Molecular dynamics simulation study of effects of key mutations in SARS-CoV-2 on protein structures

Jerome Rumdon Lon[a, b, #], Binbin Xi[a, #], Bingxu Zhong[a], Yiyuan Zheng[a], Pei Guo[a], Zixi Chen[a], Ruoran Qiu[a], Siqing Zhang[a], Hongli Du[a, *]

[a] School of Biology and Biological Engineering, South China University of Technology, Guangzhou, China

[b] Institute of Synthetic Biology, Shenzhen Institute of Advanced Technology, Shenzhen, China

[#] Equal contribution

[*] Correspondence: hldu@scut.edu.cn

Summary

SARS-CoV-2 has been spreading rapidly since 2019 and has produced large-scale mutations in the genomes. Differences in gene sequences may lead to changes in protein structure and traits, which would have a great impact on the epidemiological characteristics. In this study, we selected the key mutations of SARS-CoV-2, including D614G and A222V of S protein and Q57H of ORF3a protein, to conduct molecular dynamics simulation and analysis on the structures of the mutant proteins. The results suggested that D614G improved the stability of S protein, while A222V enhanced the ability of protein to react with the outside environment. Q57H enhanced the structural flexibility of ORF3a protein. Our findings could complete the mechanistic link between genotype--phenotype--epidemiological characteristics in the study of SARS-CoV-2. We also found no significant changes in the antigenicity of S protein, ORF3a protein and their mutants, which provides reference for vaccine development and application.

Subject areas

Virology; Molecular biology; Molecular dynamics; Bioinformatics

Introduction

In the winter of 2019, a new coronavirus, SARS-Cov-2, came into view, and the resulting COVID-19 quickly swept the world. By 0:00 on January 11, 2021 (GMT+8:00), the total number of diagnosed patients has exceeded 90.27 million. Covid-19 has been a major blow to public health, social order and economic development around the world.

As the virus spreads, it accumulates a large number of genomic mutations that alter the structure of proteins, which would increase the complexity of the virus subpopulation and pose a threat to prevention and control measures such as virus detection, drug development and the practical application of vaccines. In the case of MERS-CoV, the I529T, D510G mutations in RBD lead to decreased affinity with the human CD26 receptor (Kim et al., 2016), but increased resistance to antibody-mediated neutralization. In the case of SARS-CoV (Meulen et al., 2006), the P462L mutation of S protein can achieve the escape of human mAb Cr3014 (Kleine-Weber et al., 2019).

The genome of severe acute respiratory syndrome-coronavirus 2 (SARS-CoV-2) is composed of ~30,000 nucleotides and contains 12 open□reading frames (ORFs). SARS-CoV-2 encodes 4 structural including the spike (S), envelope (E), membrane (M) and nucleocapsid (N) proteins and 22 nonstructural proteins (nsp) which are nsp1–16 encoded by ORF1ab and 6 accessory proteins, *i.e.* ORF3a, 6, 7a, 7b, 8, and 9b (Chan et al., 2020). Many mutations have also been present in SARS-CoV-2. Recently, the United Kingdom, South Africa, Nigeria and other places have reported the emergence and prevalence of mutant strains (Tegally et al., 2020; Leung et al., 2021; Mastriani et al., 2021; Shaibu et al., 2021). More disturbingly, a SARS-CoV-2

variant that achieved in vitro escapement from highly neutralizing COVID-19 convalescent plasma has been reported (Andreano et al., 2020). Therefore, real-time monitoring of the epidemic trends of key mutations and haplotypes in evolution, and exploring the influence of mutations on virus phenotypes and the relationship between mutations and epidemiological characteristics are of great significance for the development of targeted prevention and control measures. The key mutations of this study were screened out from the previous research, revealed the D614G and A222V in S (Spike) protein and Q57H in ORF3a as the key mutations(Bai et al., 2020; Xi et al., 2020).

The D614G mutation began spreading in Europe in early February, which has been reported to significantly increase the infectivity of SARS-CoV-2 in continuous cell lines(Hou et al., 2020; Korber et al., 2020; Plante et al., 2020; Xi et al., 2020; Yurkovetskiy et al., 2020; Zhang et al., 2020). The A222V variant, first reported from Spain in March, is rapidly spread since July across Europe and accounted for about 16.3% of all GISAID submitted genomes and the Q57H variant appeared around February 7, 2020 and accounted for about 16.7% of all GISAID submitted genomes (accessed Nov 29, 2020) (Xi et al., 2020). These imply A222V variant and Q57H variant may also play an important role in increasing infectivity.

The transmembrane protein, the S protein, mediates coronavirus entry into host cells. Because of its important role in virus transmission (Tortorici and Veesler, 2019), the study of S protein has always been the focus of coronavirus-related research. The protein forms a homotrimer on the surface of the virus, containing the S1 subunit that binds to the host cell receptor and the S subunit that is involved in the fusion of the viral membrane to the cell membrane. A series of studies have shown that two subunits of the S protein of coronavirus bind non-covalently in a prefusion conformation, while host proteases lyse the S protein during viral entry into the cell. This modification has been proposed to activate the protein for membrane fusion by irreversible conformational changes (Millet and Whittaker, 2015; Kirchdoerfer et al., 2016; Walls et al., 2016; Walls et al., 2020).

Because the S protein is exposed to the surface and plays an important role in transmission, it is considered to be the primary target of post-infection neutralizing antibodies and is the focus of treatment and vaccine design. Due to its important role in viral mechanism and prevention measures, the mutation of S protein is likely to have an impact on the phenotype and epidemiological characteristics of the virus, and even the function of antibodies.

ORF3a, one of the ion channels of SARS-CoV-2, is a dimer with three transmembrane helices in each, located in the lipid nanodisk (Kern et al., 2020). Its coding region has been reported to be highly conserved in bat coronaviruses including SARS-CoV-1 and related bat coronaviruses. In mouse models of SARS-CoV-1 infection, ORF3a genomic deletions reduced viral titers and morbidity, making ORF3a a potential target for the treatment of SARS (Issa et al., 2020). Therefore, the structural change of ORF3A caused by the mutation is likely to influence the phenotype and thus affect the epidemiological characteristics of SARS-CoV-2.

In this study, we have performed molecular dynamic (MD) simulations to evaluate the effect of D614G single-site mutation and D614G&A222V double-site mutation on the structural dynamics of S protein, as well as the effect of Q57H mutation on the structural dynamics of ORF3a. The results showed that D614G improved the stability of S protein, while A222V improved the ability of protein to react with the outside environment. Q57H enhanced the flexibility of ORF3a protein, which affected the pore size and internal resistance of the channel. In addition, we also analyzed the antigenicity of D614G, D614G&A222V and Q57H variants, it was found that the change of structure had no significant effect on the antigenicity of the protein molecules. Taken together, our results provide new insights into the in-depth studies of the relationship of "genotype-phenotypic-epidemiological characteristics" of SARS-CoV-2, which would help to better understand the virus and address the complexities of COVID-19 prevention, diagnosis, and treatment.

Results

D614G improved the stability of S protein, while A222V improved the ability of protein to react with the outside environment

Based on the reported cryo-electron microscopy structure of S protein (PDB ID: 6X6P), we constructed D614G and D614G&A222V of which sequences differ from the wide-type (WT) S protein by changing the aspartic acid to glycine at the 614th amino acid site, and changing both the aspartic acid to glycine at the 614th amino acid site as well as the alanine to valine at the 222th amino acid site (Figure 1A). We exploited 100-ns MD simulations of the D614G, D614G&A222V and wide-type S proteins. To verify the convergence of MD simulation equilibrium, we analyzed the root mean square deviations (RMSD) throughout the MD process (Figure 2A), after excluding the initial effect of model relaxation due to the introduction of the force field in the first half of simulation(Pereira et al., 2019). The RMSD of wild-type as well as D614G and D614G&A222V mutants appeared to be stable in the latter 50ns, revealing a good convergence was achieved for each system. The latter 50ns could be considered as the stable period. Interestingly, the RMSD of these two mutants were found to be slightly smaller than that of the wide-type S protein especially in the latter 25 ns (Figure 2A). For the two mutants, their RMSD fluctuated within a small range and did not show an obvious difference in the latter 25 ns.

To further investigate the structural fluctuations and flexibilities of the wide-type, D614G and D614G&A222V, we further evaluated the detailed residual atomic fluctuations by computing the root mean square fluctuations (RMSF) of protein Cα atoms (Figure 3A). It should be noted that the RMSF of highly dynamic terminal residues are not shown. For S protein, RMSF analysis showed that the overall trend of flexibility of WT, D614G, D614G&A222V did not change too much, and the mutation at position 222 led to a steep increase in local structure flexibility, which was noteworthy.

The structural compactness of the wide-type, D614G and D614G&A222V mutants of

S protein was studied by evaluating the radius gyration (R_g) which provides information for the overall size of the structure. It was found that the wide type has the largest R_g value followed by D614G&222V, and D614G has the smallest R_g value (Figure 4A), suggesting that the structure of D614G is more compact than D614G&A222V and the wide type. These structures suggest that the D614G&222V of S protein is more compact than WT and more stretchable relative to D614G.T

Analysis of SASA as a surface accessible to proteins provides information about the ability of proteins to interact with other molecules. The whole molecular SASA of wide-type, D614G and D614G&A222V mutants of S protein was analyzed(Figure 5A). The results showed that under stable state, the SASA of wide type was significantly higher than that of mutant D614G, and was similar to but slightly higher than that of D614G&A222V, and the two fluctuated within a small range.

We also performed the hydrogen bond analysis, which showed that the average number of hydrogen bonds in the wild-type was lower than that of D615G and D614G was lower than that of D614G&A222V(Figure 6A) for S protein. These results indicate that the S protein mutants, especially the mutants D614G&A222V interact more with the solvent, the stability is higher than that of the wild-type of S protein.

Q57H enhanced the flexibility of ORF3a protein

In the latter 50ns, RMSD of Q57h was significantly higher than WT, but the values of both were stable in a small range(Figure 2B). Q57H and WT of ORF3a can be considered to be stable in the analysis stage.

The RMSF analysis revealed that the mutant Q57H was more flexible than WT on the overall trend of ORF3a, while the mutant Q57H was less flexible at the position 57(Figure 3B). Although there is a local decline in RMSF at the mutant residue, this indicates that the conversion of amino acid from glutamine(Q) to histidine(H) at position 57 reduces the flexibility of the protein nearby that residue. However, the significant increase in total RMSF suggests that local amino acid changes affect the

whole protein conformation and its properties.

