Evolving Infection Paradox of SARS-CoV-2: Fitness Costs Virulence?

A. S. M. Rubayet Ul Alam[1#], Ovinu Kibria Islam[1#], Md. Shazid Hasan[1], Mir Raihanul Islam[2], Shafi Mahmud[3], Hassan M. Al Emran[4], Iqbal Kabir Jahid[1], Keith A. Crandall[5], M. Anwar Hossain[6,7*]

1 Department of Microbiology, Jashore University of Science and Technology, Jashore-7408, Bangladesh

2 BRAC James P Grant School of Public Health, BRAC University, Bangladesh

3 Genetic Engineering and Biotechnology, University of Rajshahi, Rajshai-6205, Bangladesh

4 Department of Biomedical Engineering, Jashore University of Science and Technology, Jashore-7408, Bangladesh

5 Computational Biology Institute and Department of Biostatistics & Bioinformatics, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA

6 Jashore University of Science and Technology, Jashore-7408, Bangladesh

7 Department of Microbiology, University of Dhaka, Dhaka-1000, Bangladesh

***Correspondence**

M. Anwar Hossain, Jashore University of Science and Technology, Jashore-7408, Bangladesh.

E-mail: hossaina@du.ac.bd, Contact: +8801708818101

 # Authors contributed equally

ABSTRACT

Background: SARS-CoV-2 is continuously spreading worldwide at an unprecedented scale and evolved into seven clades according to GISAID where four (G, GH, GR and GV) are globally prevalent in 2020. These major predominant clades of SARS-CoV-2 are continuously increasing COVID-19 cases worldwide; however, after an early rise in 2020, the death-case ratio has been decreasing to a plateau. G clade viruses contain four co-occurring mutations in their genome (C241T+C3037T+C14408T: RdRp.P323L+A23403G:spike.D614G). GR, GH, and GV strains are defined by the presence of these four mutations in addition to the clade-featured mutation in GGG28881-28883AAC:N. RG203-204KR, G25563T:ORF3a.Q57H, and C22227T:spike.A222V+C28932T-N.A220V+G29645T, respectively. The research works are broadly focused on the spike protein mutations that have direct roles in receptor binding, antigenicity, thus viral transmission and replication fitness. However, mutations in other proteins might also have effects on viral pathogenicity and transmissibility. How the clade-featured mutations are linked with viral evolution in this pandemic through gearing their fitness and virulence is the main question of this study.

Methodology: We thus proposed a hypothetical model, combining a statistical and structural bioinformatics approach, endeavors to explain this infection paradox by describing the epistatic effects of the clade-featured co-occurring mutations on viral fitness and virulence.

Results and Discussion: The G and GR/GV clade strains represent a significant positive and negative association, respectively, with the death-case ratio (incidence rate ratio or IRR = 1.03, $p < 0.001$ and IRR= 0.99/0.97, $p < 0.001$), whereas GH clade strains showed no association with the Docking analysis showed the higher infectiousness of a spike mutant through more favorable binding of G614 with the elastase-2. RdRp mutation p.P323L significantly increased genome-wide mutations ($p<0.0001$) since more expandable RdRp (mutant)-NSP8 interaction may accelerate replication. Superior RNA stability and structural variation at NSP3:C241T might impact upon protein or RNA interactions. Another silent 5'UTR:C241T mutation might affect translational efficiency and viral packaging. These G-featured co-occurring mutations might increase the viral load, alter immune responses in host

and hence can modulate intra-host genomic plasticity. An additional viroporin ORF3a:p.Q57H mutation, forming GH-clade, prevents ion permeability by cysteine (C81)-histidine (H57) inter-transmembrane-domain interaction mediated tighter constriction of the channel pore and possibly reduces viral release and immune response. GR strains, four G clade mutations and N:p.RG203-204KR, would have stabilized RNA interaction by more flexible and hypo-phosphorylated SR-rich region. GV strains seemingly gained the evolutionary advantage of superspreading event through confounder factors; nevertheless, N:p.A220V might affect RNA binding.

Conclusion: These hypotheses need further retrospective and prospective studies to understand detailed molecular and evolutionary events featuring the fitness and virulence of SARS-CoV-2.

Key words

SARS-CoV-2, COVID-19, Infection Paradox, Fitness, Virulence, Clades, Co-occurring mutations

Highlights

- We speculated an association of particular SARS-CoV-2 clade with death rate.
- The polymerase mutant virus can speed up replication that corresponds to higher mutations.
- The impact on viral epistasis by evolving mutations in SARS-CoV-2.
- How the virus changes its genotype and circulate with other types given the overall dynamics of the epidemics?

- Human intervention seems to work well to control the viral virulence. This hygiene practice will control the overall severity of the pandemic situation as recommended by the WHO. Our work has given the same message but explain with the dominant co-occurring mutations.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has caused the COVID-19 pandemic since the beginning of 2020 [1]. This highly contagious virus spread in 213 countries infecting over 11.3 million people including half a million deaths within the first six months [2]. SARS-CoV-2 has possessed some extraordinary attributes that make it extremely infectious: high replication rate, large burst size, high stability in the environment, strong binding efficiency of S protein receptor-binding domain (RBD) with human angiotensin-converting enzyme 2 (ACE2) receptor, and additional furin cleavage site in S protein [3-5]. In addition to those, it has proofreading capability ensuring high-fidelity replication [6]. The virus contains four major structural proteins: spike glycoprotein (S), envelope (E), membrane (M), and nucleocapsid (N) proteins along with 16 nonstructural proteins (NSP1 to NSP16) and seven accessory proteins (ORF3a, ORF6, ORF7a, ORF7b, ORF8a, ORF8b, and ORF10) [7,8].

Mutational spectra within the SARS-CoV-2 genome [9,10], spike protein [11], RdRp [12], ORF3a [13], and N protein [14] are reported where Islam et al. (2020) and Rahman et al. (2020) studied the possible impact of mutations on the virulence of a strain [10]. SARS-CoV-2 has been classified into seven major clades, such as G, GH, GR, GV, S, V, and L by GISAID based on the dominant mutations [15]. Yin [16] reported the leader sequence mutation 241C > T is co-occurring with three other mutations, 3037C > T (NSP3: C318T), 14408C > T (RdRp: p.P323L), and 23403A > G (S: p.D614G). GISAID referred these co-occurring mutation containing viruses as clade G, a viral strain over the wild-type [17,18]. These clade G or lineage B.1 viruses were dominant in Europe [17] and the east coast of USA [19] during the earlier stages of the pandemic that further spread in Southeast Asia [20,21] and Oceania [18]. Remarkably, this mutation variant is reported to be more transmissible [17] and was speculated to cause high mortality in USA [22]. The GR clade or lineage B.1.1 was classified with additional trinucleotide mutations at 28881-28883 (GGG>AAC) creating two consecutive amino acid changes, R203K and G204R, in N protein [18]. The GR strains are now the dominant type causing more than one-third of infection cases globally [15]. Another derivative of G clade is GH or lineage B.1.3-B.1.66, B.1.351, characterized by an additional ORF3a:p.Q57H mutation frequently found in the USA and Europe [18,23]. The newest variant GV or lineage

B.1.177 is responsible for COVID-19 cases in Europe featuring an A222V mutation in the S protein with other mutations of the clade G. N: A220, ORF10: V30L, and three other synonymous mutations, namely T445C, C6286T, and C26801G, are also found in this clade [24].

The most frequently observed mutation is D614G of S protein [25], which has direct roles in receptor binding, and immunogenicity, thus viral immune-escape, transmission, and replication fitness. The role of other dominant mutations remains largely underestimated. However, the ORF3a:p.Q57H [23] and RdRp:p.P323L [26] have recently been investigated while the effect of NSP3: C318T, 5'UTR: C241T, N:p.RG203-203KR, and N:p.A220V is still being overlooked. Mutations in other proteins could also have effects on viral pathogenicity and transmissibility. How the clade-featured mutations are linked with viral evolution in this pandemic through gearing their fitness and virulence is the main question of this study.

Different mutations may work independently or through epistatic interaction, thus it is difficult to spot exactly when a single mutation or co-occurring mutations become dominant in populations through evolutionary fitness. Several co-occurring mutations identified in SARS-CoV-2 may have complementary roles in changing its virulence and the infection dynamics of COVID-19 pandemic. Yet, many questions remain: When are 'co-occurring mutations' first observed? Why are particular clades, GV and GR, outcompeting their speculative ancestor G clade? What are the impacts of these mutations on protein structures and what are their functional roles? Do these mutant proteins interact together? Is there any role of the coevolved 'silent' mutations? Do these mutations have any impact on viral fitness and virulence? We attempt to answer these questions by *in silico* molecular insights of SARS-CoV-2 mutants and possible interactions of coevolved proteins.

This study aims to identify the prevalence of dominant co-occurring mutations of G, GH, GR, and GV strains circulating in all over the world, correlation of the clades with death-case ratio, probable individual and/or synergistic impact of those mutants upon virulence in terms of viral entry and fusion, evasion of host cell lysis, replication rate, ribonucleoprotein stability, protein-protein interactions, translational capacity, and ultimately the combined effect on fitness and virulence.

2. Materials and Methods

2.1 Retrieval of Sequences and Mutation Analyses

This study analyzed 225,526 high-coverage (<1% Ns and <0.05% unique amino acid mutations) and complete (>29,000 nucleotide) genome sequences with specified collection date from a total of 3,16,166 sequences submitted to GISAID until January 03, 2021.We sifted the sequences generated from the non-human host out from the dataset. The Wuhan-Hu-1 (Accession ID- NC_045512.2) [27] isolate was used as the reference genome.

A python script was used to partition a significant part of the dataset into two subsets based on the RdRp: C14408T mutation and estimated the genome-wide variations (single nucleotide changes) for each strain. For the genome-wide mutation analysis, a total of 37,179 sequences (RdRp wild type or 'C' variant: 9,815; and mutant or 'T' variant: 27,364) were analyzed from our dataset. The frequency of mutations was tested for significance with the Wilcoxon signed-rank test between RdRp 'C' variant and 'T' variant using IBM SPSS statistics 25.

Random effect poisson regression model was performed in STATA v13.0 to identify the association between death-case ratio and different clade strains (G, GH, GR, and GV); both unadjusted and adjusted incidence risk ratio (IRR) were estimated where time was introduced as a panel variable [28].

2.2 Epidemiological Data Analysis and Time Plot Generation

In this study, we report the prevalence of these dominant clades in 2020, both individually and in combination, with disease progression and deaths allowing us to infer increasing fitness of the SARS-CoV-2. A weekly-based time plot of G, GH, GR, and GV clade frequencies with infection and death-cases was generated from 23 December 2019 until January 3, 2021 that counted a total of 54 weeks (supplementary table S1). The total number of infections and deaths by weeks were extracted from the WHO 2019-nCoV situation reports. The case-death ratio was estimated by dividing the number of deaths of a particular week by the number of cases identified in the earlier week based on the conservative assumption of a one-week interval between diagnosis and death [29].

Regional time plots of those clades were also generated monthly (from January to June) with frequencies of new infections, deaths, and death-case ratio based on the available data on WHO situation reports [30]. The studied sequences were divided into six regions; Europe (n= 145,254), Americas (n= 48,014), Eastern Mediterranean (n=3,103), Southeast Asia (n= 4,134), West Pacific (n= 16,974), and Africa (n= 2,740).

2.3 Stability, Secondary and Three-Dimensional Structure Prediction Analyses of S, RdRp, ORF3a, and N Proteins

DynaMut [31] and FoldX 5.0 [32,33] were used to determine the stability of both wild and mutant variants of N, RdRp, S, and ORF3a proteins. PredictProtein [34] was utilized for analyzing and predicting the possible secondary structure and solvent accessibility of both wild and mutant variants of those proteins. The SWISS-MODEL homology modeling webtool [35] was utilized for generating the three-dimensional (3D) structures of the RdRp, S, and ORF3a protein using 7c2k.1.A, 6xr8.1.A, and 6xdc.1.A PDB structure as the template, respectively. Modeller v9.25 [36] was also used to generate the structures against the same templates. I-TASSER [37] with default protein modeling mode was employed to construct the N protein 3D structure of wild and mutant type since there was no template structure available for the protein. The built-in structural assessment tools (Ramachandran plot, MolProbity, and Quality estimate) of SWISS-MODEL were used to check the quality of generated structures.

2.4 Molecular Docking and Dynamics of RdRp-NSP8 and S protein-Elastase2 Complex

Determination of the active sites affected by binding is a pre-requisite for docking analysis. We chose 323 along with the surrounding residues (315-324) of RdRp and the residues 110 to 122 of NSP8 monomer as the active sites based on the previously reported structure [38]. The passive residues were defined automatically where all surface residues were selected within the 6.5°A radius around the active residues. The molecular docking of the wild and predicted mutated RdRp with the NSP8 monomer from the PDB structure 7C2K was performed using the HADDOCKv2.4 to evaluate the interaction [39]. The binding affinity of the docked RdRp-NSP8 complex was predicted using the PRODIGY [40]. The number and specific interfacial contacts (IC) for each of the complexes were identified.

The human neutrophil elastase (hNE) or elastase-2 (PDB id: 5A0C) was chosen for docking of the S protein, based on earlier reports [41]. Here we employed CPORT [42] to find out the active and passive protein-protein interface residues of hNE. The S protein active sites were chosen based on the target region (594-638) interacting with the elastase-2. The passive residues of S protein were defined automatically as mentioned for RdRp-NSP8 docking analysis. Afterward, we individually docked wild (614D) and mutated (614G) S protein with the hNE using HADDOCK 2.4. The binding affinity of the docked complexes, as well as the

number and specific interfacial contacts (IC), were predicted as performed after RdRP-NSP8 docking. We employed HDOCK server [43] with specifying the active binding sites residues for predicting the molecular docking energy.

The structural stability of the protein complex and their variations were assessed through YASARA Dynamics software package [44]. The AMBER14 force field [45] was used for these four systems, and the cubic simulation cell was created. The TIP3P (at 0.997 g/L^{-1}, 25C, and 1 atm) water solvation model was used, and steepest gradient energy minimization techniques by simulated annealing method was used for initial energy minimization process. The hydrogen bond network system was optimized. The simulation cell was extended 20Å in all cases of the protein to move freely [46]. The PME or particle mesh Ewald methods was applied to calculate the long-range electrostatic interaction by a cut off radius of 8Å [47]. The Berendsen thermostat was applied to maintain the temperature of the simulation cell. The simulation system was neutralized by the addition of 0.9% NaCl, pH 7.4, and 310K temperature. The time step of the simulation was set as 1.25fs [46]. The simulation trajectories were saved after every 100ps. Finally, the molecular dynamics simulation was conducted for 100ns to analyze root mean square deviation (RMSD), root mean square fluctuation (RMSF), radius of gyration (Rg), solvent accessible surface area (SASA), hydrogen bond [48-51].

2.5 Mutational Analysis of Transmembrane Domain 1 of ORF3a and SR-domain of N protein

The complete genome of 12 pangolin derived coronavirus strains, as well as 38 bat, civet and human SARS-CoVs, were downloaded from GISAID and NCBI, respectively for the mutational comparison between the SARS-CoV and SARS-CoV-2. We mainly targeted transmembrane domain 1 (TM1), which covers 41 to 63 residues, of ORF3a to find the identical mutation and scan overall variation in TM1. In the case of the N protein, CoVsurver of GISAID was used to study the frequent mutations other than our target ones (RG→KR: 203-04) to bolster our speculation on the change and subsequent effects of those surrounding variations on the functions of the protein. A generalized comparison between SARS-CoV and SARS-CoV-2 reference sequences was performed to identify mutation in the SR-rich region that will help to postulate on N protein functions of novel coronavirus based on previous related research on SARS-CoV.

2.6 Analyzing RNA Folding prediction of 5´UTR and NSP3

The Mfold web server [52] was used with default parameters to check the folding pattern of RNA secondary structure in the mutated 5'UTR, synonymous leader (T445C) and NSP3 regions (C3037T). The structure of complete mutant 5'UTR (variant 'T') was compared with the wild type (variant 'C') secondary pattern as mentioned in the Huston, et al. [53]. Since the wild type (variant 'C' at 318th nucleotide) RNA structure of the NSP3 was not available in the literature, we generated the structure of mutant (variant 'T') to predict the RNA folding in the Mfold web server. From the Mfold web server, we also estimated free energy change (ΔG) for wild and mutant NSP3 RNA fold.

3. RESULTS AND DISCUSSION

We represented the global scenario of SARS-CoV-2 infection by the G, GH, GR, and GV clade strains and estimated the association between the clade strains and death-case ratio. Afterwards, the possible effects of the nine mutations in S, RdRp, ORF3a, N, 5'UTR, leader protein, and NSP3 were discussed with associated results. Whereas researches on molecular docking of the spike protein [54,55] and RdRp [56,57] in search of potential drug targets is a continuous process, our study approached in a unique way to dock spike with elastase-2 and RdRp with NSP8 to satisfy our purpose. The overall epistatic interactions of the mutant proteins and/or RNA was then depicted (Figure 1) with appropriate explanation. Finally, we endeavored to postulate using theoretical evolutionary theory how the virus might be changing virulence and fitness.

3.1 Global Emergence of Dominant Co-occurring Mutations

Analysis of the SARS-CoV-2 genome sequences has indicated that 241C > T, 3037C > T, 14408C > T, and 23403A > G mutations were discretely identified among different viruses in China on 24th January of 2020. These four mutations together in a single virus was first detected in England on 3rd February 2020 (Table 1). Since then, those mutations were found to cooccur with other mutations, thus formed clade G, GH, GR, and GV, and have become the most dominant variants in other regions of the world (Figure 2b), for example, escalating to 85% in May 2020 in Southeast Asia [20]. The G clade strain circulated predominantly (30%, n=828) in Africa, whereas the GH, GR, and GV clades have become more recognized in the Americas, Western Pacific and Europe, respectively (Table 1).

Our weekly based time-plot has depicted a gradual increase of G, GH, and GR viruses altogether since the 10th week (24 February – 2 March 2020) recording a sudden jump to 42% in that week from a mere 12% of the previous week. The global COVID-19 cases exponentially increased from the 10th week with only 7,806 cases and infected almost

500,000 people with ca. 35,000 deaths in just seven weeks while the G, GH, and GR strains reached 80% (supplementary table S1). The death rate elevated at its peak (14%) in mid of March (week 12) and gradually decreased to below 2% in early September. Correspondingly, the number of GR strains is increasing from the 10th week to 32nd week with a small fluctuation where the G and GH strains each maintained a static ratio between ~20-30%. GV clade strains initiated into the population from 30th week and became most dominant strain replacing GR clade strains in just 9 weeks (Figure 2a).

The geographical distribution plot of the G, GH, GR, and GV clades with infection and death-cases delineates that the new infections began to rise exponentially with the increase of coevolved mutant variants in all regions except the West Pacific area (Figure 2b). The West Pacific region, which includes East Asian countries as well, identified a very low number of infection cases and deaths per million (Figure 2b). Europe and America have a high rate of infections as well as a high percentage of those variants. The reason may be linked with α1-antitrypsin (AAT) deficiency, which is very rare in East Asia unlike Europe and North America [58]. The AAT allele deficiency facilitates entry of the 614G subtype into the host cells and accelerates the spread of G, GH, GR, and GV clades. Our analysis has also found that the proportion of strains containing 23403A > G mutation was 25% in East Asia and >75% in Europe and America in first 6 months of this year (data not shown).

The death-case ratio was decreasing globally while the GR mutants were increasing until August (Figure 2a). Since September, GV clade strains were increasing while the death-case ratio remained low (2%). To examine the association between clade strains and death-case ratio, both unadjusted and adjusted incidence risk ratio were estimated. In the adjusted model, G, GR, and GV clades were found to be significantly associated with death-case ratio in both models. If G clade strains increase by one percentage point then death-case ratio would be expected to increase by a factor of 1.03, while holding all other variables in the model constant (IRR = 1.03, 95% CI 1.01 – 1.06). However, GR and GV clades were inversely related with death-case ratio. One percentage point increment in GR clade led to reduction to death-case ratio by a factor of 0.99, while holding all other variables in the model constant (IRR = 0.98, 95% CI 0.97 – 0.99). And then again, If GV clade strains increase by one percentage point then death-case ratio would be expected to decrease by a factor of 0.97, while holding all other variables in the model constant (IRR = 0.97, 95% CI 0.96 – 0.98).

Like other SARS-CoV-2 studies [59-61], this statistical analysis also suffers from some limitations in dealing with genomic and calculating death-case ratio data. The death-case ratio is believed to be underestimated because of the inadequate number of tests capacity and

asymptomatic SARS-CoV-2 cases in the general population. Moreover, fewer mutation patterns are uploaded from underdeveloped or developing countries like African and Sub-Saharan countries which might lead to spatial biasedness in the analysis. Therefore, the global epidemiological scenario of different clades was explored to mitigate this problem.

Regional monthly data depicted a similar increase of GR and GV strains while the death ratio was decreasing in studied regions with some rare exceptions. In Europe, GR strains were predominant from April to August, while GV strains became predominant from September. In the Americas a high abundance (over 50%) of GH clade strains was found from March to December. In Eastern Mediterranean, Africa and Southeast Asian region there was an increased rise of GR strains in May, while in Western Pacific region GR strains became predominant in June. The increase of G strains and the death-case ratio at the same time was observed in most regions, except Eastern Mediterranean and Southeast Asia, where a limited number of sequence data were produced. (Figure 2b).

