Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Infection, Genetics and Evolution 97 (2022) 105128

Contents lists available at ScienceDirect

Infection, Genetics and Evolution

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

Research paper

Genomic diversity and molecular dynamics interaction on mutational variances among RB domains of SARS-CoV-2 interplay drug inactivation

Md Bashir Uddin [a,1], Emran Hossain Sajib [b,1], Syeda Farjana Hoque [c], Md. Nazmul Islam Bappy [b], Fazle Elahi [d], Arabinda Ghosh [e], Samuel Muhit [f], Mohammad Mahmudul Hassan [g], Mahmudul Hasan [c], Ramachandran Chelliah [d], Se Jin Park [h], Tamanna Jahan Mony [h], Deog-Hwan Oh [d,*], Syed Sayeem Uddin Ahmed [i,*]

[a] Department of Medicine, Sylhet Agricultural University, Sylhet 3100, Bangladesh
[b] Faculty of Biotechnology and Genetic Engineering, Sylhet Agricultural University, Sylhet 3100, Bangladesh
[c] Department of Pharmaceuticals and Industrial Biotechnology, Sylhet Agricultural University, Sylhet 3100, Bangladesh
[d] Department of Food Science and Biotechnology, College of Agriculture and Life Sciences, Kangwon National University, Chuncheon, Gangwon-do 24341, Republic of Korea
[e] Microbiology Division, Department of Botany, Gauhati University, Assam 781014, India
[f] School of Health Sciences, Central Michigan University, Mount Pleasant, MI 48859, United States
[g] Department of Physiology, Biochemistry and Pharmacology, Chattogram Veterinary and Animal Sciences University, Chattogram 4225, Bangladesh
[h] School of Natural Resources and Environmental Sciences, Kangwon National University,Chuncheon 24341, Gangwon-do, Republic of Korea
[i] Department of Epidemiology and Public Health, Sylhet Agricultural University, Sylhet 3100, Bangladesh

ARTICLE INFO

Keywords:
SARS-CoV-2
Phylogenomic diversity
Mutational analysis
S protein
Remdesivir
Molecular dynamics simulation

ABSTRACT

The scientific community has been releasing whole genomic sequences of SARS-CoV-2 to facilitate the investigation of molecular features and evolutionary history. We retrieved 36 genomes of 18 prevalent countries of Asia, Europe and America for genomic diversity and mutational analysis. Besides, we studied mutations in the RBD regions of Spike (S) proteins to analyze the drug efficiency against these mutations. In this research, phylogenenetic analysis, evolutionary modeling, substitution pattern analysis, molecular docking, dynamics simulation, etc. were performed. The genomic sequences showed >99% similarity with the reference sequence of China.TN93 + G was predicted as a best nucleotide substitution model. It was revealed that effective transition from the co-existing SARS genome to the SARS-CoV-2 and a noticeable positive selection in the SARS-CoV-2 genomes occurred. Moreover, three mutations in RBD domain, Val/ Phe367, Val/ Leu 382 and Ala/ Val522, were discovered in the genomes from Netherland, Bangladesh and the USA, respectively. Molecular docking and dynamics study showed RBD with mutation Val/Leu382 had the lowest binding affinity with remdesivir. In conclusion, the SARS-CoV-2 genomes are similar, but multiple degrees of transitions and transversions occurred. The mutations cause a significant conformational change, which are needed to be investigated during drug and vaccine development.

1. Introduction

Coronavirus disease (COVID-19), an infectious disease caused by a novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) that emerged from Wuhan province in China (Huang et al., 2020a; Li et al., 2020a). It may trigger a respiratory tract infection, which affects both upper and lower respiratory systems. It spreads mainly through person-to-person contact. SARS-CoV-2 can be spread by infected secretions such as saliva and respiratory secretions, as well as by respiratory droplets, through direct, indirect, or near contact with infected people (Liu et al., 2020a; Chan et al., 2020; Burke et al., 2020; Hamner et al., 2020; Ghinai et al., 2020). It can also be spread through the air during medical procedures that produce aerosols ("aerosol generating procedures") (WHO, 2020). Infections may be mild or fatal. Fever, coughing,

* Corresponding authors.
E-mail addresses: deoghwa@kangwon.ac.kr (D.-H. Oh), ahmedssu.eph@sau.ac.bd (S.S.U. Ahmed).
[1] These authors contributed equally.

https://doi.org/10.1016/j.meegid.2021.105128
Received 11 August 2021; Received in revised form 16 October 2021; Accepted 25 October 2021
Available online 6 November 2021

shortness of breath, difficulty breathing, vomiting, chills (sometimes with shaking), body aches, headache, sore throat, congestion/runny nose, loss of smell or taste, nausea, and diarrhea are the most common symptoms, but it can also lead to pneumonia, respiratory failure, heart attacks, liver problems, septic shock, and death. More than 200 countries have been affected worldwide with 4,579,667 deaths until September 05, 2021 from this virus. Till now no potential therapy has been found to improve the conditions of patients in any clinical trials (Sanders et al., 2020), though many clinical trials are being conducted on COVID-19 around the world (Wu et al., 2020; Liu et al., 2020b).

