1N1 in comparison to other H1N1 strains, by Molecular Dynamics (MD) simulations. HA is the host receptor binding protein that has a key role in virus-host interaction. The dynamic behavior of HA of two Influenza A (H1N1) variants and their mutants at residue position 149 were studied comparatively between the structures to search for the effect of this position between the two variants (K149 and R149) and the mutations (K149R and R149K) with the emphasis on the Receptor Binding Domain (RBD) of HA. The residue 149 is chosen considering its importance due to being a discriminating position between classical (R149) and 2009 pandemic (K149) swine viruses, although it is directly associated with neither antigenic nor sialic binding sites. Residue 149 is also identified as one of the hinge positions in both of the variants within their wild and mutant types, implying an allosteric dynamic affect. The position 149 is found to be highly affecting the dynamics of RBD at 130 loop and 220 loop that includes the receptor binding site (RBS) residues 134-138 and 221-228 as well as antigenic binding sites (Sa and Ca), which is particularly strong for K149. The structures of K149 and R149K displaying more restricted mobility lead to a higher cooperativity and efficient global HA behavior. Overall, the results imply that allosteric dynamic effect of 149 may affect virulence as a result of altering dynamic behavior in RBD of HA.

ÖZET

H1N1 INFLUENZA A VİRÜSÜNÜN HEMAGLÜTİNİN PROTEİNİNİN DİNAMİK YAPISI

Bir virüsün konak hücreye bağlanma proteinin konak hücre ve antijen rezidülerinin dinamiği, mutasyonlar ile immün tanınmasından kaçacak şekilde değişebilmektedir. Bu çalışmanın ana amacı, 2009 yılı Influenza A pandemik virüsü Hemaglütinin (HA) proteininin Moleküler Dinamik (MD) simülasyonlarını kullanarak diğer H1N1 soyları ile karşılaştırmalı olarak dinamik yapısının incelenmesidir. HA konak reseptör hücreye bağlanma proteinidir ve aynı zamanda virüs-konak etkileşiminde önemli bir rol oynar. İnfluenza A virüs HA proteininin iki varyasyonu ve bunların rezidü 149'da mutasyon uygulanmış yapılarının dinamik davranışları, reseptör bağlanma bölgesi üzerine yoğunlaşarak ve bu mutasyonun yapılar üzerindeki karşılıklı etkileri araştırılarak çalışıldı. Rezidü 149, direk konak hücre bağlanma ya da antijen rezidüsü olmamasına rağmen klasik (R149) ve 2009 pandemik (K149) domuz gribi arasında ayırt edilen bir pozisyonda olmasından kaynaklı öneminden dolayı seçilmiştir. Rezidü 149 doğal yapı (wild-type) ve mutant yapı dinamiği koordine eden dayanak (hinge) noktalarından biri olarak belirlenmiştir. Rezidü 149 reseptör bağlanma bölgesinde yer alan 130 (rezidü 134-138) ve 220 (221-228) döngülerinin ve aynı zamanda antijen bağlanma bölgelerinin (Sa ve Ca) dinamiğini alosterik olarak etkileyebilmektedir. Bu etki K149 varyasyon yapılarında daha kuvvetli olarak ortaya çıkmaktadır. Daha kısıtlı bir hareketlilik gösteren K149 ve R149K yapıları, daha yüksek kooperativiteye ve etkili global bir HA davranışına yol açmaktadır. Genel olarak sonuçlar 149' un HA'nın konak hücre reseptörlerine bağlanma dinamiğini etkileyerek viral etkinliği arttırmada önemli rol oynayan etkenlerden biri olduğunu göstermektedir.

TABLE OF CONTENTS

LIST OF FIGURES

LIST OF TABLES

LIST OF SYMBOLS

a	Acceleration of the particle
C-α	Alpha Carbon
dt	Timestep
dr_i	Positional derivative
F	Force exerted on the particle
m	Mass of the particle
N	Number of atoms
nm	Nanometer
ns	Nanosecond
r_i	Atomic coordinates of the particle i
t	Time
V	Potential energy
α	Alpha
Å	Angstrom

LIST OF ACRONYMS/ABBREVIATIONS

3D	Three dimensional
A	Alanine
Arg	Arginine
C	Carbon
CC	Cross-correlations
E	Glutamic acid
G	Glycine
HA	Hemagglutinin
His	Histidine
K	Lysine
L	Leucine
MD	Molecular dynamics
MSF	Mean square fluctuations
MOE	Molecular Operating Environment
NA	Neuraminidase
P	Proline
PDB	Protein Data Bank
R	Arginine
RBD	Receptor binding domain
RBS	Receptor binding site

RMSD	Root mean square deviation
S	Serine
SA	Sialic acid
T	Threonine
VMD	Visual molecular dynamics

1. INTRODUCTION

1.1. Biological Background

Viruses are small (10-400 nm) acellular, non-living, infectious biological organisms that need host cells to deliver their genetic information and increase their number; therefore viral entry into the host cell is needed (Grove and Marsh, 2011). Virus entry to the cell is not directly achieved; viruses need to use their specific cell surface molecules, namely the receptor binding sites for the entry. Thus, receptor binding domain (RBD) is a target domain for the antibodies to bind so as to block the cell infection.

Viruses enter the cells by first interacting with the host cell receptors. Non-enveloped viruses penetrate into the cell membrane either by lysing a membrane or by creating a pore-like structure. The projections or indentations in the capsid surface of the non-enveloped viruses are the binding structures for the receptors (Smith and Helenius, 2004). For the enveloped viruses, interaction of the enveloped virus binding proteins with the receptors at the host cell surface is needed. Enveloped virus first binds to the receptor at the surface of the target cell, and then the fusion of the viral envelope with the cell membrane occurs (Morizono and Chen, 2011).

The virus is taken into the plasma membrane by the fusion, and then conformational changes occur causing the viral genome to be uncoated into the cytoplasm of the host cell. The next step is the release of the nucleocapsids and initiation of the transcription (Schneider-Schaulies, 2000).

1.2. Influenza A Virus

Influenza A Virus is an enveloped (-) ssRNA virus (having size of 80-120 nm) belonging to the Orthomyxoviruses family, and Influenza has led to severe diseases that has different variants within years. Influenza A Virus contains different components of which can be seen in Figure 1.1. The abbreviations in the Figure corresponds to Hemagglutinin for HA and Neuraminidase for NA. Influenza A viruses are divided into sixteen H (H1-H16) and nine N (N1-N9) subtypes depending on their antigenicity. Only three HA strains; H1, H2, and H3, are defined as infective to humans. Influenza virus

needs a cell surface protein to bind to the host cell, since it is an enveloped virus (Mudhakir and Harashima, 2009).

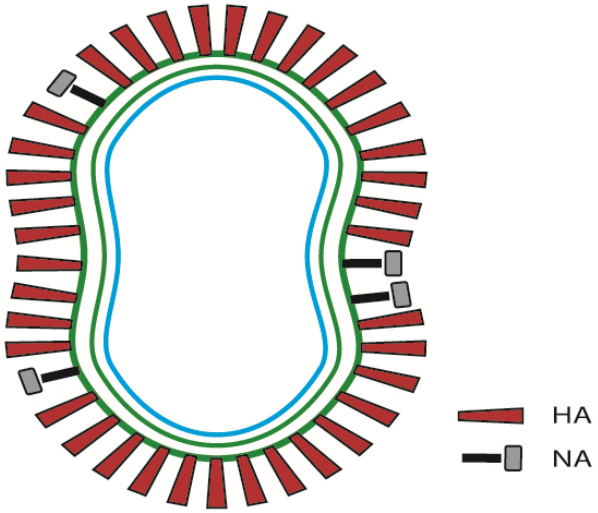

Figure 1.1. Cartoon representation of Influenza Virus.

1.3. Hemagglutinin (HA)

Hemagglutinin (HA) is the receptor binding glycoprotein of Influenza virus and Hemagglutinin (HA) itself consists of two subunits; HA1 and HA2. HA1 is responsible for virus binding and HA2 for fusion. Influenza virus binds to cells via an interaction between HA1 and sialic acid (SA) on the glycoproteins of the host cell. The representation of the virus binding through Hemagglutinin protein is given in Figure 1.2.

Influenza A Virus Hemagglutinin protein is a trimeric glycoprotein, composed of six chains. The trimer structures are H, I, J, K, L and M chains of 1930 H1 Influenza Virus Hemagglutinin (PDB ID: 1RVT), and A, B, C, D, E and F chains of 2009 H1N1 Influenza Virus Hemagglutinin (PDB ID: 3LZG) in their wild and mutant types are simulated in the present work. Table 1.1 shows the number of residues in each chain with corresponding chain identifiers.

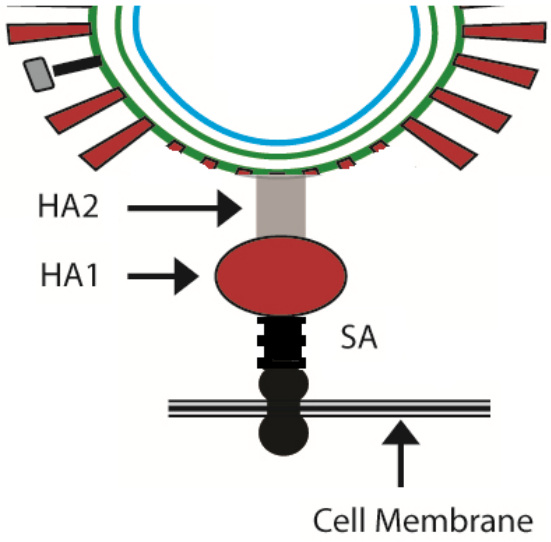

Figure 1.2. Influenza Virus Hemagglutinin binding to host cell sialic acid receptor.

Table 1.1. Number of residues in each chain for each structure.

1RVT					
H: 5-327	I: 501-660	J: 5:327	K: 501-660	L: 5-327	M: 501-660
3LZG					
A: 11-325	B: 1-175	C:11-325	D:1-171	E:11-325	F:1-171

1.4. Literature Survey

The crystal structures of the two Influenza A Virus Hemagglutinin proteins used in this study are obtained from Protein Data Bank; 1930 H1 Influenza Virus Hemagglutinin (HA) (PDB ID: 1RVT) (A/swine/Iowa/15/1930) and 2009 H1N1 Influenza Virus Hemagglutinin (HA) (PDB ID: 3LZG) (H1 A/California/04/2009 H1N1 strain). The two proteins are related due to their individual similarity to 1918 pandemic Influenza Virus (Tumpey *et al*., 2004 and Sriwilaijaroen and Suzuki, 2012).

Hemagglutinin (HA) is a trimeric glycoprotein consisting of three host receptor binding sites and five antigenic binding sites. The host receptor binding site (RBS) regions are 130 loop residues 134-138), 190 helix (residues 188-194) and 220 loop (221-228). The

antibodies have the ability to neutralize viral infectivity of HA by preventing its interaction with sialic acid receptors on the host cell (Edwards and Dimmock, 2001 and Hensley *et al*., 2009). Antigenic binding sites that the antibodies bind to neutralize the virus, is sited around the receptor binding pocket of HA. The antigenic binding sites that are determined on HA can be listed as follows; Sa (residues 128–129, 156–160, 162–167), Sb (residues 187–198), Ca_1 (residues 169–173, 206–208, 238–240), Ca_2 (residues 140–145, 224– 225), and Cb (residues 74–79) (Skehel and Wiley, 2000).

Antigenic and receptor binding sites may be distinct regions on the viruses. Recent studies have shown the effect of the mutations at antigenic sites and at some other sites distant from the receptor binding sites (RBS) on receptor binding properties (Soundararajan *et al*., 2011). Nevertheless, the molecular and structural underpinnings of this allosteric interaction have not been completely understood.

Figures 1.3 and 1.4 show the trimer Hemagglutinin protein, revealing host receptor binding residues as green colored. Influenza Virus Hemagglutinin needs to interact with sialic acid on the glycoproteins of the host cell, and the binding occurs at the top of the HA molecule. The distance of the site, where the neutralizing antibody binds, to the host receptor binding site is revealed in the complex structure of HA in PDB ID 1QFU for 1930 swine virus and in PDB ID 3UBE for 2009 swine virus (Skehel and Wiley, 2000). The PDB structure 3UBE is the sialic acid bound form of 3LZG protein.

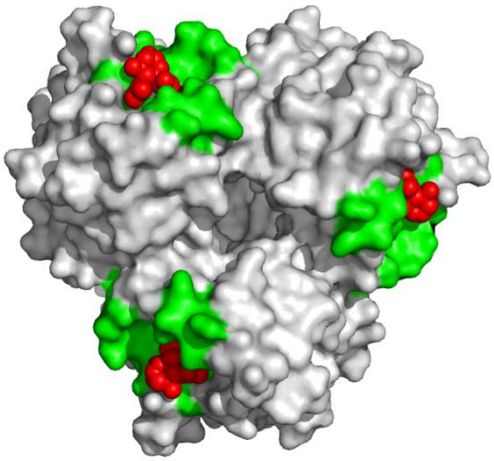

Figure 1.3. 3UBE protein with RBS’s and sialic acids highlighted (RBS: green, SA: red).

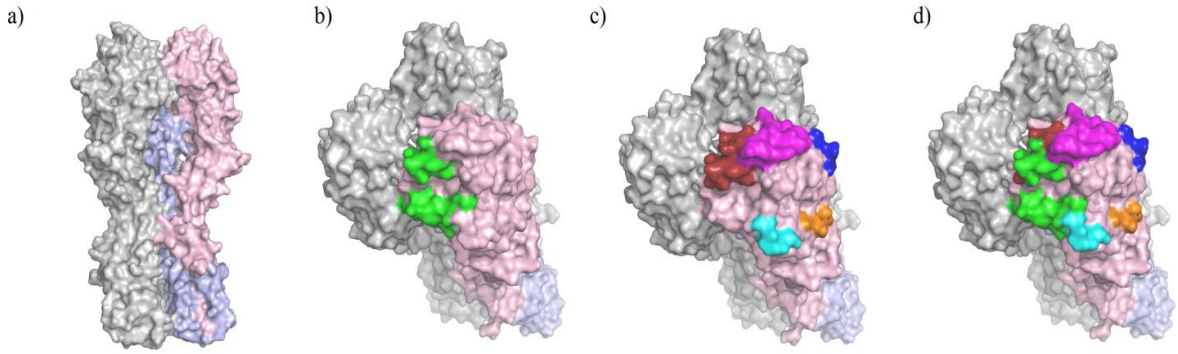

Figure 1.4. 3D structure of the trimer 1RVT Hemagglutinin protein showing RBS: green and antigenic binding residues (Cb: orange, Sa: magenta, Sb: brown, Ca_1: blue, Ca_2: cyan) shown as surface in PYMOL.

Protein's function can be controlled by structural and dynamic perturbations, such as binding to a small ligand, another protein, change in pH etc., at allosteric sites far away from the functional site. These sites have the ability to activate or inactivate the protein's functions. Antibodies; in this case, if not aiming the receptor binding sites directly, can be thought as the ligands binding to the receptors allosterically and causing conformational changes leading to differences in the dynamics of binding sites in the host receptor binding domain (Gao and Jacobson, 2006).

Substitutions of amino acids at antigenic sites during infection are associated with antigenic change. In addition to that, there are other residues that are determined to have an effect on the binding specificity of the HA. The effect of the mutation on the binding specificity of HA investigated in many studies (Hensley *et al*., 2009 and Meroz *et al*., 2009). Previous studies have shown that mutations on RBD have effect on structural conformation of HA protein. It is known that changes in hemagglutinin receptor binding site affect influenza binding to the host receptor by differentiating the recognition of sialic acids by the receptors of influenza virus.

Various experimental and computational studies concentrated on increase of virulence or change in binding specificity of HA as a result of mutation in RBD (Ye *et al*.,

2010 and Skehel and Wiley, 2000 and Nicholls *et al*., 2008 and Hensley *et al*., 2009 and Martin *et al*., 1998). There are experimental studies exploring the effect of mutation upon host cell binding as well as binding specificity of the receptors (Kumari *et al*., 2007 and Zou *et al*., 2013 and Rogers *et al.*, 1983). The 1918 variant HA is also investigated in terms of receptor binding specificity as a result of amino acid substitution (Glaser *et al*., 2005). The study of (Gamblin *et al*., 2004) on 1930 swine virus in complex with human receptor reveals potentially important residues in the RBS; in 190 helix (190, 193) and 220 loop (222, 225). An additional experimental study on H3 subtype of Influenza HA shows that mutation at residue 226 (220 loop RBS) may change the sialic acid preference on the host cell (Nicholls *et al*., 2008).

On the other hand, computational methods for examination of the HA are also available in the literature (Isın *et al*., 2002 and Soundararajan *et al*., 2011 and Tharakaraman *et al*., 2013). The changes and also similarities in the sequences of various HA structures are given in several studies (Gamblin *et al*., 2004). In sequence analysis of HA protein, conserved and differentiating residues of 1930 swine virus HA and 2009 swine virus HA were identified (Tharakaraman *et al*., 2013 and Sriwilaijaroen and Suzuki, 2012). Effect of mutation for different HA variants are studied computationally in various studies (Gamblin *et al*., 2004 and Nicholls *et al*., 2008). The most discriminating residues are studied and given in a ranked order, proposing the possible changes in the host receptor binding of the virus as well (Meroz *et al*., 2011). In the work of (Meroz *et al*., 2011), residue 149 is found as one of the important discriminating positions in the sequence analysis which is also found effective on the virulence of HA protein upon mutation. Evolving out of the latter study, HA proteins of 1930 and 2009 swine viruses are considered in the present work. Residue 149 is chosen as the position to apply mutation and observe its effect on the dynamics of HA. The changes in the structural conformation and dynamics of RBD as a result of R149K (on 1RVT) and K149R (on 3LZG) mutations with respect to wild type H1N1 HA are studied.

2. MATERIALS AND METHODS

2.1. Influenza A Virus Hemagglutinin Protein

The main focus of this study is to investigate the dynamic behavior of HA which is the host receptor binding protein of the virus by Molecular Dynamics (MD) simulations. The two variants of HA and their mutants at residue position 149 were simulated and the dynamic behavior was analyzed comparatively between the structures to search for the effect of mutation with the emphasis on the RBD of HA. Residue 149 of HA1 for 1RVT is Arginine (R) while it is Lysine (K) for 3LZG. The mutations introduced are R149K and K149R.

2.2. Molecular Dynamics (MD) Simulations

Molecular Dynamics is a computational method that produces snapshots of the system along the simulation time, from which the time dependent conformational changes of the system can be observed (Jiao and Parker, 2012). In Molecular Dynamics Simulations, time dependent behavior of the system is calculated. MD simulations provide detailed information on the fluctuations and conformational changes of proteins and nucleic acids.

Newton's equation of motion is used in MD Simulations; the equation relates the derivative of the potential energy to the positional change as a function of time. Newton's equation of motion is solved for each atom in the protein with respect to the chosen force field. A single atom is affected by the potential energy function of every atom in the system.

$$F = ma \tag{2.1}$$

where F is the force exerted on the particle, m is the mass of the particle and a is the acceleration of the particle.

The force can be also written as the gradient of the potential energy,

$$F = -\nabla V \tag{2.2}$$

By combining 2.1 and 2.2, equation in 2.3 is obtained;

$$F_i = -\frac{dV}{dr_i} = m_i \frac{d^2 r_i}{dt^2} \tag{2.3}$$

where r_i is the atomic coordinates of the particle i, t is time and V is the potential energy of the system.

Newton's equation of motion can then relate the derivative of the potential energy to the changes in position as a function of time. MD simulation of a protein includes three stages: minimization, equilibration and analysis of the trajectory. Energy minimization is required for starting dynamics simulations and obtaining initial parameters of the system. Before all, the building step in VMD is required (Humphrey *et al*., 1996). The steps of MD simulation are given below;

- Build the system in VMD (PDB and PSF)

(i) Solvation step

(ii) Ionization step

(iii) Minimization step using NAMD

- Simulate the system by NAMD
- Analyze the MD simulation trajectory.