Researchers around the globe are now trying to explore the factors associated with variable mortality rates due to the infection by SARS-CoV2 in different regions of the world. A recent report identified a Neanderthal-derived haplotype in chromosome 3 which is prevalent in South Asia as a substantial contributor to COVID-19 risk for hospitalized patients along with other risk factors [62]. Also, a haplotype inherited from Neandertals on chromosome 12 was found at ~25 to 30% in Eurasia and the Americas, can reduce relative risk of becoming severely ill with COVID-19 by ~22% [63]. But other reports say that mortality associated to COVID-19 in Indian and south Asian subcontinent is lower than in the west [64]. In-hospital mortality was also found to be considerably lower in Asia than Europe and America [65]. Several factors such as the virulence of the pathogen, host factors like innate immunity, genetic diversity in immune responses, epigenetic factors, genetic polymorphisms of ACE2 receptors, micro RNAs and universal BCG vaccination, along with environmental factors may have contributed in low mortality in Indian region [66]. In European countries, mortality rates were observed to vary from region to region [67]. In Africa, where the Neanderthal risk haplotype is almost absent, youthful population, and weather conditions might have also contributed to low morbidity and mortality of COVID-19 [68]. No association between the SARS-CoV-2 variants and mortality rates was observed in the Eastern Mediterranean Region [69]. However, another report showed that, GR, GH and L clade viruses are predominant in countries with higher deaths and GR clade showed higher prevalence among severe/deceased patient [70]. Besides these host, pathogen and environment associated

345 factors, human and social intervention like political decision making, scientific advice, and
346 health system and public health capacity can also contribute to variable COVID-19 mortality
347 rates in different regions of the world [71].

348 **3.2 S Protein D614G Mutation Favors Elastase-2 Binding**

349 Korber et al. (2020) linked the S:p.D614G mutation with viral spread or transmission
350 capacity rather than virus infectivity or virulence [17] whereas Becerra□Flores et al. (2020)
351 approached to demonstrate a potential link of D614G with virus pathogenicity [22], using
352 same data and method alike Korber et al. (2020). A preview published by Grubaugh et al.
353 (2020) stated those major points as unclear without any concrete interpretations, and
354 questioned on the correlation of D614G mutation with virus spread or transmissibility.
355 Dearlove et al. (2020) acknowledged that this 614G variant may constitute an exception to
356 their conclusion on the arising of other mutations of the virus due to genetic drift [72]. Rausch
357 et al. (2020) [73] also stressed upon the possible effect of D614G on viral fitness through
358 revisiting the points raised by Dearlove et al. (2020). Recently, Plante et al. (2020) [74] and Hou
359 et al. (2020) [75] reported that D614G mutation alters virus fitness by enhancing viral load in
360 COVID-19 patients and may increase transmission in human, which was speculated from the
361 increasing transmission of mutant-type viruses in hamsters.

362 S: p.D614G mutation may facilitate the exposure of the cleavage domains of S1-S2
363 and S2' to proteases, elastase-2, Furin, or TMPRSS2, as discussed by Eaaswarkhanth, et al. [76].
364 Another explanation of the more efficient cellular entry for D614G variant is the possibly
365 breakage of hydrogen bonds that modulates the interactions between S protein protomers and
366 ACE2 receptor binding [17,75,77]. Moreover, G614 mutation may increase S protein stability and
367 participate in N-linked glycosylation at N616 [41]. Several recent experiments have suggested
368 that mutated (G_{614}) protein contains a novel serine protease cleavage site at 615-616 that is
369 cleaved by host elastase-2, a potent neutrophil elastase playing important roles in
370 inflammatory diseases of human, more efficiently than wild protein, D_{614}
371 [25,41,58,78]. Bhattacharyya et al. (2020) hypothesized that the neutrophil elastase level at the site
372 of infection will facilitate the host cell entry for G_{614}. We explained here, with the aid of
373 structural bioinformatics, how the mutated protein gets the advantage over wild type with a
374 single mutation.

375 This study found interesting structural features of the S protein while comparing and
376 superimposing the wild (D_{614}) over mutant (G_{614}). The secondary structure prediction and

surface accessibility analyses showed that there was a slight mismatch at the S1-S2 junction (^{681}PRRAR↓S^{686}) where serine at 686 (S^{686}) was found covered in G_{614} and exposed to the surface in D_{614}. However, S^{686} in both G_{614} and D_{614} were exposed to an open-loop region to have possible contact with the proteases (supplementary Figure S1). Further investigation on the aligned 3D structures showed no conformational change at the Furin or TMPRSS2 cleavage site (Figure 3c). We observed no structural variation in the surrounding residues of the protease-targeting S1-S2 site (Figure 3c), which eliminated the assumption of Phan [79]. The predictive 3D models and structural assessment of D_{614} and G_{614} variants also confirmed that the cleavage site at 815-16 of S2 subunit (^{812}PSKR↓S^{816}) or S2' [3,80] had no structural and surface topological variation (Figure 3d-e). Rather, the superimposed 3D structures suggested a conformational change in the immediate downstream region (^{618}TEVPVAIHADQLTPT^{632}) of the 614th position of G_{614} that was not observed in D_{614} variants (Figure 3a-b).

The elastase-2 restrictedly cut valine at 615, due to its valine-dependent constriction of catalytic groove [81]. The present sequence setting surrounding of G_{614} (P6-^{610}VLYQGV↓NCTEV^{620}-P'5) showed a higher enzymatic activity over the D_{614}, which cannot be completely aligned with previous works on the sequence-based substrate specificity of elastase-2 [82]. However, the first misaligned residue of the superimposed G_{614}, located at the P'4 position (T^{618}), may also be important for binding with the elastase-2, and further down the threonine (T) at 618, the residues may affect the bonding with the respective amino acids of the protease. This changed conformation at the downstream binding site of G_{614} may help overcome unfavorable adjacent sequence motifs in the mutated S protein as the elastase-2 substrate. Therefore, the simultaneous and/or sequential processing of the mutated S protein by TMPRSS2/Furin/Cathepsin and elastase-2 facilitates a more efficient SARS-CoV-2 entry into the host cells and cell-cell fusion [41,58,78].

This study further observed the possible association of the S protein with elastase-2 and found an increased binding affinity in case of G_{614} (Table 3). Hence, the active sites of the mutated protein interacted efficiently with more amino acids of elastase-2 (Table 4), possibly providing a better catalytic activity as shown by Hu, et al. [41]. The mutation may have changed the structural configuration of the elastase-2 cleavage site in a way that the enzyme is facing less challenge to get near to the cutting site of the altered protein (Figure 3a-b and 4). The efficient cleaving of this enzyme, although located in an upstream position of the S1-S2 junction, may assist in releasing S1 from S2 and change the conformation in a way to get later cleaved at the S2' site, and then help in the fusion process [83,84]. Mutated spike protein

and elastase-2 complex was more flexible than the wild one, and the interactions with enzyme was also different as shown in RMSD deviation between the complexes (Figure 5).

Besides, this G^{614} amino acid replacement may have a destabilization effect on the overall protein structure (Table 4 and Figure 3a-b). The deformed flexible region at or near G^{614} is the proof of that destabilizing change (Figure 3f and supplementary figure S3). Zhang et al. [85] explained less S1 shedding through more stable hydrogen bonding between Q^{613} (glutamine) and T^{859} (threonine) of protomer due to greater backbone flexibility provided by G^{614}. On the other hand, increasing the number of RBD up conformation or increasing the chance of 1-RBD-up conformation due to breakage of both intra-and inter-protomer interactions of the spike trimer and symmetric conformation will give a better chance to bind with ACE2 receptor and can also increase antibody-mediated neutralization [86]. The S1 will release from S2 more effectively in G614 protein due to introduction of glycine that will break hydrogen bond in between the D614 and the neighboring protomer T859 amino acid [86,87]. Our analyses have provided the *in silico* proof of this point by showing that the mutated protein was less stable than the wild type protein by missing a hydrophobic interaction with Phe^{592} (Figure 3g-h). This new result accorded well with the later consensus as explained in Weismann et al. (2020) that G614 will increase its overall flexibility. Therefore, the overall structural change may assist the mutated S protein by providing elastase-2 a better binding space and attachment opportunity onto the cleavage site, thus providing a more stable interaction with the S2 domain that increases the credibility of an efficient infection (Figure 1).

3.3 Increased Flexibility of RdRp-NSP8 Complex: Compromise Proof-Reading Efficiency with Replication Speed

The binding free energy (ΔG) of the RdRp-NSP8 complexes have been predicted as -10.6 and -10.5 $Kcalmol^{-1}$, respectively, in wild (P_{323}) and mutant (L_{323}) type that suggests less strong interaction for mutant protein (Table 4). The number of contacts made at the interface (IC) per property and interacting amino acids increased between L_{323} and NSP8 (Table 3 and 4). This increasing contacts due to higher hydrogen bonds (Figure 5d) made the complex of mutated RdRp and NSP8 less flexible. Our analyses have identified that proline (P^{323}) or leucine (L^{323}) of RdRp can interact with the aspartic acid (D^{112}), cysteine (C^{114}), valine (V^{115}), and proline (P^{116}) of NSP8 (Figure 6). RdRp binds with NSP8 in its interface domain (residues alanine:A^{250} to arginine:R^{365}), forming positively charged 'sliding poles' for RNA

exit and enhance the replication speed probably by extending the RNA-binding surface on NSP8 [88,89]. Molecular dynamics of the mutated RdRp-NSP8 complex supported this by showing a more expanded surface area in interacting sites (Figure 5b) and maintained integrity throughout the simulation (Figure 5c). Another study reported that the NSP8 binding sites on RdRp and the RNA exit tunnel were comparatively neutral and conserved [90]. Besides, a zinc ion in the conserved metal-binding motif (H^{295}, C^{301}, $C^{306,}$ and C^{310}), close to the 323 residue, is responsible for maintaining the integrity of the RdRp architecture [90,91]. We did not find any interaction of NSP8 with the zinc-binding residues of RdRp proteins (Table 3 and Figure 6). Therefore, the P323L mutation within this conserved site of the RdRp interface domain may only affect the RdRp-NSP8 interaction without changing metal binding affinity.

Recent reports on structural analyses revealed that the RdRp:p.P323L led to the stabilization of the protein structure [92 93]. The results from the six state-of-the-art tools used in this study have suggested that mutant (L_{323}) protein cannot be concluded as 'stable' from the ambiguous ΔΔG values (Table 4), rather the interaction with the adjacent amino acid [94] mainly defined the stability. Although the hydrophobic amino acid (L^{323}) embedded in the L_{323} protein, our secondary structure analysis indicated the presence of only loop structure in 323 position for both wild and mutant types (Figure S1). However, Chand et al. (2020) reported the turn structures of 323 and 324 shifted into five sheets at positions 321, 322, 323, 324 and 327 to bury leucine. The superimposed 3D structures solved this contradiction about secondary structure by showing that there was no deviation in loop/turn structure of mutant protein (supplementary Figure S4).

Overall, the mutation at 323 position, to some extent, stabilized the L_{323} structure, made the protein more rigid, binds less strongly with the NSP8, and thus expanded the interacting region with NSP8. These variations may together increase the replication speed by exiting the processed RNA genome from the RdRp groove structure more swiftly (Figure 1). The increasing replication speed might be due to the perturbation of interaction between RdRp and NSP8 [88,94], or less possibly, the complex tripartite interactions (RdRp, NSP8, and NSP14) responsible for the speculated decrease of proof-reading efficiency [6]. Thus, RdRp mutants might increase the mutation rate by a trade-off between high replication speed and low fidelity of the mutant polymerase [95]. Another possibility could be the lower proof-reading efficiency of NSP14 that was not linked to the replication speed [6]. Analysis of our study sequences revealed that the frequency of mutation (median=8) in L_{323} mutants (n=27,364) is significantly higher ($p<0.0001$) than the frequency (median=6) of wild-type (P_{323}) strains

(n=9,815). This increased mutation rate may play a vital role in genetic drifts and provide next generations a better adaptation to adverse environments.

3.4 Q57H Substitution in ORF3a Viroporin: the Roles of Decreased Ion Permeability

This study has found that the replacement of glutamine (Q^{57}) with positively charged histidine (H^{57}) at 57 position of ORF3a transmembrane region 1 (TM1) does not change secondary transmembrane helical configuration (supplementary Figure S1), and aligned 3D structures have also shown no variation of TM1 in the monomeric state (Figure 7a). The mutant (H_{57}) protein has a non-significant increase in structural stabilization and a minimal decrease in molecular flexibility (Table 4 and supplementary Figure S5). This is because of the weak ionic interaction of C_α with the sulfur atom of cysteine (C^{81}) in TM2 and hydrogen bond of terminal N_ζ of lysine (61K) with one of the endocyclic nitrogens of H^{57} (Figure 7b).

Selection analysis predicted the accumulation of those mutations as a result of pervasive positive selection with an increasing trend [96]. One possibility of this positive selection of histidine over glutamine might be the role of H^{57} (TM1) in increasing constrictions wherein the diagonal C^{81} (TM2) may assist through the ionic interaction (Figure 7b). Notably, the Q^{57} in wild-type protein forms the major hydrophilic constriction within the ORF3a channel pore [97]. Thus, further favorable increasing constrictions within the H_{57} protein channel pore and the replacement of charge-neutral Q with a positive-charged H in the selectivity filter may reduce the passing of positive ions such as Ca^{2+}, $Na^{+,}$ and K^+ by either electrostatic repulsion or blocking [98-101]. This speculation for ORF3a mutant protein was supported by another study that showed the reduction of ion permeability of Na^{2+} and Ca^{2+} through the H_{57}, however, that decrease was not found statistically significant ($p>0.05$) [97].

The decrease intracellular concentration of cytoplasmic Ca^{2+} ions potentially reduces caspase-dependent apoptosis of the host cell [102], mainly supporting viral spread without affecting replication [23] as shown in Figure 1. Moreover, the ORF3a can drive necrotic cell death [103] wherein the permeated ions [104] and the insertion of ORF3a viroporin into lysosome [105] play vital roles. The H_{57} mutant may thus decrease pathogenicity and symptoms during the early stages of the infection, i.e., reducing 'cytokine storm' in the host [106]. Besides, ORF3a was proved to affect inflammasome activation, viral release, and cell death, as shown by Castaño-Rodriguez, et al. [107] that the deletion of ORF3a reduced viral load and morbidity in animal models.

Even though similar proteins of ORF3a have been identified in the sarbecovirus lineage infecting bats, pangolins, and humans [108], only one pangolin derived strain from 2017 in Guangxi, China contains H^{57} residue as shown by our mutation analyses (supplementary Figure S6) and also reported by Kern, et al. [97]. However, bat or civet did not contain this mutation in TM1 and the flanking vicinity was not identical; whereas, the TM1 (^{41}LPFGWLIVGVALLAVFQSASKII^{63}) does not have amino acid replacements for the strains of SARS-CoV-2 and pangolin-coronaviruses (Figure S5a-b). The presence of this mutation in pangolin could be an accidental case or might explain its impact on modulating host-specific immune response, which needs functional experimental verification. A possible explanation behind that presence might be the more accustomed nature of the virus towards reverse transmission by being less virulent, i.e., from human to other animals, as observed in recent reports [109,110].

3.5 N Protein Mutation: Augmenting Nucleocapsid Stability and Exerting Miscellaneous Effects

Our study has observed that the combined mutation (N: p.RG203-204KR) causes no conformational change in secondary and 3D structures (Figure S1 and Figure 8, respectively) of the conserved SR-rich region of the LKR (supplementary Figure S7), but there is a minor alteration in the degree of buried or exposed site (Figure 3). This result contradicted the prediction of [111] about the change in the length and arrangements of the alpha-helix in the SR-rich region. The superimposed 3D structures showed structural deviation, rather at ^{231}ESKMSGKGQQQQGQTVT^{247} of the LKR (Figure 9), corresponding to the high destabilization of the $KR_{203-204}$ protein (Table 3). On the other hand, A220V mutation in the N protein of the GV clade showed a slightly more stable formation of the mutated N protein with no change in the chemical properties (Table 4), that might affect RNA binding affinity [112].

Impedance to form particular SR-motif due to RG→KR mutation might disrupt the phosphorylation catalyzed by host glycogen synthase kinase-3 [113]. Similar hypo-phosphorylation events could arise due to the conversion of serine to nonpolar or neutral amino acids ($L^{188/194/197}$, I^{193}, and N^{202}), as represented in supplementary Table S2 and [114]. Consequently, the low phosphorylation level after entering into the cell should unwind the viral ribonucleoprotein (RNP) in a slower but more organized fashion that might have an impact upon translation and immune-modulation [115-117]. In $KR_{203-204}$, replacement of a glycine with lysine that may increase the nucleocapsid (N protein- RNA complex) stability by

forming stronger electrostatic and ionic interactions due to increased positive charge [118,119]. Besides, the more disordered orientation of the associated LKR [115] and highly destabilizing property of $KR_{203\text{-}204}$ may assist in the packaging of a stable RNP [112,120]. These interactions and impact upon mutations are depicted in Figure 1.

N protein also utilizes the dynamic nature of the intrinsically disordered linker region (LKR) that controls its affinity towards M protein, self-monomer, 5'UTR, and cellular proteins [121-123]. The phosphorylation at the LKR site may play an essential role to regulate these interactions [118]. It was speculated that $KR^{203\text{-}204}$ attained more selective advantage [96] over the other mutations of N protein (Table S2), probably because of stronger RNA binding and synchronized hypo-phosphorylation.

3.6 Silent Mutations may not be Silent

The C241T of 5'UTR (untranslated region), a single nucleotide change, or 'silent' mutation, located at the UUCGU pentaloop part of the stem-loop region (SLR5B) has a potential role in viral packaging [124]. This pentaloop of 5'UTR remains unchanged and maintains a particular structure [122,125]. The RNA secondary structural analysis in our study predicted that there is no change in the 241T structure (supplementary Figure S8a). However, the silent mutation in the loop region upstream to the ORF1a start codon (266-268 position) may be involved in differential RNA binding affinity to the ribosome and translational factors [126].

In the case of multi-domain NSP3 (papain-like protease), we have observed superior stability of the RNA after gaining of the synonymous mutation 3037C<T (C318T) where wild and mutant RNA structure has -151.63 and -153.03 Kcal/mol, respectively (Supplementary Figure S8b-c). A more stable secondary structure of (+)-ssRNA as observed in the mutant NSP3 protein corresponds to the slower translational elongation. It contributes to a range of abnormalities resulting in low translation efficiency, which affects posttranslational modifications as a part of protein regulation [127]. Because silent mutations have an impact on the ribosome occupancy time depending on the structural modification and mRNA stability which guide co-translational folding kinetics of a protein [127]. This silent mutation is located within the flexible loop of the NSP3 ubiquitin-like domain 1 (Ubl1). In SARS-CoV, Ubl1 was reported to bind with single-stranded RNA containing AUA patterns and interact with the nucleocapsid (N) protein [128,129]. Besides, Ubl1 was likely to bind with several signature repeats in 5′-UTR in SARS-CoV-2 genome [130]. Figure 1 represents the overall possible scenario due to these silent mutations.

Change in T445C in leader protein may not cause any change in expression or others since the structure and energy are same -172.34 kcal/mol. The change C6286T is at between in the NSP3 region whereas the C26801G is in between the Nucleic acid-binding (NAB) domain and betacoronavirus specific marker (βSM) domain. The change C26801G is in the TM3 of virion membrane.

3.7 Epistatic Effects of Co-occurring Mutations: Increasing Viral Fitness Costing Virulence?

The co-occurring mutations, as defined by the presence of simultaneous multi-site variations in the same or different proteins or in the genome [115], have provided new insights into the dynamic epistatic network by employing differential molecular interactions. The synergistic effects of the mutations were reported to control viral fitness and virulence, as observed in Influenza and Ebola virus outbreak [131-136]. For instance, the detrimental effect of R384G on influenza A fitness was overcome by the co-occurring mutation E375G [132,135], and co-occurring mutations at the antigenic sites of influenza hemagglutinin can also drive viral evolution [134,136]. The correlation of the co-occurring GP-L mutations affect Ebola virus virulence and thus case fatality rate [133]. The viral fitness determined by the efficiency of viruses to spread throughout the population and infect more new victims [137,138]. A general definition of virulence as proposed by Geoghegan et al. (2018) is 'the harm caused by pathogen infection, particularly in terms of host morbidity and mortality' [139]. Evolution generally drives a pathogen to gain more fitness while reducing the virulence [139] during the natural course of mutations [140]; nevertheless, there is contradiction between short-term [140] and long-term evolution [139] of virus. However, the virulence can also increase during the course of the outbreak as observed in HIV in 2002 [141]. Here we used death-case ratio and the number of infection cases as equivalent to virulence and fitness, respectively.

Between two important G clade-featured co-occurring mutations, the p:D614G of the S protein and p:P323L of RdRp, we observed no interlinked functional relationship (Figure 1). The former assists mainly by more rapid entering into the host cells with an efficient elastase-2 activity and higher aggressiveness of G_{614} mutant is related to elastase-2 or human neutrophil elastase (hNE) concentration during inflammation [142]. Besides, higher levels of functional S protein observed in G614 strains can increase the chance of host-to-host transmission [85]. The later one may boost up the replication by a faster RNA processing (exiting) that can open up the avenue to generate strains with significantly ($p<0.0001$) higher number of mutations. This increased mutation rate in L_{323} mutants can surpass the constant

proof-reading fidelity [143], with an average of ~8 (range 0-45) mutations per strain (n=37,179), and evolve a greater number of variants in a population (Figure 1). Nevertheless, SARS-CoV-2 has still low genetic diversity, and these mutations mostly are the product of genetic drift with no or unknown effect on virus virulence and fitness [53, 54], and without possessing a deleterious effect on viral fitness, a mutation fixed early on in the epidemic is unlikely to be lost.