Coronaviruses (CoVs) are single-stranded RNA, enveloped and pathogenic virus (Xu et al., 2020). In contrast to the previously described SARS-CoV (2002) and Middle East respiratory syndrome coronavirus (MERS-CoV, 2013), SARS-CoV-2 is more pathogenic (Naqvi et al., 2020). About 82% of SARS-CoV-2 genome had identity with SARS-CoV and MERS-CoV, and basic enzymes and structural proteins had >90% sequence identity. (Naqvi et al., 2020). Therefore, a common pathogenesis mechanism can be found to follow the common therapeutic targeting proteins. Spike (S), envelope (E), membrane (M), and nucleocapsid (N) proteins are the four important structural proteins of SARS-CoV-2. These proteins have a high degree of sequence similarity with the corresponding proteins of SARS-CoV and MERS-CoV (Naqvi et al., 2020). These CoVs vary primarily in their host entry mechanism, implying that changes in the residual composition of S-glycoprotein can determine host entry.

It is the spike (S) protein that SARS-CoV-2 uses to bind with the host cell-surface receptor, Angiotensin Converting Enzyme 2 (ACE-2), during host cell entry (Lau and Peiris, 2005), and the same receptor is used by SARS-CoV and CoV-NL63 to enter the host cells (Chen, 2020; Chen et al., 2020). However, SARS-CoV-2 S protein is longer than SARS-CoV S protein, and its receptor binding region is completely different according to the first studies published (Lu et al., 2020). S protein has an extracellular N-terminus, a transmembrane (TM) domain anchored in the viral membrane, and a short intracellular C-terminal segment with a size of 180–200 kDa (Bosch et al., 2003). The length of SARS-CoV-2 S protein is 1273 amino acid that consists of a signal peptide (amino acids 1–13) located at the N-terminus, the S1 subunit (14–685 residues), and the S2 subunit (686–1273 residues). First, the S1 domain binds with the ACE2 receptor which is followed by the fusion of S2 domain with membrane (Bertram et al., 2013; Hoffmann et al., 2020). The S protein binds to the host receptor through the S1 subunit's receptor-binding domain (RBD) (Zhang et al., 2020). It is the TM protease serine 2 (TMPRSS2), a type 2 TM serine protease, that is located on the host cell membrane activate the S protein and promotes virus entry into the cell after the binding of S proteins to the receptor (Hoffmann et al., 2020; Du et al., 2007). The SARS-CoV-2 S protein, which is involved in receptor recognition, attachment, and host cell entry, is highly conserved across all human coronaviruses (HCoVs). These essential functions made it an important target for COVID-19 vaccine and therapeutic research (Huang et al., 2020b; Yan et al., 2020).

In the present study, genomic sequences of SARS-CoV-2 from 18 Asian, European and American countries were collected and analysed. Multiple bioinformatics approaches were conducted to study evolutionary origin and analyze mutations in the SARS-CoV-2 genomes of those countries. Molecular evolutionary modeling was performed to study the evolutionary antiquities between the genomes. Among the studied genomes, we searched mutations in receptor binding domains (RBD) of S-proteins which we considered in homology modeling approach for predicting three-dimensional structures. Then, the validated protein molecule was used for molecular docking purpose with remdesivir. Remdesivir is an antiviral drug that has been used against COVID-19, which mode of action in inhibiting the disease is unclear, but it is predicted that it may inhibit the entry of the virus or interfere with the function of viral RNA dependent RNA polymerase (Eastman et al., 2020). Because of the unavailability of any known SARS-CoV-2 entry inhibitor, remdesivir was used in molecular docking purpose as a model drug in order to comprehend the effects of mutations in RBD region. Further, molecular dynamics (MD) simulation was conducted for thorough analysis of mutations in the S-protein, and its probable effects on drug binding. Mutational analysis was basically performed based on binding interactions, binding sites and dynamic behavior. This analysis would provide a profound knowledge for the discovery of drugs and the development of vaccines against SARS-CoV-2.

2. Results

2.1. Phylogenomic diversity analysis

A total of 36 genome sequences along with reference sequence (NC_045512.2) of SARS-CoV-2 were collected from the NCBI database (Supplementary file 1). MUSCLE and BLAST results showed that all the sequences were highly similar to the reference sequence (Supplementary file 2). Additionally, CLUSTAL-O and MAFFT reported the presence of mutations in all the sequences of SARS-CoV-2. ML method was used for the phylogenetic tree construction with the highest log likelihood (−41,753.00). The percentage replicates in which taxa clustered together are shown next to the branches (Fig. 1), which presents the evolution of SARS-CoV-2 in the different countries of Asia, Europe, and America. Estimation of evolutionary history was done by the ME method. To show the evolutionary history of the taxa, the consensus tree inferred from 100 replicates (Fig. 2). Nevertheless, the branches corresponding to partitions that were replicated in less than 50% of bootstrap replicates, however, were collapsed.The evolutionary distances were determined using the number of base substitutions per site as a unit of measurement.