2.2.1. Pre-process of the Proteins Before MD Simulations

The first information that is obtained for a protein (residue information with atomic coordinates, chain identifiers) is from the PDB text file. At the first step of the pre-process,

the PDB text file is investigated to search for the alternative coordinates for any residue; and if there is, those alternatives which are designated with A, B, C etc. next to the residue number are deleted. Additionally, ligands and metals if any are also deleted from the structure.

After the residues having more than one coordinate are deleted, the rest are opened in MOE (Molecular Operating Environment) and protonation step is applied. For this, wild type of the proteins are opened in MOE; and at 310 K temperature, protonation is occurred. Protonation step is necessary for determination of the protonated states of the Histidine (HIS) residues; whether it is HSD, HSE or HSP depending on the location of the side chain. The determined states are inserted with the corresponding residues in PDB text file and saved separately. From that point, protonated PDB's are used.

The step after this pre-process is the PSF file formation in VMD, VMD has a built in function, PSF Builder under Extensions/Modeling. In the PSF Builder function, CHARMM22 (topology file for proteins) topology file is used; a topology file contains all possible conformations of atoms, bonds and angles. Then PSF file which is protein specific is generated.

The work in this thesis requires mutation on a specific residue of the protein; for mutation VMD built in function Mutate Residue under Extensions/Modeling is used. The wild type PDB and PSF files are browsed and the 149^{th} residues corresponding to Arginine (R) for 1RVT and it Lysine (K) for 3LZG is mutated to Lysine (K) and Arginine (R) respectively. This tool generates new PSF file for mutant types.

After this step, water box formation is accomplished in VMD. Mutant and wild type proteins are separately opened in VMD and Solvation Box tool is used. Box padding sizes are determined as 10 Å in cubic size for each coordinate axis and water is used as solvation material, then the systems are tried to be neutralized by using Na^+ and Cl^- ions with VMD 1.9.1. After the preparation steps are done, minimization step is followed.

2.2.2. Energy Minimization Step

In the minimization step, ionized proteins in water box are used as input files (PDB and PSF files). Each monomeric system is minimized for 10000 steps at 310 K temperature using Langevin Dynamics; and for trimer structures it is determined as 20000. The minimization step is performed to obtain restart coordinates and velocities for the simulation.

2.2.3. Simulation Parameters for the System

Molecular dynamics simulations were performed using NAMD 2.8. After the minimization step is performed, equilibration step is started to simulate the protein in a given time and observe the dynamic behavior. Both the monomeric and trimer structures are simulated for 100 ns (50,000,000 steps) with their mutant and wild types of each variant. Therefore, a total number of 8 are simulated with NAMD 2.8 and resulting trajectories are analyzed.

The analysis of the trajectories starts with alignment of the protein with respect to the initial frame of the trajectory which corresponds to the minimized structure at the beginning of 100 ns simulations. Then aligned proteins are further analyzed by calculating RMSD and MSF. The dynamic behaviors of the structures are observed from MSF and cross-correlation analysis; therefore calculation procedures are simply summarized in the following.

2.2.4. Root Mean Square Deviation (RMSD) Calculation Procedure

Root Mean Square Deviation is the representation of the conformational change of the residues within time. The first step of RMSD analysis is the alignment of the backbone atoms of each conformation to the backbone atoms of the initial minimized conformation of the trajectory. Each structure is aligned, and then RMSD is calculated for the overall time span of MD simulation; giving one RMSD value for each time step or frame. The equation is,

$$RMSD(t_i) = \sqrt{\frac{\sum_{i=1}^{N}(r_i(t_i) - r_i(t_0))^2}{N}} \quad (2.4)$$

where N is number of atoms, r_i is the position of atom i at time t_i, t_0 is the initial time step (minimized structure).

2.2.5. Mean Square Fluctuation (MSF) Calculation Procedure

The physical meaning of the mean square fluctuations (MSF) is the mobility of each residue; the extent of the change of the position of each residue in time from its mean position. MSF calculation is performed on the equilibrated part of the MD trajectory considering only the C-α atoms of minimized structures. The minima and maxima regions of the MSF plots give idea about the conformational structure of the proteins.

$$MSF(i) = \frac{\sum_{t=1}^{T}(r_i(t) - \overline{r_i})^2}{T} \quad (2.5)$$

where T is equilibration time, r_i is the position of atom i at time$t$, $\overline{r_i}$ is the average position.

2.2.6. Cross-Correlation Calculation Procedure

The correlation between residue fluctuations (C-α atoms) of the structure is given as cross-correlation maps in Figure 3.17 for trimer structures. The cross-correlations are normalized correlations and vary within the range [-1, 1], from color red to blue; from the most postive to most negative correlations.

$$MSF(i) = \frac{\sum_{t=1}^{T}(r_i(t) - \overline{r_i})^2}{T} \quad (2.6)$$

where C_{ij} is cross-correlation between the residues i and j, ΔR_i and ΔR_j are the fluctuations of position vectors of residues i and j.

3. RESULTS AND DISCUSSION

Influenza A Virus leads to a severe disease that has different variants within years. The types that are being investigated in this thesis are 1930 H1 Influenza Virus Hemagglutinin (HA) (PDB ID: 1RVT) (A/swine/Iowa/15/1930) and 2009 H1N1 Influenza Virus Hemagglutinin (HA) (PDB ID: 3LZG) (H1 A/California/04/2009 H1N1 strain). The two proteins are related due to their individual similarity to 1918 pandemic Influenza Virus. 1918 pandemic Influenza Virus is in close genetic relationship to swine influenza virus (Sw/Iowa/30, H1N1), and 1918 pandemic virus has also similar HA sequence pattern to 2009 pandemic virus (Tumpey *et al*., 2004 and Sriwilaijaroen and Suzuki, 2012). Despite having arisen at different years, both swine viruses caused a pandemic in 1930 and in 2009. Both proteins are receptor binding proteins of swine flu epidemics and they also have similar protein sequences. The BLAST results of FASTA sequences the two proteins show 83% identity in their HA1 regions (FASTA result can be seen in the Appendix B).

Residue 149 of HA1 1RVT is Arginine (R) while it is Lysine (K) for 3LZG. Residue 149 differs from its ancestors first time at 3LZG (swine pandemic in 2009). In tertiary structure of the proteins, 149^{th} residue is sited in close contact to Ca_2 antigenic site (residue 140-145) and Cb antigenic site (residue 74-79) of the Hemagglutinin protein. In recent study, residue 149 was predicted to be a discriminating position between classical and pandemic swine viruses and mutation at this position is announced to change the binding of the virus to erythrocytes (Meroz *et al*., 2011).

The main focus of this study is to investigate the dynamic behavior of HA which is the host receptor binding protein of the virus by Molecular Dynamics (MD) simulations. The two variants of HA and their mutants at residue position 149 were simulated and the dynamic behavior was analyzed comparatively between the structures to search for the effect of mutation with the emphasis on the RBD of HA. The residue 149 is chosen considering its importance due to being a key position between classical and pandemic swine viruses.

HA functions as a trimer. The molecular simulations were performed for the trimer HA and also for the monomeric HA. Although the monomeric state is not a functional state, it is investigated here to analyze the extent of the intrinsic dynamics and effect of

trimerization on the RBD behavior. The Results and Discussion section below is mainly divided into two sections; the analysis of the MD simulations for the monomeric structure and for the trimer structure.

3.1 Monomeric Hemagglutinin

This subsection presents the results of the 100 ns MD simulations of the monomeric HA. The monomer structure is the H and I chains (HA1 and HA2 domains) of 1930 H1 Influenza Virus Hemagglutinin (PDB ID: 1RVT), and A and B chains (HA1 and HA2 domains) of 2009 H1N1 Influenza Virus Hemagglutinin (PDB ID: 3LZG) in their wild and mutant types. The mutation is performed on residue 149 for both variants; R149K is implemented on 1RVT wild type while the reverse, which is K149R, is implemented on 3LZG wild type.

3.1.1. Root Mean Square Deviation (RMSD)

Root Mean Square Deviation (RMSD) provides a measure for the representation of the conformational change of the structure within the simulation time window from the initial X-ray crystal structure.

In Figure 3.1, RMSD results are plotted for each variant. It is observed that mutant 3LZG has the highest RMSD value that is over 7 Å. The average RMSD value of the last 30 ns of the conformations is calculated, and given in Table A.1 in Appendix A. Table A.1 (in Appendix A) shows that mutant types of both variants show higher RMSD values than the wild type for the time span of 70 ns to the end of simulation time. This behavior can also be observed in the plots given in Figure 3.1.

Further, RMSD results show the equilibrated phase for the simulated protein in a given simulation time. This region is taken as 30 ns time step. The further analysis is thus taken from 30 ns to 100 ns time span of overall simulation time.

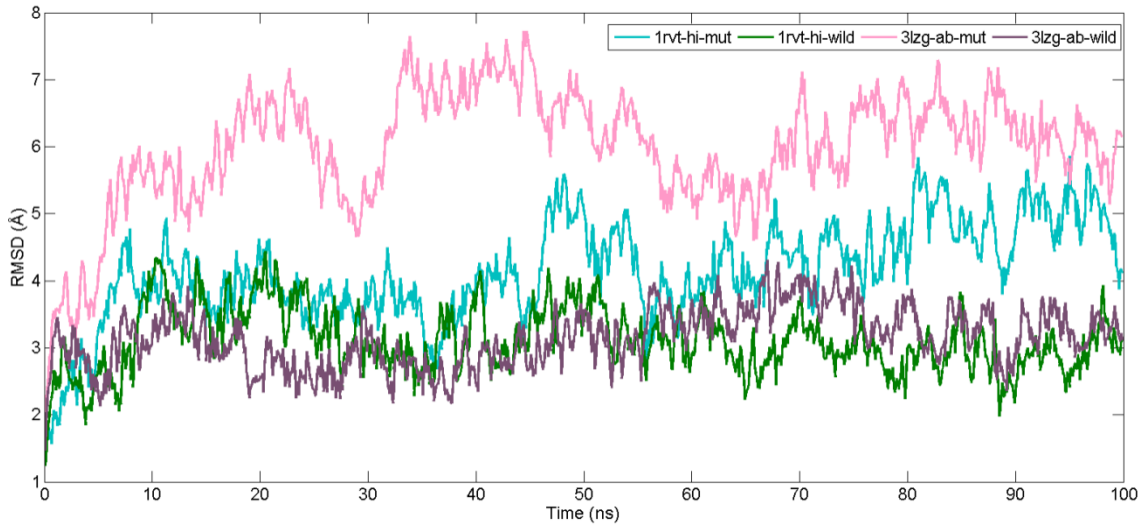

Figure 3.1. RMSD values for monomeric 1RVT and monomeric 3LZG.

If the structures are aligned in PYMOL, it would be easy to reveal the regions that lead to significant changes in the conformations with increasing RMSD values from the initial state. For this reason, last frame of each protein is obtained in VMD from the trajectory considering only the backbone atoms, and then aligned to the original PDB file. The original PDB files were modified; only chains constituting the monomeric subunit are taken into consideration in alignment.

Figures 3.2 and 3.3 show the alignment results for 1RVT and 3LZG in their wild and mutant types with respect to the original PDB files. Last snapshot of each trajectory is aligned to its original PDB file. PYMOL Alignment Tool gives RMSD value after aligning the two given structures. Here, it can be seen that mutant type of each variant differs from their initial original PDB file more than the wild type. RMSD values confirm this behavior as mutant type RMSD is higher (Table A.2 in Appendix A).

Finally, the mutant type of each variant are aligned to their own wild type at the end of simulations in Figure 3.4. Figure 3.4 shows that 1RVT mutant and wild type structures stay at close positions at the end of their monomeric simulations except at the tail of the protein. On the other hand, 3LZG is distinctly different in its mutant and wild types and moving apart; which is also seen from its high RMSD value in Table A.2 in Appendix A. The divergence is mostly observed at residues 65 to 100 of HA2 domain as well as in its RBD domain.

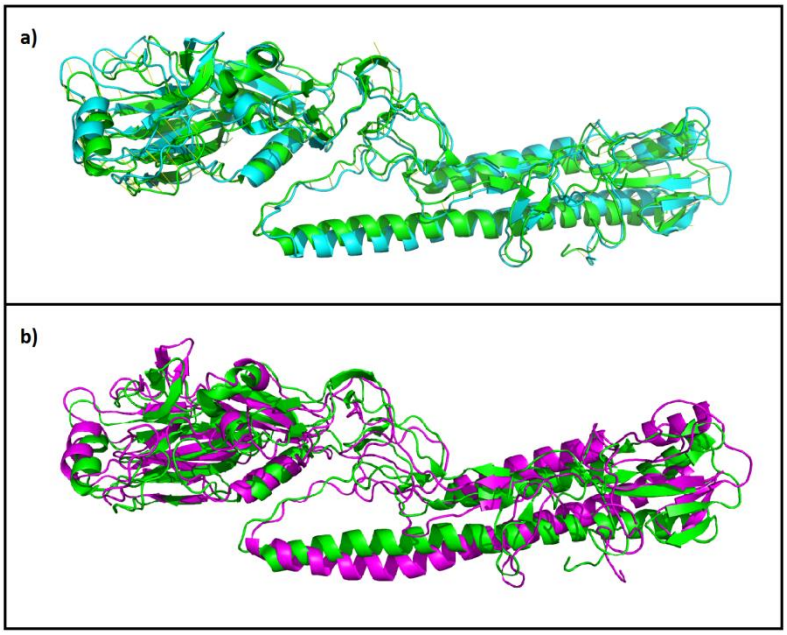

Figure 3.2. Alignment of the mutant and wild type 1RVT monomeric structures at 100 ns MD simulation to 1RVT_HI PDB file (a) wild type, (b) mutant type (green: 1RVT_HI, cyan: wild type, magenta: mutant type).

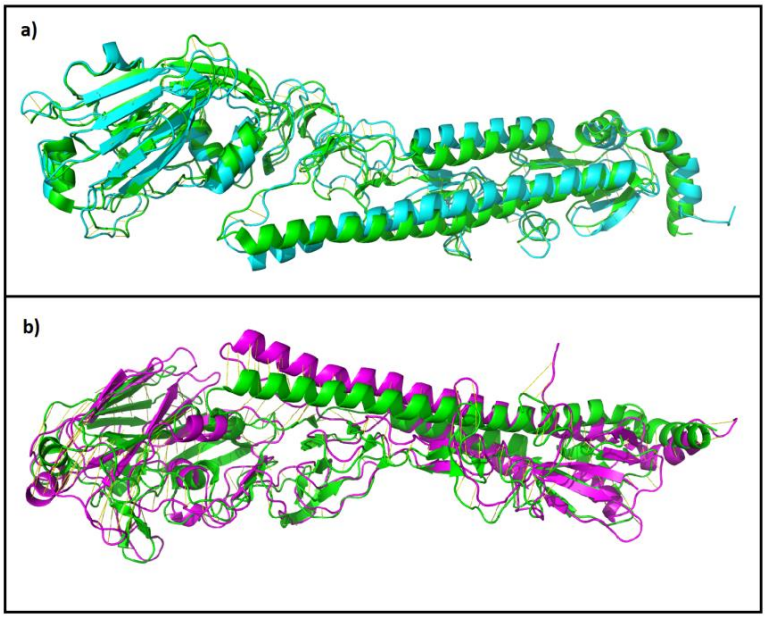

Figure 3.3. Alignment of the mutant and wild type 3LZG monomeric structures at 100 ns MD simulation to 3LZG_AB PDB file (a) wild type, (b) mutant type (green: 3LZG_AB, cyan: wild type, magenta: mutant type).

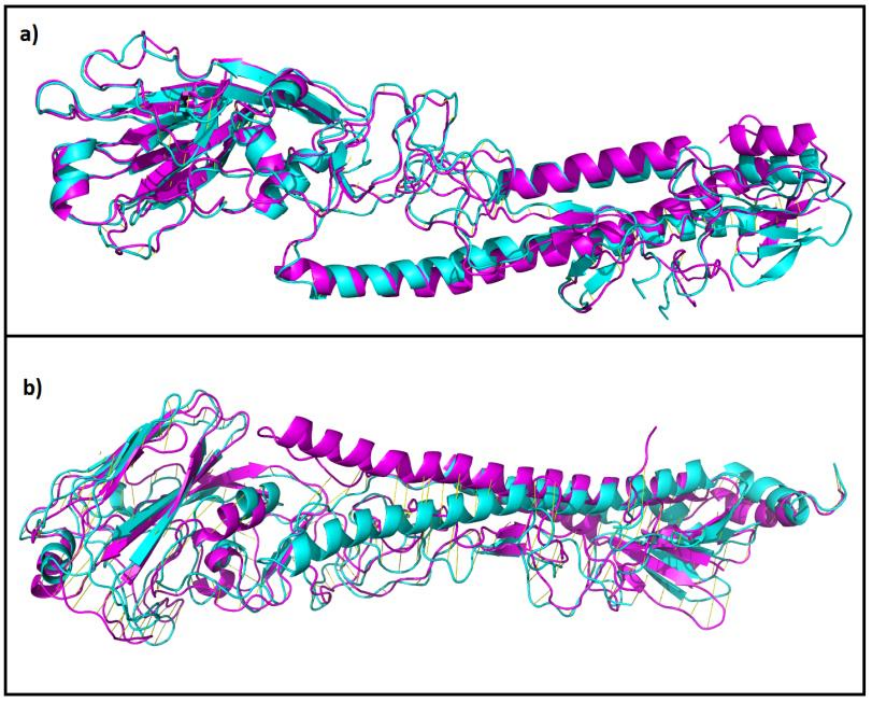

Figure 3.4. Alignment of the mutant and wild type monomeric structures at 100 ns MD simulation (a) 1RVT, (b) 3LZG (cyan: wild type, magenta: mutant type).

The difference of the conformation of monomeric structures to the original PDB files and their RMSD results may not help to deduce a reasonable change due to the fact that the structure acts as a trimer.

Further, the changes will be discussed with residue fluctuations below.

3.1.2. Mean Square Fluctuation (MSF)

The physical meaning of the mean square fluctuations (MSF) is the mobility of each residue; the extent of the change of the position of each residue in time from its mean position. MSF calculation is performed on the equilibrated part of the MD trajectory considering only the C-α atoms of minimized structures.

Figure 3.5 displays MSF values of residues of monomeric structures; 1RVT at the top and 3LZG at the bottom respectively. The residue numbers are written in accordance with the actual residue numbers from HA1 and HA2 domains.

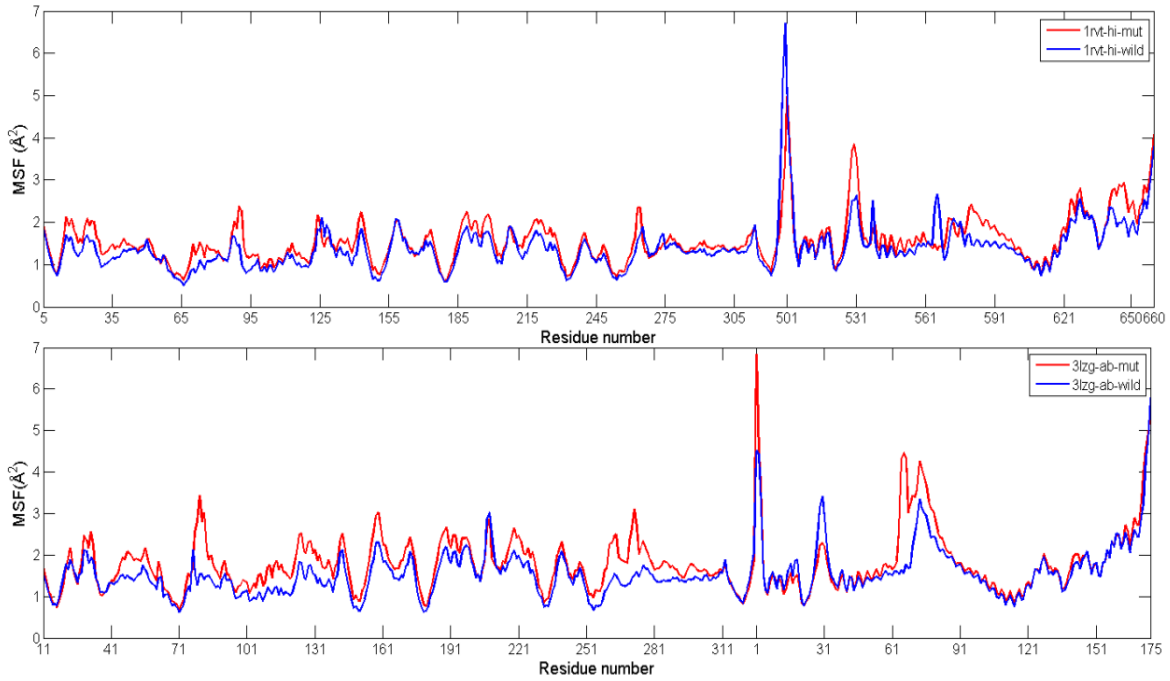

Figure 3.5. MSF values for the monomeric structure of 1RVT at the top, 3LZG at the bottom with their mutant and wild type structures (red: mutant type, blue: wild type).