The analysis of R_g showed that the Q57H mutant was slightly larger than WT(Figure 4B). Which suggest the WT of ORF3a is slightly more compact than that of Q57H mutants. The difference of protein volume and tightness reflects the degree of contact between the protein and the external environment to a certain extent, and the difference between WT and Q57H of ORF3a also implies the difference of their response ability to the external environment to a certain extent.

For ORF3a protein, the SASA of WT and Q57H are similar in stable state, fluctuating in a small range and Q57H is slightly higher than WT(Figure 5B), suggests that the wild-type ORF3a protein is less accessible. The difference of SASA could affect the ability of the wild-type or mutant of these proteins to interact with other molecules.

The average number of wild-type hydrogen bonds was higher than that of Q57H(Figure 6B). These results indicate that the wild-type protein of ORF3a interact more with the solvent, and their stability is higher than thatQ57H mutants.

No significant change in the antigenicity of S protein, ORF3a protein and their mutants was found

The development of vaccine is one of the most important means of long-term prevention and control of the virus. Antigenicity is an indicator of the ability of a peptide to bind to its complementary antibodies (van Regenmortel, 2001). The antigenicity was analyzed by Ellipro (Ponomarenko et al., 2008), a structure-based web-tool that implements Thornton's method and, together with a residue clustering algorithm, the MODELLER program and the Jmol viewer. The antigenicity of S protein and ORF3a protein was analyzed respectively (Figure 7). The results showed that the antigenicity of S protein and ORF3a protein fluctuated highly in different local regions, but there was no significant change in the overall trend. This indicated that these mutations did not change the ability of S protein and ORF3a protein of SARS-CoV-2 to react with the antibodies.

Discussion

S protein is located on the surface of SARS-CoV-2, and it has been indicated that the S protein of SARS-CoV-2 has a characteristic Furin cleavage site at the boundary between S1/S2 subunits.(Madu et al., 2009; Millet and Whittaker, 2015; Walls et al., 2020) S protein mediates the entry of SARS-CoV-2 into cells through ACE2. The amino acid variation in the SARS-CoV-2 region will lead to the change of virus infection characteristics, which is related to the effective transmission of SARS-CoV-2 in human population.

ORF3a protein is a large conductance cation channel of SARS-CoV-2, which is moderately selective for K+ and Na+ and can exist in the form of trimer, tetramer and potential high-order oligomer.(Kern et al., 2020) ORF3a has been shown to be associated with virulence, infectivity, ion channel formation and viral release.(Issa et al., 2020) The function of ORF3a is conserved in common mutations, but the pore of the channel is altered by variation(Kern et al., 2020).

Mutation is one of the main mechanisms by which viruses are constantly changing due to genetic selection. Although most point mutations are neutral and do not alter the protein the gene codes for, a few favorable mutations can give an evolutionary advantage to viruses (Wu et al., 2020). The D614G substitution in the gene encoding the spike protein emerged in late January or early February 2020, and emerged in Europe around 2020 February 22 (Korber et al., 2020). Since then, the D614G variant quickly became popular in Europe and by June 2020 became the dominant form of the virus circulating globally (Korber et al., 2020; Xi et al., 2020; World Health Organization, 2021). The D614G mutation has been reported to change the S protein conformation and enhance the efficiency of protease cleavage at the S1/S2 subunit junctions, which may be one of the reasons for promoting its infection to the host and improving the transmission efficiency of SARS-CoV-2 (Gobeil et al., 2021). Studies on the D614G mutant of S protein at the atomic simulation level indicate that D614G

has stronger adaptability and higher transmissible carbon skeleton (Omotuyi et al., 2020). A222V has recently been reported as a rapidly increasing genotype (Bartolini et al., 2020; Ward et al., 2021). Q57H of ORF3a has been reported that it shows an opposing association with decreased deaths and increased cases per million (Oulas et al., 2021), It also causes significant structural changes that affect the interaction between ORF3a and other proteins (Wu et al., 2020). From the level of molecular simulation, our study explored the change of protein traits when D614G and A222V were coupled for S protein, and the influence of Q57H on protein traits for ORF3a protein. In general, we have conducted in-depth studies on the changes of the characteristics of these two proteins and their mutants at the molecular level, which provides reference information for explaining the phenomena found in other studies.

The effect of amino acid mutations on proteins can be thoroughly analyzed using molecular dynamics simulations. Molecular dynamics simulation is an important tool in the study of biomolecules, which has been widely used in the study of protein stability, conformational change and other fields. The purpose of molecular dynamics simulation is to reproduce the real behavior of proteins in their links and calculate their trajectories over time. During this process, the detailed information of protein conformational changes and fluctuations can be used to evaluate structural parameters such as stability and flexibility.(Krebs and Mesquita, 2016) Changes in protein structural parameters may result in pathological phenotypes(Kumar et al., 2014; Khan et al., 2016). In this study, the S protein, ORF3a protein and their variants of SARS-CoV-2 were simulated by molecular dynamics, aiming to study the effects of the D614G and A222V mutations of S protein and the Q57H mutation of ORF3a protein on the structure and function of SARS-CoV-2 and the relationship between them and the epidemiological characteristics.

During the analysis phase in this study, RMSD values were observed to be in a stationary state, indicating that the structure fluctuated around a stable mean conformation. Therefore, it is meaningful to evaluate the fluctuation of local structure.(Martínez, 2015; Pereira et al., 2019)

During the stable period (the latter 50ns) of the simulation process, the two proteins and their mutants showed different properties. For S protein, the RMSD (based on the analysis of Cα) of D614G(0.4894485) and D614G&A222V (0.5085670) has a certain decline compared with wild type (0.5376845), this suggests that the mutation of the residue located on No. 614 to glycine(G) from aspartic acid (D) promoted S protein stability (Arooj et al., 2020), the stability of the ascension is proportional to the effectiveness (Zhang et al., 2020). This is also consistent with current epidemiological data. This phenomenon was also verified in hydrogen bond analysis, since the stability of protein structure is determined by the number of interactions (Pikkemaat et al., 2002), and the average number of hydrogen bonds between the mutant D614G (2348.866627) and D614G&A222V (2354.032793) in the stable state was higher than that of the wild type (2311.066187). The difference between the two mutants showed another phenomenon, although the number of hydrogen bonds in the D614G&A222V mutants was higher than that in the D614G mutants, the RMSD was lower. This suggests that the increase in hydrogen bonding not only promotes the stability of the protein, but also affects other properties. The analysis of SASA (habitual accessible surface areas) provides evidence to explain this phenomenon. Compared with D614G (1237.652733), D614G&A222V (1264.366661) has a higher value of SASA, which indicates that D614G&A222V has a higher protein interaction ability (Joshi et al., 2017) while retaining the stability brought by the mutation at 614, and is more likely to react with the outside environment. This further enhanced the efficiency of transmission, which is consistent with epidemiological data.

In the case of ORF3a, an increase in the RMSD of the mutant Q57H relative to the wild-type and a decrease in the mean hydrogen bond number suggest a partial loss of stability. However, from the increase of Rg, we can know that the Q57H mutant of ORF3a results in largen in volume, and the increase of SASA indicates the increase of ORF3a protein's reaction ability to the external environment (Joshi et al., 2017). For the flexibility, although the mutation of amino acid No. 57 reduced the local flexibility near the residue, the overall RMSF increased compared with the wild type,

indicating that the flexibility of the whole ORF3a protein increased, so its channel pore size expanded and its internal space resistance decreased (Zhang et al., 2019). In general, the performance of ORF3a as a channel protein has been improved, which may be a beneficial mutation for SARS-CoV-2, which is consistent with the epidemiological data.

Our study provides a mechanistic hypothesis for the change in transmissibility caused by mutations in the SARS-CoV-2 genotype. Although these mutations and their consequences are disturbing, fortunately, our analysis of the antigenicity of S proteins and ORF3a suggests that these mutations do not significantly alter their ability to react with antibodies. The prevention and control of SARS-CoV-2 is a long process. With the emergence of more and more mutants, the mechanism of "genotype -- phenotype -- epidemiological characteristics" caused by mutations will become one of the key directions of future research. Our study provides some reference for the continuous exploration of this problem, and also provides information from the perspective of molecular dynamics for the deeper understanding of SARS-CoV-2.

Limitations of the Study

In this study, we used MD simulation to explore the molecular mechanism by which key mutations in SARS-CoV-2 affect transmission. Due to the fluctuation of molecular simulation calculation, the calculation result is bound to have some residuals. MD provides a unique perspective at the molecular level for relevant studies, and the epidemiological trend is consistent with the results of MD. However, the simulation results can only be more accurate after the verification of relevant experiments. Computers cannot fully simulate the movement of real proteins, which is the limitation of this project.

Resource Availability

Lead Contact

Further information and requests for resources and reagents should be directed to and will be fulfilled by the Lead Contact, Hongli Du (hldu@scut.edu.cn)

Materials Availability

This study did not generate new unique reagents.

Data and Code Availability

Raw data tables for RMSD, RMSF, RG, SASA, H-Bond, and Antigenicity are provided in the supplementary information. The simulation data is stored in the HPC hardware housed at the Tianhe-2 and can be shared with the corresponding author on request.

Methods

The cryo-electron microscopic structures of the wide type S protein (PDB ID: 6X6P) and ORF3a (PDB ID: 7KJR) were obtained from the Protein Data Bank (Berman et al., 2002). The key mutation information of S protein and ORF3a protein was previously studied by Xi et al 2020. The sequence information comes from the A chain of the two proteins. The point mutation of protein was conducted by MOE2019.

Using the structures obtained from PDB as initial templates, we induced mutations in Molecular Operating Environment 2019 to obtain S protein mutants D614G, D614G&A222V and ORF3a protein mutant Q57H. For each system, the MD simulations were performed using Gromacs 2018.3, and the Gromacs 54A7 force field was selected (Huang et al., 2011). All structures were solvated with SPCE (Berendsen et al., 1987; Zielkiewicz, 2005) water molecules in a cubic simulation box, and Na^+ and Cl^- ions were added to neutralize the system. We first performed a 5000-step energy minimization using the steepest descent method, followed by three processes in the MD simulation: NVT, NPT, and production. The NVT ensemble

(constant particle number, volume, and temperature) in 300 K for 100 ps, followed by 100 ps in the NPT ensemble (constant particle number, pressure, and temperature) at 1 atm. The above simulations are all locational constraints, in which all the bonds are confined to the protein to relax the water around the protein and reduce the entropy of the system. Select Parrinello□Rahman as the barometric regulator and V-Resale as the thermostat.