2.2. Mutational analysis and molecular evolutionary modeling of SARS-CoV-2

The evaluation was done for the best ML DNA/protein models for nucleotide substitutions. There were 24 nucleotide substitution models in total for ML fits (Table 1). All of them, the TN93 + G model was found as the best nucleotide substitution pattern model, with the lowest BIC scores (Table 1). Then, the test for homogeneity of substitution patterns among genome sequences was performed. Significant *P*-values were considered less than 0.05 and labeled with yellow color. The estimates of the disparity index per site were displayed for every sequence pair above the diagonal (Table 2). Next, the ML substitution matrix was estimated to observe the probabilities of substitution that lead to the most likely model of evolution, from which molecular evolution of SARS-CoV-2 can be determined (Table 3). Relative values of instantaneous r were considered during evaluation, and A = 29.90%, T/U = 32.12%, C = 18.36%, and G = 19.62% were the nucleotide frequencies. A tree topology was automatically computed for estimating ML values. The maximum Log-likelihood for this computation was −41,753.003.

ML estimation of transition/transversion bias (R) was also calculated, which was 3.00, to analyze the DNA evolution pattern and estimation of sequence distances throughout the genomes. For this computation, the maximum Log likelihood was −42,674.229, and the nucleotide frequencies were A = 25.00%, T/U = 25.00%, C = 25.00%, and G = 25.00%. Between the sequences, Codon-based test of neutrality was also conducted. The probability of rejecting the null hypothesis of strict-neutrality (dN = dS) (below diagonal) is illustrated in Table 4. Values of $P < 0.05$ were considered significant (Table 4). Tajima's neutrality D test statistic based on allele frequencies is a common method to identify the selection and nucleotide polymorphisms in the genome sequences. This test based on allele frequencies is useful to detect the total number of genes that go through the selection. Thus, Tajima's neutrality test was also performed which values are represented in Supplementary Table 1.