The S protein and RdRp mutations, albeit seemingly unrelated, can cumulatively escalate the infectiousness of the virus as a result of higher viral load and shorter burst time. This rapid within-host replication might be directly correlated with the virulence, in turn, morbidity and mortality rate [144]. This study also showed that co-occurring mutations, by acting together, may benefit viral populations by incrementing the ability to produce a diverse viral population (Figure 1), that might be able to adapt more quickly in adverse climatic condition, evade the immune response, and survive within different selective pressure [145,146]. The S protein mutation is also related to human allelic variation [58], and may get an advantage in particular populations.

NSP3 is a scaffolding protein for the replication-transcription complex, and the possible change in its structure may affect the overall dynamics of viral replication [128,129]. P323L mutation of RdRp may change binding affinity to the Ubl1 region of NSP3 [147] (Figure 1). The p.C241T mutation, on its own, may affect the transcription and viral packaging, although we have not found any possible association of it with the S, RdRp, and NSP3 mutant proteins.

The mutant N protein may have an impact on viral replication and transcription, like other coronaviruses [123], through the binding with NSP3 protein that is linked to RdRp centered replication complex. This N protein can affect the membrane stability by yet uncharacterized interaction with the M protein, which should ultimately produce more stable virion particles [148-150]. A stronger N protein-RNA complex provokes slower intracellular immune response [116], and at the same time, can remain highly contagious and aggressive because of the concurrent presence of G clade-featured S protein and RdRp mutations (Figure 1). In the case of mutant ORF3a protein, we have not found any report that correlates with other co-occurring mutations. H_{57} mutant, possibly linked to the mild or asymptomatic cases, may allow the silent transmission and increase the chance of viral spread by lowering the

636 activation of inflammatory response (Figure 1), such as reduced viral particle release and
637 cytokine storm [23,151,152].

638 The fittest viral strains will dominate in a population considering other selective
639 parameters associated with the virulence [139,144]. Public health interventions were able to
640 create a selective pressure to make a strain less virulent and highly competitive [153,154], which
641 might increase the chance of viral transmission [155]; however, there was also argument against
642 this established theory [156]. The SARS-CoV-2 with multiple clades and variants needs to be
643 more efficient to maintain the delicate evolutionary trade-off between fitness and virulence.
644 A recent publication on COVID-19 demonstrated a reduced viral load in the upper respiratory
645 system for a particular genotype [157], which may in turn influence viral transmission and
646 spread. We proposed here a hypothesis (model) on SARS-CoV-2 infection as to how the
647 fitness of the major clade strains might cost virulence.

648 The presence of mostly G strains in the early pandemic period (starting weeks of
649 March) might be linked to a higher mortality rate (Figure 2a). G strains were continuously
650 present in any region at a particular level (~20% among these clades), which could be the
651 reason for maintaining a balanced high mortality rate. This stage of the pandemic might be
652 linked to the initial lack of awareness, hygienic practice, and social distancing that led to a
653 large number of transmissions. The high viral load and quick immune response will spread
654 the G strains more efficiently in a crowded area as shown in the case of Africa (Figure 1)
655 where the lockdown and other social interventions were not very strict [158]. The G strains thus
656 spread unprecedentedly from Europe to the other parts of the world. Being more
657 transmissible alone can also cause problem by putting a greater strain on hospitals and
658 leading to a sharper spike in deaths.

659 The GH strains were mainly restricted to the USA and partly Eastern Mediterranean
660 (Fig. 1), and mostly spread by cryptic [159] and pre/a-symptomatic transmission [160]. These
661 variants might trade-off virulence by a slower release of virions, and in exchange, benefited
662 from the induction of low immune response in asymptomatic hosts. They were more fit at
663 that time, in theory, when the people were dealing with the pandemic in panic. The GH type
664 might be able to maximize their transmission by residing within the host unknowingly and
665 spread at ease. We can not nullify that there could be no significant link of the mutation with
666 virulence or death-case ratio since the effect of ORF3a was not found significant in wet lab
667 experiement and in our statistical analysis as well.

The GR strains possibly attained an advantage over G and GH by a more orchestrated, delicately balanced synergistic effects on replication and transmission fitness. These epistatic effects could have increased the fitness by hiding the virus from immunity and increasing stability in the environment, i.e., more transmissible through air and surface. The asymptomatic patient infected with GR strains, similar to GH, would have a weaker immune response and shed the virus for a longer period [152]. Hamed et al. (2020) [70] reported a result against this hypothesis with a statistical approach using GISAID data; however only 4.75% (27,608/ 581,120 on February 20, 2021) sequence data contains patient status, that will definitely misinterpret true scenario.

GV strains featuring an A222V mutation in the S protein with other mutations of the clade G was reported to have a less pronounced effect due to its structural position and have no role in antibody binding, thus founder effect might be the cause of its spreading within Europe and the age group bias towards young adults [24]. There was hence probably no effect of the mutations on the viral transmission, severity, and escaping antibody [24,161], rather superspreading events after lifting up of travel restriction in Europe and lack of effective containment could be the main factors for the rapid spread and establishment of this clade. There was, however, speculation about the effect of this mutation on immunogenicity since computational analysis predicted its location within T-cell target region [162,163]. A plausible link of the mutation with viral re-infection in some patients was of prime concern and demands further study [164]. How the more stabilized linker region of the mutated N protein due to A220V mutation might give an edge for the GV strains in terms of association with lower death-case ratio but seemingly higher infection number by interacting with RNA is a question for further study. Together, the mutations of GV strains can affect in an unknown way in transmission and spread that cannot be captured alone in terms of spike A222V mutation.

Lineage B.1.1.7, a variant under GR clade, consists of 17 mutations (14 nonsynonymous and 3 deletions) in spike, N, ORF1ab, and ORF8 proteins, as well as, 6 silent mutations [165,166]. This lineage shows a 50% more transmissibility, realistic possibility in increased risk of death, and is rapidly dominating over other variants in more than 75 countries. The mutations on the spike protein might play most crucial role both in increasing transmission fitness and a slightly reduced neutralization to antibody [167,168]. Other mutations may still play roles in viral replication and pathogenesis through different epistatic interactions. B.1.351 lineage of clade GH also showed a faster spread and higher load of virus

in swabs, as well as, higher binding affinity to ACE2 receptor and escape from neutralizing antibody due to three mutations (K417N, E484K and N501Y) at key sites of the RBD [169]. This lineage has mutations on the E, N, and ORF1a proteins, functions of which are yet to be determined [170]. Similar important mutational package in RBD with known biological importance to B.1.351 lineage, except for K417T and several others in the ORF1a, N, ORF8, and ORF3a, was found in P.1 (descendent of B.1.1.28) lineage of GR clade [171,172]. B.1.258 lineage under G clade contained N439K that has higher binding affinity with new contact point to ACE2 and can evade antibody-mediated immunity[173,174]. The mutations in proteins other than spike of these new variants could have effects that needs further investigation. Recently, the variants are supplanting their original clade strains throughout the world due to RBD mutation effects, which makes correlating the clade with case fatality rate difficult and evolutionary theory complex. Due to these variants, recent data of January suggests an increase of GR strains where GH and GV are similar in different regions and worldwide in a weekly basis (data not shown). However, the death-case ratio is following the previous trends of around 2% in the weeks after 3rd January.

The GH and GR strains have seemingly arisen during the initial phase of the pandemic with a limited frequency, and at that time, have not attained the necessary viral fitness. Both strains have evolved from Europe and Australia after spreading quietly alongside the vigorous G strain. After the public awareness and other social interventions, these strains, especially GR have probably increased in fitness. Although Hu and Riley [159] have recently reported the co-circulation of both G and GH strains correlating with mortality in different states, how the association among multiple clades would modulate the evolutionary dynamics is now a burning question. With the new data of GV strains as shown in our study, the previous analysis by Hodcroft et al. (2020) [24] can be updated to recent time that may change result of the association. Our analysis has considered the evolutionary trade-off between virulence and sustainable fitness of these strains based on the global emergence of the strains, their death-case ratio, structural stability, and predicted molecular mechanisms. The relation of the host-pathogen interactions, host allelic variations, and host immunity needs to correlate with virulence by in vivo studies.

4. CONCLUSION

The course of COVID-19 pandemic was continuing, although the death rate was gradually decreasing in 2020. Our study hypothesized that this paradoxical scenario was

related to the effect of the dominant co-occurring mutations. The significant association of death-case ratio with GR clade mutant could be linked with the signature mutation in nucleocapsid protein. GH clade mutant might contribute to cryptic transmission by ORF3a-mediated low immune response in asymptomatic hosts. GV strains have spread quickly by superspreading events, but we speculated a role of more stable linker region of N protein on low virulence. G clade mutants, however, was speculated to assist in frequent transmissions through severe symptoms lead to the isolation of infected persons restricting further spreading and significantly increased the risk of death by a factor of 1.03. Therefore, it can be speculated that the fitness of SAR-CoV-2 virus may increase in terms of survivability and transmission, while the next phase of the COVID-19 pandemic may continue with less virulent mutants. However, further investigations are required to strengthen this hypothesis. Effective management of the infected patients and social distancing to prevent viral transmission should be the first priority that may prevent the emergence of more virulent strains.

Data Availability

All the sequence data were taken from the GISAID (https://www.gisaid.org/) and RCSB PDB (https://www.rcsb.org/) as mentioned in the methodology section. We provide all the necessary information such as accession numbers, date-based data source for helping readers and reviewers to check the authencity of the work.

Acknowledgments

We would like to acknowledge the team at GISAID for creating SARS-CoV-2 global database. The funding of the research was provided by Jashore University of Science and Technology. We appreciate to the Microbial Genetics and Bioinformatics Laboratory of University of Dhaka for the support of the high-performance computer access. We thanked M. Shaminur Rahman and Spencer Mark Mondol for their technical assistance in protein structure analyses.

Biographical Note:

A. S. M. Rubayet Ul Alam is a lecturer in the Department of Microbiology, Jashore University of Science and Technology. His research mainly focuses on molecular biology, data analysis, and bioinformatics.

Ovinu Kibria Islam is an assistant professor in the Department of Microbiology, Jashore University of Science and Technology. His research activity is focused on molecular biology and big data analysis.

Md. Shazid Hasan is a lecturer in the Department of Microbiology, Jashore University of Science and Technology. His research activity is focused on molecular biology and bioinformatics.

Mir Raihanul Islam is the Senior Research Associate of BRAC James P Grant School of Public Health, BRAC University who is expert in statistical analysis and epidemiological study and is currently working on many active researches related to public health.

Shafi Mahmud is working as a thesis student in the department of Genetic Engineering and Biotechnology, University of Rajshahi, Rajshai-6205, Bangladesh.

Hassan M. Al Emran is an assistant professor in the Department of Biomedical Engineering, Jashore University of Science and Technology, Jashore-7408, Bangladesh. His research mainly focuses on clinical microbiology and infectious diseases.

Dr. Iqbal Kabir Jahid is a professor and chairman in the Department of Microbiology, Jashore University of Science and Technology. His research mainly focuses on molecular biology.

Keith A. Crandall is a professor in the Department of Biostatistics & Bioinformatics and the director of Computational Biology Institute, Director of Genomics Core, Milken Institute School of Public Health, The George Washington University, Washington, DC, USA.

M. Anwar Hossain is the director of Genome Center, and vice chancellor of Jashore University of Science and Technology. He is an expert in molecular biology, virology and vaccine development.

REFERENCES

1 Lu, R. *et al*. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. *The Lancet* **395**, 565-574 (2020).

2 WHO. Coronavirus disease (COVID-19) Situation Report – 168 *WHO situation Reports* (2020).

3 Hoffmann, M., Kleine-Weber, H. & Pöhlmann, S. A multibasic cleavage site in the spike protein of SARS-CoV-2 is essential for infection of human lung cells. *Molecular Cell* (2020).

4 Hoque, M. N., Chaudhury, A., Akanda, M. A. M., Hossain, M. A. & Islam, M. T. Genomic diversity and evolution, diagnosis, prevention, and therapeutics of the pandemic COVID-19 disease. *PeerJ* **8**, e9689 (2020).

5 Petersen, E. *et al.* Comparing SARS-CoV-2 with SARS-CoV and influenza pandemics. *The Lancet infectious diseases* (2020).

6 Romano, M., Ruggiero, A., Squeglia, F., Maga, G. & Berisio, R. A Structural View of SARS-CoV-2 RNA Replication Machinery: RNA Synthesis, Proofreading and Final Capping. *Cells* **9**, 1267 (2020).

7 Rahman, M. S. *et al.* Epitope-based chimeric peptide vaccine design against S, M and E proteins of SARS-CoV-2 etiologic agent of global pandemic COVID-19: an in silico approach. *PeerJ* **8**, e9572 (2020).

8 Yoshimoto, F. K. The Proteins of Severe Acute Respiratory Syndrome Coronavirus-2 (SARS CoV-2 or n-COV19), the Cause of COVID-19. *The protein journal* **39**, 198-216, doi:10.1007/s10930-020-09901-4 (2020).

9 Callaway, E. The coronavirus is mutating-does it matter? *Nature* **585**, 174-177 (2020).

10 Islam, M. R. *et al.* Genome-wide analysis of SARS-CoV-2 virus strains circulating worldwide implicates heterogeneity. *Scientific reports* **10**, 1-9 (2020).

11 Rahman, M. S. *et al.* Comprehensive annotations of the mutational spectra of SARS□CoV□2 spike protein: a fast and accurate pipeline. *Transboundary and emerging diseases* (2020).

12 Eskier, D., Karakülah G, Suner A, Oktay Y. RdRp mutations are associated with SARS-CoV-2 genome evolution. *PeerJ* **8**, 9587, doi:https://doi.org/10.7717/peerj.9587 (2020).

13 Hassan, S. S., Choudhury, P. P., Basu, P. & Jana, S. S. Molecular conservation and Differential mutation on ORF3a gene in Indian SARS-CoV2 genomes. *Genomics* (2020).

14 Shaminur Rahman, M. *et al.* Evolutionary dynamics of SARS□CoV□2 nucleocapsid (N) protein and its consequences. *Journal of medical virology*.

15 Shu, Y. & McCauley, J. GISAID: Global initiative on sharing all influenza data–from vision to reality. *Eurosurveillance* **22**, 30494 (2017).

16 Yin, C. Genotyping coronavirus SARS-CoV-2: methods and implications. *Genomics* **112**, 3588-3596, doi:10.1016/j.ygeno.2020.04.016 (2020).

17 Korber, B. *et al.* Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* **182**, 812-827. e819 (2020).

18 Mercatelli, D. & Giorgi, F. M. Geographic and Genomic Distribution of SARS-CoV-2 Mutations. *Frontiers in Microbiology* **11**, doi:10.3389/fmicb.2020.01800 (2020).

19 Brufsky, A. Distinct Viral Clades of SARS□CoV□2: Implications for Modeling of Viral Spread. *Journal of medical virology* (2020).

20 Islam, O. *et al.* Emergence of European and North American mutant variants of SARS-CoV-2 in Southeast Asia. (2020).

21 Ul Alam, A. R., Rafiul Islam, M., Shaminur Rahman, M., Islam, O. K. & Anwar Hossain, M. Understanding the possible origin and genotyping of first Bangladeshi SARS□CoV□2 strain. *Journal of Medical Virology* (2020).

22 Becerra□Flores, M. & Cardozo, T. SARS□CoV□2 viral spike G614 mutation exhibits higher case fatality rate. *International Journal of Clinical Practice* (2020).

23 Issa, E., Merhi, G., Panossian, B., Salloum, T. & Tokajian, S. SARS-CoV-2 and ORF3a: Nonsynonymous Mutations, Functional Domains, and Viral Pathogenesis. *Msystems* **5** (2020).

24 Hodcroft, E. B. *et al.* Emergence and spread of a SARS-CoV-2 variant through Europe in the summer of 2020. *medRxiv* (2020).
25 Grubaugh, N. D., Hanage, W. P. & Rasmussen, A. L. Making sense of mutation: what D614G means for the COVID-19 pandemic remains unclear. *Cell* (2020).
26 Pachetti, M. *et al.* Emerging SARS-CoV-2 mutation hot spots include a novel RNA-dependent-RNA polymerase variant. *Journal of Translational Medicine* **18**, 1-9 (2020).
27 of the International, C. S. G. The species Severe acute respiratory syndrome-related coronavirus: classifying 2019-nCoV and naming it SARS-CoV-2. *Nature Microbiology* **5**, 536 (2020).
28 Dalgaard, P. in *Introductory statistics with R* 259-274 (Springer, 2008).
29 Wang, W., Tang, J. & Wei, F. Updated understanding of the outbreak of 2019 novel coronavirus (2019□nCoV) in Wuhan, China. *Journal of medical virology* **92**, 441-447 (2020).
30 WHO. Coronavirus disease (COVID-19) Weekly Epidemiological Update and Weekly Operational Update. *WHO situation Reports* (2020).
31 Rodrigues, C. H., Pires, D. E. & Ascher, D. B. DynaMut: predicting the impact of mutations on protein conformation, flexibility and stability. *Nucleic acids research* **46**, W350-W355 (2018).
32 Schymkowitz, J. *et al.* The FoldX web server: an online force field. *Nucleic acids research* **33**, W382-W388 (2005).
33 Delgado, J., Radusky, L. G., Cianferoni, D. & Serrano, L. FoldX 5.0: working with RNA, small molecules and a new graphical interface. *Bioinformatics* **35**, 4168-4169 (2019).
34 Rost, B., Yachdav, G. & Liu, J. The predictprotein server. *Nucleic acids research* **32**, W321-W326 (2004).
35 Waterhouse, A. *et al.* SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic acids research* **46**, W296-W303 (2018).
36 Webb, B. & Sali, A. Comparative protein structure modeling using MODELLER. *Current protocols in bioinformatics* **54**, 5.6. 1-5.6. 37 (2016).
37 Roy, A., Kucukural, A. & Zhang, Y. I-TASSER: a unified platform for automated protein structure and function prediction. *Nature protocols* **5**, 725-738 (2010).
38 Wang, Q. *et al.* Structural basis for RNA replication by the SARS-CoV-2 polymerase. *Cell* **182**, 417-428. e413 (2020).
39 Van Zundert, G. *et al.* The HADDOCK2. 2 web server: user-friendly integrative modeling of biomolecular complexes. *Journal of molecular biology* **428**, 720-725 (2016).
40 Xue, L. C., Rodrigues, J. P., Kastritis, P. L., Bonvin, A. M. & Vangone, A. PRODIGY: a web server for predicting the binding affinity of protein–protein complexes. *Bioinformatics* **32**, 3676-3678, doi:10.1093/bioinformatics/btw514 (2016).
41 Hu, J. *et al.* The D614G mutation of SARS-CoV-2 spike protein enhances viral infectivity. *bioRxiv* (2020).
42 de Vries, S. J. & Bonvin, A. M. CPORT: a consensus interface predictor and its performance in prediction-driven docking with HADDOCK. *PloS one* **6**, e17695 (2011).
43 Yan, Y., Tao, H., He, J. & Huang, S.-Y. The HDOCK server for integrated protein–protein docking. *Nature protocols* **15**, 1829-1852 (2020).
44 Land, H. & Humble, M. S. in *Protein Engineering* 43-67 (Springer, 2018).