Finally, a 100-ns production run was performed at 300 K. LINCS algorithm was used to constrain all the bonds, and the cutoff distances of 12 Å for the long-range electrostatic through the Particle Mesh Ewald (PME) method . A time step of 0.002 ps was selected in the simulation without constant force being applied.

The antigenicity information is analyzed using the Ellipro (Ponomarenko et al., 2008) server with the default parameters.

The biological macromolecules presented in this paper were drawn with PyMol V2.3.2, and the statistical charts were drawn with Origin 2018.

Acknowledgements

This work was supported by the National Key R&D Program of China (2018YFC0910201), the Key R&D Program of Guangdong Province (2019B020226001), and the Science and the Technology Planning Project of Guangzhou (201704020176).

Author contributions

J.R.L. designed the experiments and analyzed the data. B.X. did the calculation of simulation. B.Z. did the calculation of antigenicity and helped finish the manuscript. Y.Z. and P.G. helped finish the manuscript. R.Q. and S.Z. helped helped with the introduction and disposal data. H.D. designed the project and helped finish the

manuscript.

Declaration of interests

The authors declare no competing financial interests.

References

Andreano, E., Piccini, G., Licastro, D., Casalino, L., Johnson, N.V., Paciello, I., Monego, S.D., Pantano, E., Manganaro, N., Manenti, A., Manna, R., Casa, E., Hyseni, I., Benincasa, L., Montomoli, E., Amaro, R.E., McLellan, J.S., Rappuoli, R., 2020. SARS-CoV-2 escape in vitro from a highly neutralizing COVID-19 convalescent plasma. bioRxiv : the preprint server for biology. https://doi.org/10.1101/2020.12.28.424451.

Arooj, M., Shehadi, I., Nassab, C.N., Mohamed, A.A., 2020. Physicochemical stability study of protein-benzoic acid complexes using molecular dynamics simulations. Amino acids 52, 1353–1362.

Bai, Y., Jiang, D., Lon, J.R., Chen, X., Hu, M., Lin, S., Chen, Z., Wang, X., Meng, Y., Du, H., 2020. Comprehensive evolution and molecular characteristics of a large number of SARS-CoV-2 genomes reveal its epidemic trends. International journal of infectious diseases : IJID : official publication of the International Society for Infectious Diseases 100, 164–173.

Bartolini, B., Rueca, M., Gruber, C.E.M., Messina, F., Giombini, E., Ippolito, G., Capobianchi, M.R., Di Caro, A., 2020. The newly introduced SARS-CoV-2 variant A222V is rapidly spreading in Lazio region, Italy.

Berendsen, H.J.C., Grigera, J.R., Straatsma, T.P., 1987. The missing term in effective pair potentials. J. Phys. Chem. 91, 6269–6271.

Berman, H.M., Battistuz, T., Bhat, T.N., Bluhm, W.F., Bourne, P.E., Burkhardt, K.,

Feng, Z., Gilliland, G.L., Iype, L., Jain, S., Fagan, P., Marvin, J., Padilla, D., Ravichandran, V., Schneider, B., Thanki, N., Weissig, H., Westbrook, J.D., Zardecki, C., 2002. The Protein Data Bank. Acta crystallographica. Section D, Biological crystallography 58, 899–907.

Gobeil, S.M.-C., Janowska, K., McDowell, S., Mansouri, K., Parks, R., Manne, K., Stalls, V., Kopp, M.F., Henderson, R., Edwards, R.J., Haynes, B.F., Acharya, P., 2021. D614G Mutation Alters SARS-CoV-2 Spike Conformation and Enhances Protease Cleavage at the S1/S2 Junction. Cell reports 34, 108630.

Hou, Y.J., Chiba, S., Halfmann, P., Ehre, C., Kuroda, M., Dinnon, K.H., Leist, S.R., Schäfer, A., Nakajima, N., Takahashi, K., Lee, R.E., Mascenik, T.M., Graham, R., Edwards, C.E., Tse, L.V., Okuda, K., Markmann, A.J., Bartelt, L., Silva, A. de, Margolis, D.M., Boucher, R.C., Randell, S.H., Suzuki, T., Gralinski, L.E., Kawaoka, Y., Baric, R.S., 2020. SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. Science 370, 1464–1468.

Huang, W., Lin, Z., van Gunsteren, W.F., 2011. Validation of the GROMOS 54A7 Force Field with Respect to β-Peptide Folding. Journal of chemical theory and computation 7, 1237–1243.

Issa, E., Merhi, G., Panossian, B., Salloum, T., Tokajian, S., 2020. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. mSystems 5.

Joshi, H., Bhatia, D., Krishnan, Y., Maiti, P.K., 2017. Probing the structure and in silico stability of cargo loaded DNA icosahedra using MD simulations. Nanoscale 9, 4467–4477.

Kern, D.M., Sorum, B., Hoel, C.M., Sridharan, S., Remis, J.P., Toso, D.B., Brohawn, S.G., 2020. Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. bioRxiv : the preprint server for biology. https://doi.org/10.1101/2020.06.17.156554.

Khan, F.I., Wei, D.-Q., Gu, K.-R., Hassan, M.I., Tabrez, S., 2016. Current updates on computer aided protein modeling and designing. International journal of biological macromolecules 85, 48–62.

Kim, Y., Cheon, S., Min, C.-K., Sohn, K.M., Kang, Y.J., Cha, Y.-J., Kang, J.-I., Han, S.K., Ha, N.-Y., Kim, G., Aigerim, A., Shin, H.M., Choi, M.-S., Kim, S., Cho, H.-S., Kim, Y.-S., Cho, N.-H., 2016. Spread of Mutant Middle East Respiratory Syndrome Coronavirus with Reduced Affinity to Human CD26 during the South Korean Outbreak. mBio 7, e00019.

Kirchdoerfer, R.N., Cottrell, C.A., Wang, N., Pallesen, J., Yassine, H.M., Turner, H.L., Corbett, K.S., Graham, B.S., McLellan, J.S., Ward, A.B., 2016. Pre-fusion structure of a human coronavirus spike protein. Nature 531, 118–121.

Kleine-Weber, H., Elzayat, M.T., Wang, L., Graham, B.S., Müller, M.A., Drosten, C., Pöhlmann, S., Hoffmann, M., 2019. Mutations in the Spike Protein of Middle East Respiratory Syndrome Coronavirus Transmitted in Korea Increase Resistance to Antibody-Mediated Neutralization. Journal of virology 93.

Korber, B., Fischer, W.M., Gnanakaran, S., Yoon, H., Theiler, J., Abfalterer, W., Hengartner, N., Giorgi, E.E., Bhattacharya, T., Foley, B., Hastie, K.M., Parker, M.D., Partridge, D.G., Evans, C.M., Freeman, T.M., Silva, T.I. de, McDanal, C., Perez, L.G., Tang, H., Moon-Walker, A., Whelan, S.P., LaBranche, C.C., Saphire, E.O., Montefiori, D.C., 2020. Tracking Changes in SARS-CoV-2 Spike: Evidence that D614G Increases Infectivity of the COVID-19 Virus. Cell 182, 812-827.e19.

Krebs, B.B., Mesquita, J.F. de, 2016. Amyotrophic Lateral Sclerosis Type 20 - In Silico Analysis and Molecular Dynamics Simulation of hnRNPA1. PloS one 11, e0158939.

Kumar, C.V., Swetha, R.G., Anbarasu, A., Ramaiah, S., 2014. Computational Analysis Reveals the Association of Threonine 118 Methionine Mutation in PMP22 Resulting in CMT-1A. Advances in bioinformatics 2014, 502618.

Leung, K., Shum, M.H., Leung, G.M., Lam, T.T., Wu, J.T., 2021. Early transmissibility assessment of the N501Y mutant strains of SARS-CoV-2 in the United Kingdom, October to November 2020. Euro surveillance : bulletin Europeen sur les maladies transmissibles = European communicable disease bulletin 26.

Madu, I.G., Roth, S.L., Belouzard, S., Whittaker, G.R., 2009. Characterization of a

highly conserved domain within the severe acute respiratory syndrome coronavirus spike protein S2 domain with characteristics of a viral fusion peptide. Journal of virology 83, 7411–7421.

Martínez, L., 2015. Automatic identification of mobile and rigid substructures in molecular dynamics simulations and fractional structural fluctuation analysis. PloS one 10, e0119264.

Mastriani, E., Rakov, A.V., Liu, S.-L., 2021. Isolating SARS-CoV-2 Strains From Countries in the Same Meridian: Genome Evolutionary Analysis. JMIR bioinformatics and biotechnology 2, e25995.

Meulen, J. ter, van den Brink, E.N., Poon, L.L.M., Marissen, W.E., Leung, C.S.W., Cox, F., Cheung, C.Y., Bakker, A.Q., Bogaards, J.A., van Deventer, E., Preiser, W., Doerr, H.W., Chow, V.T., Kruif, J. de, Peiris, J.S.M., Goudsmit, J., 2006. Human monoclonal antibody combination against SARS coronavirus: synergy and coverage of escape mutants. PLoS medicine 3, e237.

Millet, J.K., Whittaker, G.R., 2015. Host cell proteases: Critical determinants of coronavirus tropism and pathogenesis. Virus research 202, 120–134.

Omotuyi, I.O., Nash, O., Ajiboye, O.B., Iwegbulam, C.G., Oyinloye, E.B., Oyedeji, O.A., Kashim, Z.A., Okaiyeto, K., 2020. Atomistic simulation reveals structural mechanisms underlying D614G spike glycoprotein-enhanced fitness in SARS-COV-2. Journal of computational chemistry 41, 2158–2161.

Oulas, A., Zanti, M., Tomazou, M., Zachariou, M., Minadakis, G., Bourdakou, M.M., Pavlidis, P., Spyrou, G.M., 2021. Generalized linear models provide a measure of virulence for specific mutations in SARS-CoV-2 strains. PloS one 16, e0238665.

Pereira, G.R.C., Da Silva, A.N.R., Do Nascimento, S.S., Mesquita, J.F. de, 2019. In silico analysis and molecular dynamics simulation of human superoxide dismutase 3 (SOD3) genetic variants. Journal of cellular biochemistry 120, 3583–3598.