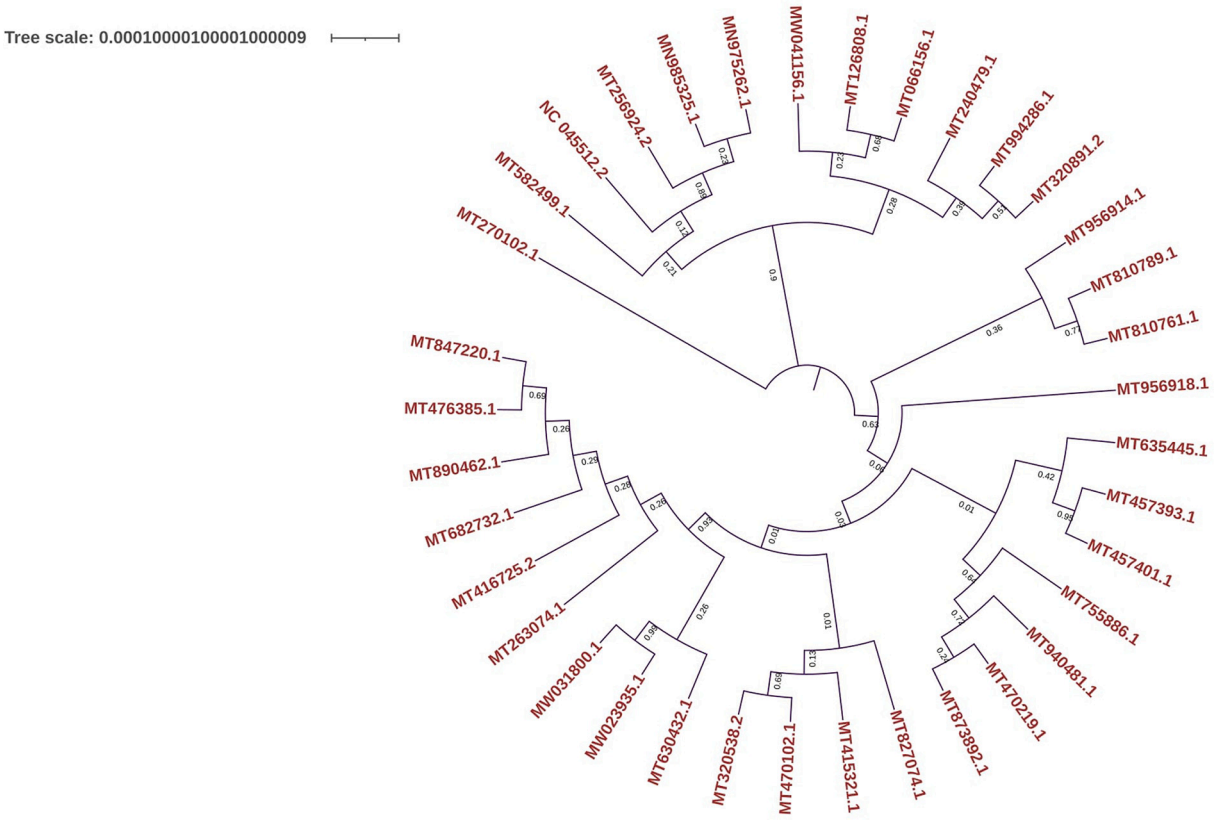

Fig. 1. Phylogenetic tree of SARS-CoV-2 constructed by using the Maximum Likelihood method. Initial tree(s) for the heuristic search were obtained automatically by applying Neighbor-Join and BioNJ algorithms to a matrix of pairwise distances estimated using the Tamura-Nei model, and then selecting the topology with superior log likelihood value. This analysis involved 36 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding.

2.3. Homology modeling and quality assessment of the models

All the 36 S-proteins of SARS-CoV-2 viruses were modeled by employing template-based homology modeling with SARS-CoV-2 spike receptor-binding domain bound with ACE2 (PDB ID: 6M0J) as template. Basically, by considering the query sequences and selected template, Swiss-Model developed the 3D structures for all the 36 receptor binding domain (RBD) regions of 36 S-proteins based on the PDB ID: 6M0J. The S-proteins that do not have mutation in receptor-binding domain regions showed 100% sequence identity with the spike receptor-binding domain region of PDB ID: 6M0J. On the other hand, the mutated receptor-binding domains of S-proteins experienced less sequence identity. Among these models, one sequence from Netherland (MT457401.1: QJS39627.1), one sequence from Bangladesh (MT847220.1: QMU94884.1) and one sequence from the USA (MT940481.1: QNL36022.1) showed mutations in the RBD regions. The residue Val367 was mutated to Phe367 (Netherland; MT457401.1: QJS39627.1), Val382 was mutated to Leu382 (Bangladesh; MT847220.1: QMU94884.1) and Ala522 was mutated to Val522 (The USA; MT940481.1: QNL36022.1) showing 99.48% sequence identity with PDB: 6M0J (Fig. 3). Remaining sequences including the reference sequence (China; NC_045512.2: YP_009724390.1) showed 100% sequence identity with PDB: 6M0J (Fig. 3).

The reference S-protein sequence (China; YP_009724390.1) was used as the representative of 33 S-proteins that did not show any mutation in RBD region. The homology model of RBD of representative S-protein revealed 0.80 as molprobity score with 96.35% in ramachandran favored region and 0.00% in ramachandran outliers region. QMEAN server showed it's qmean score as −1.66. Moreover, the result of ramachandran plot analysed by Procheck revealed 90.5% residues in most favored, 9.5% residues in additional allowed and 0.0% residues in disallowed region. Overall quality factor validated by ERRAT was 92.697 for this representative protein (Fig. 4).

Besides, the models of Netherland (MT457401.1: QJS39627.1), Bangladesh (MT847220.1:QMU94884.1), and the USA (MT940481.1: QNL36022.1) showed molprobity scores 0.85, 0.80 and 1.18 with 95.83%, 96.35% and 95.31% in ramachandran favored regions, respectively (Figs. 5, 6, 7). Qmean scores were − 1.69 (Netherland; MT457401.1: QJS39627.1), −1.66 (Bangladesh; MT847220.1: QMU94884.1) and − 1.67 (The USA; MT940481.1: QNL36022.1) (Figs. 5, 6, 7). Procheck server showed 91.7% (Netherland; MT457401.1: QJS39627.1), 90.5% (Bangladesh; MT847220.1: QMU94884.1) and 91.7% (The USA; MT940481.1: QNL36022.1) residues in most favored regions, while other residues were in additional allowed regions. Furthermore, overall quality factors were 91.573, 92.697 and 92.179 for Netherland (MT457401.1: QJS39627.1), Bangladesh (MT847220.1:QMU94884.1), and the USA (MT940481.1: QNL36022.1), respectively. Besides, the amino acid distributions of all the models were represented by comparing expected and observed structure (Supplementary Figs. 1, 2, 3, & 4).

2.4. Molecular docking and binding site identification

The modeled proteins and remdesivir (ligand) were utilized for molecular docking to evaluate the binding affinity and compare binding energies to detect the effect of mutation in RBD regions of S-proteins.