45 Dickson, C. J. *et al.* Lipid14: the amber lipid force field. *Journal of chemical theory and computation* **10**, 865-879 (2014).
46 Krieger, E. & Vriend, G. New ways to boost molecular dynamics simulations. *Journal of computational chemistry* **36**, 996-1007 (2015).
47 Krieger, E., Nielsen, J. E., Spronk, C. A. & Vriend, G. Fast empirical pKa prediction by Ewald summation. *Journal of molecular graphics and modelling* **25**, 481-486 (2006).
48 Chowdhury, K. H. *et al.* Drug Repurposing Approach against Novel Coronavirus Disease (COVID-19) through Virtual Screening Targeting SARS-CoV-2 Main Protease. *Biology* **10**, 2 (2021).
49 Pramanik, S. K. *et al.* Fermentation optimization of cellulase production from sugarcane bagasse by Bacillus pseudomycoides and molecular modeling study of cellulase. *Current Research in Microbial Sciences* **2**, 100013 (2021).
50 Swargiary, A., Mahmud, S. & Saleh, M. A. Screening of phytochemicals as potent inhibitor of 3-chymotrypsin and papain-like proteases of SARS-CoV2: an in silico approach to combat COVID-19. *Journal of Biomolecular Structure and Dynamics*, 1-15 (2020).
51 Yan, H. & Yu, C. Repair of full-thickness cartilage defects with cells of different origin in a rabbit model. *Arthroscopy: The Journal of Arthroscopic & Related Surgery* **23**, 178-187 (2007).
52 Zuker, M. Mfold web server for nucleic acid folding and hybridization prediction. *Nucleic acids research* **31**, 3406-3415 (2003).
53 Huston, N. C., Wan, H., Tavares, R. d. C. A., Wilen, C. & Pyle, A. M. Comprehensive in-vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *BioRxiv* (2020).
54 Rane, J. S., Chatterjee, A., Kumar, A. & Ray, S. (ChemRxiv, 2020).
55 Souza, P. F., Lopes, F. E., Amaral, J. L., Freitas, C. D. & Oliveira, J. T. A molecular docking study revealed that synthetic peptides induced conformational changes in the structure of SARS-CoV-2 spike glycoprotein, disrupting the interaction with human ACE2 receptor. *International journal of biological macromolecules* **164**, 66-76 (2020).
56 Elfiky, A. A. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): A molecular docking study. *Life sciences*, 117592 (2020).
57 Aftab, S. O. *et al.* Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach. *Journal of translational medicine* **18**, 1-15 (2020).
58 Bhattacharyya, C. *et al.* SARS-CoV-2 mutation 614G creates an elastase cleavage site enhancing its spread in high AAT-deficient regions. *Infection, Genetics and Evolution*, 104760 (2021).
59 Backhaus, A. Common Pitfalls in the Interpretation of COVID-19 Data and Statistics. *Intereconomics* **55**, 162-166 (2020).
60 Sáez, C., Romero, N., Conejero, J. A. & García-Gómez, J. M. Potential limitations in COVID-19 machine learning due to data source variability: A case study in the nCov2019 dataset. *Journal of the American Medical Informatics Association* **28**, 360-364 (2021).
61 L.R. Lopes Junior, S. A., M. Qazvini COVID-19: Statistics, Ratios and understanding limitations of data available. *ICAT: Data Limitations sub-group* (2020).
62 Zeberg, H. & Pääbo, S. The major genetic risk factor for severe COVID-19 is inherited from Neanderthals. *Nature* **587**, 610-612 (2020).

63 Zeberg, H. & Pääbo, S. A genomic region associated with protection against severe COVID-19 is inherited from Neandertals. *Proceedings of the National Academy of Sciences* **118** (2021).

64 Jain, V. K., Iyengar, K., Vaish, A. & Vaishya, R. Differential mortality in COVID-19 patients from India and western countries. *Diabetes & Metabolic Syndrome: Clinical Research & Reviews* **14**, 1037-1041 (2020).

65 Goel, S. *et al.* Clinical characteristics and in-hospital mortality for COVID-19 across the Globe. *Cardiology and Therapy* **9**, 553-559 (2020).

66 Samaddar, A., Gadepalli, R., Nag, V. L. & Misra, S. The enigma of low COVID-19 fatality rate in India. *Frontiers in Genetics* **11**, 854 (2020).

67 Villani, L., McKee, M., Giraldi, L., Ricciardi, W. & Boccia, S. Comparison of deaths rates for COVID-19 across Europe. (2020).

68 Njenga, M. K. *et al.* Why is there low morbidity and mortality of COVID-19 in Africa? *The American journal of tropical medicine and hygiene* **103**, 564-569 (2020).

69 Omais, S. O., Kharroubi, S. O. & Zaraket, H. No association between the SARS-CoV-2 variants and mortality rates in the Eastern Mediterranean Region. *medRxiv* (2021).

70 Hamed, S. M., Elkhatib, W. F., Khairallah, A. S. & Noreddin, A. M. Global dynamics of SARS-CoV-2 clades and their relation to COVID-19 epidemiology. (2020).

71 McKee, M., Gugushvili, A., Koltai, J. & Stuckler, D. Are populist leaders creating the conditions for the spread of COVID-19?; Comment on "A scoping review of populist radical right parties' influence on welfare policy and its implications for population health in Europe". *International journal of health policy and management* (2020).

72 Dearlove, B. *et al.* A SARS-CoV-2 vaccine candidate would likely match all currently circulating variants. *Proceedings of the National Academy of Sciences* **117**, 23652-23662 (2020).

73 Rausch, J. W., Capoferri, A. A., Katusiime, M. G., Patro, S. C. & Kearney, M. F. Low genetic diversity may be an Achilles heel of SARS-CoV-2. *Proceedings of the National Academy of Sciences* **117**, 24614-24616 (2020).

74 Plante, J. A. *et al.* Spike mutation D614G alters SARS-CoV-2 fitness. *Nature*, 1-9 (2020).

75 Hou, Y. J. *et al.* SARS-CoV-2 D614G variant exhibits efficient replication ex vivo and transmission in vivo. *Science* (2020).

76 Eaaswarkhanth, M., Al Madhoun, A. & Al-Mulla, F. Could the D614 G substitution in the SARS-CoV-2 spike (S) protein be associated with higher COVID-19 mortality? *International Journal of Infectious Diseases* (2020).

77 Omotuyi, I. O. *et al.* Atomistic simulation reveals structural mechanisms underlying D614G spike glycoprotein□enhanced fitness in SARS□COV□2. *Journal of computational chemistry* **41**, 2158-2161 (2020).

78 Korber, B. *et al.* Tracking changes in SARS-CoV-2 Spike: evidence that D614G increases infectivity of the COVID-19 virus. *Cell* (2020).

79 Phan, T. Genetic diversity and evolution of SARS-CoV-2. *Infection, genetics and evolution* **81**, 104260 (2020).

80 Belouzard, S., Chu, V. C. & Whittaker, G. R. Activation of the SARS coronavirus spike protein via sequential proteolytic cleavage at two distinct sites. *Proceedings of the National Academy of Sciences* **106**, 5871-5876 (2009).

81 Perona, J. J. & Craik, C. S. Structural basis of substrate specificity in the serine proteases. *Protein Science* **4**, 337-360 (1995).

82 Fu, Z., Thorpe, M., Akula, S., Chahal, G. & Hellman, L. T. Extended cleavage specificity of human neutrophil elastase, human proteinase 3, and their distant

ortholog clawed frog PR3—three elastases with similar primary but different extended specificities and stability. *Frontiers in Immunology* **9**, 2387 (2018).

83 Li, F. Structure, function, and evolution of coronavirus spike proteins. *Annual review of virology* **3**, 237-261 (2016).

84 Walls, A. C. *et al.* Tectonic conformational changes of a coronavirus spike glycoprotein promote membrane fusion. *Proceedings of the National Academy of Sciences* **114**, 11157-11162 (2017).

85 Zhang, L. *et al.* SARS-CoV-2 spike-protein D614G mutation increases virion spike density and infectivity. *Nature communications* **11**, 1-9 (2020).

86 Weissman, D. *et al.* D614G spike mutation increases SARS CoV-2 susceptibility to neutralization. *Cell host & microbe* **29**, 23-31. e24 (2021).

87 Wrapp, D. *et al.* Cryo-EM structure of the 2019-nCoV spike in the prefusion conformation. *Science* **367**, 1260-1263 (2020).

88 Hillen, H. S. *et al.* Structure of replicating SARS-CoV-2 polymerase. *Nature*, 1-6 (2020).

89 Yin, W. *et al.* Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir. *Science* (2020).

90 Kirchdoerfer, R. N. & Ward, A. B. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. *Nature communications* **10**, 1-9 (2019).

91 Gao, Y. *et al.* Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* **368**, 779-782 (2020).

92 Chand, G., Banerjee A, Azad GK. Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure. *PeerJ* **8**, e9492, doi:https://doi.org/10.7717/peerj.9492 (2020).

93 Begum, F., Mukherjee, D., Das, S., Thagriki, D., Tripathi, P. P., Banerjee, A. K., & Ray, U. Specific mutations in SARS-CoV2 RNA dependent RNA polymerase and helicase alter protein structure, dynamics and thus function: Effect on viral RNA replication. *biorxiv.org* (2020).

94 Chand, G. B., Banerjee, A. & Azad, G. K. Identification of novel mutations in RNA-dependent RNA polymerases of SARS-CoV-2 and their implications on its protein structure. *bioRxiv* (2020).

95 Eskier, D., Karakülah, G., Suner, A. & Oktay, Y. RdRp mutations are associated with SARS-CoV-2 genome evolution. *bioRxiv* (2020).

96 Pond, S. Genomic diversity and divergence of SARS-CoV-2/COVID-19 from GISAID. (2020).

97 Kern, D. M. *et al.* Cryo-EM structure of the SARS-CoV-2 3a ion channel in lipid nanodiscs. *BioRxiv* (2020).

98 Malasics, A. *et al.* Protein structure and ionic selectivity in calcium channels: Selectivity filter size, not shape, matters. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1788**, 2471-2480 (2009).

99 Naranjo, D., Moldenhauer, H., Pincuntureo, M. & Díaz-Franulic, I. Pore size matters for potassium channel conductance. *Journal of General Physiology* **148**, 277-291 (2016).

100 Stephens, R. F., Guan, W., Zhorov, B. S. & Spafford, J. D. Selectivity filters and cysteine-rich extracellular loops in voltage-gated sodium, calcium, and NALCN channels. *Frontiers in Physiology* **6**, 153 (2015).

101 Suárez-Delgado, E. & Islas, L. D. Ion Channels: A novel origin for calcium selectivity. *Elife* **9**, e55216 (2020).

102 Kondratskyi, A., Kondratska, K., Skryma, R. & Prevarskaya, N. Ion channels in the regulation of apoptosis. *Biochimica et Biophysica Acta (BBA)-Biomembranes* **1848**, 2532-2546 (2015).

103 Yue, Y. *et al.* SARS-Coronavirus Open Reading Frame-3a drives multimodal necrotic cell death. *Cell death & disease* **9**, 1-15 (2018).

104 Pinton, P., Giorgi, C., Siviero, R. & Zecchini, E. Rizzuto R. *Bcl-2 and Ca* **2**, 1409-1418 (2006).

105 Nieva, J. L., Madan, V. & Carrasco, L. Viroporins: structure and biological functions. *Nature Reviews Microbiology* **10**, 563-574 (2012).

106 Ren, Y. *et al.* The ORF3a protein of SARS-CoV-2 induces apoptosis in cells. *Cellular & molecular immunology*, 1-3 (2020).

107 Castaño-Rodriguez, C. *et al.* Role of severe acute respiratory syndrome coronavirus viroporins E, 3a, and 8a in replication and pathogenesis. *MBio* **9** (2018).

108 Boni, M. F. & Lemey, P. Evolutionary origins of the SARS-CoV-2 sarbecovirus lineage responsible for the COVID-19 pandemic. *Nature microbiology*, doi:10.1038/s41564-020-0771-4 (2020).

109 Halfmann, P. J. *et al.* Transmission of SARS-CoV-2 in domestic cats. *New England Journal of Medicine* (2020).

110 Shi, J. *et al.* Susceptibility of ferrets, cats, dogs, and other domesticated animals to SARS–coronavirus 2. *Science* **368**, 1016-1020 (2020).

111 Ayub, M. I. Reporting Two SARS-CoV-2 Strains Based on A Unique Trinucleotide-Bloc Mutation and Their Potential Pathogenic Difference. *Preprints* doi:10.20944/preprints202004.0337.v1 (2020).

112 Chang, C.-k., Hou, M.-H., Chang, C.-F., Hsiao, C.-D. & Huang, T.-h. The SARS coronavirus nucleocapsid protein–forms and functions. *Antiviral research* **103**, 39-50 (2014).

113 Tylor, S. *et al.* The SR-rich motif in SARS-CoV nucleocapsid protein is important for virus replication. *Canadian journal of microbiology* **55**, 254-260 (2009).

114 Rahman, M. S. *et al.* Evolutionary dynamics of SARS□CoV□2 nucleocapsid protein and its consequences. *Journal of medical virology* (2020).

115 Järvelin, A. I., Noerenberg, M., Davis, I. & Castello, A. The new (dis) order in RNA regulation. *Cell Communication and Signaling* **14**, 9 (2016).

116 Kikkert, M. Innate immune evasion by human respiratory RNA viruses. *Journal of innate immunity* **12**, 4-20 (2020).

117 Kopecky-Bromberg, S. A., Martínez-Sobrido, L., Frieman, M., Baric, R. A. & Palese, P. Severe acute respiratory syndrome coronavirus open reading frame (ORF) 3b, ORF 6, and nucleocapsid proteins function as interferon antagonists. *Journal of virology* **81**, 548-557 (2007).

118 McBride, R., Van Zyl, M. & Fielding, B. C. The coronavirus nucleocapsid is a multifunctional protein. *Viruses* **6**, 2991-3018 (2014).

119 Sokalingam, S., Raghunathan, G., Soundrarajan, N. & Lee, S.-G. A study on the effect of surface lysine to arginine mutagenesis on protein stability and structure using green fluorescent protein. *PloS one* **7**, e40410 (2012).

120 Haynes, C. & Iakoucheva, L. M. Serine/arginine-rich splicing factors belong to a class of intrinsically disordered proteins. *Nucleic acids research* **34**, 305-312 (2006).

121 Carlson, C. R., Asfaha, J. B., Ghent, C. M., Howard, C. J, Hartooni, N., Safari, M, ... & Morgan, D. O. Phosphoregulation of Phase Separation by the SARS-CoV-2 N Protein Suggests a Biophysical Basis for its Dual Functions. *Molecular Cell* **11**, 025, doi:https://doi.org/10.1016/j.molcel.2020.11.025. (2020).

122 Schuster, N. A. Using the nucleocapsid protein to investigate the relationship between SARS-CoV-2 and closely related bat and pangolin coronaviruses. *BioRxiv* (2020).
123 de Haan, C. A. & Rottier, P. J. Molecular interactions in the assembly of coronaviruses. *Advances in virus research* **64**, 165-230 (2005).
124 Rangan, R. *et al.* RNA genome conservation and secondary structure in SARS-CoV-2 and SARS-related viruses: a first look. *Rna* **26**, 937-959 (2020).
125 Huston, N. C. *et al.* Comprehensive in vivo secondary structure of the SARS-CoV-2 genome reveals novel regulatory motifs and mechanisms. *Molecular cell* **81**, 584-598. e585 (2021).
126 Kristofich, J. *et al.* Synonymous mutations make dramatic contributions to fitness when growth is limited by a weak-link enzyme. *PLoS genetics* **14**, e1007615 (2018).
127 Mitra, S., Ray, S. K. & Banerjee, R. Synonymous codons influencing gene expression in organisms. *Research and Reports in Biochemistry* **6**, 57 (2016).
128 Hurst, K. R., Ye, R., Goebel, S. J., Jayaraman, P. & Masters, P. S. An interaction between the nucleocapsid protein and a component of the replicase-transcriptase complex is crucial for the infectivity of coronavirus genomic RNA. *Journal of virology* **84**, 10276-10288 (2010).
129 Hurst, K. R., Koetzner, C. A. & Masters, P. S. Characterization of a critical interaction between the coronavirus nucleocapsid protein and nonstructural protein 3 of the viral replicase-transcriptase complex. *Journal of virology* **87**, 9159-9172 (2013).
130 Serrano, P. *et al.* Nuclear magnetic resonance structure of the N-terminal domain of nonstructural protein 3 from the severe acute respiratory syndrome coronavirus. *Journal of virology* **81**, 12049-12060 (2007).
131 Du, X. *et al.* Networks of genomic co-occurrence capture characteristics of human influenza A (H3N2) evolution. *Genome research* **18**, 178-187 (2008).
132 Rimmelzwaan, G. *et al.* Full restoration of viral fitness by multiple compensatory co-mutations in the nucleoprotein of influenza A virus cytotoxic T-lymphocyte escape mutants. *Journal of general virology* **86**, 1801-1805 (2005).
133 Deng, L. *et al.* Network of co-mutations in Ebola virus genome predicts the disease lethality. *Cell research* **25**, 753-756 (2015).
134 Chen, H., Zhou, X., Zheng, J. & Kwoh, C.-K. Rules of co-occurring mutations characterize the antigenic evolution of human influenza A/H3N2, A/H1N1 and B viruses. *BMC medical genomics* **9**, 69 (2016).
135 Rimmelzwaan, G. F., Kreijtz, J. H., Bodewes, R., Fouchier, R. A. & Osterhaus, A. D. Influenza virus CTL epitopes, remarkably conserved and remarkably variable. *Vaccine* **27**, 6363-6365 (2009).
136 Lyons, D. M. & Lauring, A. S. Mutation and epistasis in influenza virus evolution. *Viruses* **10**, 407 (2018).
137 Crow, J. F. *Basic concepts in population, quantitative, and evolutionary genetics*. (WH Freeman and Company, 1986).
138 Di Giallonardo, F. & Holmes, E. C. Viral biocontrol: grand experiments in disease emergence and evolution. *Trends in microbiology* **23**, 83-90 (2015).
139 Geoghegan, J. L. & Holmes, E. C. The phylogenomics of evolving virus virulence. *Nature Reviews Genetics* **19**, 756-769 (2018).
140 Berngruber, T. W., Froissart, R., Choisy, M. & Gandon, S. Evolution of virulence in emerging epidemics. *PLoS Pathog* **9**, e1003209 (2013).
141 Pantazis, N. *et al.* Temporal trends in prognostic markers of HIV-1 virulence and transmissibility: an observational cohort study. *The Lancet HIV* **1**, e119-e126 (2014).
142 Mohamed, M. M., El-Shimy, I. A. & Hadi, M. A. (BioMed Central, 2020).

143 Smith, E. C., Sexton, N. R. & Denison, M. R. Thinking outside the triangle: replication fidelity of the largest RNA viruses. *Annual Review of Virology* **1**, 111-132 (2014).
144 Skjesol, A. *et al.* IPNV with high and low virulence: host immune responses and viral mutations during infection. *Virology journal* **8**, 1-14 (2011).
145 Pfeiffer, J. K. & Kirkegaard, K. Increased fidelity reduces poliovirus fitness and virulence under selective pressure in mice. *PLoS Pathog* **1**, e11 (2005).
146 Vignuzzi, M., Stone, J. K., Arnold, J. J., Cameron, C. E. & Andino, R. Quasispecies diversity determines pathogenesis through cooperative interactions in a viral population. *Nature* **439**, 344-348 (2006).
147 Lei, J., Kusov, Y. & Hilgenfeld, R. Nsp3 of coronaviruses: Structures and functions of a large multi-domain protein. *Antiviral research* **149**, 58-74 (2018).
148 Schoeman, D. & Fielding, B. C. Coronavirus envelope protein: current knowledge. *Virology journal* **16**, 1-22 (2019).
149 Escors, D., Ortego, J., Laude, H. & Enjuanes, L. The membrane M protein carboxy terminus binds to transmissible gastroenteritis coronavirus core and contributes to core stability. *Journal of virology* **75**, 1312-1324 (2001).
150 J Alsaadi, E. A. & Jones, I. M. Membrane binding proteins of coronaviruses. *Future Virology* **14**, 275-286 (2019).
151 Bai, Y. *et al.* Presumed asymptomatic carrier transmission of COVID-19. *Jama* **323**, 1406-1407 (2020).
152 Quan-Xin, L. *et al.* Clinical and immunological assessment of asymptomatic SARS-CoV-2 infections. *Nature medicine* (2020).
153 Dennis, C. (Nature Publishing Group, 2001).
154 Dieckmann, U., Metz, J. A. & Sabelis, M. W. *Adaptive dynamics of infectious diseases: in pursuit of virulence management*. Vol. 2 (Cambridge University Press, 2005).
155 Day, T., Gandon, S., Lion, S. & Otto, S. P. On the evolutionary epidemiology of SARS-CoV-2. *Current Biology* (2020).
156 Ebert, D. & Bull, J. J. Challenging the trade-off model for the evolution of virulence: is virulence management feasible? *Trends in microbiology* **11**, 15-20 (2003).
157 Lorenzo-Redondo, R. *et al.* A clade of SARS-CoV-2 viruses associated with lower viral loads in patient upper airways. *EBioMedicine* **62**, 103112 (2020).
158 Abbas, K. *et al.* Routine childhood immunisation during the COVID-19 pandemic in Africa: a benefit–risk analysis of health benefits versus excess risk of SARS-CoV-2 infection. *The Lancet Global Health* (2020).
159 Hu, Y. & Riley, L. W. Dissemination and co-circulation of SARS-CoV2 subclades exhibiting enhanced transmission associated with increased mortality in Western Europe and the United States. *medRxiv* (2020).
160 He, X. *et al.* Temporal dynamics in viral shedding and transmissibility of COVID-19. *Nature medicine* **26**, 672-675 (2020).
161 McCallum, M. *et al.* N-terminal domain antigenic mapping reveals a site of vulnerability for SARS-CoV-2. *bioRxiv* (2021).
162 Zhang, B.-z. *et al.* Mining of epitopes on spike protein of SARS-CoV-2 from COVID-19 patients. *Cell research* **30**, 702-704 (2020).
163 Mateus, J. *et al.* Selective and cross-reactive SARS-CoV-2 T cell epitopes in unexposed humans. *Science* **370**, 89-94 (2020).
164 To, K. K.-W. *et al.* Coronavirus disease 2019 (COVID-19) re-infection by a phylogenetically distinct severe acute respiratory syndrome coronavirus 2 strain confirmed by whole genome sequencing. *Clinical Infectious Diseases* (2020).