Pikkemaat, M.G., Linssen, A.B.M., Berendsen, H.J.C., Janssen, D.B., 2002. Molecular dynamics simulations as a tool for improving protein stability. Protein engineering 15, 185–192.

Plante, J.A., Liu, Y., Liu, J., Xia, H., Johnson, B.A., Lokugamage, K.G., Zhang, X.,

Muruato, A.E., Zou, J., Fontes-Garfias, C.R., Mirchandani, D., Scharton, D., Bilello, J.P., Ku, Z., An, Z., Kalveram, B., Freiberg, A.N., Menachery, V.D., Xie, X., Plante, K.S., Weaver, S.C., Shi, P.-Y., 2020. Spike mutation D614G alters SARS-CoV-2 fitness. Nature. https://doi.org/10.1038/s41586-020-2895-3.

Ponomarenko, J., Bui, H.-H., Li, W., Fusseder, N., Bourne, P.E., Sette, A., Peters, B., 2008. ElliPro: a new structure-based tool for the prediction of antibody epitopes. BMC bioinformatics 9, 514.

Shaibu, J.O., Onwuamah, C.K., James, A.B., Okwuraiwe, A.P., Amoo, O.S., Salu, O.B., Ige, F.A., Liboro, G., Odewale, E., Okoli, L.C., Ahmed, R.A., Achanya, D., Adesesan, A., Amuda, O.A., Sokei, J., Oyefolu, B.A.O., Salako, B.L., Omilabu, S.A., Audu, R.A., 2021. Full length genomic sanger sequencing and phylogenetic analysis of Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) in Nigeria. PloS one 16, e0243271.

Tegally, H., Wilkinson, E., Giovanetti, M., Iranzadeh, A., Fonseca, V., Giandhari, J., Doolabh, D., Pillay, S., San, E.J., Msomi, N., Mlisana, K., Gottberg, A. von, Walaza, S., Allam, M., Ismail, A., Mohale, T., Glass, A.J., Engelbrecht, S., van Zyl, G., Preiser, W., Petruccione, F., Sigal, A., Hardie, D., Marais, G., Hsiao, M., Korsman, S., Davies, M.-A., Tyers, L., Mudau, I., York, D., Maslo, C., Goedhals, D., Abrahams, S., Laguda-Akingba, O., Alisoltani-Dehkordi, A., Godzik, A., Wibmer, C.K., Sewell, B.T., Lourenço, J., Alcantara, L.C.J., Pond, S.L.K., Weaver, S., Martin, D., Lessells, R.J., Bhiman, J.N., Williamson, C., Oliveira, T. de, 2020. Emergence and rapid spread of a new severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) lineage with multiple spike mutations in South Africa.

Tortorici, M.A., Veesler, D., 2019. Structural insights into coronavirus entry. Advances in virus research 105, 93–116.

van Regenmortel, M.H., 2001. Antigenicity and immunogenicity of synthetic peptides. Biologicals : journal of the International Association of Biological Standardization 29, 209–213.

Walls, A.C., Park, Y.-J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020.

Structure, Function, and Antigenicity of the SARS-CoV-2 Spike Glycoprotein. Cell 183, 1735.

Walls, A.C., Tortorici, M.A., Bosch, B.-J., Frenz, B., Rottier, P.J.M., DiMaio, F., Rey, F.A., Veesler, D., 2016. Cryo-electron microscopy structure of a coronavirus spike glycoprotein trimer. Nature 531, 114–117.

Ward, D., Higgins, M., Phelan, J.E., Hibberd, M.L., Campino, S., Clark, T.G., 2021. An integrated in silico immuno-genetic analytical platform provides insights into COVID-19 serological and vaccine targets. Genome medicine 13, 4.

World Health Organization, 2021. SARS-CoV-2 Variants. World Health Organization. https://www.who.int/csr/don/31-december-2020-sars-cov2-variants/en/. Accessed 28 January 2021.

Wu, S., Tian, C., Liu, P., Guo, D., Zheng, W., Huang, X., Zhang, Y., Liu, L., 2020. Effects of SARS-CoV-2 mutations on protein structures and intraviral protein-protein interactions. Journal of medical virology. https://doi.org/10.1002/jmv.26597.

Xi, B., Jiang, D., Li, S., Lon, J.R., Bai, Y., Lin, S., Hu, M., Meng, Y., Qu, Y., Huang, Y., Liu, W., Du, H., 2020. Real-time monitoring epidemic trends and key mutations in SARS-CoV-2 evolution by an automated tool. https://doi.org/10.1101/2020.12.24.424271.

Yurkovetskiy, L., Wang, X., Pascal, K.E., Tomkins-Tinch, C., Nyalile, T., Wang, Y., Baum, A., Diehl, W.E., Dauphin, A., Carbone, C., Veinotte, K., Egri, S.B., Schaffner, S.F., Lemieux, J.E., Munro, J., Rafique, A., Barve, A., Sabeti, P.C., Kyratsous, C.A., Dudkina, N., Shen, K., Luban, J., 2020. SARS-CoV-2 Spike protein variant D614G increases infectivity and retains sensitivity to antibodies that target the receptor binding domain. bioRxiv : the preprint server for biology. https://doi.org/10.1101/2020.07.04.187757.

Zhang, L., Jackson, C.B., Mou, H., Ojha, A., Rangarajan, E.S., Izard, T., Farzan, M., Choe, H., 2020. The D614G mutation in the SARS-CoV-2 spike protein reduces S1 shedding and increases infectivity. bioRxiv : the preprint server for biology. https://doi.org/10.1101/2020.06.12.148726.

Zhang, W., Yang, T., Zhou, S., Cheng, J., Yuan, S., Lo, G.V., Dou, Y., 2019. Molecular Dynamics Simulation of Transmembrane Transport of Chloride Ions in Mutants of Channelrhodopsin. Biomolecules 9.

Zielkiewicz, J., 2005. Structural properties of water: comparison of the SPC, SPCE, TIP4P, and TIP5P models of water. The Journal of chemical physics 123, 104501.

Figure

Figure 1 The structure of S protein, ORF3a and mutants

A. The structure of ORF3a protein(trimer), the structure of single chain of ORF3a and the structure of key residues(mutants are translucent); B. The structure of S protein(trimer), the structure of single chain of S and the structure of key residues(mutants are translucent)