Molecular interactions of remdesivir with the modeled RBD of representative S protein from China and with the mutated RBDs from the USA, Netherlands and Bangladesh are displayed in Fig. 8. The RBD of S protein of China complexed with remdesivir displayed well accommodation in the binding pocket with binding energy $\Delta G = -4.5$ kcal/mol (Fig. 8A). Binding pocket displayed the formation of conventional

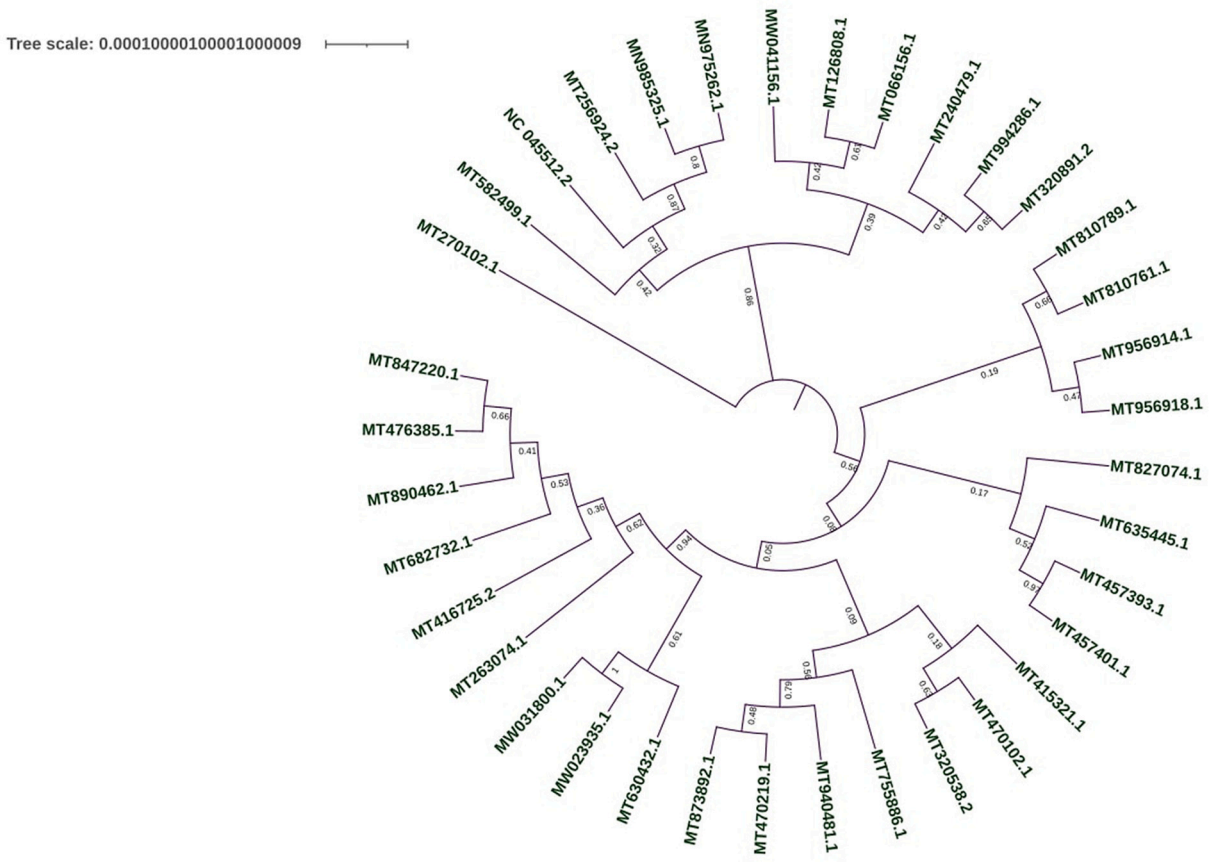

Fig. 2. The evolutionary history of SARS-CoV-2 was analysed by the Minimum Evolution method. The Neighbor-joining algorithm used to generate the initial tree. This analysis involved 36 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. All ambiguous positions were removed for each sequence pair (pairwise deletion option).

hydrogen bonds with Arg 355, Lys 462 and Ser 514 (Fig. 8B). While the other weak interactions were taken place such as pi-alkyl and van der Waal's with the ligand and binding site residues to attain a good binding. While remdesivir with the mutated RBD of S protein of the USA displayed loose binding at the binding cavity with much lower binding energy $\Delta G = -3.2$ kcal/mol (Fig. 8C). ASN 343 and ASP 364 are the principal amino acid residues of binding cavity in the USA forming conventional hydrogen bonds (Fig. 8D). On the other hand, remdesivir displayed high binding at the cavity of the mutated RBD of S protein from Netherlands (Fig. 8E). The free energy of binding was found to be $\Delta G = -8.71$ kcal/mol and the higher binding energy contributed by major amino acid residues viz.Asp 428, Thr 430, Gly 431, Phe 515 and Leu 517. In addition, contribution to high binding energy was also contributed by van der Waal's interaction by the binding site residues (Fig. 8F). In contrast to these, remdesivir did not respond well to the mutated RBD of S protein of Bangladesh and displayed the outward movement of the ligand (Fig. 8G) and Arg 355 and Phe 464 contributed in conventional hydrogen bond formation with the ligand (Fig. 8H). The approximation of binding of ligand to the S proteins from molecular docking studies widen the further analysis in MD simulation for better understanding of dynamic properties of S protein and remdesivir interaction.