165 Volz, E. *et al.* Transmission of SARS-CoV-2 Lineage B. 1.1. 7 in England: Insights from linking epidemiological and genetic data. *medRxiv*, 2020.2012. 2030.20249034 (2021).
166 Muik, A. *et al.* Neutralization of SARS-CoV-2 lineage B. 1.1. 7 pseudovirus by BNT162b2 vaccine–elicited human sera. *Science* (2021).
167 McCarthy, K. R. *et al.* Recurrent deletions in the SARS-CoV-2 spike glycoprotein drive antibody escape. *Science* (2021).
168 Kemp, S. *et al.* Neutralising antibodies drive Spike mediated SARS-CoV-2 evasion (medRxiv). *bioRxiv* (2020).
169 Tegally, H. *et al.* 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. *medRxiv* (2020).
170 grinch. B.1.351_South African lineage defined by new variant 501Y.V2 - A more detailed description of the lineage. *global report investigating novel coronavirus haplotypes* (2021).
171 grinch. P.1_Brazilian lineage with variants of biological significance E484K, N501Y and K417T. *global report investigating novel coronavirus haplotypes* (2021).
172 Faria, N. R. *et al.* Genomic characterisation of an emergent SARS-CoV-2 lineage in Manaus: preliminary findings. *Virological* (2021).
173 Zhou, W. *et al.* N439K variant in spike protein may alter the infection efficiency and antigenicity of SARS-CoV-2 based on molecular dynamics simulation. *bioRxiv* (2020).
174 Thomson, E. C. *et al.* Circulating SARS-CoV-2 spike N439K variants maintain fitness while evading antibody-mediated immunity. *Cell* (2021).

Figure 1. Schematic diagram of SARS-CoV-2 replication in cell showcasing the related to S, N, ORF3a, RdRp, NSP3 and 5'-UTR based epistatic interactions.

The replication cycle starts with the ACE2 receptor binding of the spike glycoprotein (S) as cornered at the top-left and finishes with the exocytosis at the top-right. The viruses which do not carry G-, GH- and/or GR-featured mutations in the S, N, ORF3a, RdRp, NSP3 and 5'-UTR are denoted as the wild type where mutants contain those. Throughout the diagram, the red and green color icons such as proteins, genome, and virion represent the wild and mutant type, respectively. For a generalized virion, we used the blue color. Although this theme is

not show the co-infection of both types, which might occur in rare occasions, we showed the comparative epistatic effects side-by-side fashion during the whole replication cycle that will make it easy to grasp. Related figure(s) for each protein are shown in the enclosed box. To mean the uncertainty or unknown effects of any mutant proteins/RNA structure, we used the 'question' mark in a pathway and on explanatory box. RdRp- RNA dependent RNA polymerase; NSP14- proof-reading enzyme of SARS-CoV-2; ER- endoplasmic reticulum; ERGIC- endoplasmic reticulum (ER) Golgi intermediate compartment.

Figure 2. Global epidemiological scenario of G, GH and GR clades.

(a) Weekly time-plot showing the percentage of G, GHGR and GV clade viruses with worldwide SARS-CoV-2 infections, death cases and death rates from 24 December, 2019 (collection date of first sequenced virus) to 19th October 2020 where the weeks are shown in X axis. As low as 9 (for 2nd week) to 9587 sequences (for 14th Week) were analyzed to determine the ratio of G, GH and GR clade (shown in right Y axis) in the sequenced viruses based on the availability of high coverage genomes submitted in GISAID until 28th of October, 2020. For analyzing the death rate for each week, the number of detected infections and death cases (left Y axis) were collected from WHO situation reports.

(b) Monthly region wise time-plot showing the percentage of G, GH and GR clades with infections, deaths and death rates from January to October of 2020 in 6 WHO regions. 68,125 sequences were analyzed for Europe, 30,395 for Americas, 14778 for West Pacific, 3050 for Southeast Asia, 1170 for Eastern Mediterranean and 1523 for Africa. Infections and death cases were collected from WHO situation reports. The data labeling and values in the X and Y axis are same to figure 2a.

Figure 3. Different structural and stability comparison of wild and mutant spike protein. Structural superposition of wild and mutant spike proteins **(a-b)**; conformation in the S1-S2 **(c)** and S2'sites **(d-e)**; representation of vibrational entropy energy change on the mutant type structure **(f)**; and interatomic interaction prediction of both wild **(g)** and mutant **(h)** types. For Figure a-e, the gray and yellow color represent the wild and mutant protein, respectively. **(a)** The downstream (617-636) of D614G in wild (green) and mutant (red) S protein was focused. Overlapping of the wild (D_{614}) and mutant (G_{614}) S protein showed conformational change in the 3D structures. **(b)** However, the conformational change are in the loop region (618-632) of the proteins thus may potentially play role in interacting with other proteins or enzymes, such as elastase-2 as we focused in this work. **(c)** No change was

found in the S1-S2 cleavage site (685-686), depicted in blue color, of the wild and mutant protein. **(d)** Surface and **(e)** cartoon (2°) structure of the superimposed wild and mutant proteins where the S2' (pink) is situated in surface region and do not show any change in accessibility in the residual loop region. **(f)** The mutant (G_{614}) protein showed higher flexibility in the G^{614} (sticks) and its surroundings (red). The intra-molecular interaction determined the overall stability of the **(g)** wild and **(h)** mutant structure where C_{β} of D^{614} (aspartic acid at 614; green stick modelled) had two hydrophobic interaction with the benzene rings. This intramolecular contacts stabilize the S protein of wild type and missing of this bond destabilize the mutant (G_{614}) protein. The mutant protein has glycine at 614 which has less chance of interacting with other neighboring amino acids due to its shorter and nonpolar R-group. The color code representing the bond type is presented in each **(g)** and **(h)**.

Figure 4. The molecular docking of wild and mutant with elastase-2. Both the (upper figure) wild (D_{614}) and (lower figure) mutant (G_{614}) version of S protein was shown in golden color whereas the elastase-2 docked to D_{614} and G_{614} in blue and green color, respectively. The enlarged views of the docked site were shown in separate boxes. **(a)** The possible docked residues (stick model) on the wild S protein (warm pink) and elastase-2 (green) are 618(Thr)-619(Glu)-620(Val) and 198(Cys)-199(Phe)-225:227 (Gly, Gly, Cys), respectively. The aspartic acid at 614 is 17.3°A far away from the valine (101), apparently the nearest amino acid of elastase-2 to the cleavage site (615-616). **(b)** The possible interacting residues (stick model) on the mutant S protein (blue) and elastase-2 (warm pink) are 614(Gly)-618(Thr)-619(Glu)-620(Val) and 101(Val)-103(Leu)-181(Arg)-222:227(Phe, Val, Arg, Gly, Gly, Cys), and 236 (Ala) respectively. In this case, the glycine at 614 is only 5.4°A far away from the valine (101), the nearest amino acid of elastase-2 to the cleavage site (615-616).

Figure 5. (a) Both the wild and mutated spike protein had lower RMSD profile till 60ns, then it rised and maintained steady state. Although the spike protein had higher degree of deviation in RMSD profile than RdRp but they did not exceed 3.0Å. The RMSD from demonstrated that mutant and wild RdRp protein complex has initial rise of RMSD profile due to flexibility. Therefore, both RdRp complexes stabilized after 30ns and maintained steady peak. The wild type RdRp complex had little bit higher RMSD peak than mutant RdRp which indicates the more flexible nature of the wild type. (b) The spike protein complex had similar SASA profile and did not change its surface volume and maintained similar trend during the whole simulation time. The higher deviation of SASA indicates that mutant and wild type RdRp had straight line but mutant structure had higher SASA profile

which indicates the protein complex had enlarged its surface area. Therefore, mutation in RdRp protein leads to more expansion of the surface area than wild types as their average SASA value had significant difference. (c) Mutated spike exhibits little more Rg profile than the wild type which correlates with the comparative labile nature of the mutant. The higher level of Rg value defines the loose packaging system and mobile nature of the protein systems. The mutant RdRp had lower level of fluctuations and maintains its integrity in whole simulation time. The wild type RdRp complexes had higher deviations and more mobility than the mutant complex. (d) Any aberration in hydrogen bond number can lead to a higher flexibility. Therefore, the mutant and wild spike protein exhibit same flexibility in terms of H-bonding. The mutant RdRp protein had more hydrogen bonding than the wild types, but they did not differentiate too much and relatively straight line was observed for the protein.

Figure 6. The molecular interaction of mutant RdRp with NSP8. The mutant (L_{323}) RdRp (pale green) and NSP8 (light blue) are interacting as shown in center of the lower figure and an enlarged view of the docked site is presented above within a box. The leucine at 323 interacted with the Asp (112), Cys (114), Val (115), and Pro (116). The wild (P_{323}) RdRp has identical docking interactions with NSP8 (table 4), thus is not presented as separate figure here.

Figure 7. The effect on transmembrane channel pore of ORF3a viroporin due to p.Q57H mutation. (a) The wild (Q_{57}) and mutant (H_{57}) ORF3a protein are presented in light gray and green color, respectively. The structural superposition displays no overall conformation change, however the histidine at 57 position of mutant ORF3a (deep blue) has slightly rotated from glutamine at same position of the wild protein (bright red). This change in rotamer state at 57 residue may influence **(b)** the overall stability of H_{57} (upper part) over Q_{57} (lower part) because of ionic interaction of histidine (green; stick model) of transmembrane domain 1 (TM1) with cysteine at 81 (yellow stick) of TM2. The color code defined different bond types is shown in inlet.

Figure 8. Structural superposition of wild and mutant N protein. The light grey color represents both wild ($RG_{203\text{-}204}$) and mutant ($KR_{203\text{-}204}$) N protein. The linker region (LKR: 183-247 amino acids) of wild ($RG_{203\text{-}204}$) and mutant ($KR_{203\text{-}204}$) are in pale yellow and warm pink color, respectively. **(a)** The aligned structures showed a highly destabilizing (Table 3) conformational change from 231 to 247 amino acids within LKR. Other regions of the N

protein, especially the SR-rich region (184-204 amino acids) where the mutations occur, do not change. **(b)** A more emphasized look into the SR-rich and mutated sites (RG203-204KR) of wild and mutant N protein represent slight deviation in the Ser (197) and Thr (198) while only glycine (green) to arginine (blue) substitution at 204 position shows changing at rotamer state. The enlarged view is shown in the bottom part.

Table 1. First occurrence, global frequencies and prevalent zones for co-occurring mutation were shown.

Mutation/Clade	**First occurrence (date, country, accession)**	**Global Frequencies (frequency, percentage)**	**Prevalent zone (zone, frequency, percentage)**
241C>T, 3037C>T, 14408C>T, 23403A>G (G clade)	03-02-20, England EPI_ISL_464302	47943 (16.57%)	Africa (828, 30.22%)
241C>T, 3037C>T, 14408C>T, 23403A>G, 25563G>T (GH clade)	05-02-20, Australia EPI_ISL_480608	61323 (21.19%)	Americas (26647, 55.50%)
241C>T, 3037C>T, 14408C>T, 23403A>G, 28881-3GGG>AAC (GR Clade)	16-02-20, England EPI_ISL_466615	95217 (32.91%)	Western Pacific (10779, 63.50%)
241C>T, 3037C>T, 14408C>T, 23403A>G, 22227C>T (GV Clade)	15-05-20, Mexico EPI_ISL_516622	65195 (22.53%)	Europe (52864, 36.39%)

Table 2. Incidence risk ratio (IRR) for different clade strains against death-case ratio.

Variables	Unadjusted			Adjusted		
	IRR	P value	95% CI	IRR	P value	95% CI
G clade	1.03	0.020	1.01 - 1.06	1.03	0.012	1.01 - 1.06
GH clade	1.01	0.171	0.99 - 1.03	1.00	0.789	0.99 - 1.02
GR clade	0.99	0.044	0.98 - 0.99	0.98	<0.001	0.97 - 0.99
GV clade	0.98	<0.001	0.97 – 0.98	0.97	<0.001	0.96 - 0.98

Table 3. The scores of HADDOCK, PRODIGY (ΔG and Kd (M) at 37.0 □) for RdRp/NSP8 and Spike-Elastase docked complex.

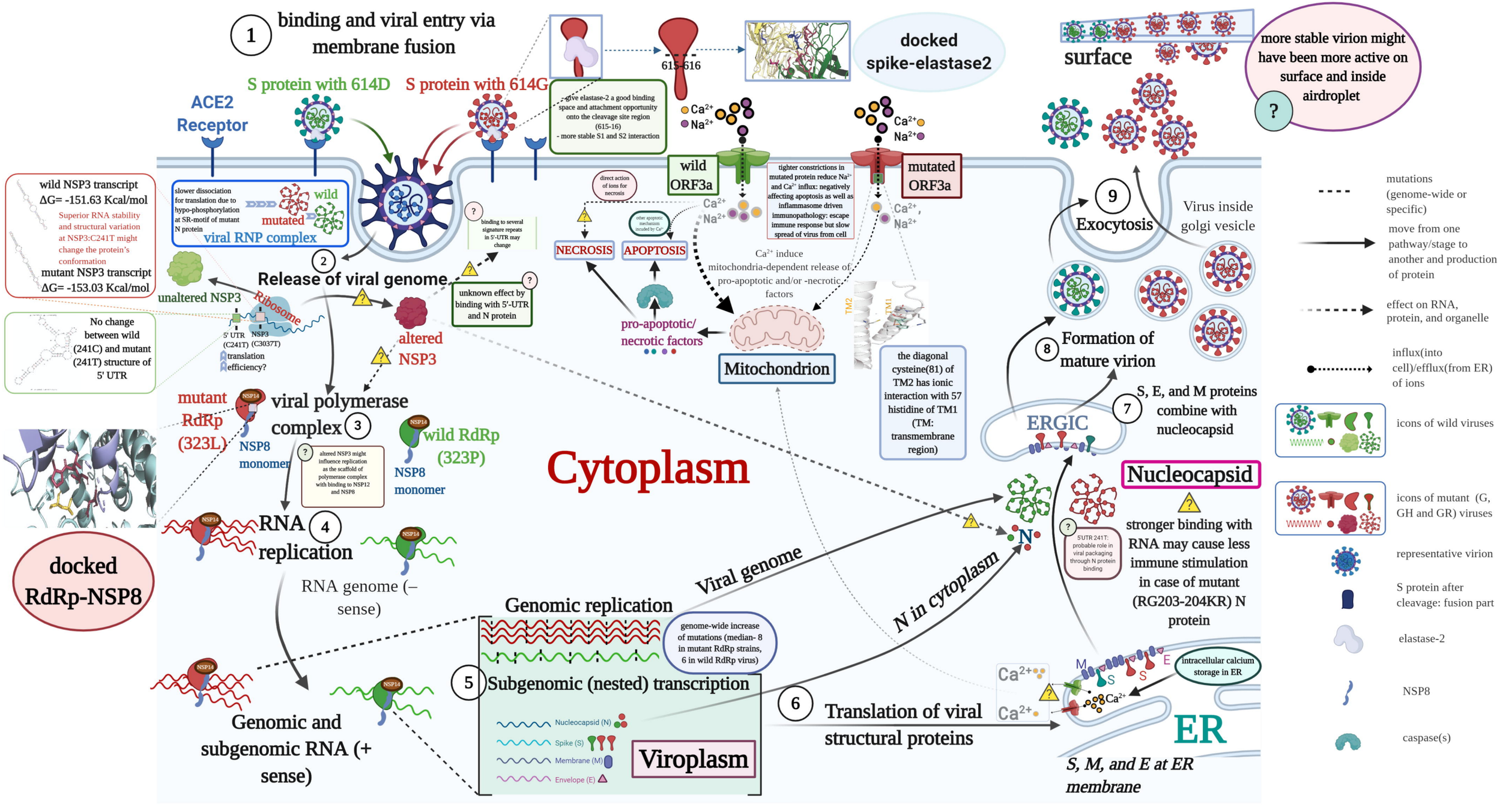
Variables	types	RdRp/NSP8	Spike-Elastase
HADDOCK score	Wild	-82.2 +/- 7.8	-43.0 +/- 8.9
	Mutant	-118.3 +/- 2.5	-61.9 +/- 4.5
ΔG (kcal mol^{-1})	Wild	-10.6	-13.3
	Mutant	-10.5	-13.7
Kd (M) at 37.0 □	Wild	3.5E^{-08}	4.5E^{-10}
	Mutant	3.9E^{-08}	2.3E^{-10}
Number of interfacial contacts (ICs) per property	Wild	charged-charged (5); charged-polar (9); charged-apolar (15); polar-polar (2); polar-apolar (16); and apolar-apolar (21)	charged-charged (17);charged-polar (22);charged-apolar (32);polar-polar (5);polar-apolar (31); and apolar-apolar (23)
	Mutant	charged-charged(5);charged-polar (16);charged-apolar (19); polar-polar (3);polar-apolar(15); apolar-apolar (23)	charged-charged (13);charged-polar (18);charged-apolar (27);polar-polar (4); polar-apolar (28) and apolar-apolar (36)
Associated amino acids of Elastase-2 with possible docking interactions (for spike) or NSP8 (for RdRp)	Wild	P323: Asp(112), Cys(114), Val(115) and Pro (116)	605 (Ser) and 607 (Gln): 36 (Arg); 618 (Thr): 199 (Phe); 619 (Glu): 199 (Phe), Cys (227); 620 (Val): 198 (Cys), 225:227(Gly, Gly, Cys)
	Mutant	P323: Asp(112), Cys(114), Val(115) and Pro (116)	614 (Gly): 101 (Val); 618 (Thr): 181 (Arg), 223-226(Val, Arg, Gly, Gly); 619 (Glu): 103 (Leu), 181(Arg), 222-225(Phe, Val, Arg, Gly), 236 (Ala); 620 (Val): 223-227 (Val, Arg, Gly, Gly, Cys)

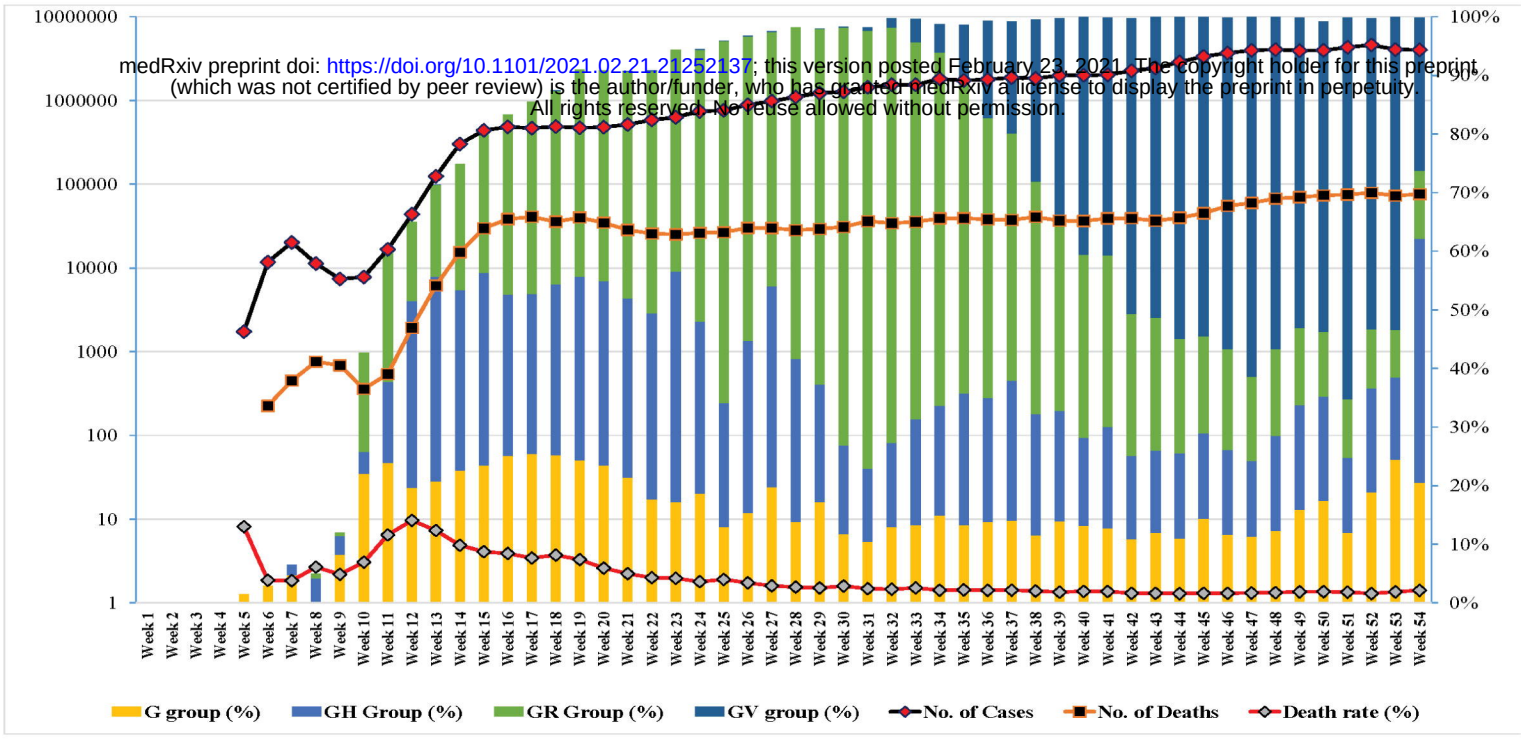
Table 4. Assess the effect of mutations on structural dynamics of NSP-12/ RDRP, Spike, NS3 and N Protein of SARS CoV-2 using DynaMut. The value of ΔΔG <0 indicates that the mutation causes destabilization and ΔΔG > 0 represents protein stabilization. For

ΔΔSVibENCoM, positive and negative value denotes the increase and decrease of molecular flexibility, respectively.