Figure 2 The RMSD of S protein, ORF3a protein and mutants

A. The RMSD of S protein and mutants; The RMSD of ORF3a protein and mutants

Figure 3 The RMSF of S protein, ORF3a protein and mutants

A. The RMSF of S protein and mutants; The RMSF of ORF3a protein and mutants

Figure 4 The Rg of S protein, ORF3a protein and mutants

A. The Rg of S protein and mutants; The Rg of ORF3a protein and mutants

Figure 5 The SASA of S protein, ORF3a protein and mutants

A. The SASA of S protein and mutants; The SASA of ORF3a protein and mutants

Figure 6 The H-bond of S protein, ORF3a protein and mutants

A. The H-bond of S protein and mutants; The H-bond of ORF3a protein and mutants

Figure 7 The antigenicity of S protein, ORF3a protein and mutants

A. The antigenicity of S protein and mutants; The antigenicity of ORF3a protein and mutants

Supplemental Information

Table S1 RMSD OF S

Table S2 RMSD OF ORF3a

Table S3 RMSF OF S

Table S4 RMSF OF ORF3a

Table S5 Rg OF S

Table S6 Rg OF ORF3a

Table S7 SASA OF S

Table S8 SASA OF ORF3a

Table S9 H-bond OF S

Table S10 H-bond OF ORF3a

Table S11 Antigenicity OF S

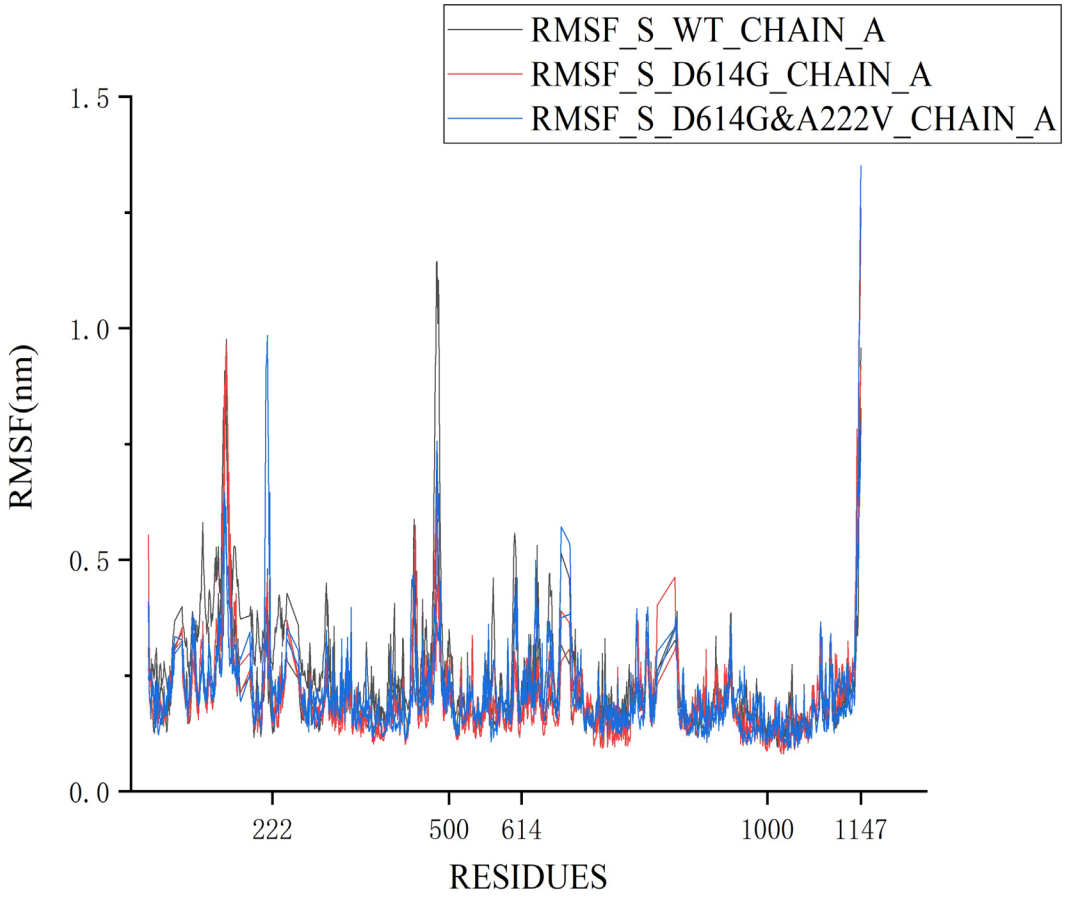
Table S12 Antigenicity OF ORF3a

Highlight

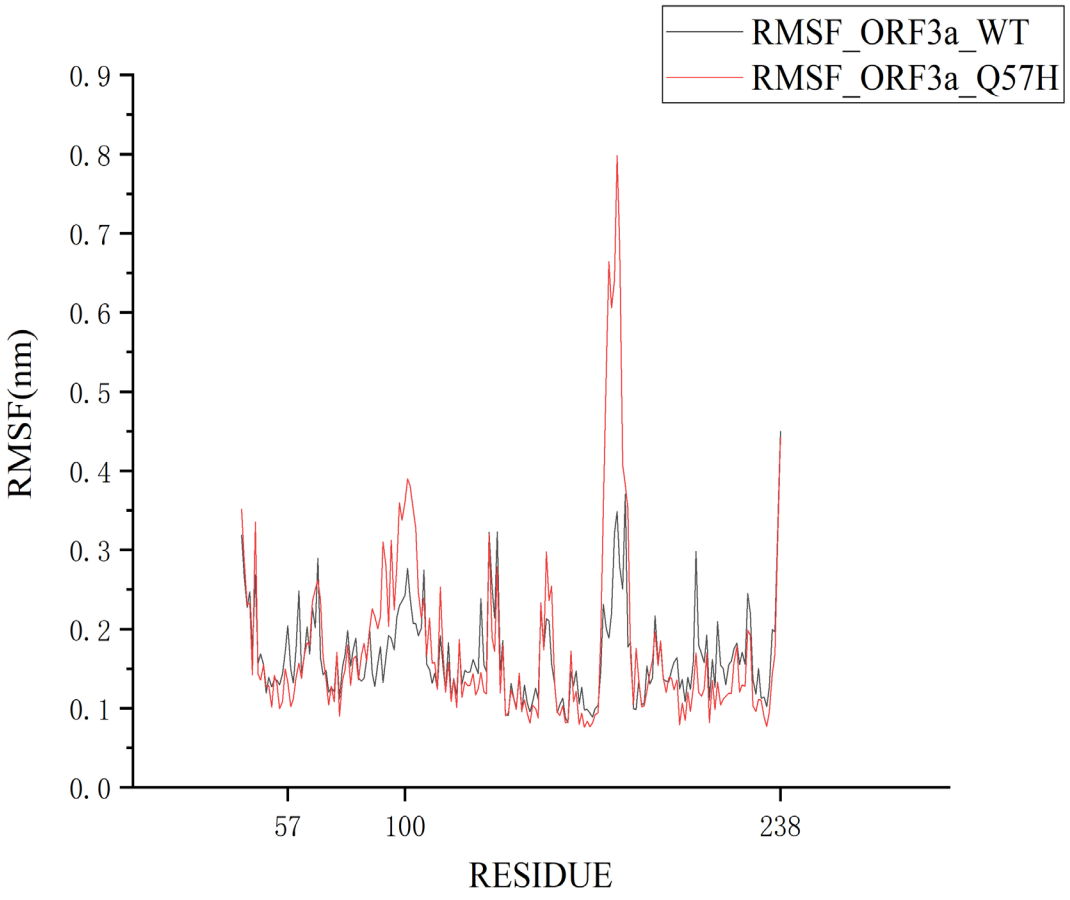
- The structures of D614G and D614G&A222V mutants of S protein, and Q57H mutant of ORF3a protein of SARS-CoV-2 were simulated by 100-ns total atomic molecular dynamics.

- D614G improved the stability of S protein, while A222V improved the ability of protein to react with the outside environment. Q57H enhanced the flexibility of ORF3a protein, which affected the pore size and internal resistance of the channel.
- The influence of differences of molecular property, which caused by mutations, in SARS-CoV-2 was analyzed from the perspective of genotype-phenotype-epidemiological characteristics.

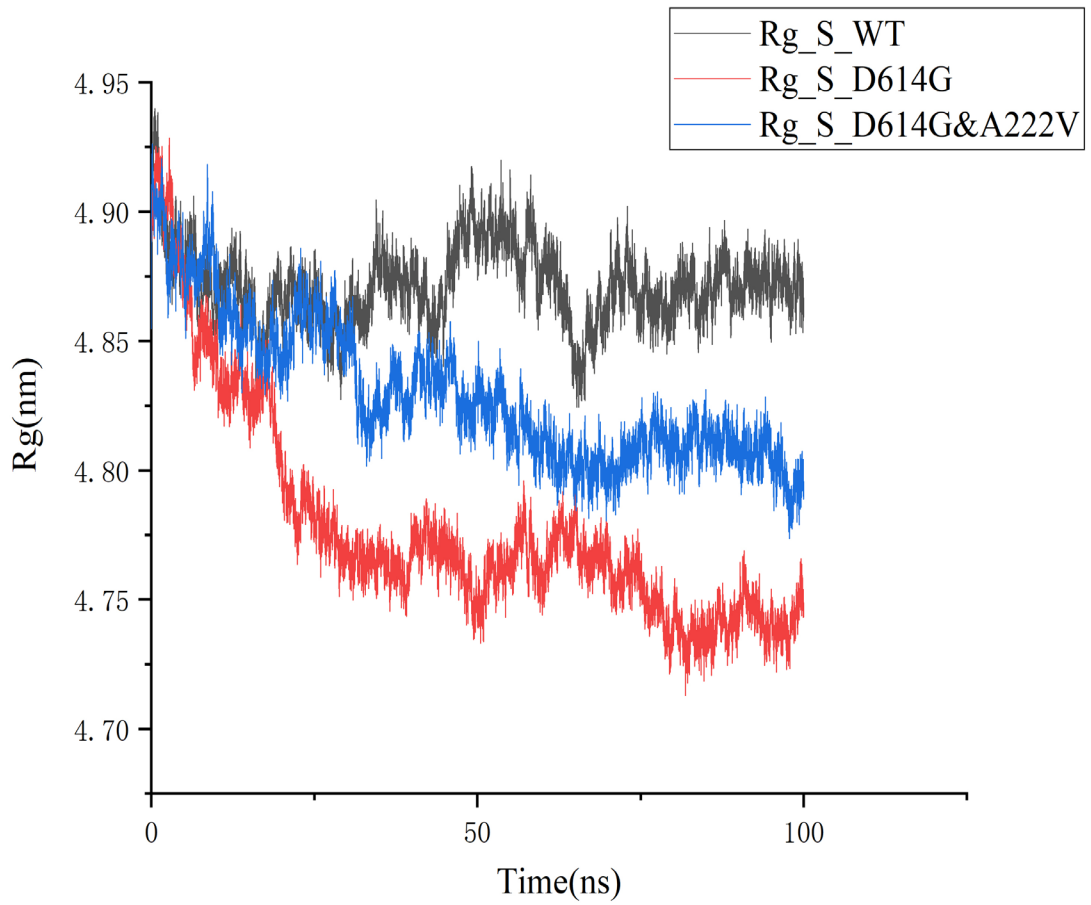
A


RMSF_S_WT_CHAIN_A
RMSF_S_D614G_CHAIN_A
RMSF_S_D614G&A222V_CHAIN_A
1.5
1.0
0.5
0.0
RMSF(nm)
222
500
614
1000
1147
RESIDUES