Figure 3.5 shows the minima and maxima points of the 1RVT and 3LZG structures. The regions of the residues corresponding to minima and maxima points of MSF values are listed in Tables A.3 and A.4. The wild type and mutant type structures mostly have the same maxima and minima positions, the exceptions are given with bold and underlined. The maxima positions because of the chain ends are not listed.

For the case of 1RVT, the results show the same trend for HA1 domain of both cases. Although the corresponding MSF values are different, the minima and maxima values are the same. 1RVT structure has higher MSF values for the mutant type at residues 15-17, 24-27, 49-50 and 87-91 than its wild type. Cb, Ca_2 and Sb antigenic sites also have higher fluctuations for the mutant type 1RVT (residues 74-79, 224-225, and 187-198, respectively). RBS residues have higher fluctuations for the mutant type (130 loop, 190 helix and 220 loop). On the other hand, the wild type 1RVT has higher fluctuation values only at HA2 domain. The residue where the mutation is applied (149) is a global minimum for both structures.

For the case of 3LZG, the results show the same trend for HA1 and HA2 domains for both cases. Despite their corresponding MSF values are different, the minima and maxima

values are the same with a difference among the mutant and wild types. 3LZG structure has higher MSF values for the mutant type at residues 29-33, 46-62, 78-91, 99-138, 156-174, 186-199, 217-228, 255-300 than its wild type. The antigenic sites Cb (74-79), Sa (128-129, 156-160, 162-167), Ca_1 (169-173) Sb (187-198), Ca_2 (224-225) overlap these areas. RBS also has higher fluctuations for the mutant type (130 loop, 190 helix and 220 loop). HA2 of the mutant type generally has higher fluctuation values then the wild type, whereas the wild type has higher fluctuation than the mutant type only at residues 30-32 of HA2 domain. The residue where the mutation is applied (149) is a global minimum for both structures. If the residues showing minima or maxima behavior belong to HA2 domain, it is denoted within parenthesis next to the residue number.

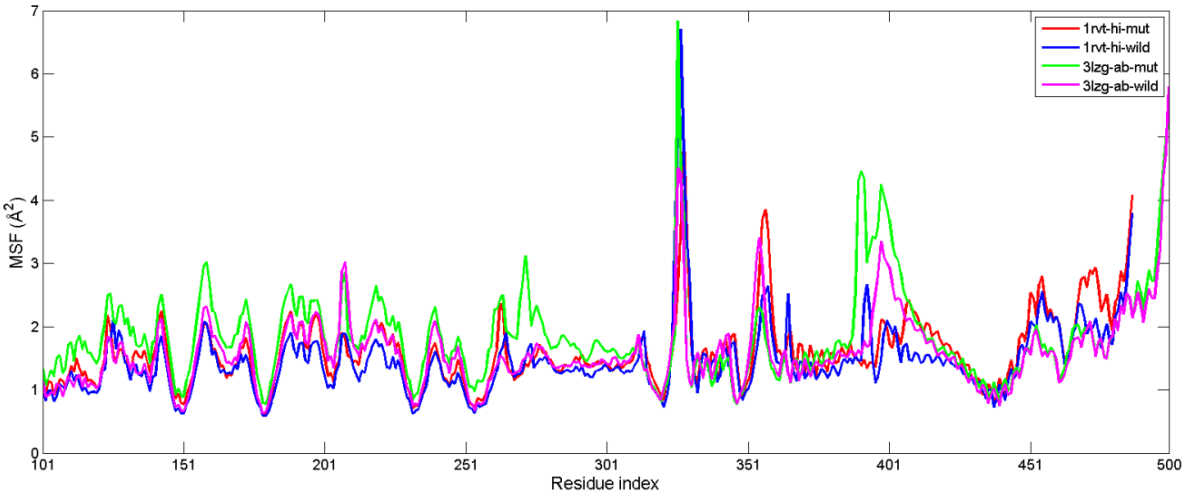

Figure 3.6. MSF values for the monomeric structure of 1RVT and 3LZG with their mutant and wild type structures for the last 70 ns analysis of 100 ns MD simulation.

The MSF values are divided into two separate parts to align the residues to be able to compare the MSF values for different variants. The four structures are considered together, and the MSF values are plotted starting from residue 100 for each case (Figure 3.6). The mutant type 3LZG has highest fluctuations at most of the residues especially at RBD residues, followed by its wild type for HA1 domain. End residues of stem region however show higher fluctuations for the mutant type 1RVT followed by its wild type. The structures show a complex behavior at HA2 domain corresponding to residues 561-591 for 1RVT and 61-91 (Residue index: 386-416) for 3LZG. Other than that, all the structures share the same global minima points.

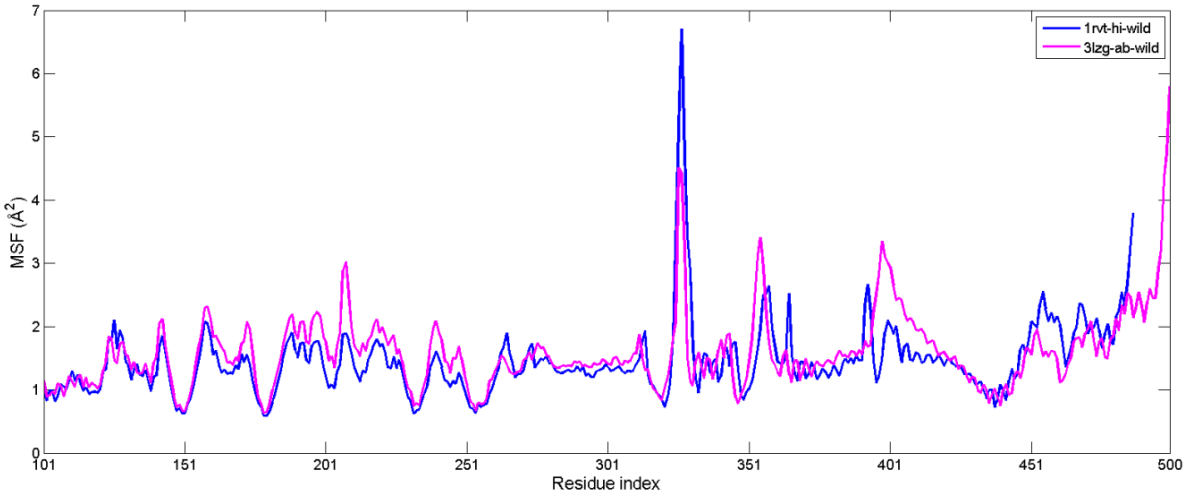

Figure 3.7. MSF values for the monomeric structure of wild type 1RVT and wild type 3LZG for the last 70 ns analysis of 100 ns MD simulation.

Figure 3.7 is obtained by plotting MSF values of residues from 100 to the end for the wild type structures of each variant. As seen, HA1 domain has the same trend for both of the structures. The minima points, the hinge regions, correspond to residues around 149-151, 178-181, 231-235 and 251-256. Residue 149 is also the mutation point.

3.1.3. Comparison of Wild and Mutant Type Structures

MSF values for most of residues are higher for the mutant than the wild type structure. This could be partly expected as the mutant cases might have needed a longer equilibration period or otherwise mobile due to an intrinsic dynamic behavior. The monomeric state was taken from the trimer structure, which may require longer simulations to deduce a behavioral difference between the wild and mutant structures. On the other hand, the chain ends (from HA1 domain to HA2 domain), at residue 325 for 3LZG and residue 327 for 1RVT, show maxima peaks in MSF values.

3.1.4. Cross Correlations (CC)

The correlation between residue fluctuations (C-α atoms) of the structure is given as cross-correlation maps in Figure 3.8 for monomeric structures (covering HA1 and HA2 units). A specific region is focused from the overall cross-correlation map and given in

Figure 3.9. The importance of this specific region is that it displays the RBD with respect to host receptor (sialic) and antigenic binding sites in a closer view. The cross-correlations are normalized correlations and vary within the range [-1, 1], from color red to blue; from the most positive to most negative correlations, representing the correlated and anti-correlated motions within the structure. These maps are based on 100 ns MD simulations and they contain information only for the first monomeric unit.

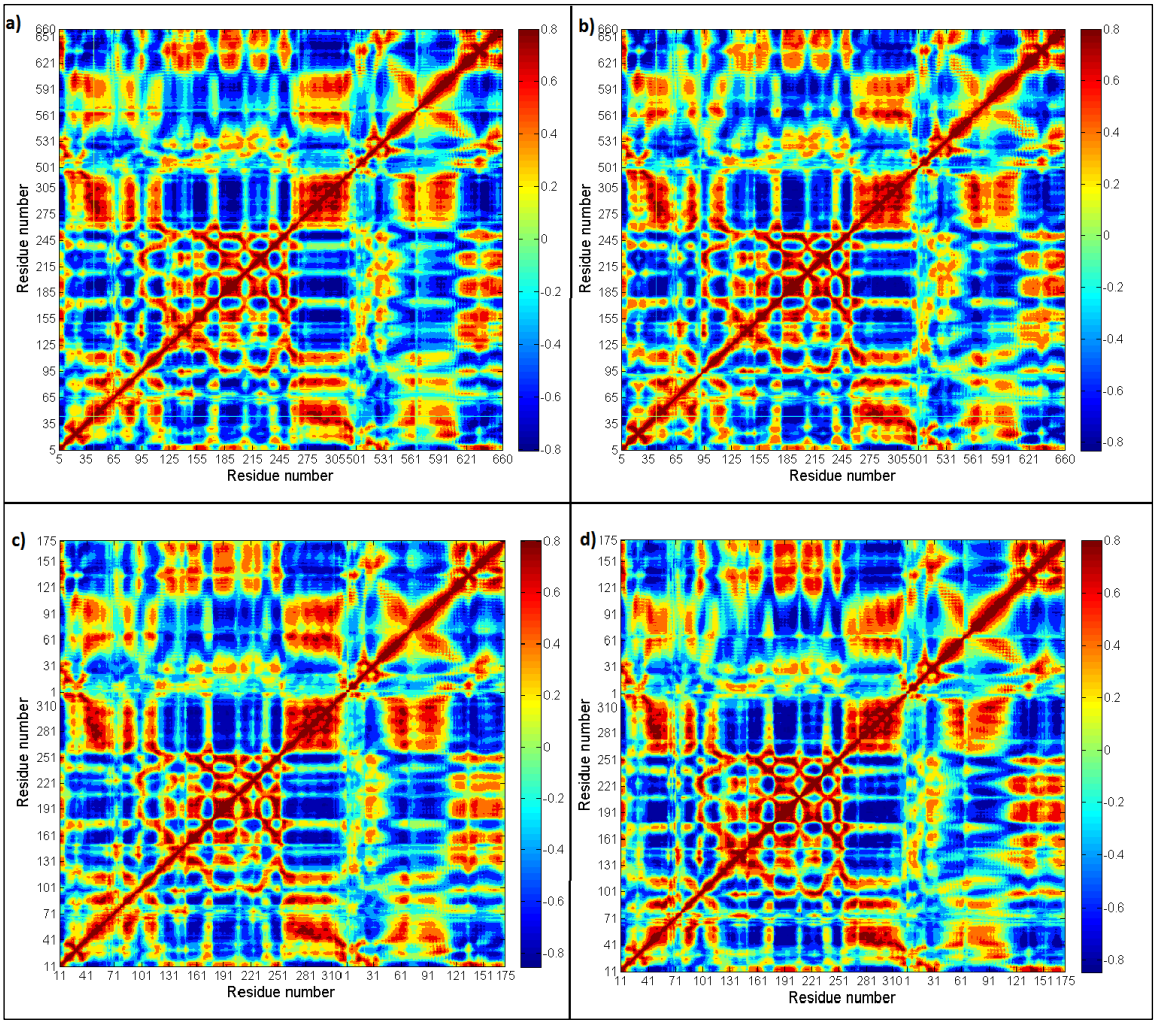

Figure 3.8. Cross-correlation maps for each monomeric structure (a) 1RVT wild type, (b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

Cross-correlation maps in Figure 3.8 show how each variant correlates with its mutant type. The relation of the wild type of one variant can also be analyzed for it relation to the mutant type of the other variant.

For 1RVT, the mutant type has higher correlations than its wild type for residues 15 to 100 of HA1 domain and residues 550 to 570 of HA2 domain with the overall structure at residues 95-125, 155-185, and 245-327 of HA1. Besides, HA2 domain of the wild type mostly has higher intra-domain correlation regions than its mutant type. The most specific correlation region is seen at the end of HA1 domain, corresponding to residues 270 to 327.

For 3LZG, the wild type has higher correlation values in its HA2 domain with the overall structure than the mutant type specifically at residues 130-160 and 270-310 of HA1 domain. The residues at the end of HA1 domain also have higher correlation values between residues 11-90 of HA1 for the wild type. Besides, the end residues of HA2 domain has high correlation values with 180-230 region of HA1 corresponding to the 190 helix and 220 loop of RBS.

Overall, the mutant type structures appear to have higher correlation values than their wild types for HA1 domain, and residues at the end of HA2 domain shows higher correlation values within residues 130-150.

Figure 3.9 includes RBS (130 loop, 190 helix, 220 loop) and the antigenic binding sites for the equilibrium region of last 70 ns analysis of a total 100 ns simulation. Figure 3.9 focuses on a specific region, the correlations of residues 130-150 and overall HA domain, between the wild and mutants type for both variants.

The correlation of residues 130-150 with the overall structure shows higher correlation of this region with the overall structure in general. However, the change in the correlations within specific domains can be observed when examined in detail. While all structure has an anti-correlated motion with the 130-150 region that also includes the 130 loop (134-138), residues 215-225 (both RBS at 220 loop and Ca_2 antigenic site), 128-129 and156-160 and 162-167 (Sa antigenic site) have high correlation with the 130 loop as well as Ca_2 antigenic site (140-145). The 220 loop (221-228) and 190 helix (188-194) also have high correlation with 130 loop (134-138). The correlation at the Ca_2 antigenic binding site (140-145) is higher for the wild type structures.

If the correlation maps are observed for the analysis of the change in the variants after the mutation, the following results can be listed.

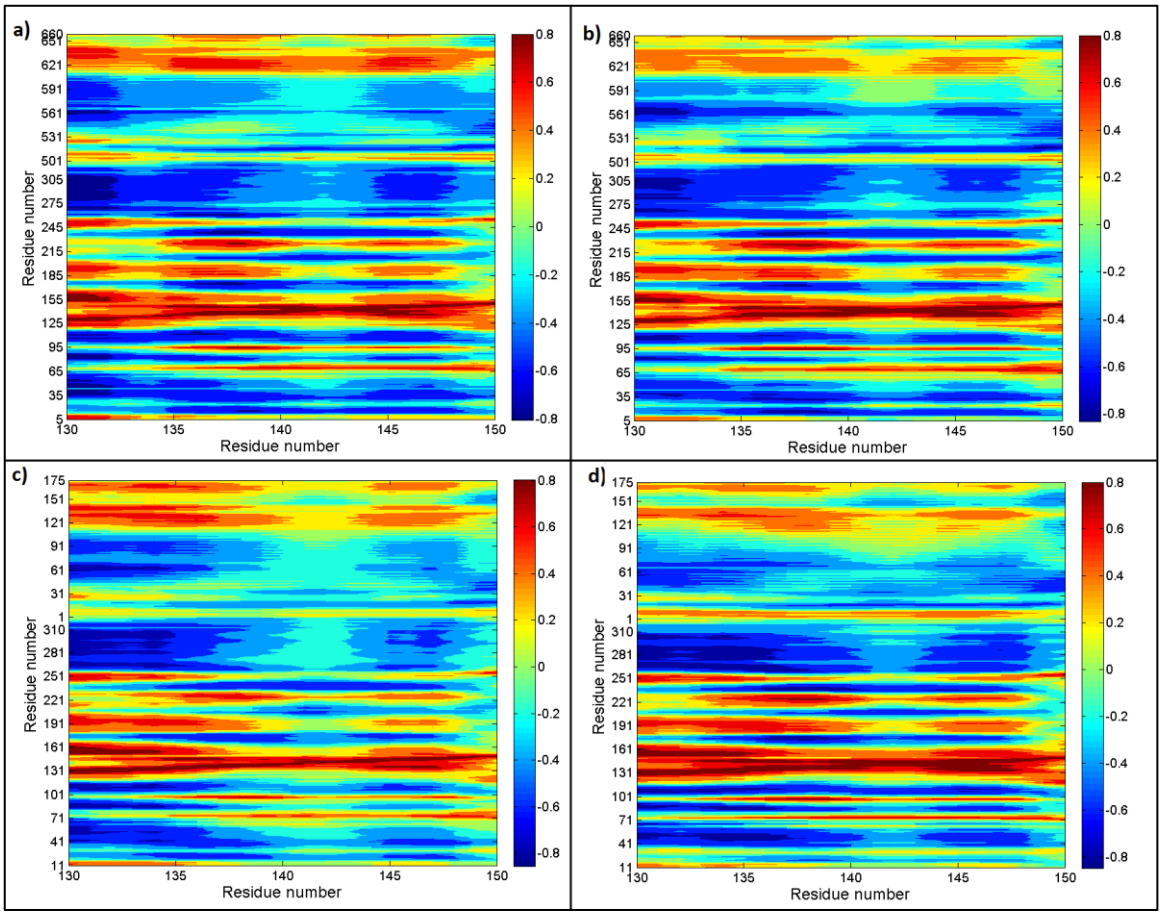

Figure 3.9. Cross-correlation maps for each monomeric structure for residues of 130-150 versus overall HA (a) 1RVT wild type, (b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

For 1RVT, the 190 helix and 220 loop have increased in correlation with 130-150 region. The same behavior is also observed for 3LZG. Additionally, the mutant type 1RVT is expected to show a correlation patters that is more likely as the wild type 3LZG, whereas the wild type 1RVT and the mutant type 3LZG are expected to behave more likely to each other. However, the monomeric subunit simulations do not fulfill the sufficient simulation time to observe such a behavior. Further and more importantly, the absence of the other subunits may lead to a limited cooperativity that is needed to result in the expected behavior. The actual Hemagglutinin structure acts as a trimer. That may lead to the conclusion that although some intrinsic behaviors in terms of fluctuations, the residue fluctuation profile, for example hinge positions, is intrinsic in monomeric state, the cooperativity might need the inter-subunit contacts. Below, we will further be able to

discuss the monomeric state with respect to the trimer dynamics, for which the results are provided below.

3.2. Trimer Hemagglutinin

This subsection is the 100 ns MD simulation analysis results of the trimer structure. The trimer structures are H, I, J, K, L and M chains of 1930 H1 Influenza Virus Hemagglutinin (PDB ID: 1RVT) and A, B, C, D, E and F chains of 2009 H1N1 Influenza Virus Hemagglutinin (PDB ID: 3LZG) in their wild and mutant types.