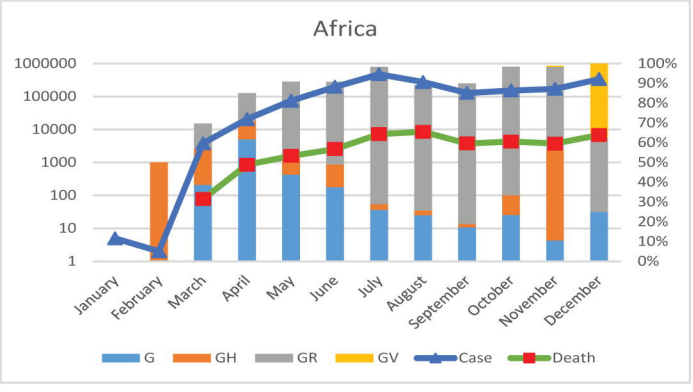
Protein Name	Mutation with position	ΔΔG DynaMut kcal/mol	ΔΔG ENCoM kcal/mol	ΔΔG mCSM kcal/mol	ΔΔG SDM kcal/mol	ΔΔG DUET kcal/mol	ΔΔG FoldX (kcal/mol)	Results*	$\Delta\Delta S_{Vib}$ ENCoM $kcal.mol^{-1}.K^{-1}$
RdRp	P323L	1.054	-0.441	-0.264	0.700	0.118	-0.733	Stabilizing	-0.551
Spike	D614G	-0.769	+0.408	-0.492	2.530	0.195	+0.289	Destabilizing	0.510
ORF3a	Q57H	0.275	-0.128	0.788	0.520	-0.464	-1.438	Stabilizing	-0.160
N Protein	RG203-04KR	-	-	-	-	-	-3.42262	Highly Destabilizing	-
N protein	A220V	0.109	0.458	-0.586	-1.460	-0.567	+1.6	Stabilizing	-0.572

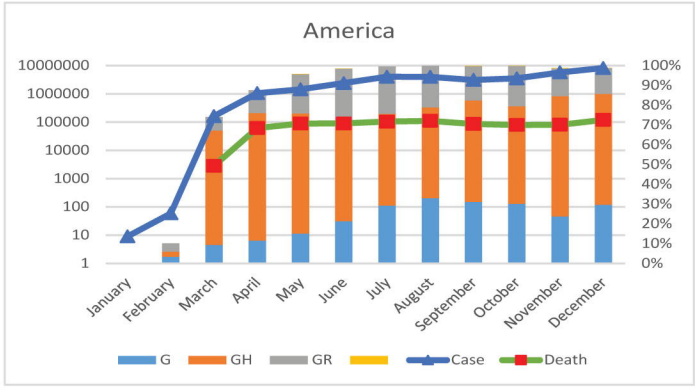
***The final result of the stability for each protein was determined based on the intra-molecular interactome analysis.**

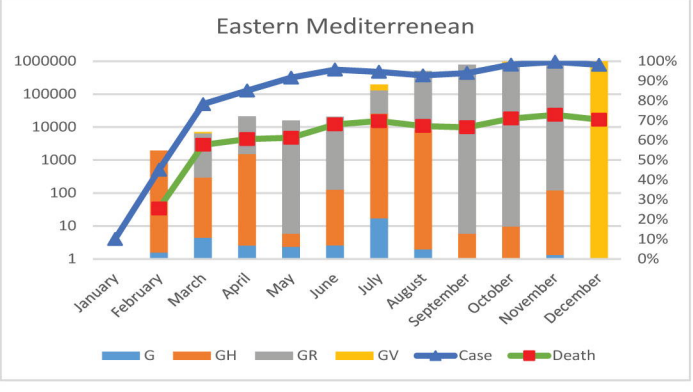

1 binding and viral entry via membrane fusion
S protein with 614D
S protein with 614G
ACE2 Receptor
615-616
docked spike-elastase2
give elastase-2 a good binding space and attachment opportunity onto the cleavage site region (615-16)
- more stable S1 and S2 interaction
surface
more stable virion might have been more active on surface and inside airdroplet
wild NSP3 transcript
ΔG= -151.63 Kcal/mol
Superior RNA stability and structural variation at NSP3:C241T might change the protein's conformation
mutant NSP3 transcript
ΔG= -153.03 Kcal/mol
No change between wild (241C) and mutant (241T) structure of 5' UTR
slower dissociation for translation due to hypo-phosphorylation at SR-motif of mutant N protein
wild
mutated
viral RNP complex
binding to several signature repeats in 5'-UTR may change
unknown effect by binding with 5'-UTR and N protein
2 Release of viral genome
unaltered NSP3
Ribosome
5' UTR (C241T)
NSP3 (C3037T)
translation efficiency?
altered NSP3
wild ORF3a
mutated ORF3a
direct action of ions for necrosis
other apoptotic mechanism incuded by Ca2+
NECROSIS
APOPTOSIS
tighter constrictions in mutated protein reduce Na2+ and Ca2+ influx: negatively affecting apoptosis as well as inflammasome driven immunopathology: escape immune response but slow spread of virus from cell
Ca2+ induce mitochondria-dependent release of pro-apoptotic and/or -necrotic factors
pro-apoptotic/ necrotic factors
Mitochondrion
TM2
TM1
the diagonal cysteine(81) of TM2 has ionic interaction with 57 histidine of TM1 (TM: transmembrane region)
9 Exocytosis
Virus inside golgi vesicle
8 Formation of mature virion
7 S, E, and M proteins combine with nucleocapsid
ERGIC
mutant RdRp (323L)
NSP8 monomer
viral polymerase complex 3
wild RdRp (323P)
NSP8 monomer
altered NSP3 might influence replication as the scaffold of polymerase complex with binding to NSP12 and NSP8
Cytoplasm
docked RdRp-NSP8
RNA replication 4
RNA genome (– sense)
Viral genome
N in cytoplasm
N
Nucleocapsid
stronger binding with RNA may cause less immune stimulation in case of mutant (RG203-204KR) N protein
5'UTR 241T: probable role in viral packaging through N protein binding
Genomic replication
genome-wide increase of mutations (median- 8 in mutant RdRp strains, 6 in wild RdRp virus)
5 Subgenomic (nested) transcription
Nucleocapsid (N)
Spike (S)
Membrane (M)
Envelope (E)
Viroplasm
Genomic and subgenomic RNA (+ sense)
6 Translation of viral structural proteins
intracellular calcium storage in ER
ER
S, M, and E at ER membrane
mutations (genome-wide or specific)
move from one pathway/stage to another and production of protein
effect on RNA, protein, and organelle
influx(into cell)/efflux(from ER) of ions
icons of wild viruses
icons of mutant (G, GH and GR) viruses
representative virion
S protein after cleavage: fusion part
elastase-2
NSP8
caspase(s)