2.5. MD simulation and MMGBSA calculations

Molecular dynamics and simulation (MD) of remdesivir bound RBD complexes of China, USA, Netherlands and Bangladesh are displayed in Fig. 9. The root mean square deviation (RMSD) of 100 ns MD simulation trajectories displayed most stable conformation of remdesivir bound RBD of S protein from Netherlands (Fig. 9A, red) complex having 0.8 Å deviation. While, China bound complex with remdesivir displayed the rmsd displacement of 1.5 Å (Fig. 9A, black), USA bound complex 1.8 Å (Fig. 9A, green) and highest displacement observed in the Bangladesh bound complex 2 Å (Fig. 9A, cyan). The rmsd plots signify the stable conformation of Remdesivir bound complex with Netherlands, USA and China except with the Bangladesh S protein. Root mean square fluctuation of the amino acid residue position of 100 ns simulation trajectories of Remdesivir bound S proteins displayed in Fig. 9B. Least fluctuation of amino acid residues observed in the Netherland S protein (Fig. 9B, black). Where the maximum fluctuations observed at the residue positions 360 and 475, ranging between 3 and 4 Å. Whereas, China protein showed the fluctuation (Fig. 9B, cyan) at the residue positions 360, 475 and 510, USA (Fig. 9B, green) at 475–480 and the highest fluctuations observed in case of Bangladesh (Fig. 9B, red) in the residue positions 430, 455, 505 and 506. Radius of gyration is the indicator of size and compactness of the protein in the ligand bound state displayed in Fig. 9C. The Rg plot of Cα-backbone displayed Netherlands S protein (Fig. 9C, red) has least fluctuations in compactness with an average of 18.6 Å form the beginning to end of the 100 ns simulation. Whereas, lowering of Rg score in China (cyan), USA (Black) and Bangladesh (green) with Rg scores 18 Å, 18.1 Å and 17.3 Å signify less compactness of the structures with remdesivir bound state.

Formation of H-bonds are important aspect to show the protein and ligand interaction throughout the simulation of 100 ns. It was observed from Fig. 10A, remedesivir forms average 5 numbers of H-bonds with the RBD of S protein of Netherlands throughout the simulation. Whereas, with China (Fig. 10B), Bangladesh (Fig. 10C) and USA (Fig. 10D) were 4, 2 and 3, respectively. The pattern of H-bonds

Table 1
Maximum Likelihood fits of 24 different nucleotide substitution models.