3.2.1. Root Mean Square Deviation (RMSD)

Root Mean Square Deviation (RMSD) provides a measure for the conformational change of the structure from the initial X-ray crystal structure within the simulation time window. RMSD results are plotted for each variant in Figure 3.10.

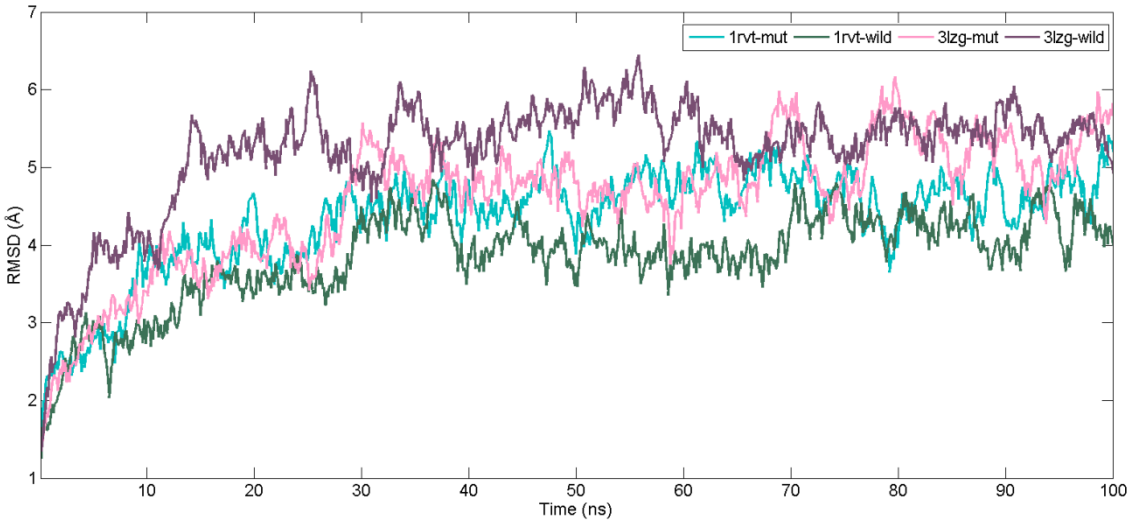

Figure 3.10. RMSD values for trimer 1RVT and trimer 3LZG.

It is observed that 3LZGs have relatively higher RMSD values compared to 1RVTs. The average RMSD values calculated are given in Table A.5 in Appendix A. Table A.5 shows that RMSD values between wild and mutant type of the same variant do not differ significantly for the last 30 ns of 100 ns MD simulations. This result of the trimer

structures differs from that of the monomeric results in which the 3LZG mutant type has the highest RMSD among all. RMSD behavior suggests the equilibrated phase for the MD trajectories to be around 30 ns for both variants. The MD simulations of 30 to 100 ns time span have thus been used for the subsequent analysis.

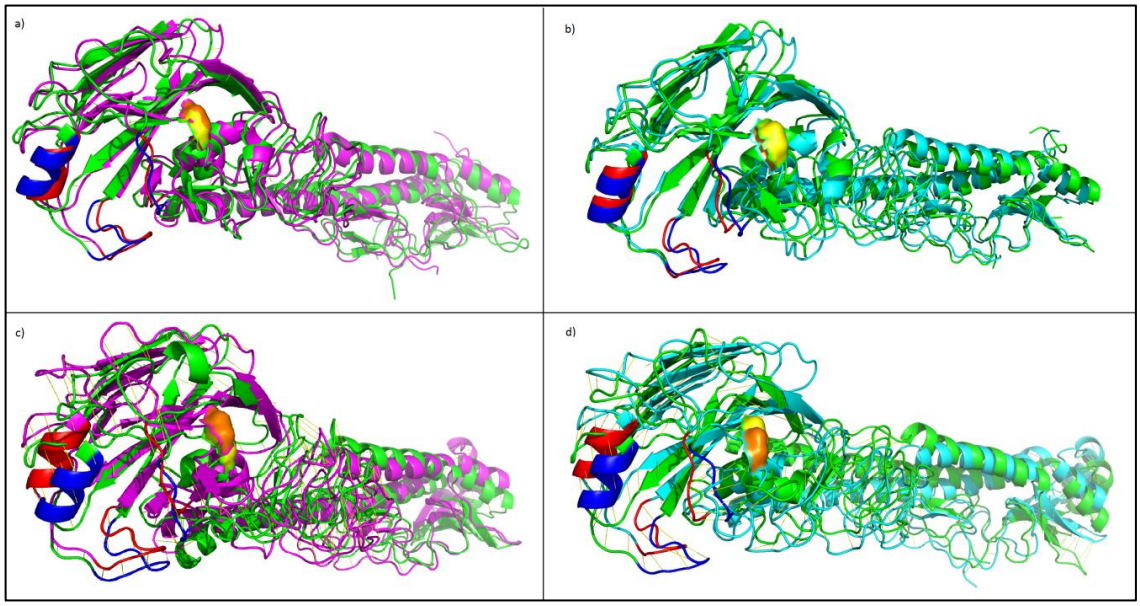

Figure 3.11. Alignment of the subunit of trimer structures to the monomeric structure at 100 ns MD simulation (a) 1RVT wild type, (b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

The alignment of the monomeric structure to one of the subunit of the trimer is made for the last snapshot of each, and RMSD between the two is calculated in PYMOL in Figure 3.11, the superimposed wild type of each variant and their mutant forms on the corresponding isolated monomeric structures are presented. Aligned structures show the directional change of each trimer structure with respect to its monomeric state. Here, the mutant and wild type 1RVT structures show high similarity between the isolated monomer and the subunit of the trimer. This can also be observed from their RMSD values in Table A.6 in Appendix A. On the other hand, the monomeric structure and the subunit of the mutant and wild type 3LZG trimer differ mostly in the HA1 chain; the region that contains the RBS residues. In Figure 3.11, green colors represent monomeric structures while cyan represents the subunit of the mutant trimer structure and magenta represents the subunit of the wild trimer structure. The RBS residues are colored red and blue for the subunit of the

trimer structure and the monomeric structure, respectively. The position 149 is colored orange and yellow for the latter respective states.

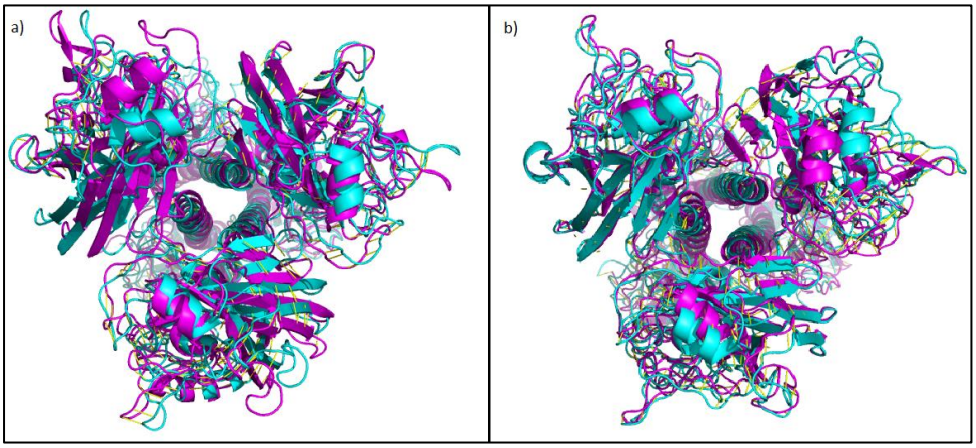

Figure 3.12. Alignment of the trimer structures to the monomeric structure at 100 ns MD simulation (a) 1RVT, (b) 3LZG (cyan: mutant type, magenta: wild type).

Figure 3.12 displays the mutant and wild type conformations of the trimer from top view at the end of 100 ns simulations. The 190 loop in mutant with respect to the wild type 1RVT leads to an inward motion, while this motion is outward for the mutant with respect to wild type 3LZG.

RMSD of the trimer structures at their last frame of the 100 ns trajectory between the mutant and wild type is also calculated and listed in Table A.7 in Appendix A. The change of each conformation with respect to its own initial position through simulation time is given in Figure 3.13. Color code for Figure 3.13 is that green denotes the initial structure (after minimization), cyan denotes snapshot at 50 ns and magenta denotes snapshot at 100 ns. There appears a twisted motion in the counter-clockwise direction with respect to its wild type for both variants when it is looked from the top view. The stem domain moves in the opposite direction to the head domain until the end of the simulation. Interestingly, this motions is uncovered by the most cooperative mode/slowest modes by the normal mode analysis carried out by (Isın *et al.*, 2002); the globular head and the tail of HA moves in counter rotations. The comparison across mutant and wild types show that the K-variant moves (and rotates) more compared to the R variant. A more cooperative behavior may result in this observation, which will be further discussed with the cross-correlation maps.

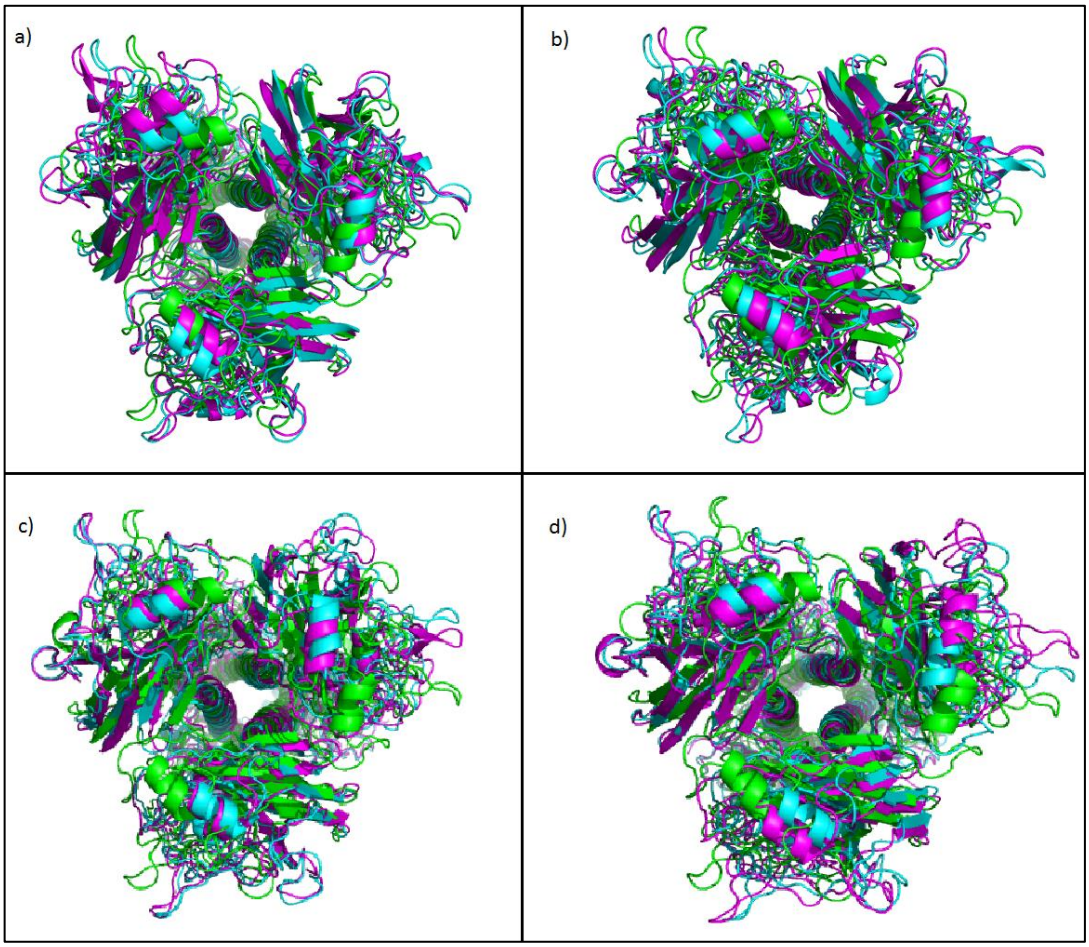

Figure 3.13. Alignment of trimer structures at time steps 50 ns and 100 ns to the initial structure (a) 1RVT wild type, (b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

3.2.2. Mean Square Fluctuation (MSF)

MSF calculation is performed on the equilibrated part of the MD trajectory considering only the C-α atoms of minimized structures. Figure 3.14 displays MSF values of residues for one of the subunits of trimer structures 1RVT and 3LZG. The residue numbers are written in accordance with the actual residue numbers from HA1 and HA2 domains. The MSF figures for whole structure (trimer) can be seen in Appendix A.

Figure 3.14 shows the minima and maxima points of the 1RVT and 3LZG structures. The regions of the residues corresponding to minima and maxima points of MSF values are listed in Tables A.8 and A.9. The wild type and mutant type structures mostly have the

same maxima and minima positions while the relative values change the exceptions are given with bold and underlined.

There is similarity between the MSF values of the mutant type 1RVT and wild type 3LZG in which they both show the same behavior at antigenic binding and the RBS residues. In Tables A.8 and A.9, antigenic binding sites are defined in parenthesis while host receptor binding sites are given italic and underlined. The antigenic sites that are being colored all belong to maxima points. The mutant type of 3LZG shows more fluctuations, however the corresponding MSF values are higher for the wild type 3LZG.

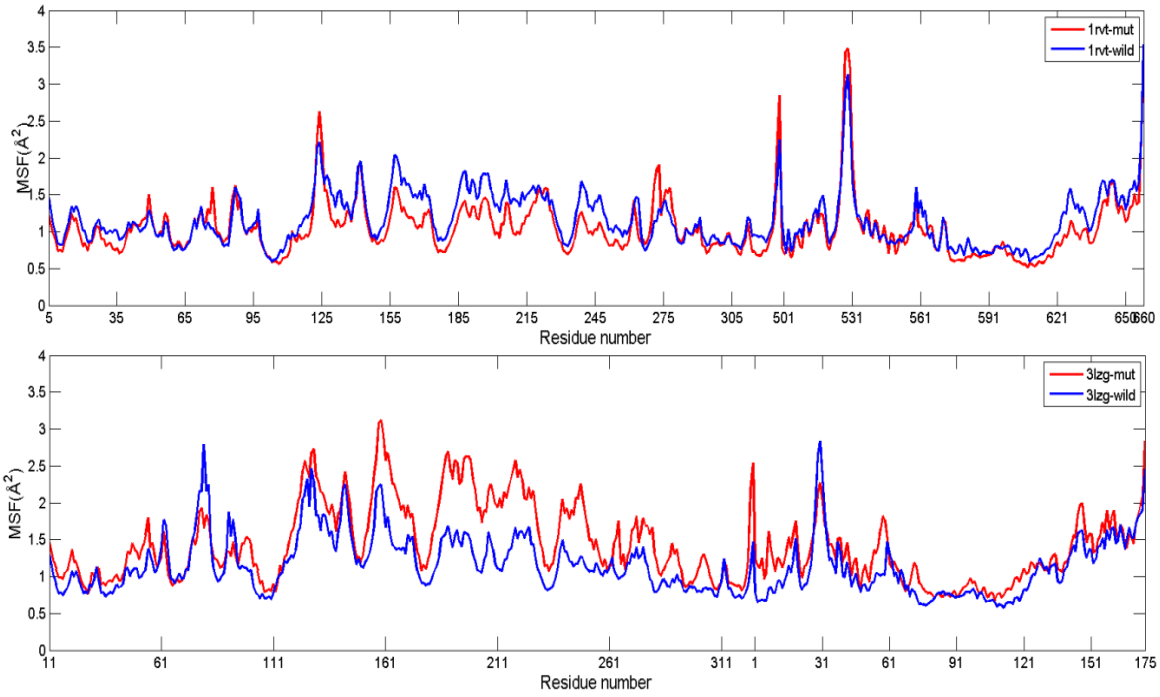

Figure 3.14. MSF values for the trimer structure of 1RVT at the top and 3LZG at the bottom (red: mutant type, blue: wild type).

For the case of 1RVT, the results show the same trend for HA1 and HA2 domain for both cases. Although their corresponding MSF values are different, the minima and maxima values are mostly at the similar residue positions. 1RVT structure has higher MSF values for the wild type with respect to the mutant type at residues 28-39, 127-138, 159-173, 175-221, 240-261, 314-322 of HA1 domain and 623-643 of HA2 domain. The antigenic binding sites and the RBS residues overlap the latter residues. Antigenic sites at Sa (128-129, 159-160, 162-167), Sb (187-198) and Ca_1 (169-173), and RBS at the 130

loop (134-138) and the 190 helix (188-194) are in these regions and have higher fluctuation for the wild type. The stem region of the wild type generally has higher fluctuation values as well. The residue position where the mutation is applied (149) is a global minimum for both structures. One important region where the mutant type has higher fluctuations is at Cb antigenic site at residues 76-79.

For the case of 3LZG, the fluctuations are highly different in magnitude between the wild and mutant types. Interestingly, here the mutant type structure has higher MSF values than its wild type at most of residue positions, such as residues 15-17, 45-55, 97-111, 156-174, 179-259 and 263-307 of HA1 domain and 14-25, 63-70 of HA2 domain. The highest MSF difference among these residues is seen at residues 156-174 and 179-259 in HA1 domain. The antigenic binding sites and RBS residues remain within these ranges. The antigenic sites at Sa (156-160, 162-167), Sb (187-198), Ca_1 (169-173, 206-208, 238-240), Ca_2 (224-225) and the RBS at 130 loop (134-138), 190 helix (188-194) and 220 loop (221-228) are in these regions and have higher fluctuation for the mutant type. The stem region of the mutant type generally has higher fluctuation values as well. The position of the mutation (149) is a minimum point for both structures. One important region where the wild type has higher fluctuation is at Cb antigenic site at residues 78-79.

3.2.3. Comparison of Wild and Mutant Type Structures

The four structures are considered together, and the MSF values are presented from residue 100 for each in Figure 3.15. The mutant type 3LZG has higher fluctuation at most of the residues especially at the antigenic and the RBS residues than its wild type. The wild type 1RVT has higher fluctuation than its mutant type. The chain ends (from HA1 domain to HA2 domain) on the other hand show high MSF values at residue 325 for 3LZG and 327 for 1RVT. This is expected as they are not connected.

MSF values for most of residues are higher for the mutant type 3LZG than its wild type structure while the situation is the opposite of this for 1RVT; wild type 1RVT has higher MSF fluctuation values than its mutant type. This can show the similarity between 3LZG mutant-1RVT wild and 3LZG wild-1RVT mutant type structures considering their MSF fluctuation behavior. Among all structures, the mutant type 3LZG has the highest fluctuations followed by the wild type 1RVT at the antigenic and the RBS residues.

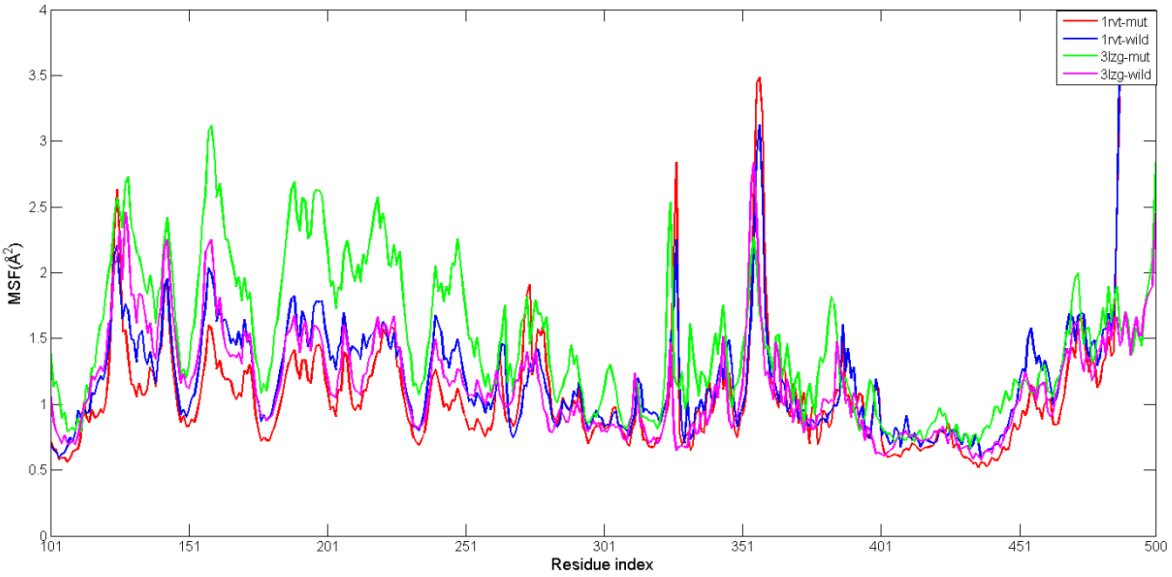

Figure 3.15. MSF values for the monomeric subunit trimer structure of 1RVT and 3LZG with their mutant and wild type structures for the last 70 ns analysis of 100 ns MD simulation.