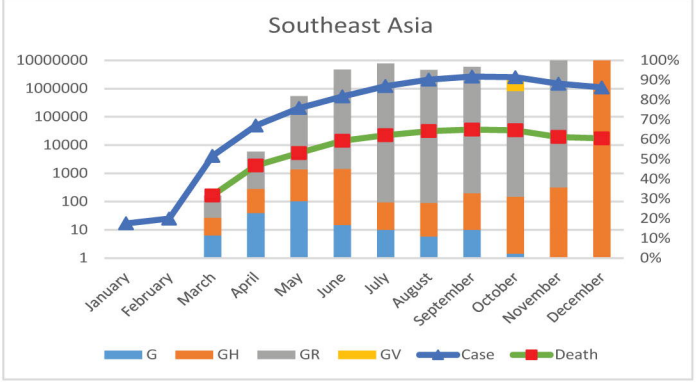


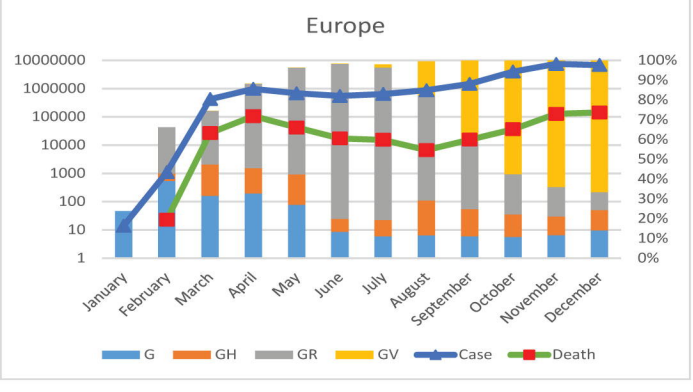
a

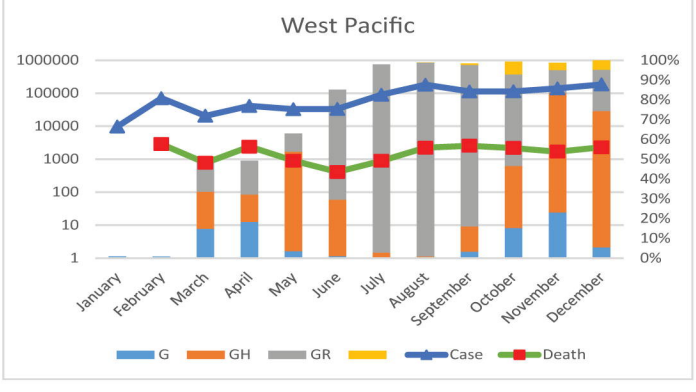




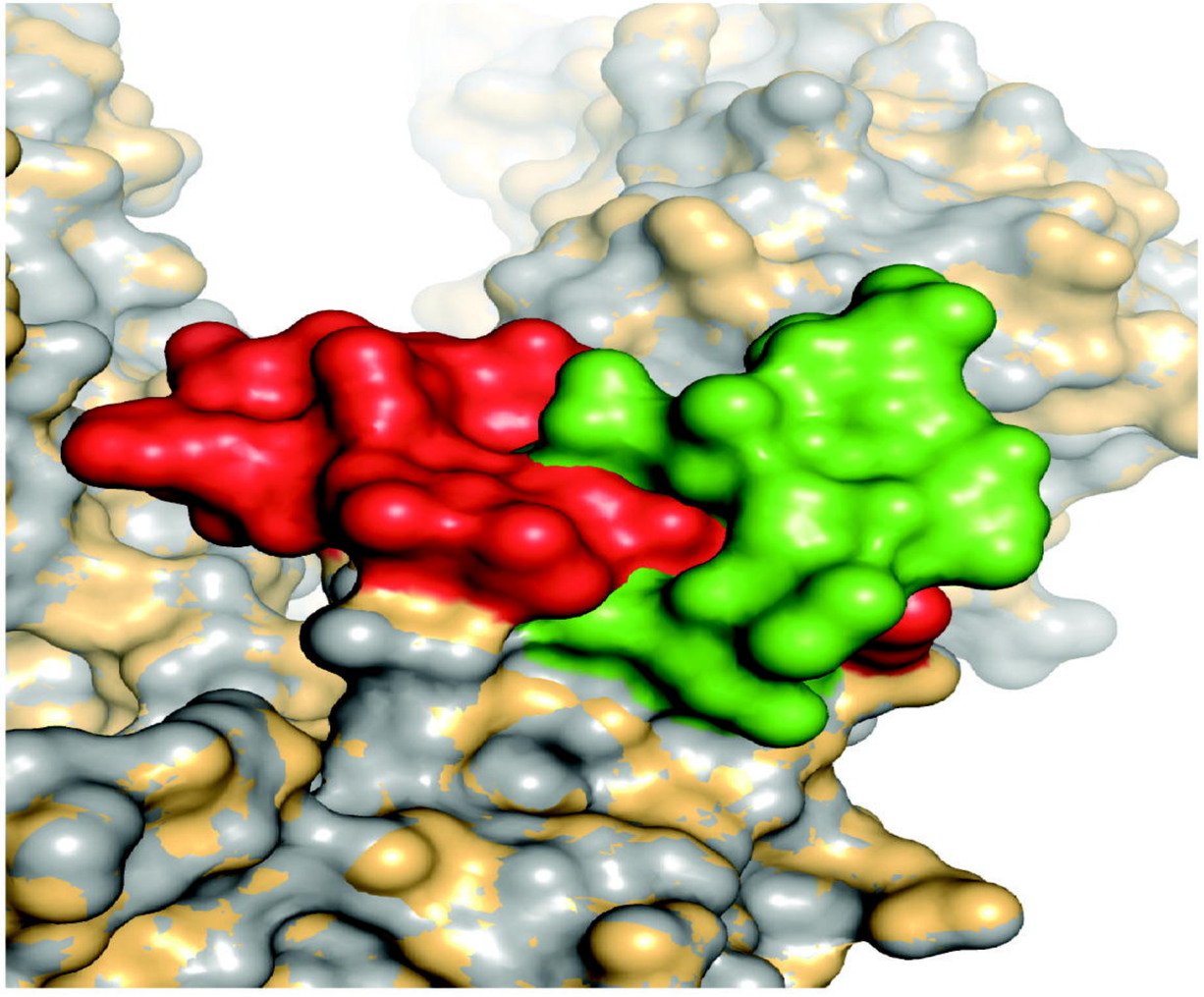


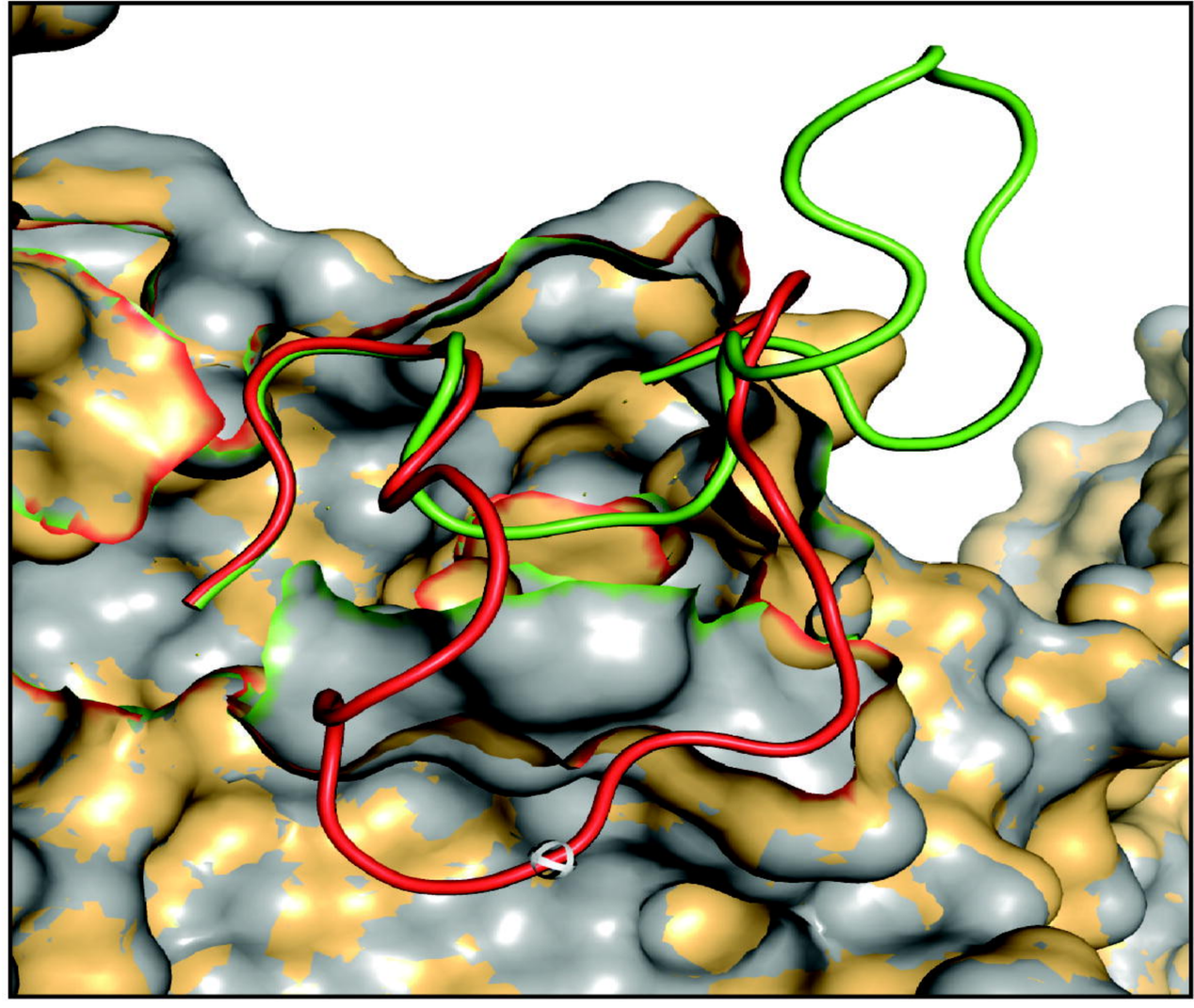




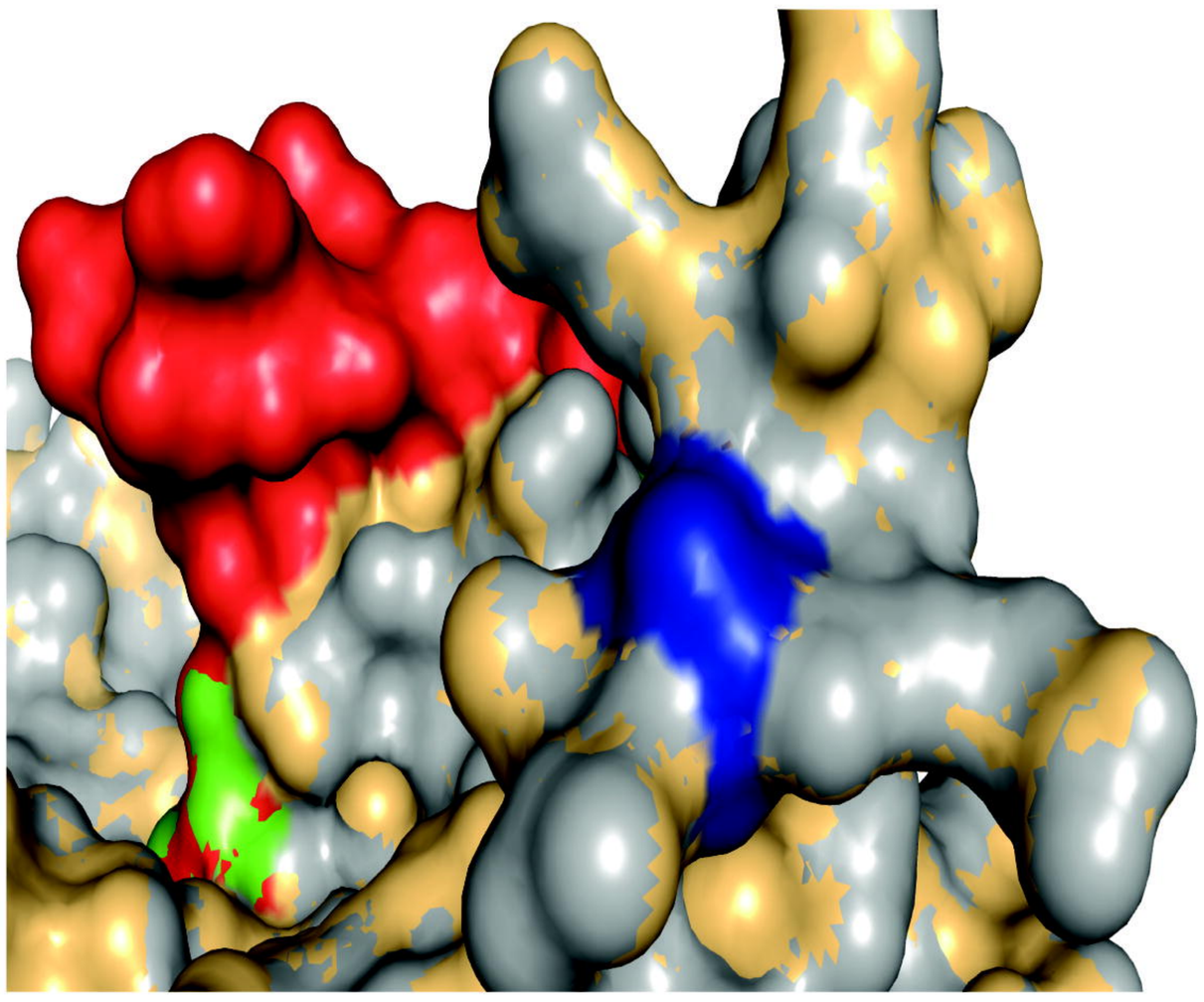


b

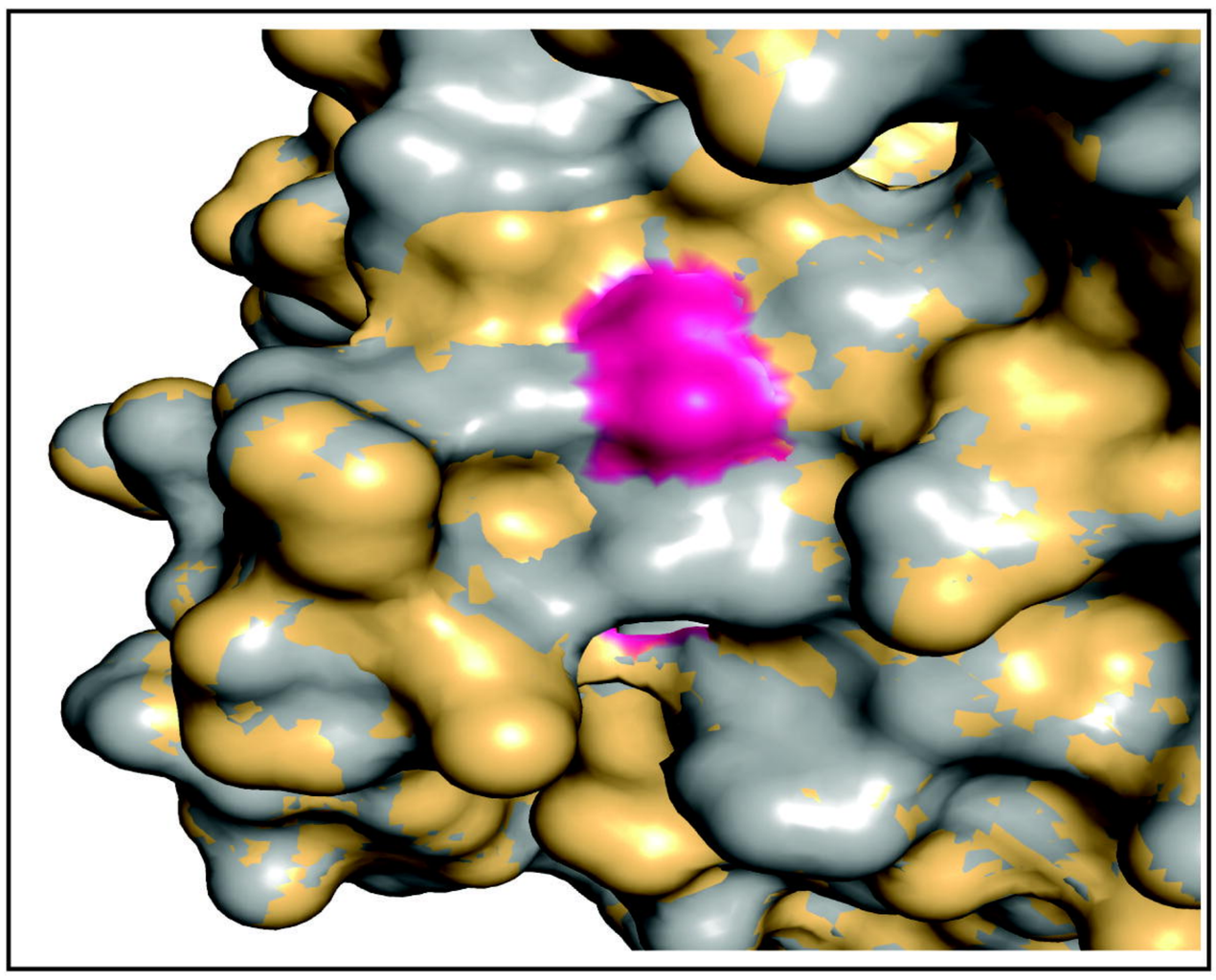




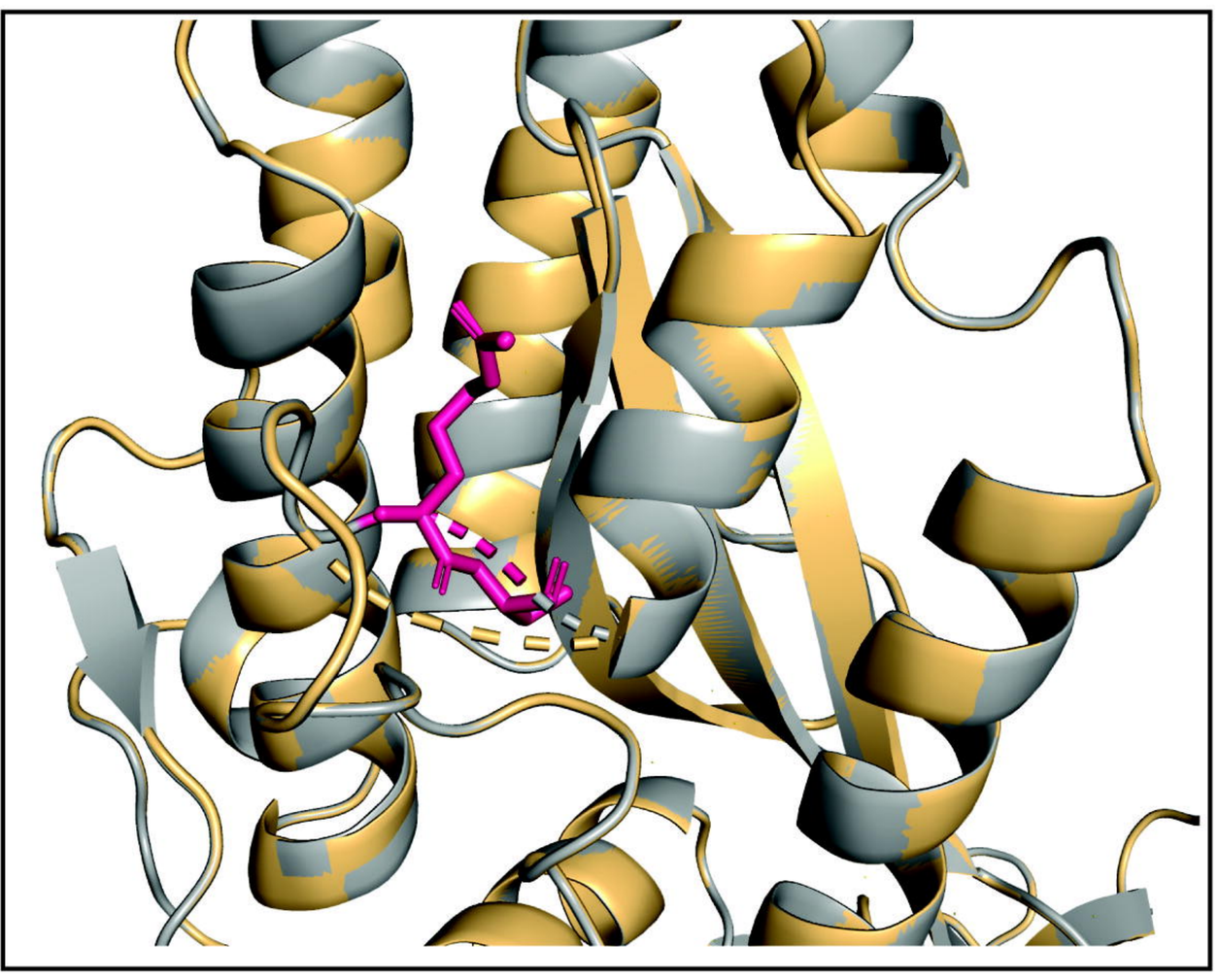
a



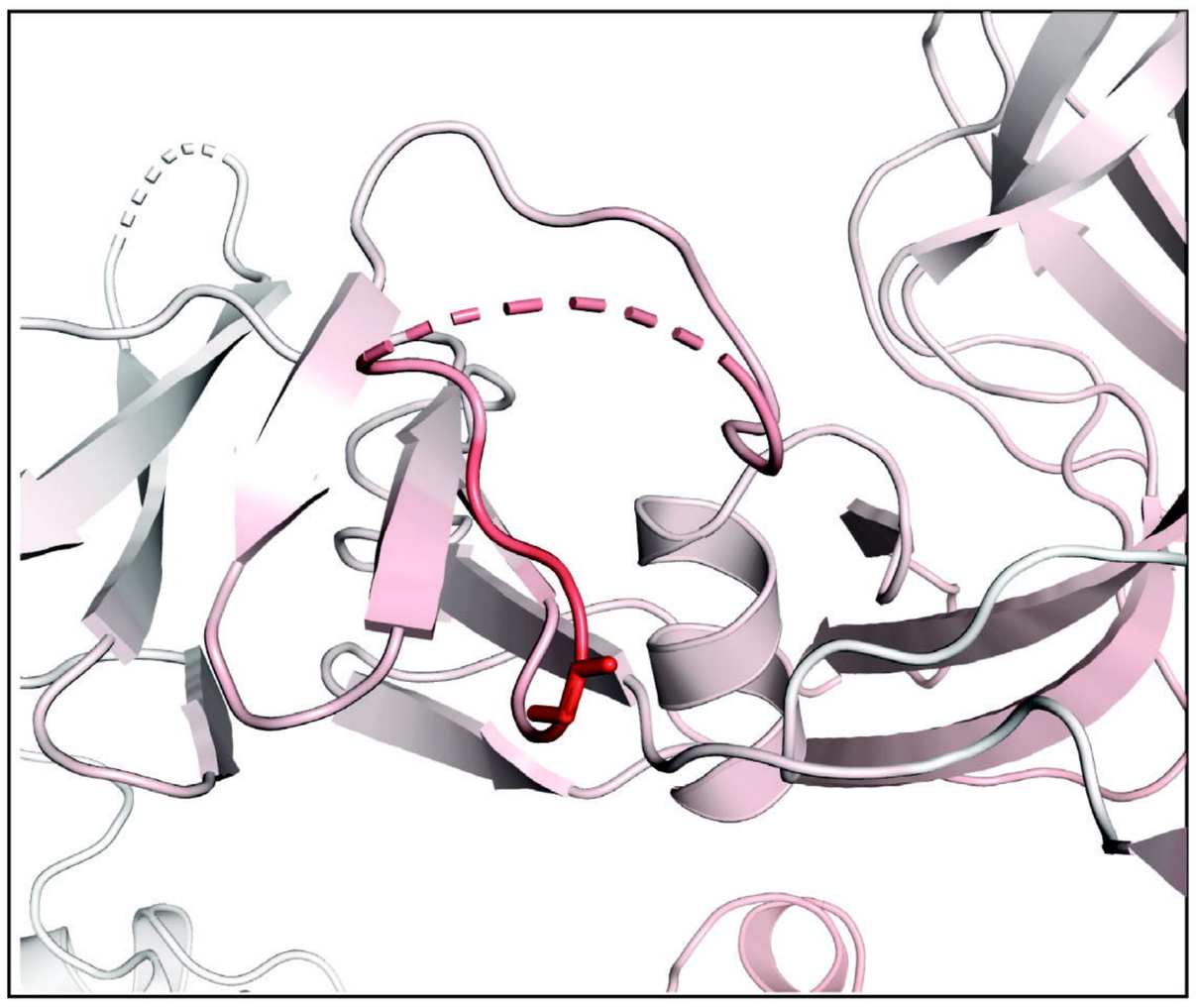
b



c



d



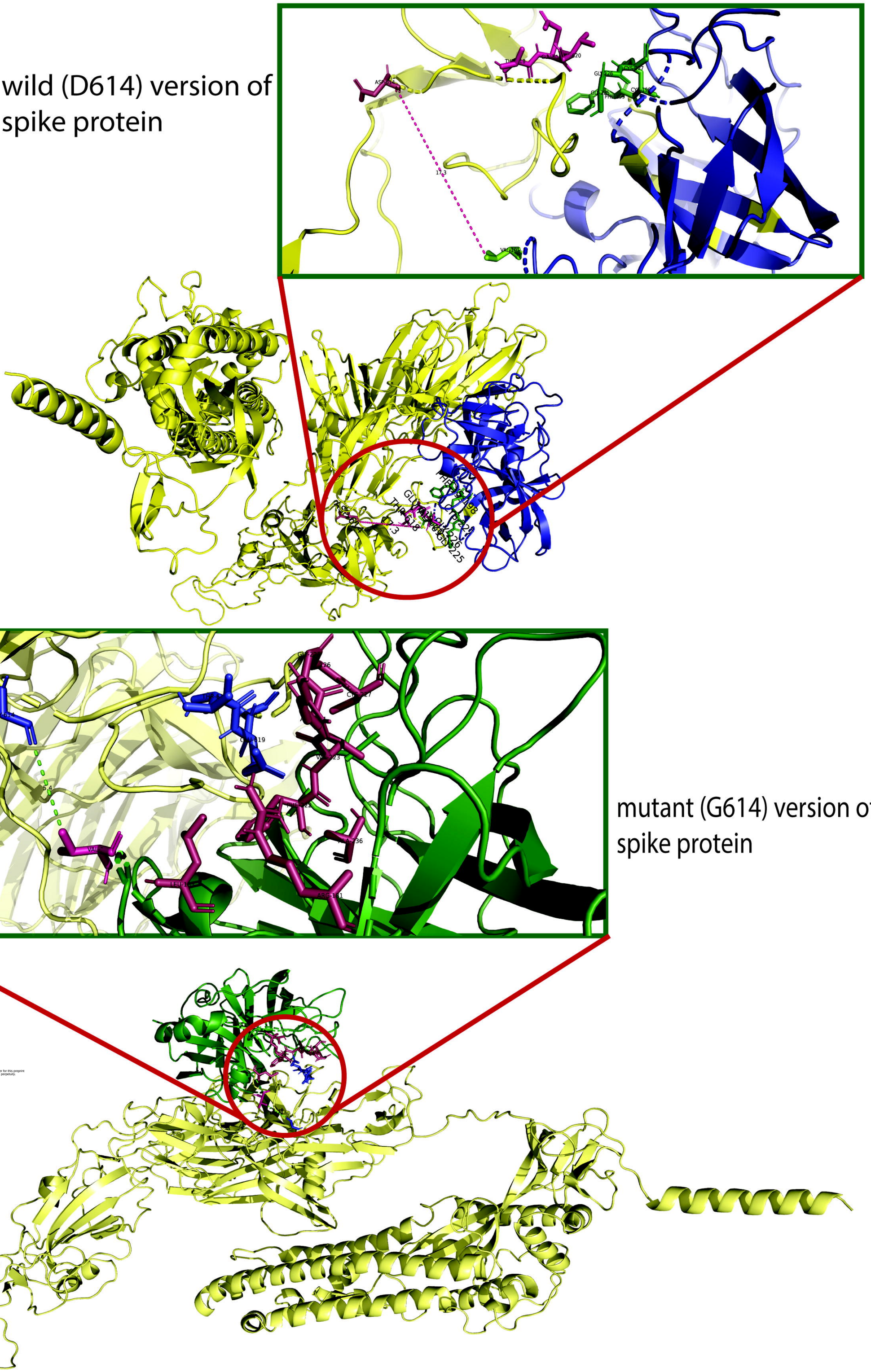
e

f

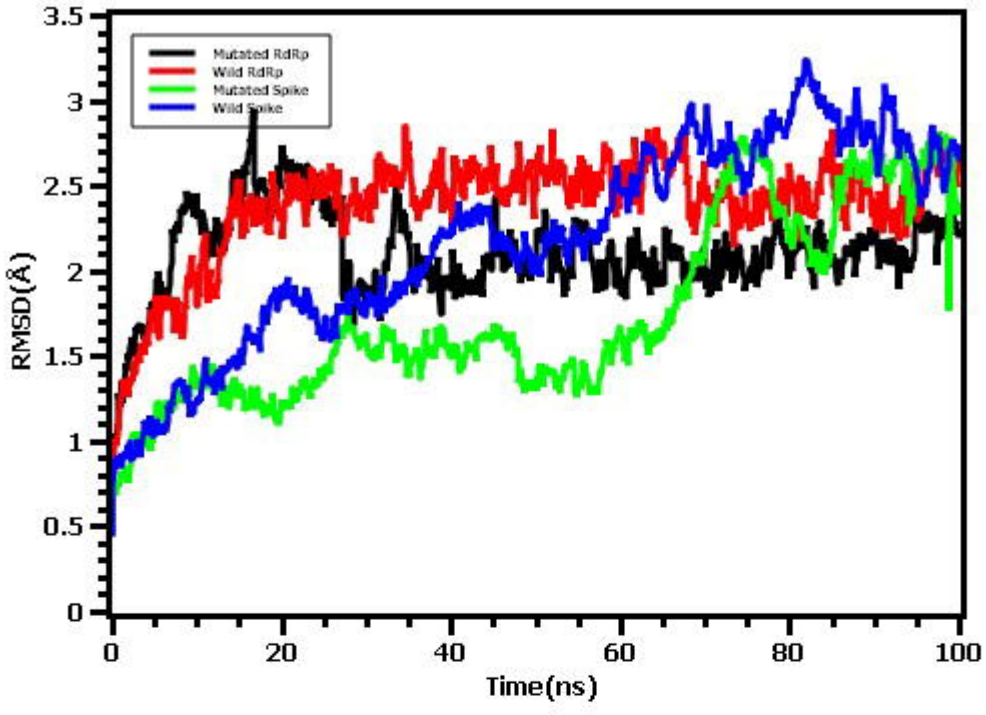
g

h

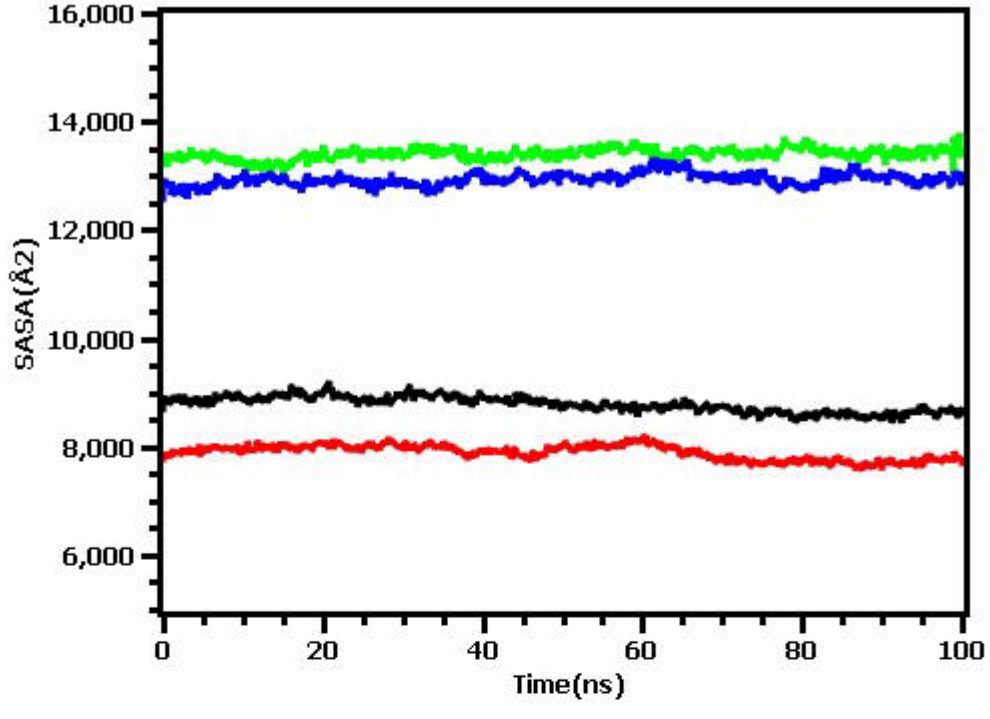
wild (D614) version of spike protein



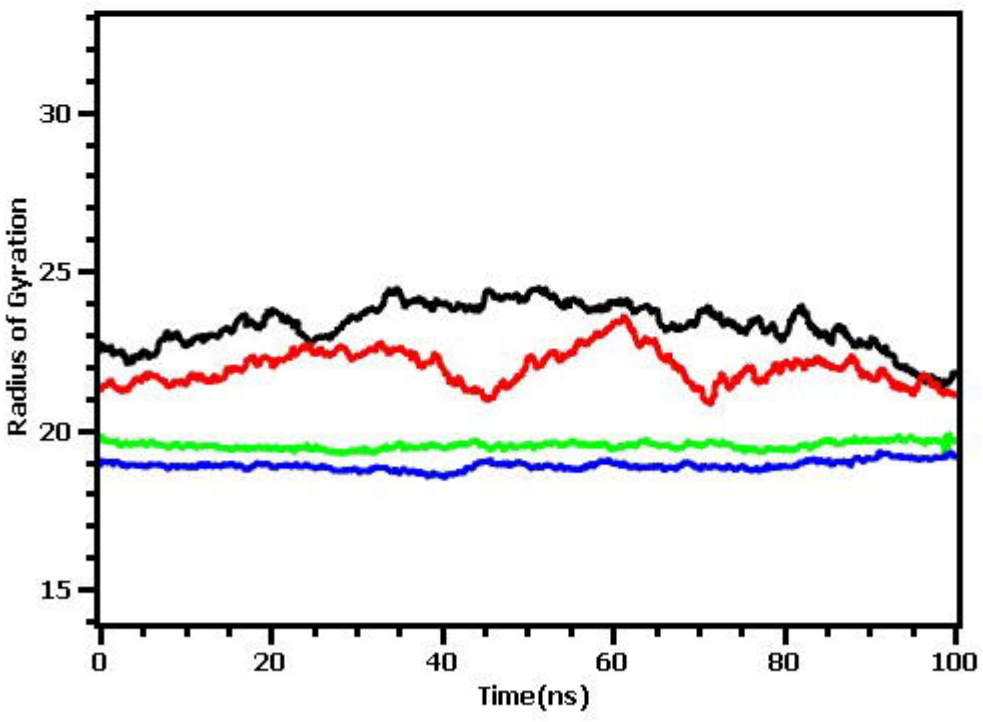
mutant (G614) version of spike protein



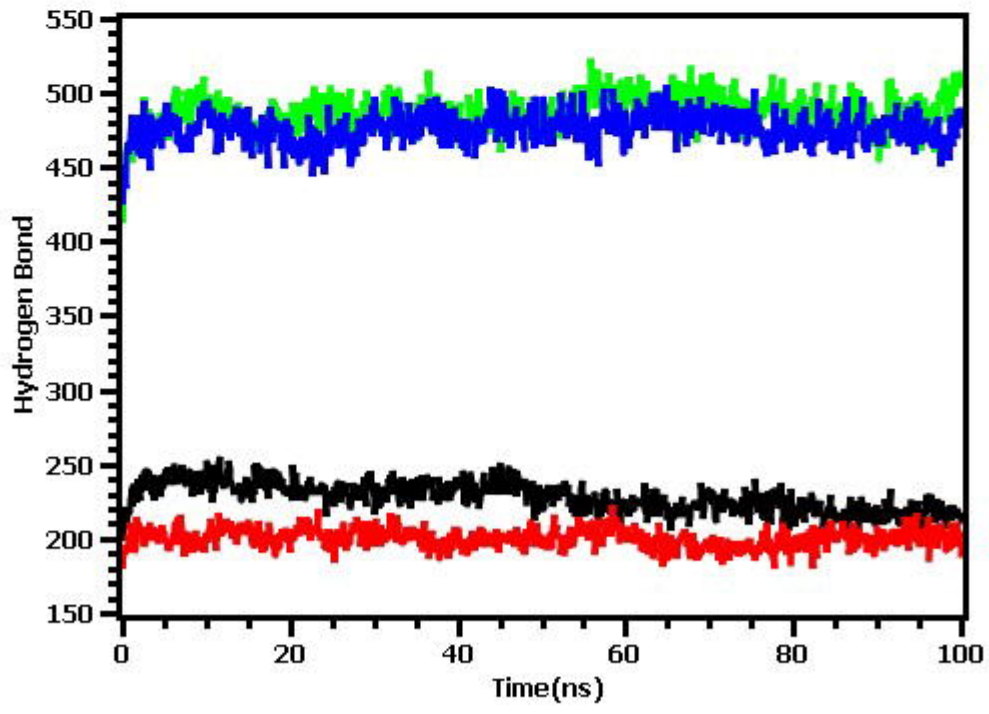
(a)