Model	Parameters	BIC	AICc	*lnL*	(+*I*)	(+*G*)	*R*	*f*(A)	*f*(T)	*f*(C)	*f*(G)	*r*(AT)	*r*(AC)	*r*(AG)	*r*(TA)	*r*(TC)	*r*(TG)	*r*(CA)	*r*(CT)	*r*(CG)	*r*(GA)	*r*(GT)	*r*(GC)
TN93 + G	75	84,536.73	83,645.255	−41,747.622	n/a	0.05	3.01	0.299	0.321	0.184	0.196	0.038	0.022	0.078	0.035	0.205	0.023	0.035	0.359	0.023	0.120	0.038	0.022
HKY + G	74	84,542.39	83,662.778	-41,757.384	n/a	0.05	3.01	0.299	0.321	0.184	0.196	0.038	0.022	0.149	0.036	0.140	0.023	0.036	0.245	0.023	0.228	0.038	0.022
HKY	73	84,545.488	83,677.773	−41,765.881	n/a	n/a	3.02	0.299	0.321	0.184	0.196	0.038	0.022	0.150	0.035	0.140	0.023	0.035	0.245	0.023	0.228	0.038	0.022
TN93 + G + I	76	84,547.920	83,644.546	−41,746.267	0.50	0.05	3.00	0.299	0.321	0.184	0.196	0.038	0.022	0.078	0.036	0.206	0.023	0.036	0.360	0.023	0.119	0.038	0.022
T92	71	84,550.245	83,706.302	−41,782.146	n/a	n/a	3.00	0.310	0.310	0.190	0.190	0.037	0.023	0.145	0.037	0.145	0.023	0.037	0.236	0.023	0.236	0.037	0.023
T92 + G	72	84,553.243	83,697.414	−41,776.702	n/a	0.05	3.00	0.310	0.310	0.190	0.190	0.037	0.023	0.144	0.037	0.144	0.023	0.037	0.236	0.023	0.236	0.037	0.023
HKY + G + I	75	84,553.465	83,661.977	−41,755.983	0.50	0.05	3.00	0.299	0.321	0.184	0.196	0.038	0.022	0.149	0.036	0.140	0.023	0.036	0.244	0.023	0.227	0.038	0.022
T92 + I	72	84,559.075	83,703.246	−41,779.618	0.50	n/a	3.00	0.310	0.310	0.190	0.190	0.037	0.023	0.144	0.037	0.144	0.023	0.037	0.236	0.023	0.236	0.037	0.023
GTR	77	84,562.802	83,647.541	−41,746.765	n/a	n/a	2.20	0.299	0.321	0.184	0.196	0.036	0.014	0.079	0.033	0.183	0.060	0.023	0.320	0.017	0.120	0.098	0.016
T92 + G + I	73	84,564.084	83,696.369	−41,775.179	0.50	0.05	2.99	0.310	0.310	0.190	0.190	0.037	0.023	0.144	0.037	0.144	0.023	0.037	0.236	0.023	0.236	0.037	0.023
GTR + G	78	84,565.783	83,638.635	−41,741.312	n/a	0.05	2.20	0.299	0.321	0.184	0.196	0.036	0.014	0.079	0.033	0.183	0.060	0.023	0.320	0.017	0.120	0.098	0.016
GTR + I	78	84,571.696	83,644.549	−41,744.269	0.50	n/a	2.20	0.299	0.321	0.184	0.196	0.036	0.014	0.079	0.033	0.183	0.060	0.023	0.320	0.017	0.120	0.098	0.016
GTR + G + I	79	84,576.596	83,637.562	−41,739.775	0.50	0.05	2.20	0.299	0.321	0.184	0.196	0.036	0.014	0.078	0.033	0.183	0.060	0.023	0.321	0.017	0.119	0.099	0.016
HKY + I	74	84,586.110	83,706.508	−41,779.249	0.50	n/a	3.01	0.299	0.321	0.184	0.196	0.038	0.022	0.149	0.036	0.140	0.023	0.036	0.245	0.023	0.228	0.038	0.022
K2	70	86,320.787	85,488.731	−42,674.361	n/a	n/a	3.00	0.250	0.250	0.250	0.250	0.031	0.031	0.188	0.031	0.188	0.031	0.031	0.188	0.031	0.188	0.031	0.031
K2 + G	71	86,323.375	85,479.432	−42,668.711	n/a	0.05	3.00	0.250	0.250	0.250	0.250	0.031	0.031	0.188	0.031	0.188	0.031	0.031	0.188	0.031	0.188	0.031	0.031
K2 + I	71	86,329.594	85,485.651	−42,671.821	0.50	n/a	3.00	0.250	0.250	0.250	0.250	0.031	0.031	0.188	0.031	0.188	0.031	0.031	0.188	0.031	0.188	0.031	0.031
K2 + G + I	72	86,333.548	85,477.719	−42,666.855	0.50	0.05	3.00	0.250	0.250	0.250	0.250	0.031	0.031	0.188	0.031	0.188	0.031	0.031	0.188	0.031	0.188	0.031	0.031
JC	69	86,388.225	85,568.055	−42,715.023	n/a	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC + G	70	86,390.798	85,558.742	−42,709.366	n/a	0.05	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC + I	70	86,397.040	85,564.984	−42,712.487	0.50	n/a	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
JC + G + I	71	86,400.984	85,557.042	−42,707.516	0.50	0.05	0.50	0.250	0.250	0.250	0.250	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083	0.083
TN93	74	107,315.839	106,436.237	−53,144.114	n/a	n/a	2.97	0.299	0.321	0.184	0.196	0.039	0.022	0.069	0.036	0.213	0.024	0.036	0.373	0.024	0.105	0.039	0.022
TN93 + I	75	107,322.588	106,431.100	−53,140.545	0.50	n/a	3.01	0.299	0.321	0.184	0.196	0.038	0.022	0.079	0.036	0.205	0.023	0.036	0.359	0.023	0.120	0.038	0.022

Description: Models with the lowest BIC scores (Bayesian Information Criterion) are considered to describe the substitution pattern the best. For each model, AICc value (Akaike Information Criterion, corrected), Maximum Likelihood value (*lnL*), and the number of parameters (including branch lengths) are also presented. Non-uniformity of evolutionary rates among sites may be modeled by using a discrete Gamma distribution (+*G*) with 5 rate categories and by assuming that a certain fraction of sites are evolutionarily invariable (+*I*). Whenever applicable, estimates of gamma shape parameter and/or the estimated fraction of invariant sites are shown.Assumed or estimated values of transition/transversion bias (*R*) are shown for each model, as well. They are followed by nucleotide frequencies (*f*) and rates of base substitutions (*r*) for each nucleotide pair.Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r* values is made equal to 1 for each model. For estimating ML values, a tree topology was automatically computed. This analysis involved 36 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. There were a total of 29,904 positions in the final dataset.
Abbreviations: TR: General Time Reversible; HKY: Hasegawa-Kishino-Yano; TN93: Tamura-Nei; T92: Tamura 3-parameter; K2: Kimura 2-parameter; JC: Jukes-Cantor./div>.

Table 2
Test of the homogeneity of substitution patterns between sequences.