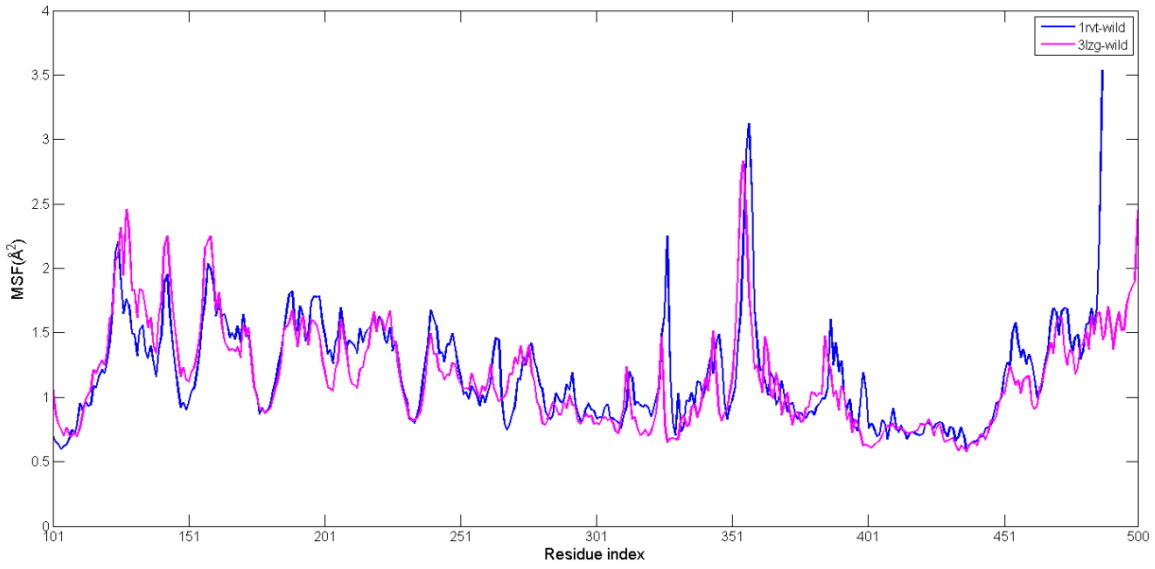

Figure 3.16. MSF values for trimer structure of wild type 1RVT and wild type 3LZG (shown for the first subunit) for the last 70 ns analysis of 100 ns MD simulation.

Figure 3.16 is obtained by plotting MSF values of residues from 101 to the end for the wild type structures of each variant. As seen, at most of the regions, the wild type structures show the same trend. The minima points, the hinge regions in other words,

correspond to the residues around 149-151, 178-181, 231-235 and 251-256. Residue 149 is also the mutation point.

3.2.4. Discussion of the MSF Values with Respect to the Sequence Variability of HA

MSF values are compared to the discriminating residue positions proposed in the work by Meroz *et al*., 2011. The discriminating residue positions between the pandemic and human swine viruses are determined and listed. Overlapping residues (MSF values of the MD trajectory with respect to discriminating residues) is given in the rank order previously defined in the literature.

Residue 149 is determined as the highest ranked discriminating residue, which is found as minima for all structures. Residue 225 is found as maxima only for wild type 3LZG. Residue 132 is found as minima only for mutant type 1RVT. Residue 188 is found as maxima for all structures whereas 1RVT structures show maxima at residue 206 and 3LZG structures show maxima at residue 208. Residue 131 is found as minima only for mutant type 1RVT. Residue 189 is found as maxima for all structures whereas residue 186 is found as maxima only for 1RVT structures. Of these residues; 149, 131, and 132 do not refer to host receptor or antigenic sites. The minima are observed at these three residues. However, residue 149 is sited close to Ca and Cb antigenic sites; and residues 131 and 132 are known to be situated in the receptor binding pocket. The maxima are observed at either antigenic (as in residues 206 and 208) or sites covering both host receptor and antigenic (as in residues 188, 189 and 225). Residue 186 is a maximum but is none of the above.

Residue 261 and 263 was also proposed as a discriminating position with residue 71 and 75 towards the end of among 20 discriminating residue list ranked. Analysis on MSF plots shows that the wild and mutant type 3LZG shows that 71 and 262 are at minima. The mutant 1RVT is a minimum point at residue 262, whereas this is not certain for the wild type.

There also exist residues appearing in the discriminating ranking (171, 226, 200), but do not show minima or maxima behavior in the MSF values. Within all minima points compared with discriminative analysis; residue 149 is the most definite minimum and is highly possible that it is an effective hinge. Principal component analysis of fluctuations would provide a more comprehensive picture for the hinge behavior of HA.

There are experimental and computational studies concentrating on mutations on HA1 domain of HA protein. A computational study represents the importance of residue 226 that mutation applied at this site (from Leucine to Glutamine) differentiates the recognition of sialic acid receptors by the Influenza A virus (Nicholls *et al*., 2008).On the other hand, in vivo mutations are also available of which E156K is found to effect Sb antigenicity and receptor binding avidity (Hensley *et al*., 2009).

Highly networked residues such as 98, 183, 153, 234 etc. are found for HA1 domain of HA (Soundararajan *et al*., 2011), within those residues 234 is found as hinge position for both of the variants in this study implying the importance of this position in the dynamics of the protein.

3.2.5. Cross Correlations (CC)

The correlation between residue fluctuations (C-α atoms) of the structure is given as cross-correlation maps in Figure 3.17 for trimer structures. The cross-correlations are normalized correlations and vary within the range [-1, 1], from color red to blue; from the most postive to most negative correlations. These maps are based on 100 ns MD simulations.

Figure 3.17 shows the overall correlation for the structures; in which it may be difficult to observe in most positions. The subunits of the trimer display similar behavior, yet some differences are observed in the strength of the couplings. This must be related to the sampling problem in this time window of MD simulations.

Considering that mutation at position 149 for 1RVT should lead to the similarity in the sequence of the wild type 3LZG and vice versa, the mutant type of one variant can be compared to the wild type of the other. If the mutant type 1RVT and the wild type 3LZG are examined together, it can be seen that cross-correlation maps show more similar behavior than those of their wild type structures.

The further analysis and visual representation of the cross-correlation maps are made in the following section.

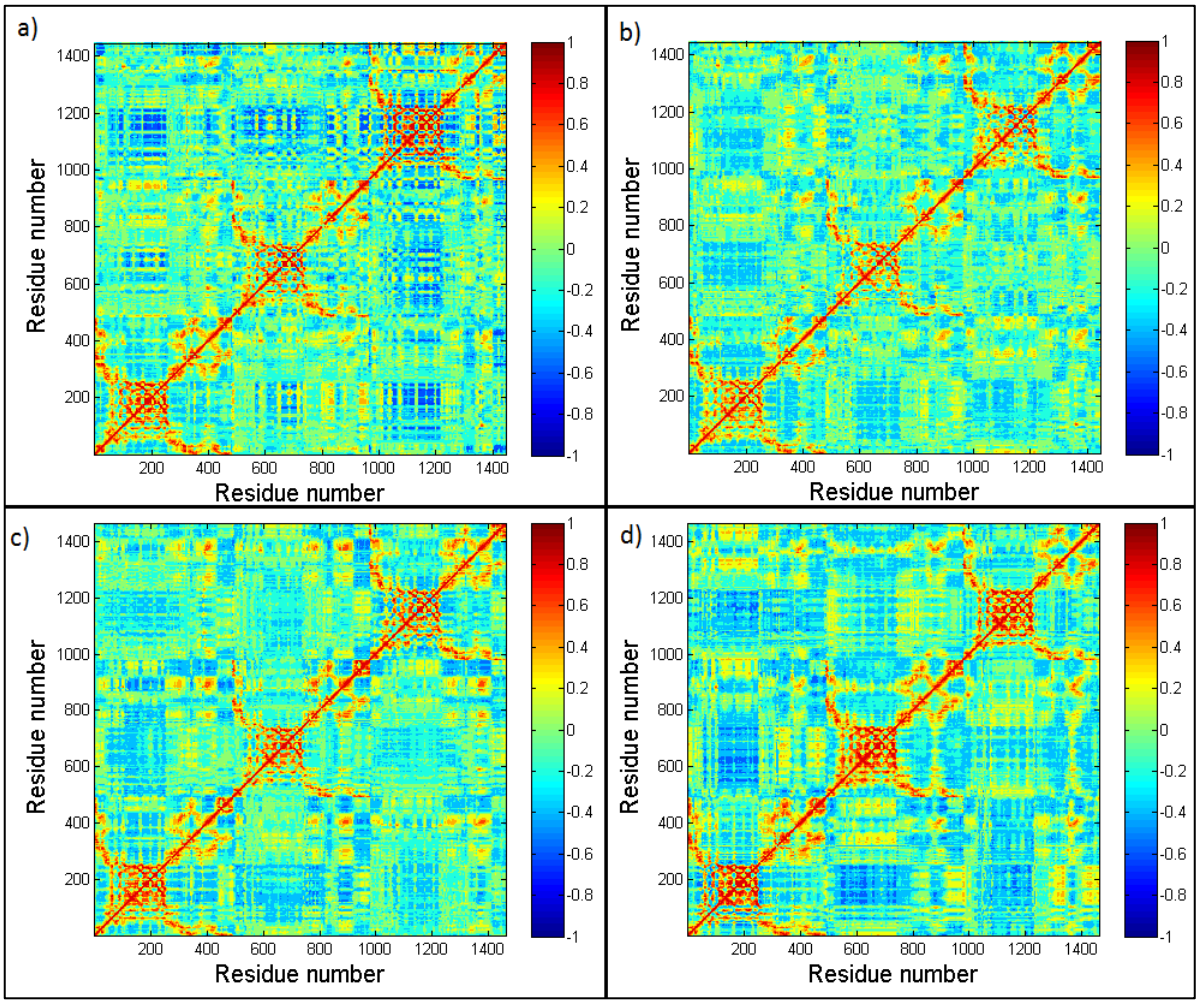

Figure 3.17. Cross-correlation maps for each trimer structure for the equilibrium region of last 70 ns analysis of 100 ns simulation (a) 1RVT wild type, (b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

Understanding the correlation within RBS and antigenic sites a more specific region is focused. First the subunit of the trimer structure is presented in Figure 3.18 and then a more specific region of residues 130-150 in Figures 3.19 and 3.20.

In Figure 3.18, first subunit of each trimer structure is presented. More specific cross-correlation maps are focused on to observe the change within RBS as well as the antigenic binding site residues of the structures.

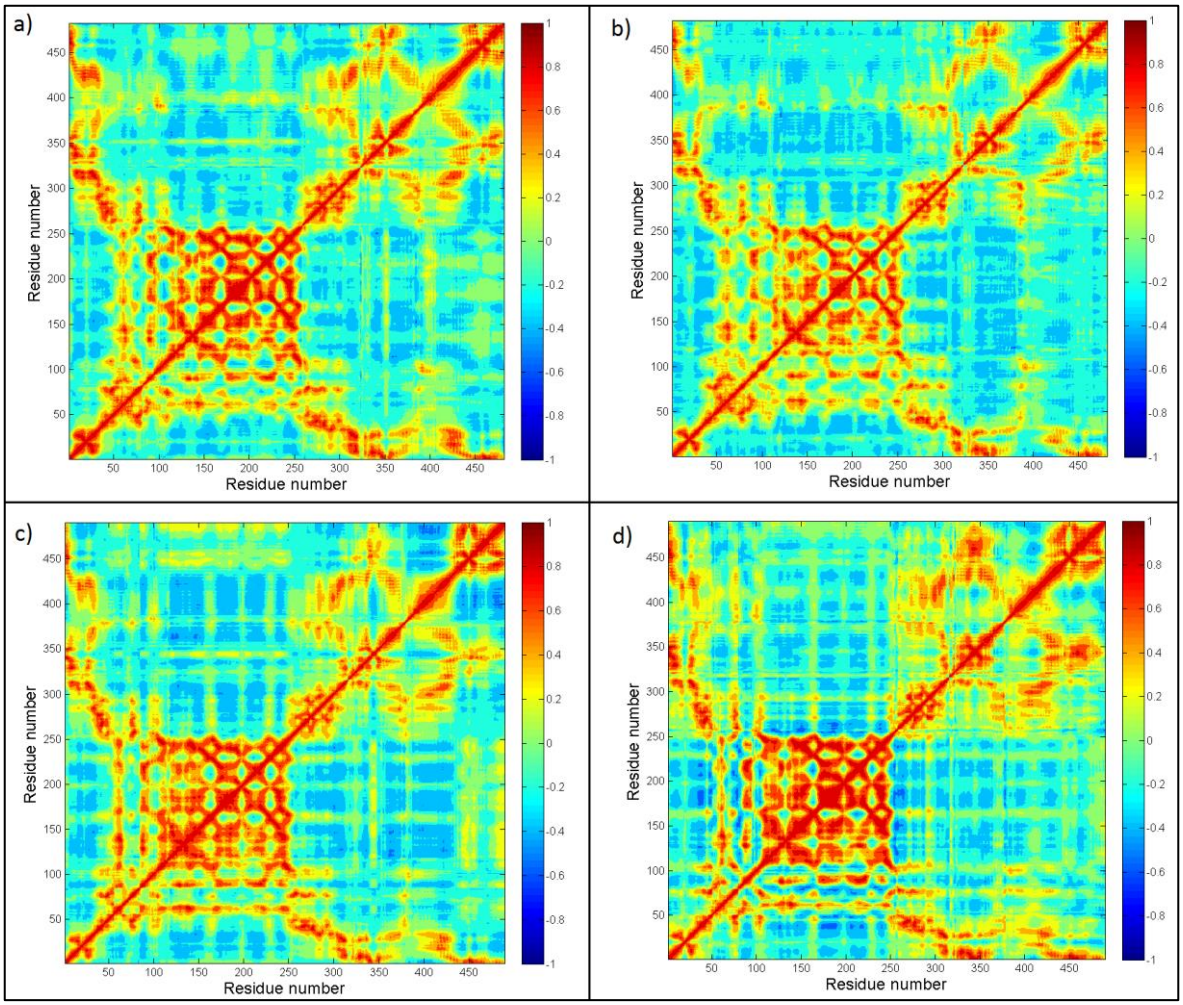

Figure 3.18. Cross-correlation maps for monomeric subsection of each trimer structure for the equilibrium region of last 70 ns analysis of 100 ns simulation (a) 1RVT wild type, (b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

The host receptor and antigenic binding sites of HA are in the HA1 domain of each subunit. Figures 3.19 and 3.20 are plotted to focus on the change within RBD of the structures for two time windows of the present MD simulations; 30-75 ns 30-100 ns, respectively. The correlation of residues 130-150 with residues 11-300 is elaborated. RBS's (130 loop, 190 helix and 220 loop) and five antigenic binding sites (Ca_1, Ca_2, Cb, Sb, Sa) are located in the HA1 domain of the HA and x axis covers 130 loop (134-138) and Ca_2 antigenic binding site (140-145).

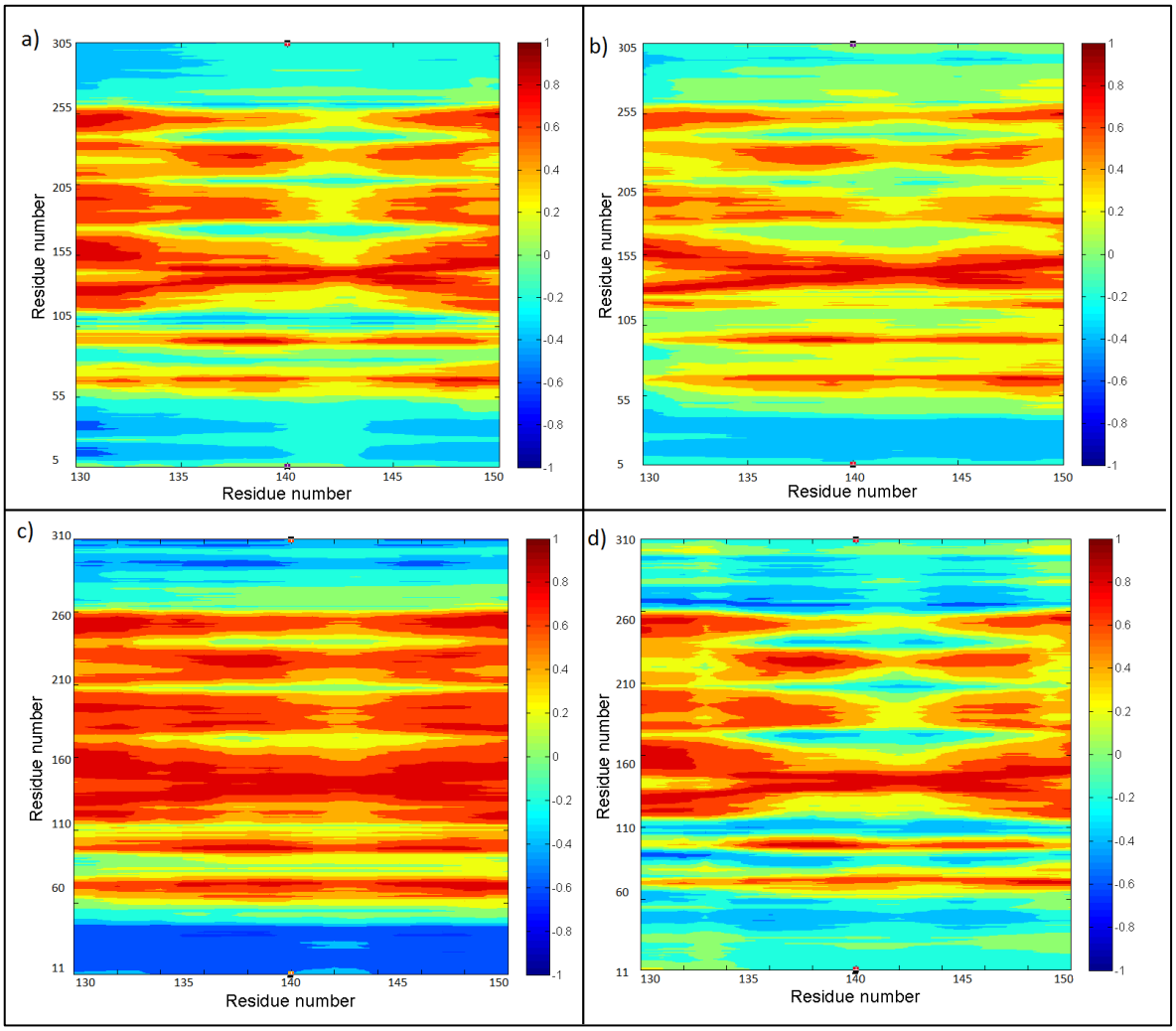

Figure 3.19. Cross-correlation maps for residues 130-150 to the first 300 residues of each trimer structure for the equilibrium region of last 45 ns analysis of 75 ns simulation (a) 1RVT wild type, b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

Cross-correlation maps reveal the change in both variants as a result of the mutation at residue position 149. The changes from wild to mutant type are listed briefly in the following for 75 ns and 100 ns MD simulations.