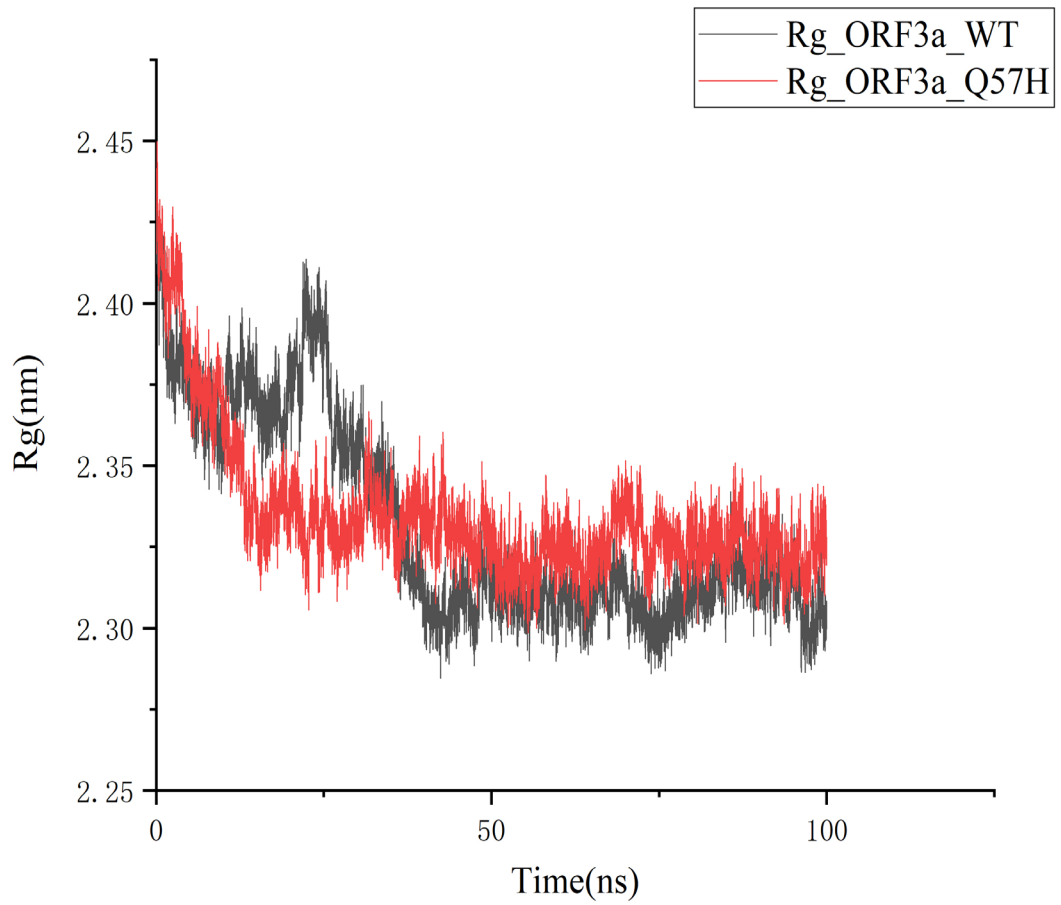
B


RMSF_ORF3a_WT
RMSF_ORF3a_Q57H
0.9
0.8
0.7
0.6
0.5
0.4
0.3
0.2
0.1
0.0
RMSF(nm)
57
100
238
RESIDUE

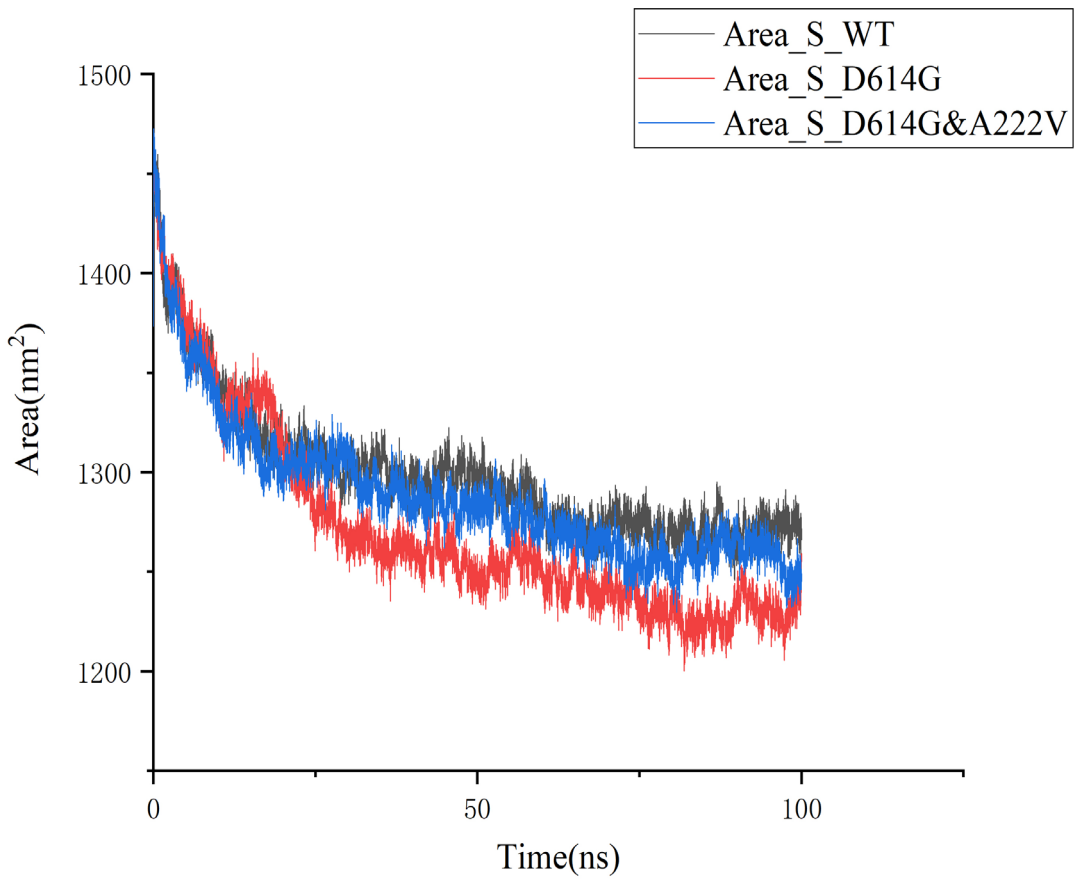
A


Rg_S_WT
Rg_S_D614G
Rg_S_D614G&A222V
4.95
4.90
4.85
4.80
4.75
4.70
Rg(nm)
0
50
100
Time(ns)

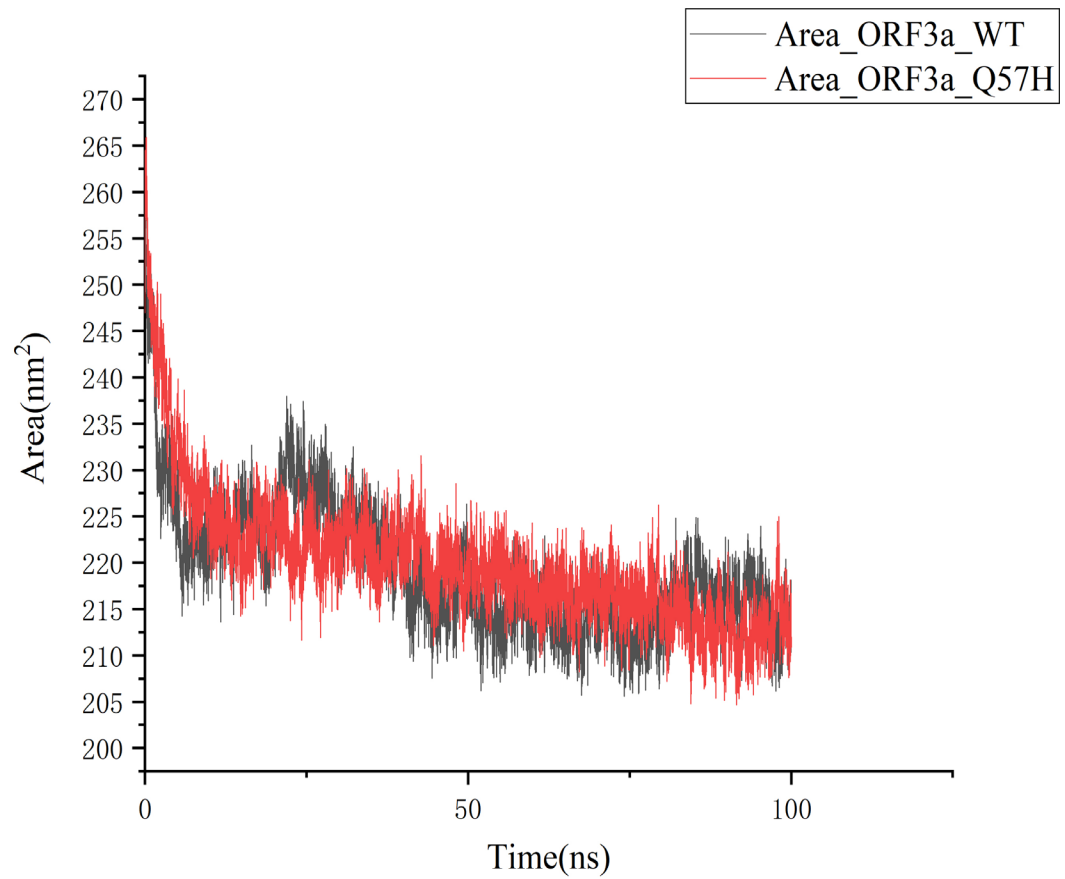
B


Rg_ORF3a_WT
Rg_ORF3a_Q57H
2.45
2.40
2.35
2.30
2.25
Rg(nm)
0
50
100
Time(ns)

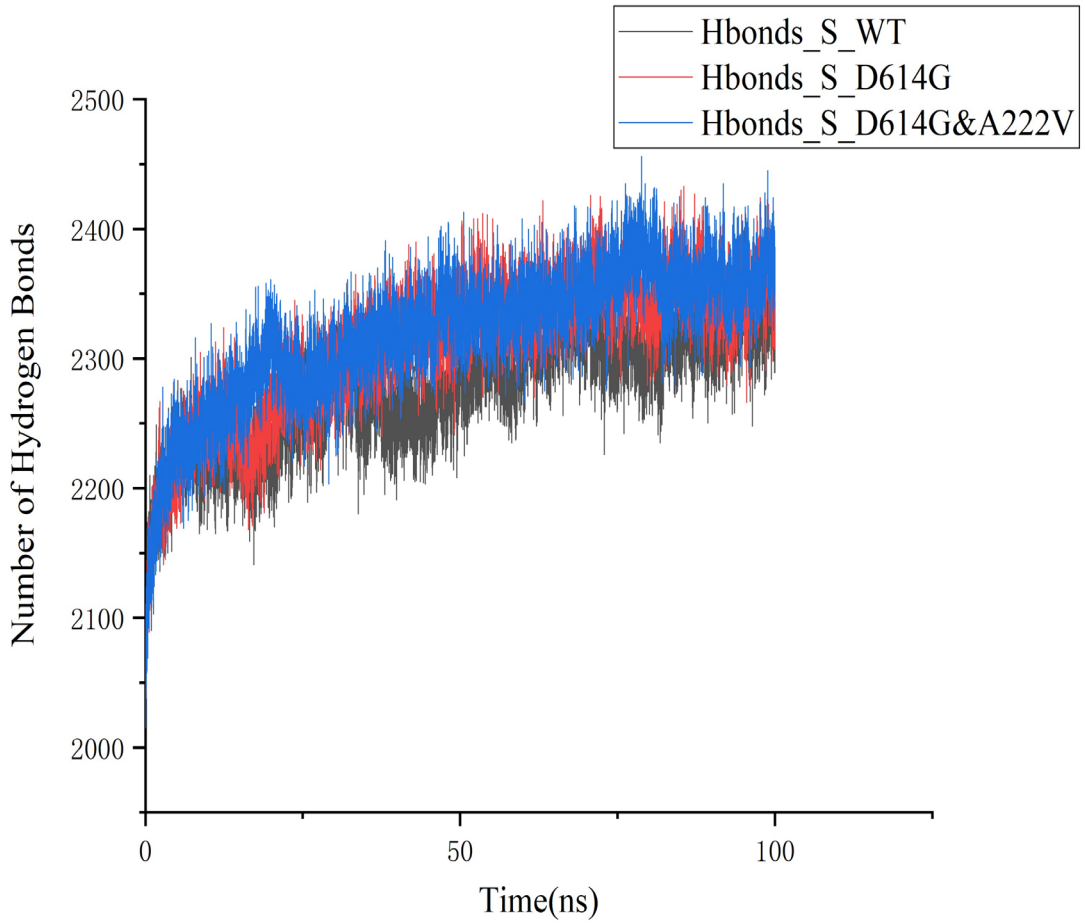
A


Area_S_WT
Area_S_D614G
Area_S_D614G&A222V
1500
1400
1300
1200
Area(nm2)
0
50
100
Time(ns)