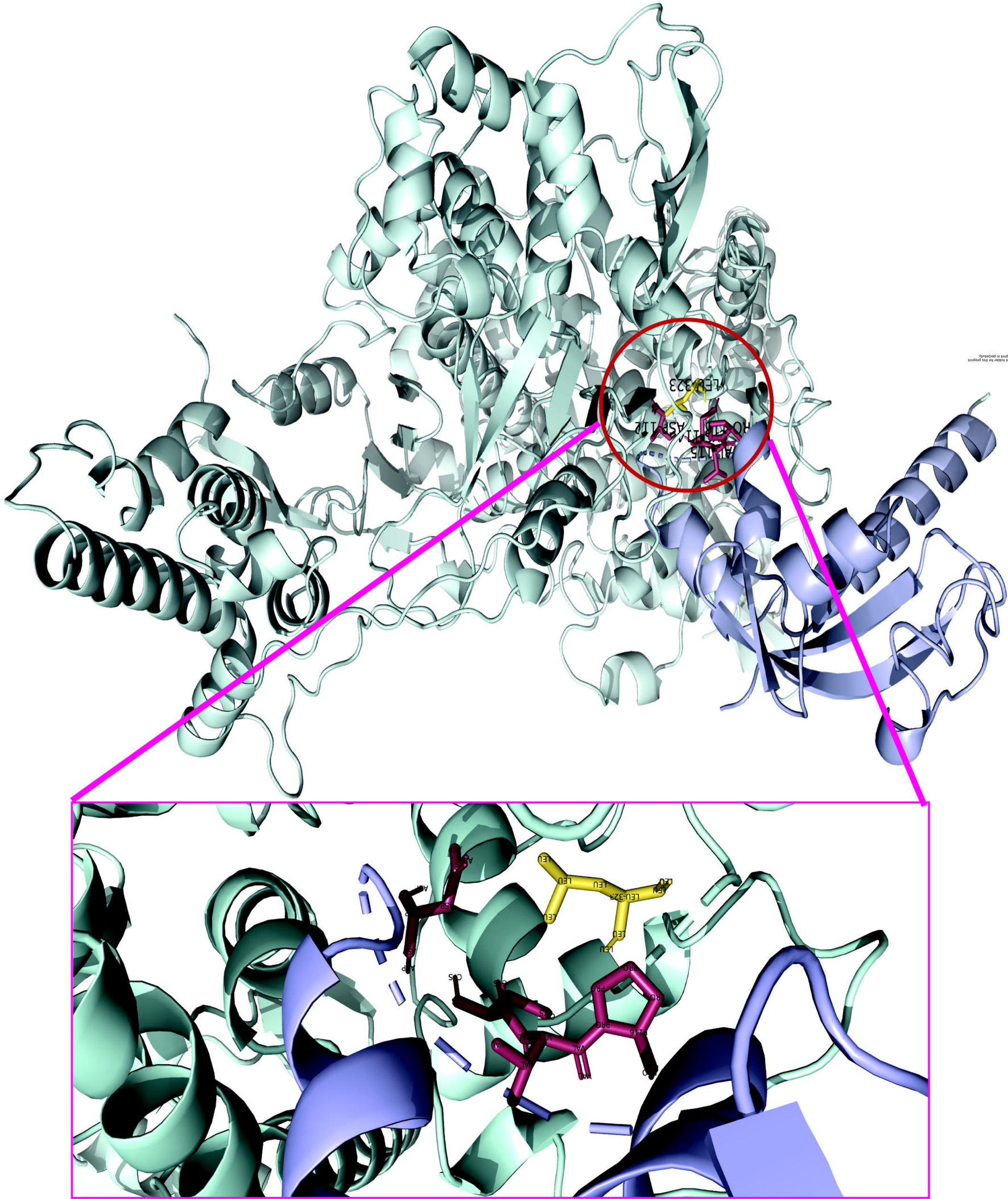
(b)

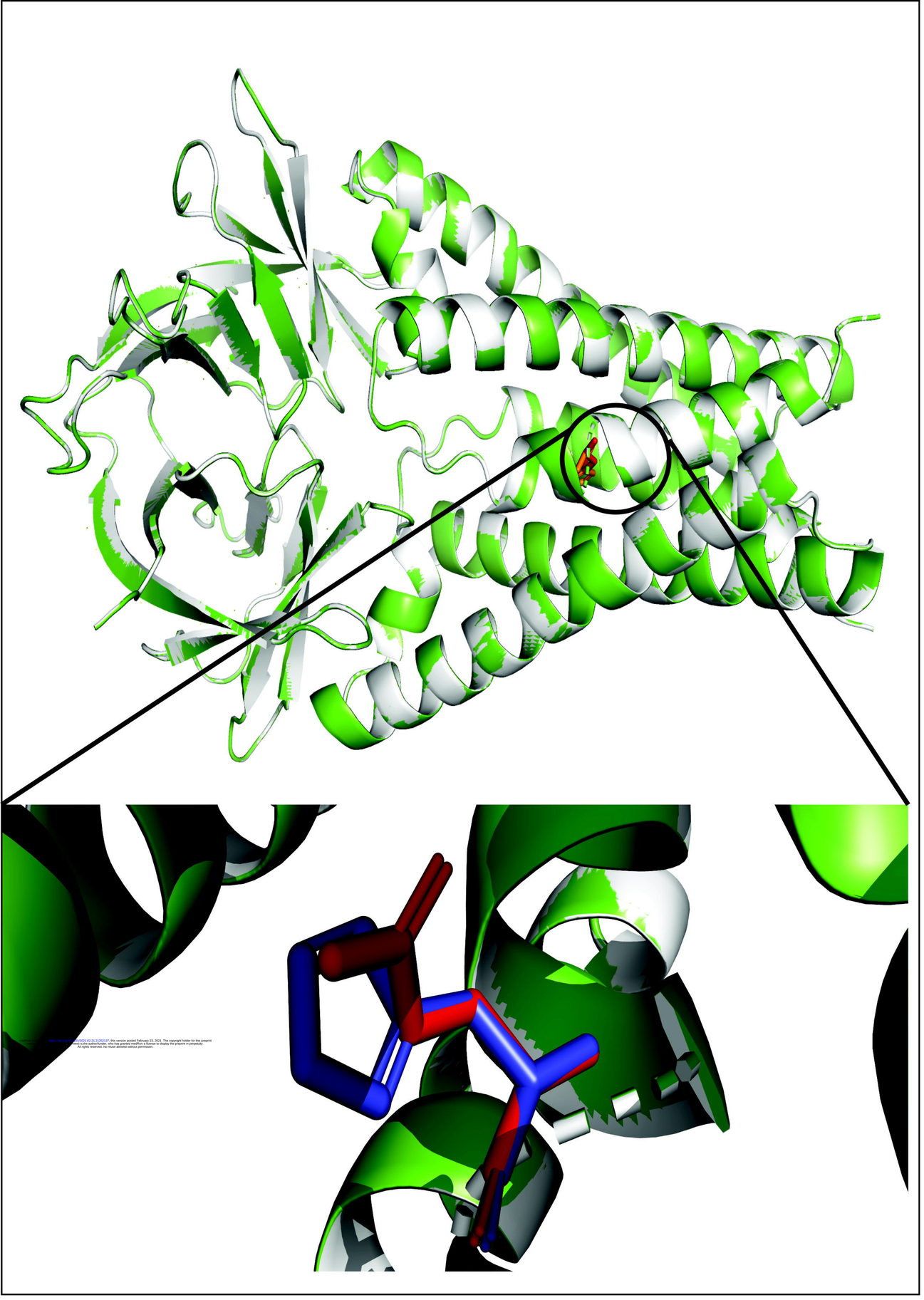


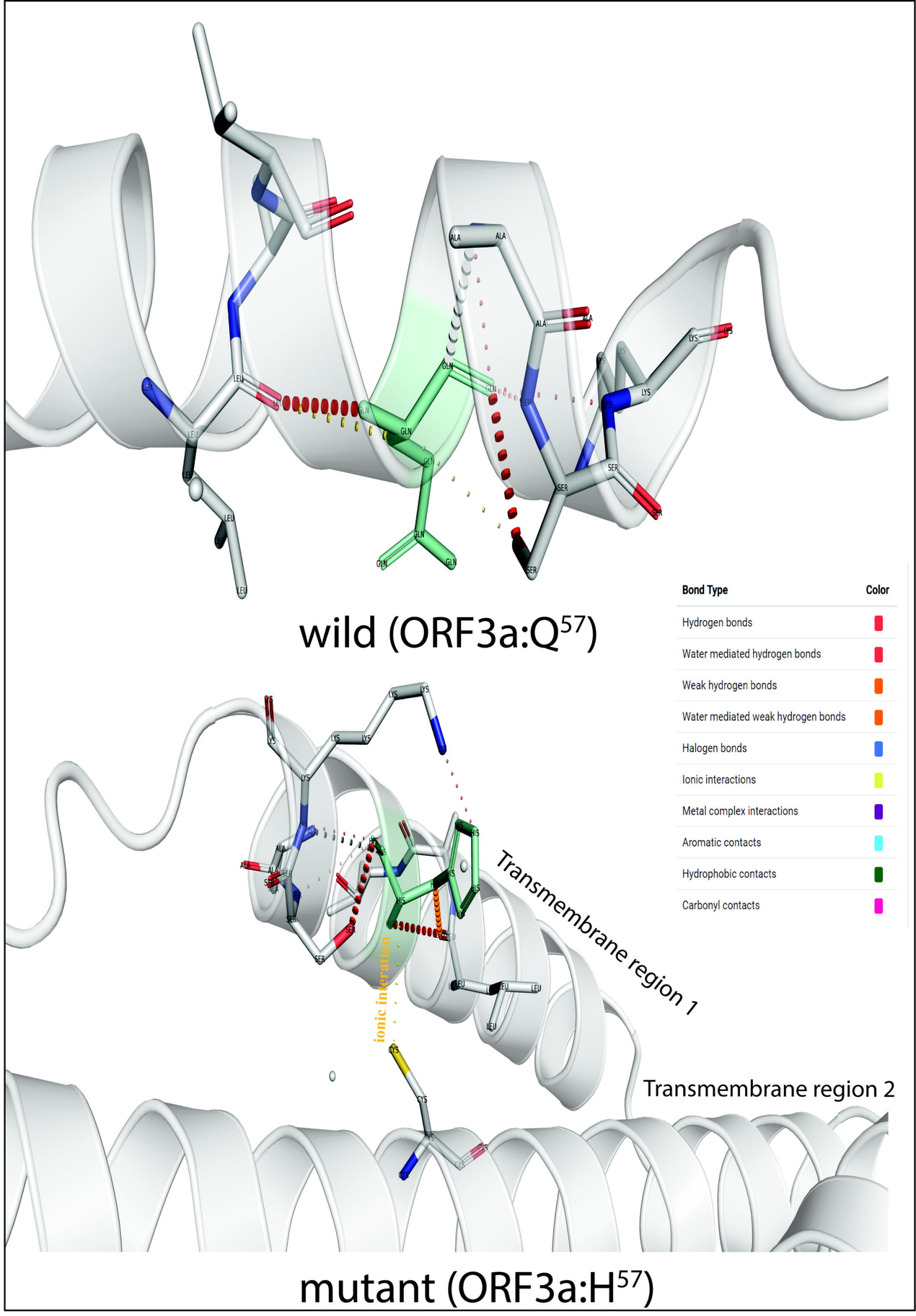
(c)



(d)

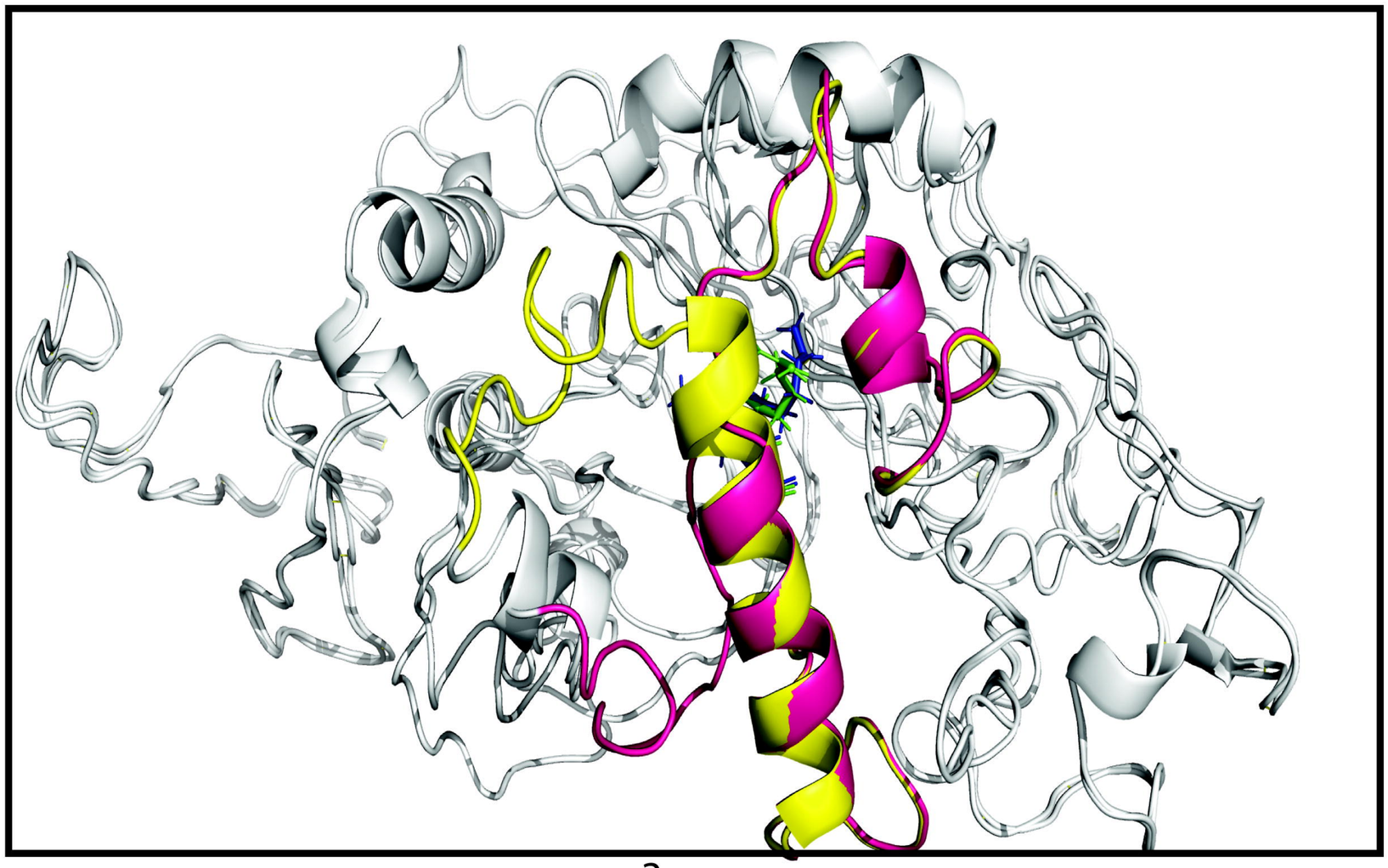




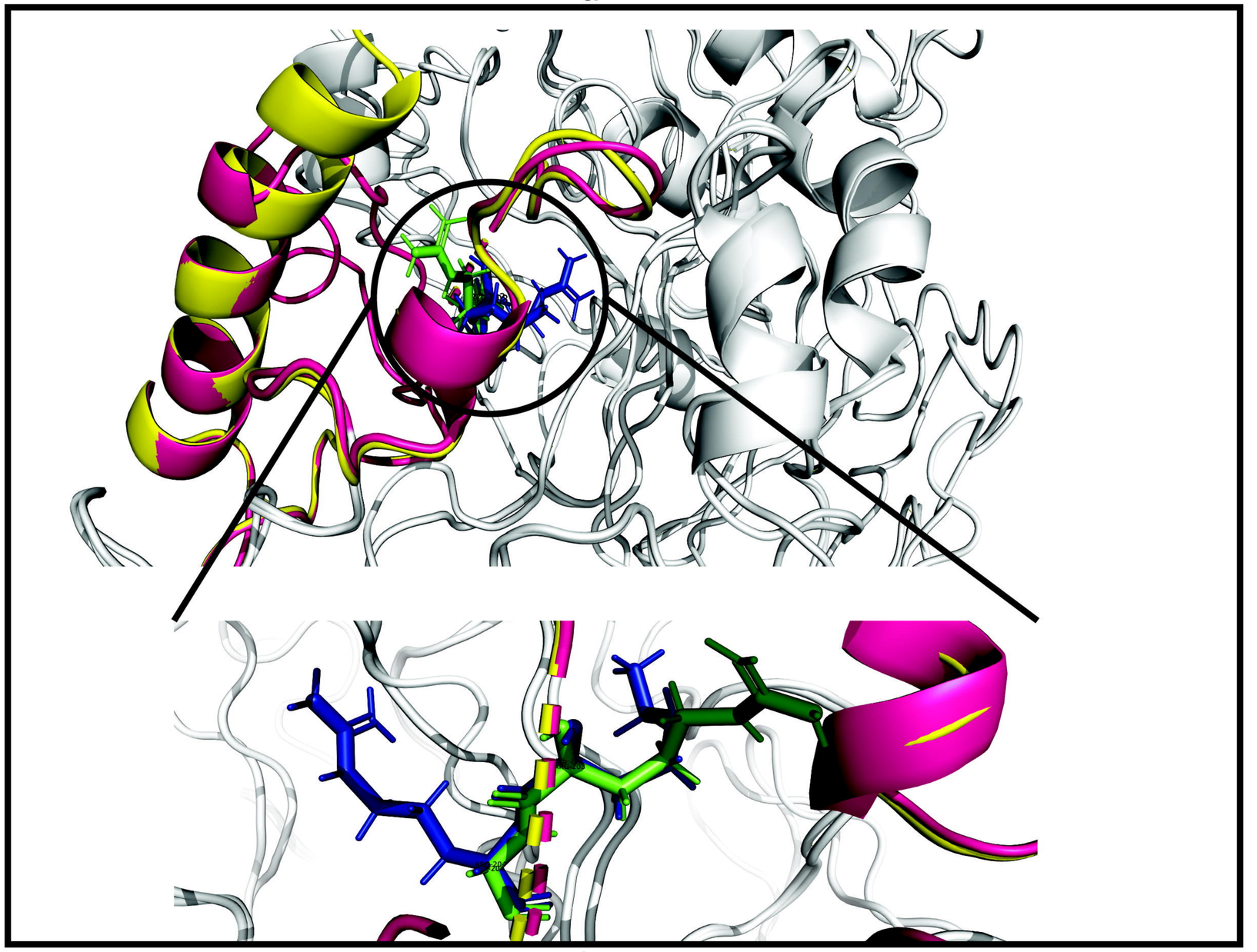

wild (ORF3a:Q57)
Bond Type
Color
Hydrogen bonds
Water mediated hydrogen bonds
Weak hydrogen bonds
Water mediated weak hydrogen bonds
Halogen bonds
Ionic interactions
Metal complex interactions
Aromatic contacts
Hydrophobic contacts
Carbonyl contacts
Transmembrane region 1
ionic interaction
Transmembrane region 2
mutant (ORF3a:H57)

a

b



a



b