For the 75 ns analysis of 1RVT protein, the correlated motion between the 130-150 region and RBS (190 helix and 220 loop) can be observed. In addition to that, antigenic site residues are highly correlated with 130-150 region as well. 3LZG protein shows correlated motion between 130-150 region and RBS (190 helix and 220 loop). The mutation leads to a decrease in the correlation of the residues for 3LZG protein especially

at antigenic binding sites at residues around 160 and 220. Additionally, an opening in the correlation maps is observed for the mutant type 3LZG between 130-135 (RBS at 130 loop) and 140-145 (Ca_2 antigenic binding site) regions.

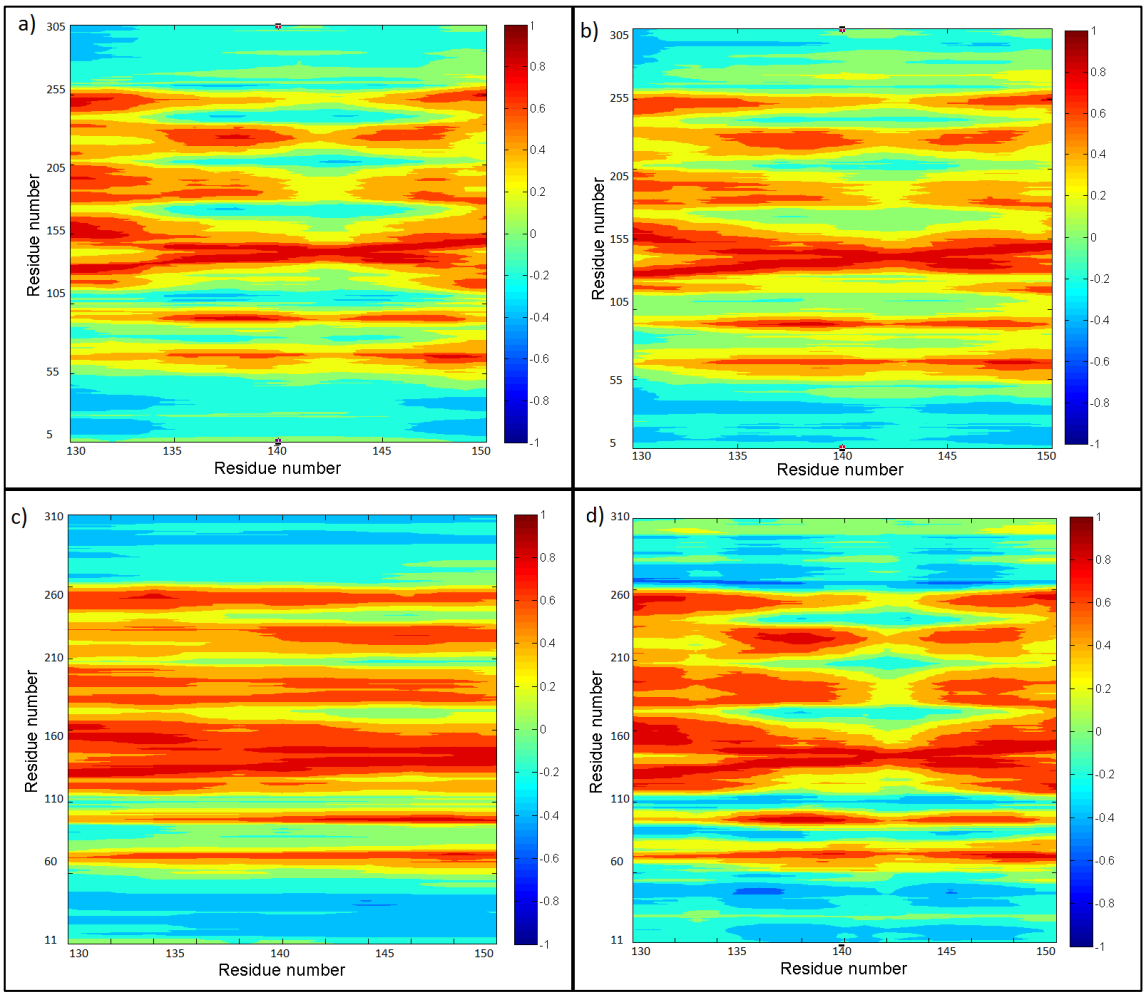

Figure 3.20. Cross-correlation maps for residues 130-150 to the first 300 residues of each trimer structure for the equilibrium region of last 70 ns analysis of 100 ns simulation (a) 1RVT wild type, b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

For the 100 ns analysis of 1RVT protein, the correlated motion between the 130-150 region and RBS (190 helix and 220 loop) can be observed. In addition to that, antigenic site residues are highly correlated with 130-150 region as well. The increase in the correlation of the mutant type with respect to the wild type is observed between 130-150 region and RBS (190 helix and 220 loop). The correlation between the Ca_2 antigenic site with 130-150 region is also increased in the mutant type. 3LZG protein shows correlated

motion between 130-150 region and RBS (190 helix and 220 loop). The correlation of the Ca_2 antigenic binding site in 130-150 region is higher for the wild type through entire structure.

As a general pattern, it is found that mutation leads to the tendency of increase in the correlations for 1RVT structure with respect to its wild type whereas the reverse is valid for 3LZG structure. The wild type 3LZG tends to show more correlated motion than its mutant type.

The residues around the receptor binding pocket have been investigated and they are found to be affecting the virus binding. Double mutations of T200A and E227A were proposed to be enhancing binding to the human receptor (Sriwilaijaroen and Suzuki, 2012). Both residues are seen as highly correlating with 130 loop in the cross-correlation maps. T200 and E227 are seen in 3LZG whereas A200 and A227 are seen in 1RVT protein.

Additionally, mutations at residues on HA such as P128S, G143R are reported to have higher binding affinity than the wild type for 1934 variant (Soundararajan *et al*., 2011). Those residues belong to Sa and Ca_2 antigenic binding sites respectively and found to be highly correlating with RBD of 1930 and 2009 variants in the cross-correlation maps. Two other mutations E119G and L71P are found to be enhancing receptor binding specificity for 1934 variant, with 130 loop and 220 loop RBS's. Those residues exist as E119 and L71 in 2009 variant, so similar results can be expected within these regions.

Highly networked residues are studied for HA1 domain of HA protein (Soundararajan *et al*., 2011). The residues that are reported to have high network scores are compared to the cross-correlation maps from trimer structures. Residue 234 among these highly networked residues is found as hinge position in addition to that it is also a highly correlated position with RBS residues in this study. Other than that, there are other highly correlated residues with RBD such as 183, 98, 250, 127 etc. as proposed highly networked in the previous study (Soundararajan *et al*., 2011).

The effect of the mutation on each variant is viewed in Figures 3.21 and 3.22 and Figures 3.23 and 3.24. The average cross-correlation values of residues 141 and 142 are used for revealing the effect of mutation. The reason of choosing these specific residues was that, they show the effect of mutation at most and they also belong to Ca_2 antigenic site. Then newly generated PDB's are colored as a spectrum with respect to the C-α atoms

and cartoon trace option is used in PYMOL. Finally, following figures are obtained both for 75 ns and 100 ns analysis.

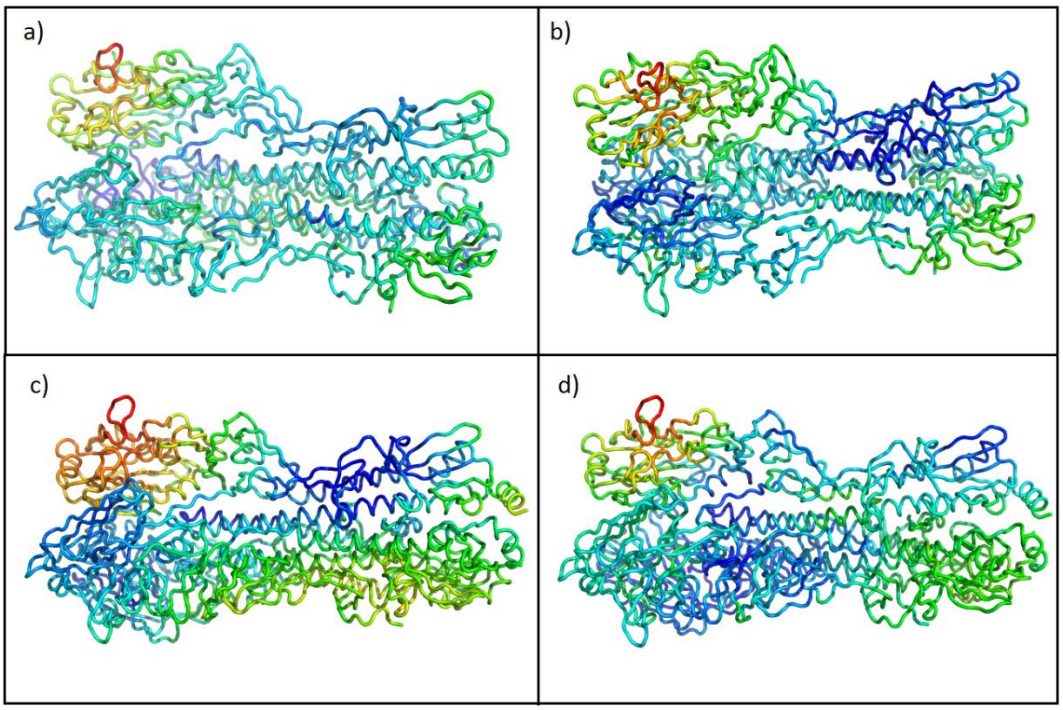

Figure 3.21. Trace shown structure of the C-α atoms for the 75 ns simulation of the structures (a) 1RVT wild type, b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

Figure 3.21 shows decreasing correlation from wild to mutant type for 3LZG protein, and increasing correlation from wild to mutant type for 1RVT protein mostly in the first monomeric unit for 75 ns analysis. The subunits other than the first one also show deviations within the wild and mutant type of each variant.

Figure 3.22 shows the correlations on the first subunit. The results show that the wild type 3LZG has higher correlations than its mutant type. On the other hand, the mutant 1RVT has slightly higher correlation that the wild type 1RVT. The mutant type 1RVT shows more negative correlation in the HA2 domain as that of the wild type 3LZG.

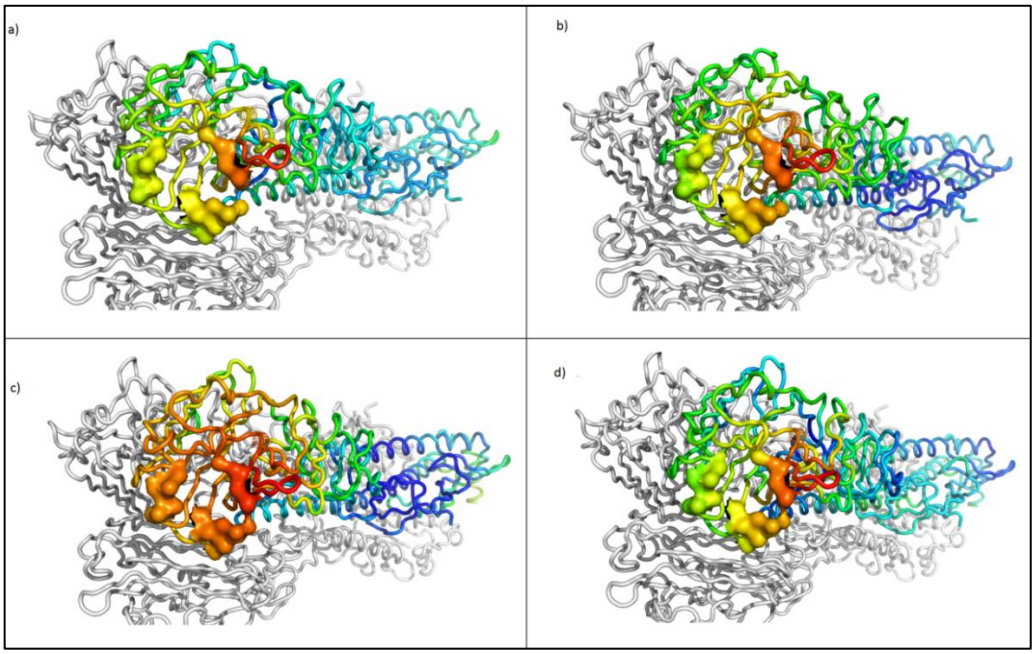

Figure 3.22. RBD of the first subunit as trace shown structure of the C-α atoms for the 75 ns simulation of the structures (a) 1RVT wild type, b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

Highest correlation is seen in 130 loop for all structures which is located at the top right position for all structures, while 190 helix shows the lowest correlation results which is located at the top left position.

Figure 3.23 shows decreasing correlation from wild to mutant type for 3LZG protein, and increasing correlation from the wild to the mutant type for 1RVT protein mostly in the first monomeric unit for 100 ns analysis.

Figure 3.24 shows the first monomeric subunit and the three RBS's in a closer view. The results show that wild type 3LZG shows more correlation than its mutant type while the mutant type 1RVT is correlated more than its wild type in their RBS's. Highest correlation is seen in the 130 loop for all structures which is located at the top right position of Figure 3.24 for all structures while 190 helix shows the lowest correlation results which is located at the top left position.

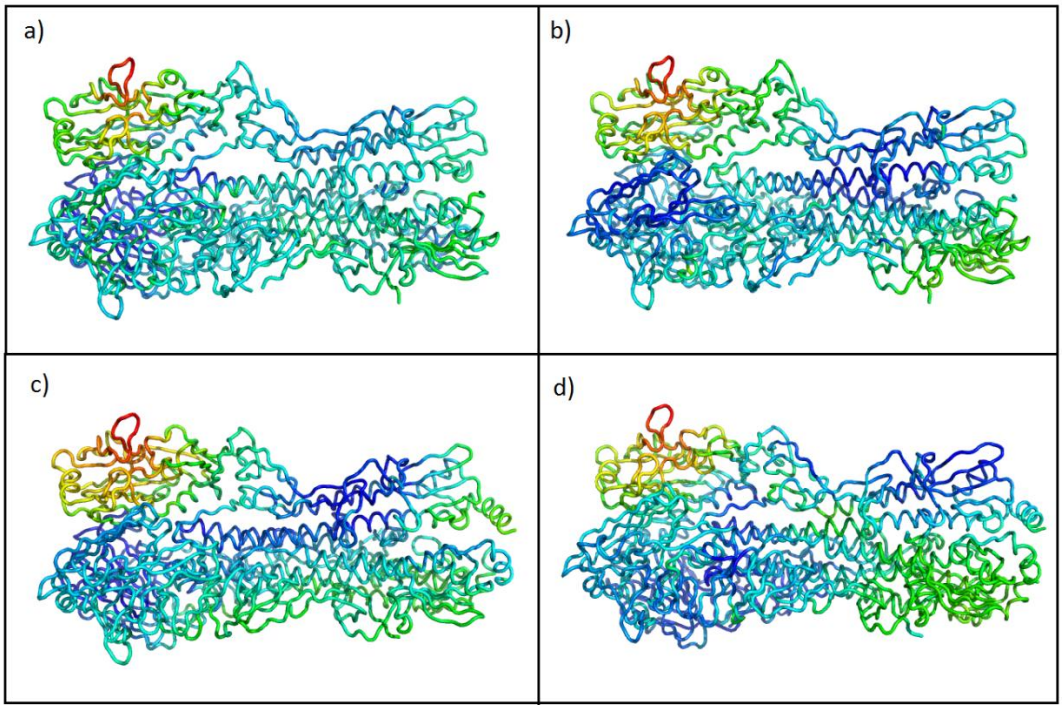

Figure 3.23. Trace shown structure of the C-α atoms for the 100 ns simulation of the structures (a) 1RVT wild type, b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

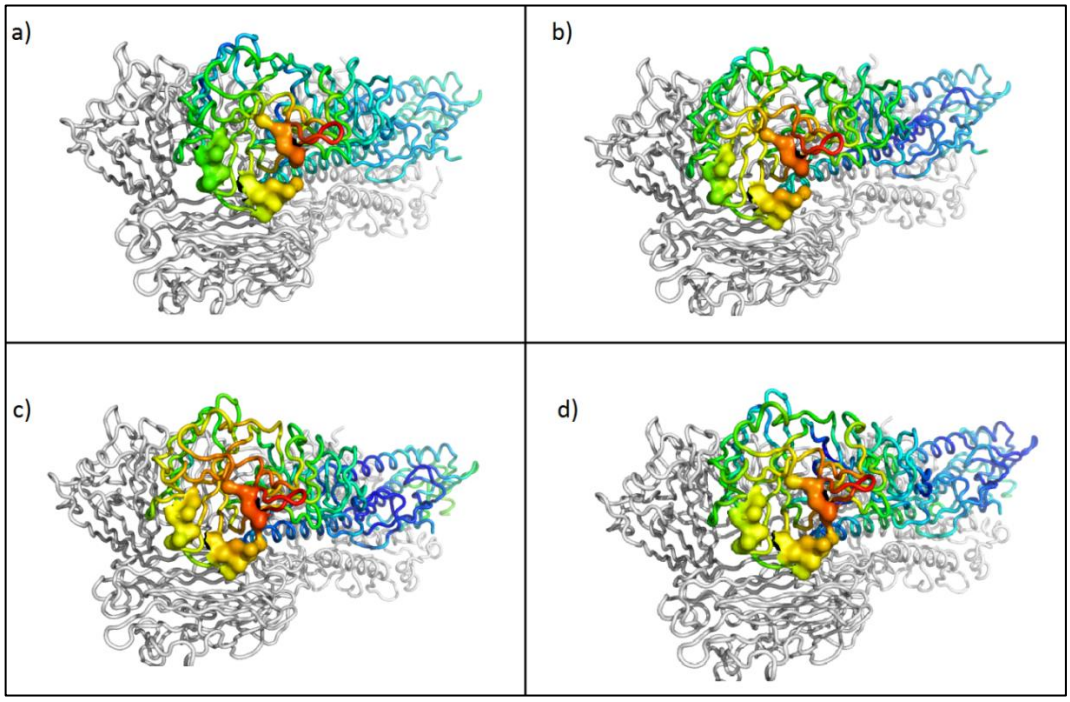

Figure 3.24. RBD of the first monomeric unit as trace shown structure of the C-α atoms for the 100 ns simulation of the structures (a) 1RVT wild type, b) 1RVT mutant type, (c) 3LZG wild type, (d) 3LZG mutant type.

The correlation values that are used to color the trace shown structures in Figures 3.21-3.24 are plotted in Figures 3.25 and 3.26. These plots represent the positive average cross-correlation values of residues 141 and 142 and show differences between the mutant and wild types of each variant for 75 ns and 100 ns analysis. The maxima points in the plots correspond to the antigenic binding sites of the structures.