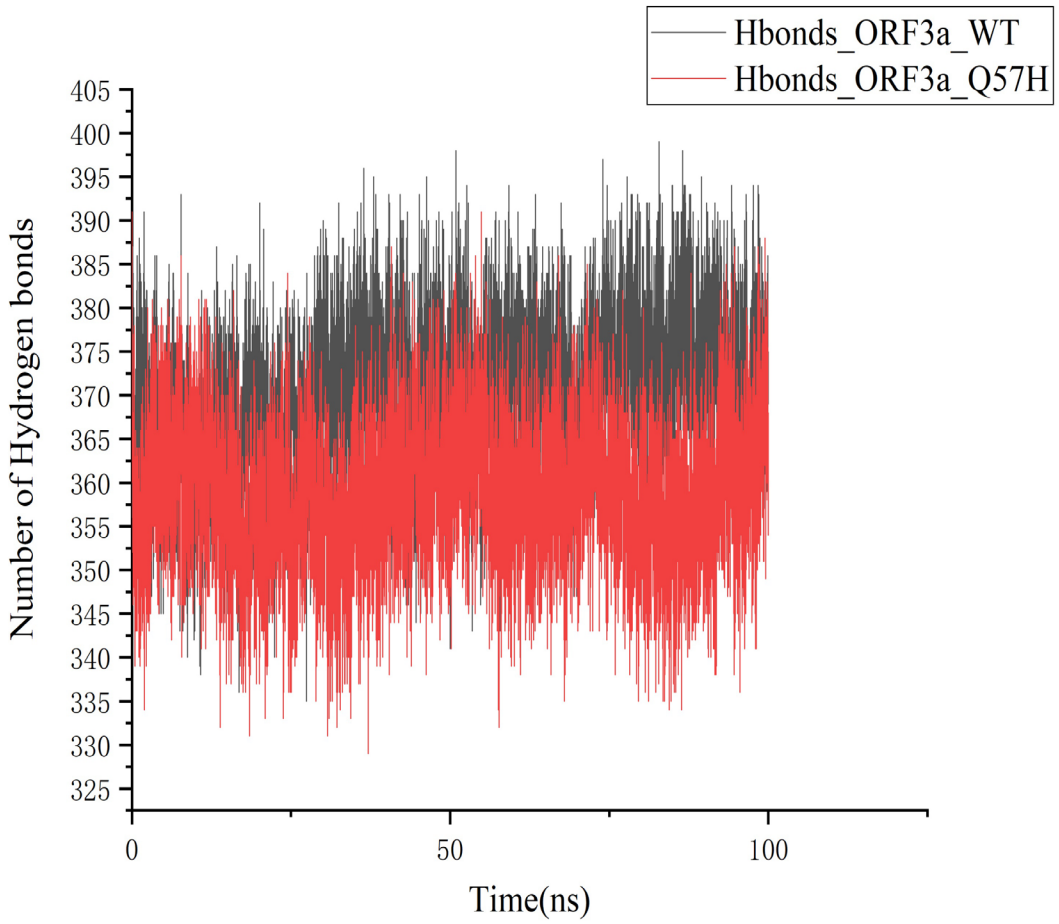
B


Area_ORF3a_WT
Area_ORF3a_Q57H
270
265
260
255
250
245
240
235
230
225
220
215
210
205
200
Area(nm2)
0
50
100
Time(ns)

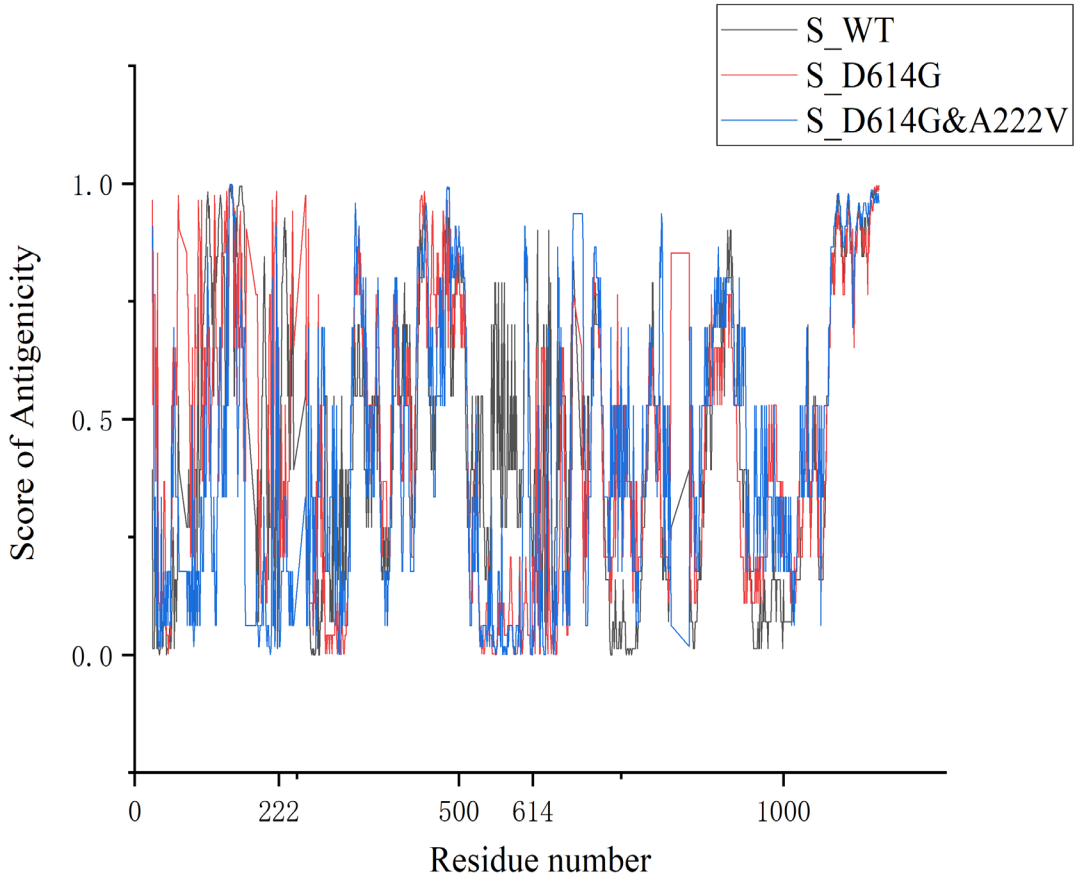
A


Hbonds_S_WT
Hbonds_S_D614G
Hbonds_S_D614G&A222V
Number of Hydrogen Bonds
2500
2400
2300
2200
2100
2000
0
50
100
Time(ns)