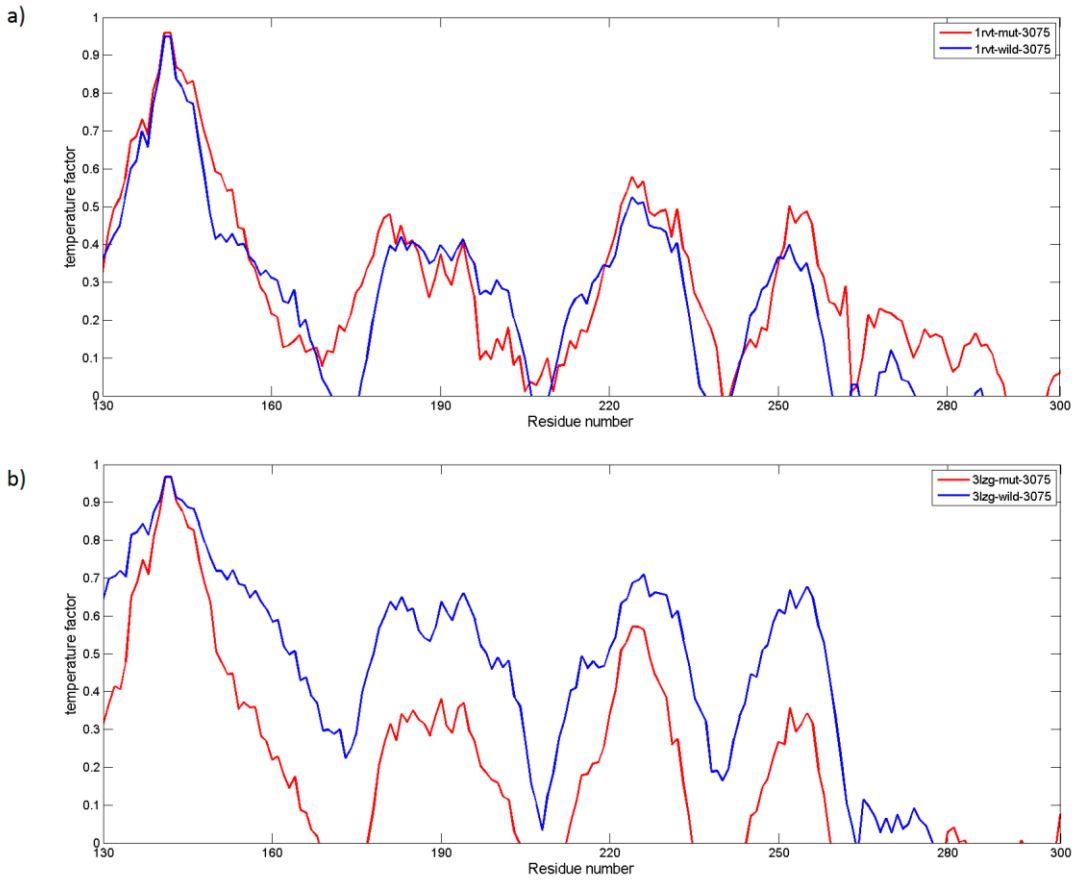

Figure 3.25. Average cross-correlation values for 75 ns analysis (a) 1RVT, b) 3LZG (red: mutant type, blue: wild type).

Figure 3.25 shows 1RVT and 3LZG proteins with their mutant and wild types for 75 ns simulation. The plots are obtained by the average cross-correlation results of residues 141 and 142. The peaks correspond to positions 190 helix (RBS, 188-194), Sb (antigenic site, 187-198), 220 loop (RBS, 221-228), Ca_2 (antigenic site, 224-225). The expected result for 1RVT is to observe higher correlated motion for the mutant type than the wild type in the RBD. The expectation is the reverse for 3LZG, the correlation at the RBD is higher for the wild type than the mutant type.

Considering the variants separately, it can be said that the mutant type 1RVT has relatively higher average correlation values at the maxima points than the wild type (170-182, 223-238, 251-290). On the other hand, the wild type 3LZG has distinctly higher values than the mutant type. The corresponding values of the antigenic sites (other than Ca_2) increase and get closer to each other for the mutant type 1RVT and the wild type 3LZG.

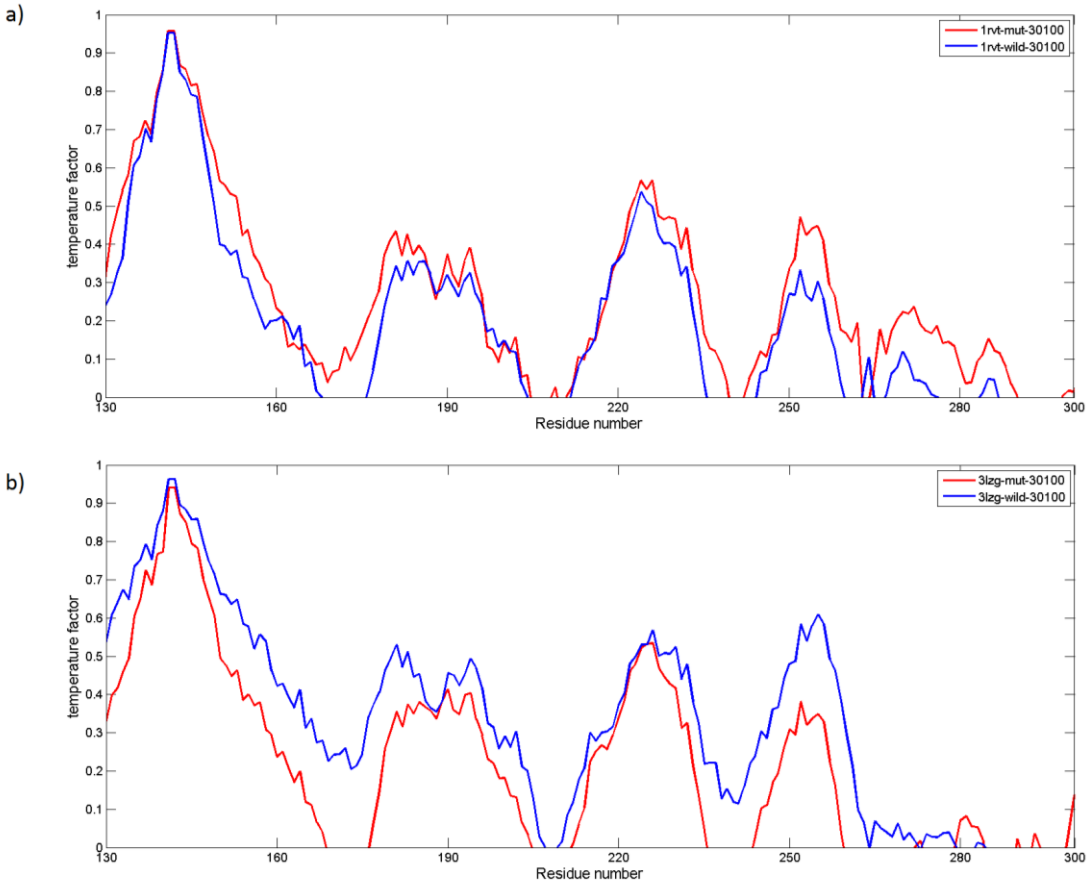

Figure 3.26. Average cross-correlation values for 100 ns analysis (a) 1RVT b) 3LZG (red: mutant type, blue: wild type).

Figure 3.26 shows 1RVT and 3LZG proteins with their mutant and wild types for 100 ns simulation. The plots are obtained by the average cross-correlation results of residues 141 and 142. Similar to Figure 3.25, the peaks correspond to positions 190 helix (RBS, 188-194), Sb (antigenic site, 187-198), 220 loop (RBS, 221-228), Ca_2 (antigenic site, 224-225).

Considering the variants separately, it can be said that the mutant type 1RVT has relatively higher average correlation values at the maxima points than the wild type (150-160, 168-185, 250-290). On the other hand, the wild type 3LZG has higher values than the mutant type. The corresponding values of the antigenic sites (other than Ca_2) increase and get closer to each other for the mutant type 1RVT and the wild type 3LZG.

4. CONCLUSIONS AND RECOMMENDATIONS

4.1. Conclusions

The analysis of dynamics of HA of two Influenza A Virus variants (A/swine/Iowa/15/1930 (PDB ID: 1RVT) and A/California/04/2009 (PDB ID: 3LZG) display that the antigenic and host receptor binding behavior of RBD could be affected by allosteric mutations. Mutations on mechanistically important sites such as hinge residues by affecting global dynamics have the possibility to lead to an antigenic drift, which is a common phenomenon for Hemagglutinin protein.

The analysis of HA as a monomer and in the context of trimer demonstrates that although the overall residue mobility profiles are similar between the two states, the cooperativity between the subunits are important to observe the dynamic difference between R149 and K149 variants.

The residue mobility profiles demonstrate that maxima points mostly correspond to host receptor and antigenic binding sites, where the minima points are the hinge positions. Minima positions at the residue fluctuations refer to restricted behavior of motion for the important residues. Hinge residues are key sites in coordinating the cooperativity in structures. Thus, any mutation in a hinge position is highly possible that affects the cooperativity of the structure; i.e. it can lead to a higher or lower cooperativity. Here, a decrease in residue mobilities lead to enhanced cooperativity of HA with the K in 149^{th} position. Thus the mutation at this position affects the intrinsic cooperativity between the antigenic and sialic binding residues of the binding protein of H1N1 Influenza Virus.

This work reveals change of cooperativity and mobility upon mutation of one residue within RBD of HA. The results serve as a case study where we could observe for the sequence, structure, dynamics and function paradigm.

4.2. Recommendations

H1N1 Influenza Virus is one of a highly investigated virus. Several mutations in the binding protein of Influenza (HA) are reported to make antigenic drift or alterations in the

binding behavior in the structure. Mutations at residue positions in RBD are known to effect virulence of the protein. In this work, residue 149 is chosen based on the importance of the position as discriminating based on the sequence analysis; this residue stays between host receptor and antigenic binding sites on RBD. There are other residues reported to be discriminating or having antigenic effect as well. As a future work, in silico mutations with respect to sequence versus dynamics analysis will be carried out to observe the dynamic effect to further suggests for experimental verification.

APPENDIX A: SUPPORTING FIGURES

Table A.1. Average RMSD values corresponding to the last 30 ns time span of the simulation time for the monomeric structure.

Structure name	Mutant type 1RVT	Wild type 1RVT	Mutant type 3LZG	Wild type 3LZG
RMSD value (Å)	4.74	2.91	6.26	3.34

Table A.1. shows that mutant types of both variants show higher RMSD values than the wild types for the time span of 70 ns to the end of simulation time. This behavior can also be observed in the plots given in Figure 3.1.

Table A.2. RMSD values between the monomeric structures to the monomeric subunit of the original PDB files after 100 ns MD simulation.

Structure		RMSD (Å)
1RVT wild type	1RVT_HI	2.62
1RVT mutant type	1RVT_HI	3.26
3LZG wild type	3LZG_AB	2.55
3LZG mutant type	3LZG_AB	5.38
1RVT mutant type	1RVT wild type	1.61
3LZG mutant type	3LZG wild type	4.46

Table A.3. MSF values of minima and maxima points of the mutant and the wild type 1RVT monomeric structures read from MSF plots (bold-underlined: different peak between wild and mutant types, italic-underlined: RBS).

WILD TYPE		MUTANT TYPE	
Minima	Maxima	Minima	Maxima
10-12	15-17	10-12	15-17
19-21	24-27	30-32	24-27
30-32	49-50	65-67	49-50
65-67	**87-88**	148-151	**90-91**
92-95	**125-128** (Sa)	178-182	**124-125**
148-151	142-144 (Ca_2)	202-204	142-144 (Ca_2)
178-182	158-160 (Sa)	212-215	158-160 (Sa)
202-204	*187-189*(Sb)	231-235	**171-174** (Ca_1)
212-215	196-199 (Sb)	251-256	*187-189* (Sb)
231-235	207-209 (Ca_1)	320-322	196-199 (Sb)
251-256	*218-222*	520-522 (HA2)	207-209 (Ca_1)
320-322	238-241 (Ca_1)	607-614 (HA2)	*218-222*
520-522 (HA2)	263-264		238-241 (Ca_1)
607-614 (HA2)	**272-274**		263-264
	313-314		313-314
	529-531 (HA2)		529-531 (HA2)
	538 (HA2)		579-581 (HA2)
	565-567 (HA2)		632-634 (HA2)
	579-581 (HA2)		**644-646** (HA2)
	632-634 (HA2)		

Table A.4. MSF values of minima and maxima points of the mutant and the wild type 3LZG monomeric structures read from MSF plots (bold-underlined: different peak between wild and mutant types, italic-underlined: RBS).

WILD TYPE		MUTANT TYPE	
Minima	Maxima	Minima	Maxima
15-18	22-24	15-18	22-24
25-27	29-33	25-27	29-33
36-39	55-56	36-39	**48-50**
67-73	124-125	67-73	55-56
148-152	142-144 (Ca_2)	148-152	**79-82**
178-181	157-160 (Sa)	178-181	124-125
202-205	172-174 (Ca_1)	202-205	142-144 (Ca_2)
212-216	*189-198* (Sb)	212-216	157-160 (Sa)
231-235	207-209 (Ca_1)	231-235	172-174 (Ca_1)
251-256	219	251-256	*189-198* (Sb)
20-23 (HA2)	239-241 (Ca_1)	20-23 (HA2)	207-209 (Ca_1)
109-118 (HA2)	30-32 (HA2)	109-118 (HA2)	218-220
136-138 (HA2)	64-66 (HA2)	136-138 (HA2)	239-241 (Ca_1)
	127-129 (HA2)		**262-264**
			271-273
			30-32 (HA2)
			64-66 (HA2)
			71-74 (HA2)
			127-129 (HA2)

Table A.5. Average RMSD values corresponding to the last 30 ns time span of the simulation time for the trimer structure.

Structure name	Mutant type 1RVT	Wild type 1RVT	Mutant type 3LZG	Wild type 3LZG
RMSD value (Å)	4.64	4.27	5.22	5.44

Table A.6. RMSD values between the isolated monomeric subunit of each trimer structure with its monomeric structure from first simulation.

Structure name	Wild type 1RVT	Mutant type 1RVT	Wild type 3LZG	Mutant type 3LZG
RMSD value (Å)	2.19	2.40	4.16	4.31

Table A.7. RMSD values for the last frame of the trajectory of each variant between the mutant type and the wild type.

Structure name	1RVT	3LZG
RMSD value (Å)	2.18	2.82

Table A.8. MSF values of minima and maxima points of the mutant and the wild type 1RVT trimer structures read from MSF plots (bold-underlined: different peak between wild and mutant types, italic-underlined: RBS).

WILD TYPE		MUTANT TYPE	
Minima	Maxima	Minima	Maxima
9-11	15-18	9-11	15-17
21-24	26-27	21-24	26-27
61-66	50	**35-37**	50
81-85	57-58	61-66	57-58
100-106	**73**	80-84	**78** (Cb)
139	88-89	102-109	87-88
148-151	98	148-153	98
177-181	124-125	176-181	125
232-235	142-143 (Ca_2)	232-236	142-143 (Ca_2)
268-270	158-159 (Sa)	**253-261**	158-159 (Sa)
282-285	*188-189 (Sb)*	266-269	*188-189 (Sb)*
294-296	196-199 (Sb)	282-285	196-199 (Sb)
309-311	207 (Ca_1)	294-297	*221-224*(Ca_2)
320-322	*218-222*	309-311	239-241 (Ca_1)
505-506 (HA2)	240-242 (Ca_1)	**317-320**	263
521-523 (HA2)	263-265	505-506 (HA2)	**273-274**
636-637 (HA2)	276-278	521-523 (HA2)	277-279
	292	**547** (HA2)	292
	313-314	**550-551** (HA2)	313-314
	518-519 (HA2)	**607-612** (HA2)	518-519 (HA2)
	529-530 (HA2)	636-638 (HA2)	529-530 (HA2)
	560 (HA2)		560 (HA2)
	572-573 (HA2)		572-573 (HA2)
	627-629 (HA2)		627-629 (HA2)
	631-633 (HA2)		634-636 (HA2)
	635-637 (HA2)		

Table A.9. MSF values of minima and maxima points of the mutant and the wild type 3LZG trimer structures read from MSF plots (bold-underlined: different peak between wild and mutant types, italic-underlined: RBS).

WILD TYPE		MUTANT TYPE	
Minima	Maxima	Minima	Maxima
15-18	21-23	15-17	21
26-28	32	27-29	32-33
35-38	55	59-60	**48**
58-59	62-63	65-71	55
65-68	**80**	86-89	62
70-71	**91**	107-112	**78-79** (Cb)
85-88	128 (Sa)	*139*	**99-100**
105-111	142-143 (Ca_2)	148-151	128-129 (Sa)
138-139	158-159 (Sa)	177-180	143 (Ca_2)
148-152	*188-189* (Sb)	204	158-159 (Sa)
177-181	**207** (Ca_1)	232-236	*188-189* (Sb)
202-204	219	**260-262**	**196-199** (Sb)
211-213	*225* (Ca_2)	**268**	218-219
232-235	**240** (Ca_1)	283-285	**248**
265-266	273	**294-296**	**265**
281-283	276	307-311	273
308-309	312	**319-321**	276
21-24	19	21-22	**289**
47	30	48	**302-304**
74-79	**60**	**108**	312-313
112		**152-153**	19
		166-167	29-30
			58-59
			72-73
			146-147
			158-161

These plots represent the average cross correlation values of residues 141 and 142; and show differences between the mutant and wild types of each variant for 75 ns and 100 ns analysis.

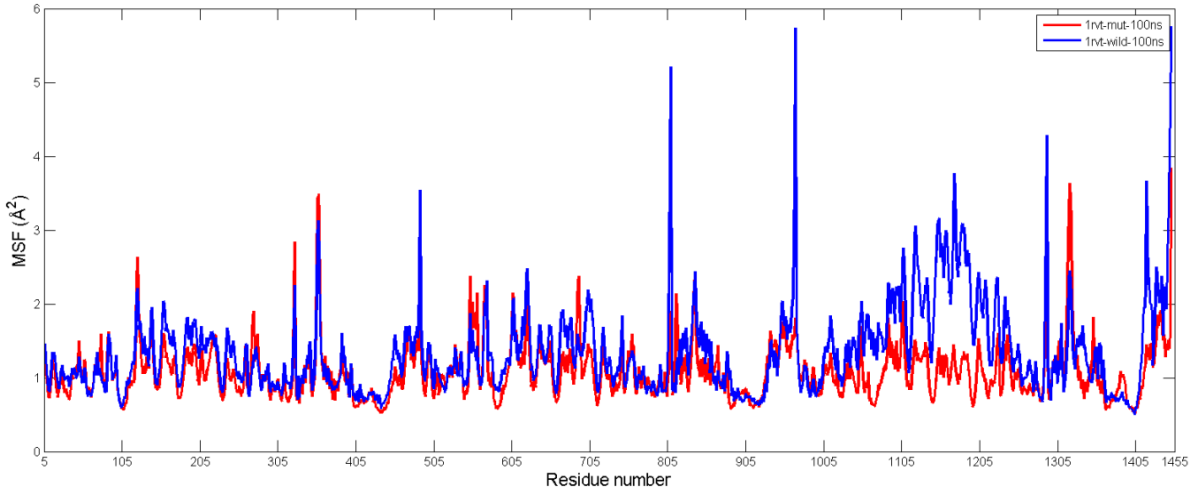

Figure A.1. MSF values for the trimer structure of 1RVT (red legend for mutant and blue for wild type).

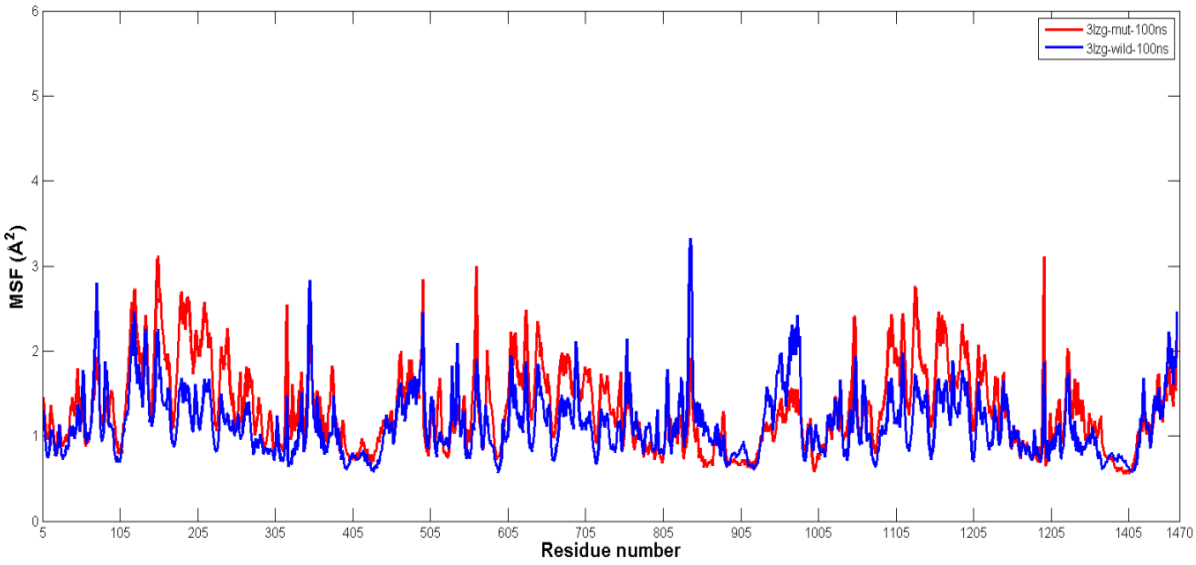

Figure A.2. MSF values for the trimer structure of 3LZG (red legend for mutant and blue for wild type).

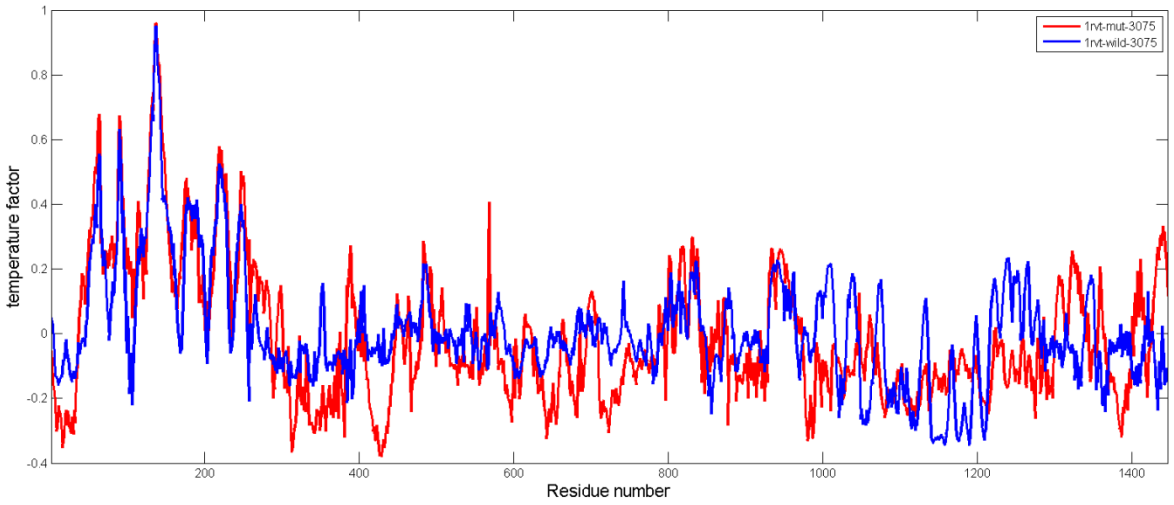

Figure A.3. Temperature factor values obtained from average cross correlation results for 1RVT protein for last 45 ns analysis of the 75 ns simulation (red: mutant type 1RVT, blue: wild type 1RVT).

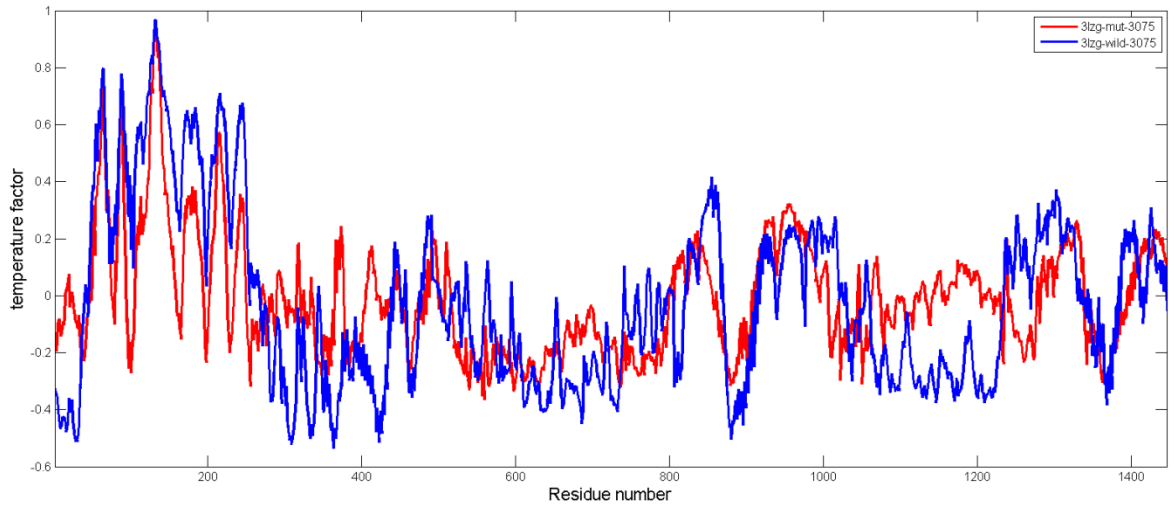

Figure A.4. Temperature factor values obtained from average cross correlation results for 3LZG protein for last 45 ns analysis of the 75 ns simulation (red: mutant type 3LZG, blue: wild type 3LZG).

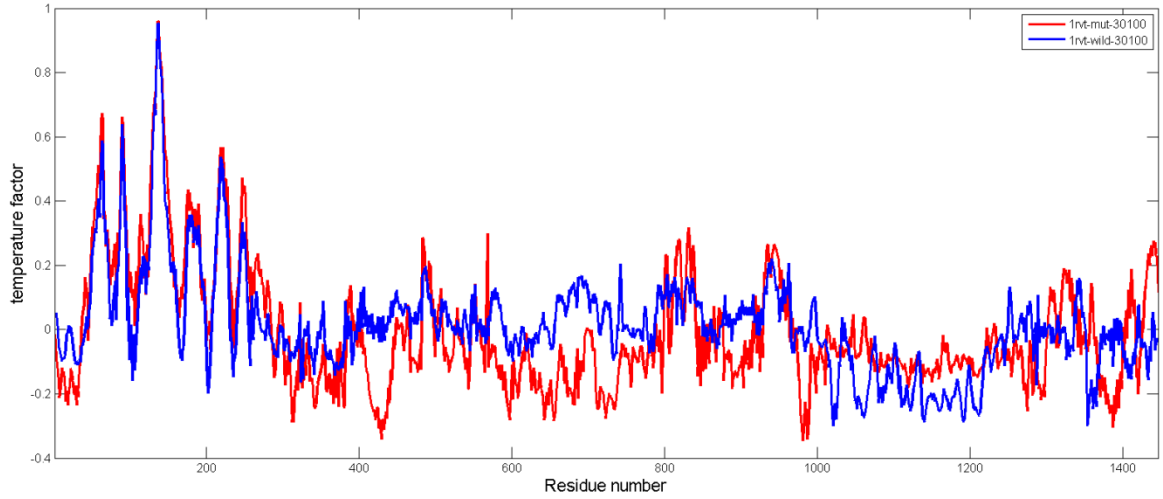

Figure A.5. Temperature factor values obtained from average cross correlation results for 1RVT protein for last 70 ns analysis of the 100 ns simulation (red: mutant type 1RVT, blue: wild type 1RVT).

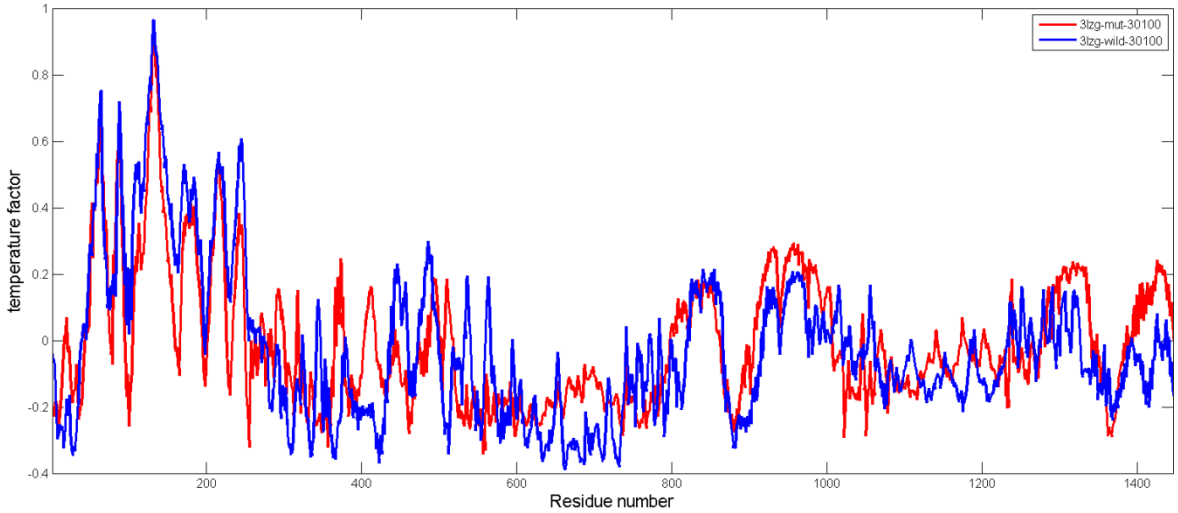

Figure A.6. Temperature factor values obtained from average cross correlation results for 3LZG protein for last 70 ns analysis of the 100 ns simulation (red: mutant type 3LZG, blue: wild type 3LZG).

APPENDIX B: FASTA RESULTS

```
RID: EFKBU6EP114


Query= 3LZG:A|PDBID|CHAIN|SEQUENCE

Length=329


                                                                      Score     E
Sequences producing significant alignments:                          (Bits)  Value

lcl|49153  1RVT:H|PDBID|CHAIN|SEQUENCE                                 587   0.0

ALIGNMENTS
>lcl|49153 1RVT:H|PDBID|CHAIN|SEQUENCE
Length=328

 Score =   587 bits (1513),  Expect = 0.0, Method: Compositional matrix adjust.
 Identities = 270/324 (83%), Positives = 297/324 (92%), Gaps = 0/324 (0%)

Query  2    GDTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDKHNGKLCKLRGVAPLHLGKCNIAG  61
            DTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLED HNGKLC+L G+APL LGKCNIAG
Sbjct  4    ADTLCIGYHANNSTDTVDTVLEKNVTVTHSVNLLEDSHNGKLCRLGGIAPLQLGKCNIAG  63

Query  62   WILGNPECESLSTASSWSYIVETPSSDNGTCYPGDFIDYEELREQLSSVSSFERFEIFPK  121
            W+LGNPEC+ L T SSWSYIVET +SDNGTCYPGDFIDYEELREQLSSVSSFE+FEIFPK
Sbjct  64   WLLGNPECDLLLTVSSWSYIVETSNSDNGTCYPGDFIDYEELREQLSSVSSFEKFEIFPK  123

Query  122  TSSWPNHDSNKGVTAACPHAGAKSFYKNLIWLVKKGNSYPKLSKSYINDKGKEVLVLWGI  181
            TSSWPNH++ +GVTAACP+AGA SFY+NL+WLVKKGNSYPKLSKSY+N+KGKEVLVLWG+
Sbjct  124  TSSWPNHETTRGVTAACPYAGASSFYRNLLWLVKKGNSYPKLSKSYVNNKGKEVLVLWGV  183

Query  182  HHPSTSADQQSLYQNADTYVFVGSSRYSKKFKPEIAIRPKVRDQEGRMNYYWTLVEPGDK  241
            HHP TS DQQSLYQNAD YV VGSS+Y ++F PEIA RPKVR Q GRMNYYWTL+EPGD
Sbjct  184  HHPPTSTDQQSLYQNADAYVSVGSSKYDRRFTPEIAARPKVRGQAGRMNYYWTLLEPGDT  243

Query  242  ITFEATGNLVVPRYAFAMERNAGSGIIISDTPVHDCNTTCQTPKGAINTSLPFQNIHPIT  301
            ITFEATGNLV PRYAFA+ R +GSGII SD PVHDC+T CQTP GAIN+SLPFQNIHP+T
Sbjct  244  ITFEATGNLVAPRYAFALNRGSGSGIITSDAPVHDCDTKCQTPHGAINSSLPFQNIHPVT  303

Query  302  IGKCPKYVKSTKLRLATGLRNIPS  325
            IG+CPKYVKSTKLR+ATGLRNIP+
Sbjct  304  IGECPKYVKSTKLRMATGLRNIPA  327
```

REFERENCES

Edwards, M. J. and N. J. Dimmock, 2001, "A Haemagglutinin (HA1)-Specific FAb Neutralizes Influenza A Virus by Inhibiting Fusion Activity", *The Journal of General Virology,* Vol. 82, No. 6, pp. 1387-1395.

Gamblin, S. J., L. F. Haire, R. J. Russell, D. J. Stevens, B. Xiao, Y. Ha, N. Vasisht, D. A. Steinhauer, R. S. Daniels, A. Elliot, D. C. Wiley and J. J. Skehel, 2004, "The Structure and Receptor Binding Properties of the 1918 Influenza Hemagglutinin", *Science (New York, N.Y.),* Vol. 303, No. 5665, pp. 1838-1842.

Gao, Z. and K. A. Jacobson, 2006, "Keynote Review : Allosterism in Membrane Receptors", *Elsevier,* Vol. 11, No. 5, pp. 191-202.

Glaser, L., J. Stevens, D. Zamarin, I. Wilson, A. Garcia-Sastre, T. M. Tumpey, C. F. Basler, J. K. Taubenberger and P. Palese, 2005, "A Single Amino Acid Substitution in 1918 Influenza Virus Hemagglutinin Changes Receptor Binding Specificity", *Journal of Virology*, Vol. 79, No. 17, pp. 11533-11536.

Grove, J. and M. Marsh, 2011, "The Cell Biology of Receptor-Mediated Virus Entry", *The Journal of Cell Biology*, Vol. 195, No. 7, pp. 1071-1082.

Hensley, S. E., S. R. Das, A. L. Bailey, L. M. Schmidt, H. D. Hickman, A. Jayaraman, K. Viswanathan, R. Raman, R. Sasisekharan, J. R. Bennink and J. W. Yewdell, 2009, "Hemagglutinin Receptor Binding Avidity Drives Influenza A Virus Antigenic Drift", *Science (New York, N.Y.),* Vol. 326, No. 5953, pp. 734-736.

Humphrey, W., A. Dalke and K. Schulten, 1996, "VMD - Visual Molecular Dynamics", *Journal of Molecular Graphics*, Vol. 14, No. 1, pp. 33-38.

Isin, B., P. Doruker and I. Bahar, 2002, "Functional Motions of Influenza Virus Hemagglutinin: A Structure-Based Analytical Approach", *Biophysical Journal,* Vol.

82, No. 2, pp. 569-581.

Jiao, Wanting and E. J. Parker, 2012, "Using a Combination of Computational and Experimental Techniques to Understand the Molecular Basis for Protein Allostery", *Advances in Protein Chemistry and Structural Biology*, Vol. 87, pp. 391-413.

Kumari, K., S. Gulati, D. F. Smith, U. Gulati, R. D. Cummings and G. M. Air, 2007, "Receptor Binding Specificity of Recent Human H3N2 Influenza Viruses", *Virology Journal*, Vol. 4, No. 42.

Martín J, S. A. Wharton, Y. P. Lin, D. K. Takemoto, J. J. Skehel, D. C. Wiley and D. A. Steinhauer, 1998, "Studies of the Binding Properties of Influenza Hemagglutinin Receptor-Site Mutants", *Virology*, Vol. 241, No. 1, pp. 101-111.

Meroz, D., S. Yoon, M. F. Ducatez, T. P. Fabrizio, R. J. Webby, T. Hertz and N. Ben-Tal, 2011, "Putative Amino Acid Determinants of the Emergence of the 2009 Influenza A (H1N1) Virus in the Human Population", *PNAS*, Vol. 108, No. 33.

Molecular Operating Environment (MOE), 2013.08; Chemical Computing Group Inc., 1010 Sherbooke St. West, Suite #910, Montreal, QC, Canada, H3A 2R7, 2013.

Morizono, K. and I. Chen, 2011, "Receptors and Tropisms of Envelope Viruses", *Current Opinion in Virology*, Vol. 1, No. 1, pp. 13-18.

Mudhakir, D. and H. Harashima, 2009, "Learning from the Viral Journey: How to Enter Cells and How to Overcome Intracellular Barriers to Reach the Nucleus", *The AAPS Journal*, Vol. 11, No. 1, pp. 65-77.

Nicholls, J. M., R. W. Y. Chan, R. J. Russell, G. M. Air and J. S. M. Peiris, 2008, "Evolving Complexities of Influenza Virus and its Receptors", *Trends in Microbiology*, Vol. 16, No. 4, pp. 149-157.

Rogers G. N., J. C. Paulson, R. S. Daniels, J. J. Skehel, I. A. Wilson and D. C. Wiley,

1983, “Single Amino Acid Substitutions in Influenza Haemagglutinin Change Receptor Binding Specificity”, *Nature*, Vol. 304, No. 5921, pp. 76-78.

Schneider-Schaulies, J., 2000, “Cellular Receptors for Viruses: Links to Tropism and Pathogenesis”, *Journal of General Virology*, Vol. 81, pp. 1413-1429.

Skehel J. J. and D. C. Wiley, 2000, “Receptor Binding and Membrane Fusion in Virus Entry: The Influenza Hemagglutinin”, *Annual Review of Biochemistry*, Vol. 69, pp. 531-569.

Smith, A. E. and A. Helenius, 2004 “How Viruses Enter Animal Cells”, *Science (New York, N.Y.)*, Vol. 304, No. 5668, pp. 237-242.

Soundararajan, V., S. Zheng, N. Patel, K. Warnock, R. Raman, I. A. Wilson, R., S. Raguram, and R. Sasisekharan, 2011, “Networks Link Antigenic and Receptor-Binding Sites of Influenza Hemagglutinin: Mechanistic Insight into Fitter Strain Propagation”, *Scientific Reports*, Vol. 1, No. 200.

Sriwilaijaroen, N., and Y. Suzuki, 2012, “Molecular Basis of The Structure and Function of H1 Hemagglutinin of Influenza Virus”, *Proceedings of the Japan Academy Series B*, Vol. 88, No. 6, pp. 226-249

Tharakaraman, K., R. Raman, N. W. Stebbins, K. Viswanathan, V. Sasisekharan, and R. Sasisekharan, 2013, “Antigenically Intact Hemagglutinin in Circulating Avian and Swine Influenza Viruses and Potential for H3N2 Pandemic”, *Scientific Reports*, Vol. 3, pp. 1822-1830.

DeLano, W. L., “*The PyMOL Molecular Graphics System”*, 2000, http://www.pymol.org/, May 2013.

Tumpey, T. M., A. García-Sastre, J. K. Taubenberger, P. Palese, D. E. Swayne, and C. F. Basler, 2004, “Pathogenicity and Immunogenicity of Influenza Viruses with Genes from the 1918 Pandemic Virus”, *Proceedings of the National Academy of Sciences of*

the United States of America, Vol. 101, No. 9, pp. 3166-3171.

Ye, J., E. M. Sorrell, Y. Cai, H. Shao, K. Xu, L. Pena, H. Song, M. Angel, R. A. Medina, B. Manicassamy, A. Garcia-Sastre, D. R. Perez, 2010, "Variations in the Hemagglutinin of the 2009 H1N1 Pandemic Virus : Potential for Strains with Altered Virulence Phenotype", *PLOS Pathogens*, Vol. 6, No. 10.

Zou, W., D. Chen, M. Xiong, J. Zhu, X. Lin, L. Wang, J. Zhang, L. Chen, H. Zhang, H. Chen, M. Chen, and M. Jin, 2013, "Insights into the Increasing Virulence of the Swine-origin Pandemic H1N1/2009 Influenza Virus", *Scientific Reports*, Vol. 3, p. 1601.