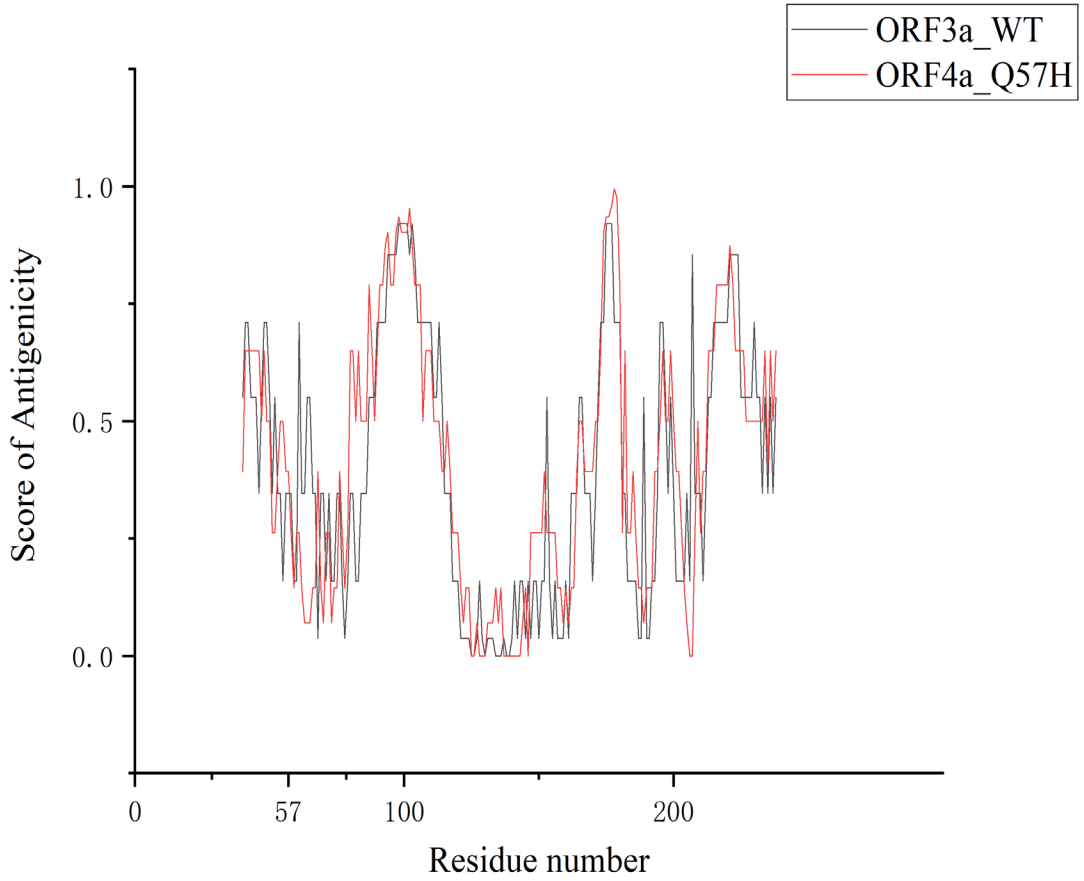
B

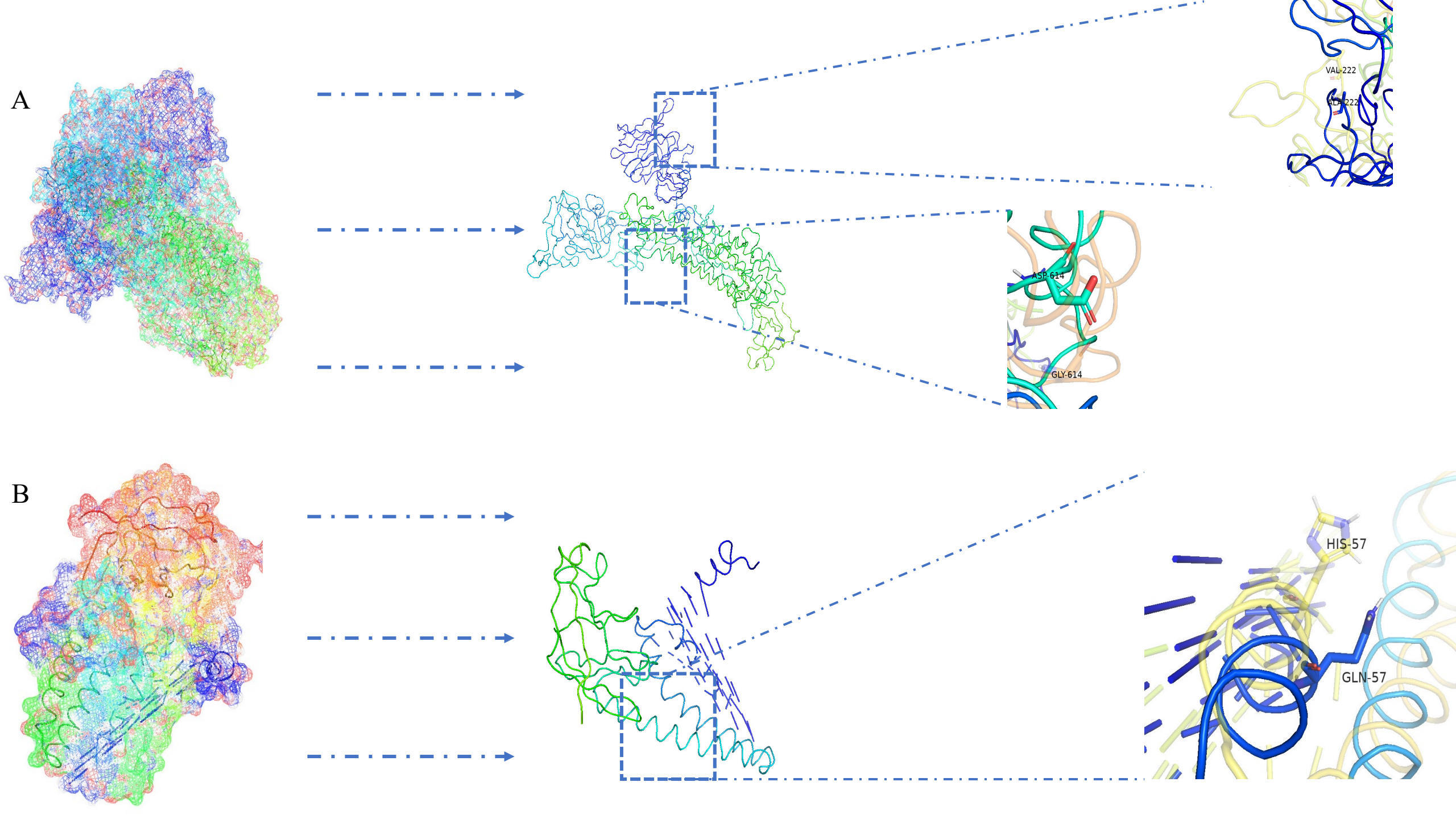

Hbonds_ORF3a_WT
Hbonds_ORF3a_Q57H
Number of Hydrogen bonds
405
400
395
390
385
380
375
370
365
360
355
350
345
340
335
330
325
0
50
100
Time(ns)

A

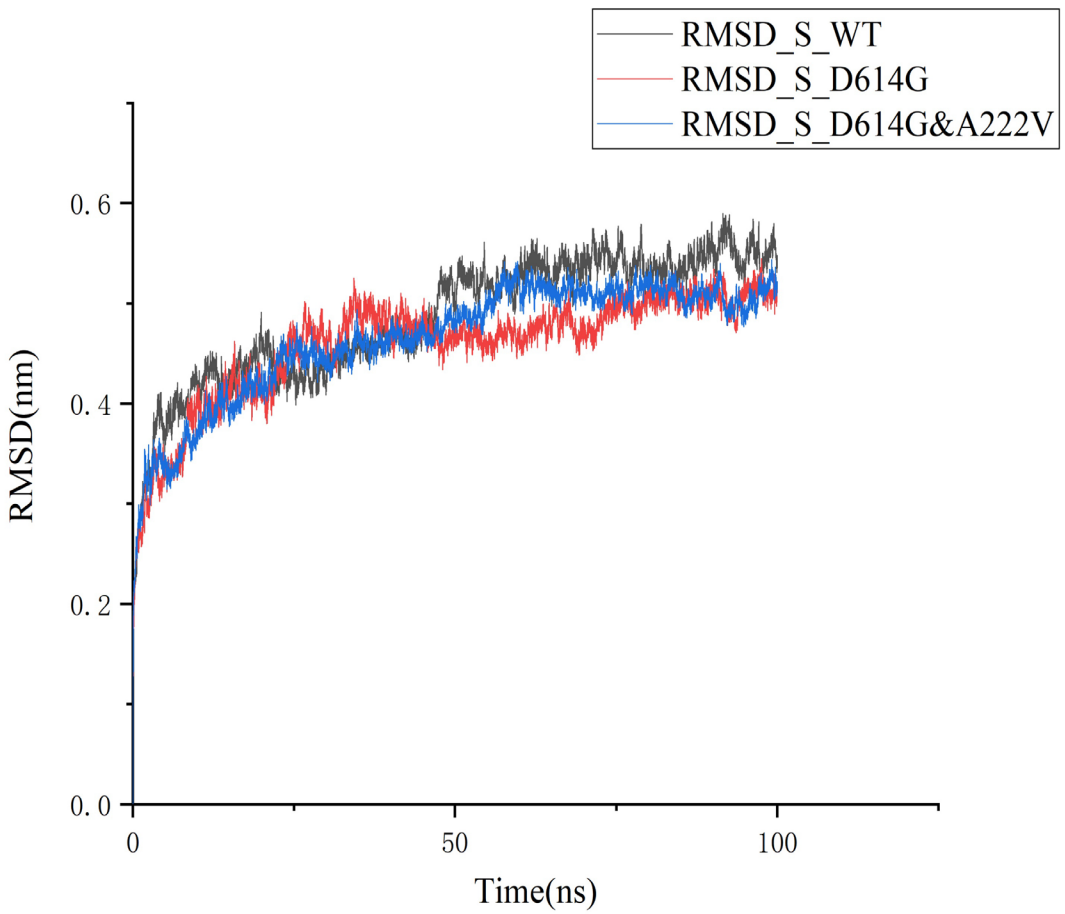

S_WT
S_D614G
S_D614G&A222V
1.0
0.5
0.0
Score of Antigenicity
0
222
500
614
1000
Residue number

B

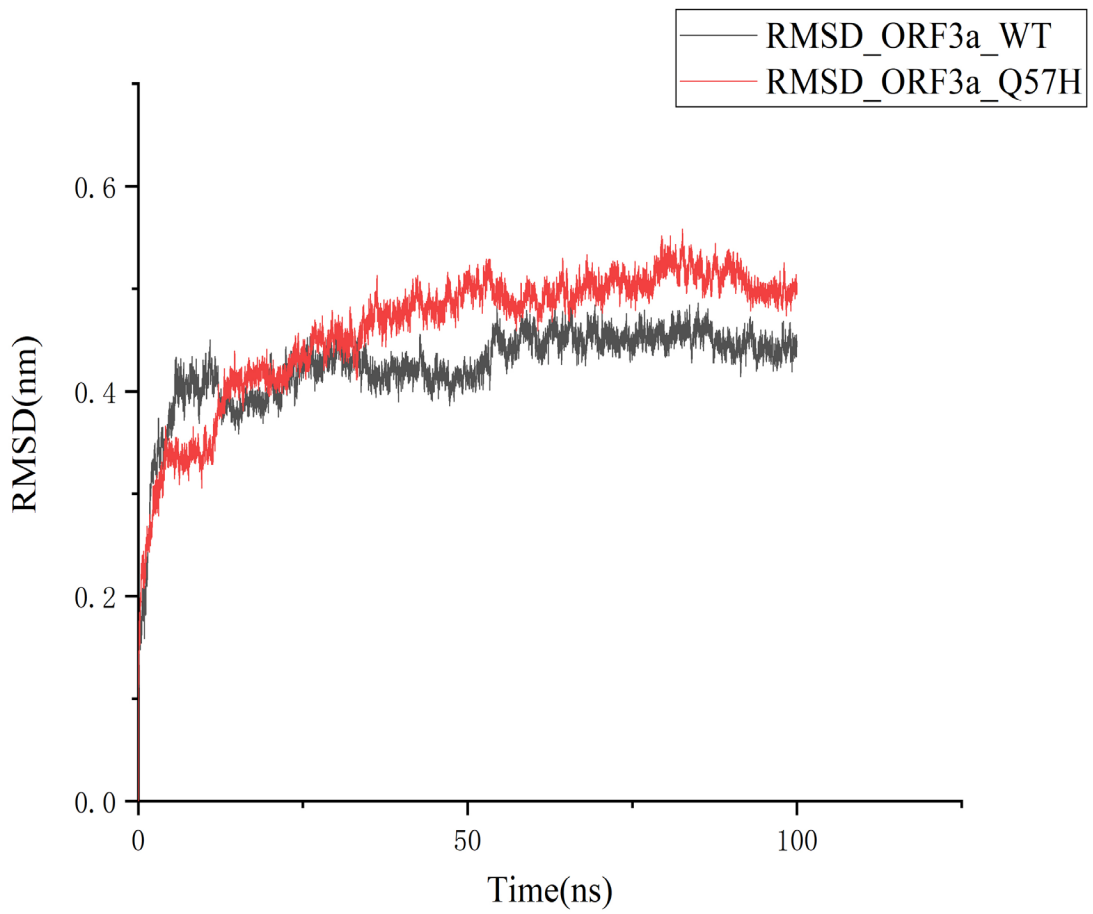

ORF3a_WT
ORF4a_Q57H
1.0
0.5
0.0
Score of Antigenicity
0
57
100
200
Residue number


A
VAL-222
ALA-222
ASP-614
GLY-614
B
HIS-57
GLN-57

A


RMSD_S_WT
RMSD_S_D614G
RMSD_S_D614G&A222V
0.6
0.4
0.2
0.0
RMSD(nm)
0
50
100
Time(ns)

B


RMSD_ORF3a_WT
RMSD_ORF3a_Q57H
0.6
0.4
0.2
0.0
RMSD(nm)
0
50
100
Time(ns)