	MT416725.2	MT415321.1	MT320891.2	MT994286.1	MT847220.1	MT476385.1	MT630432.1	MT755886.1	MT240479.1	MW031800.1	NC_045512.2	MN975262.1	MT890462.1	MT635445.1	MT956918.1	MT956914.1	MT320538.2
MT416725.2																	
MT415321.1	0.10800 00000																
MT320891.2	1.00000 00000	0.09200 00000															
MT994286.1	1.00000 00000	0.33200 00000	1.00000 00000														
MT847220.1	0.25600 00000	0.00000 00000	0.26400 00000	0.22400 00000													
MT476385.1	1.00000 00000	0.08400 00000	1.00000 00000	1.00000 00000	0.11400 00000												
MT630432.1	1.00000 00000	0.14800 00000	1.00000 00000	1.00000 00000	0.12400 00000	1.00000 00000											
MT755886.1	0.25200 00000	0.01800 00000	0.04200 00000	0.07600 00000	0.17400 00000	0.16600 00000	0.34600 00000										
MT240479.1	1.00000 00000	0.39000 00000	0.31000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000									
MW031800.1	1.00000 00000	0.00400 00000	0.27800 00000	0.16600 00000	1.00000 00000	0.17000 00000	0.38000 00000	0.34600 00000	1.00000 00000								
NC_045512.2	0.06000 00000	0.04600 00000	0.02600 00000	1.00000 00000	0.00000 00000	0.12000 00000	0.04600 00000	0.00000 00000	0.00200 00000	0.00200 00000							
MN975262.1	0.30600 00000	1.00000 00000	0.36600 00000	1.00000 00000	0.02800 00000	1.00000 00000	0.40600 00000	0.05000 00000	0.29200 00000	0.05400 00000	1.00000 00000						
MT890462.1	0.12400 00000	0.08600 00000	1.00000 00000	1.00000 00000	0.00000 00000	0.11400 00000	0.00800 00000	0.00800 00000	1.00000 00000	0.00000 00000	1.00000 00000	1.00000 00000					
MT635445.1	0.07800 00000	1.00000 00000	0.10600 00000	0.30400 00000	0.00000 00000	0.05400 00000	0.17000 00000	0.03600 00000	0.36000 00000	0.00200 00000	0.04400 00000	1.00000 00000	0.09400 00000				
MT956918.1	0.22800 00000	1.00000 00000	0.08200 00000	0.24600 00000	0.00600 00000	0.14400 00000	0.38600 00000	0.18000 00000	1.00000 00000	0.04000 00000	0.01400 00000	1.00000 00000	0.05600 00000	1.00000 00000			
MT956914.1	1.00000 00000	1.00000 00000	0.26400 00000	0.35200 00000	0.11400 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.23000 00000	0.05200 00000	1.00000 00000	0.19800 00000	1.00000 00000	1.00000 00000		
MT320538.2	0.08000 00000	1.00000 00000	0.02600 00000	0.19000 00000	0.01400 00000	0.06000 00000	0.16200 00000	0.24800 00000	0.32200 00000	0.03200 00000	0.02200 00000	0.37200 00000	0.02400 00000	1.00000 00000	1.00000 00000	1.00000 00000	
MT470102.1	0.19600 00000	1.00000 00000	0.08000 00000	0.24600 00000	0.00400 00000	0.13800 00000	0.39800 00000	0.19200 00000	1.00000 00000	0.04400 00000	0.01400 00000	1.00000 00000	0.05200 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000
MW041156.1	1.00000 00000	1.00000 00000	0.31000 00000	1.00000 00000	0.22600 00000	1.00000 00000	1.00000 00000	0.38600 00000	1.00000 00000	0.28200 00000	0.08800 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000
MT873892.1	1.00000 00000	0.43400 00000	0.15600 00000	0.32400 00000	0.06000 00000	1.00000 00000	1.00000 00000	0.36200 00000	1.00000 00000	0.19600 00000	0.01400 00000	0.36200 00000	0.10800 00000	0.43000 00000	1.00000 00000	1.00000 00000	1.00000 00000
MT682732.1	0.08200 00000	0.03600 00000	1.00000 00000	1.00000 00000	0.00000 00000	0.21800 00000	0.01000 00000	0.00800 00000	1.00000 00000	0.00400 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.01400 00000	0.01200 00000	0.12600 00000	0.01600 00000
MT066156.1	0.22400 00000	0.41200 00000	0.42600 00000	1.00000 00000	0.03600 00000	1.00000 00000	0.29600 00000	0.02200 00000	0.18600 00000	0.02200 00000	0.42200 00000	1.00000 00000	1.00000 00000	0.41000 00000	0.18600 00000	0.31000 00000	0.14600 00000
MT582499.1	1.00000 00000	0.10400 00000	1.00000 00000	1.00000 00000	0.11400 00000	1.00000 00000	1.00000 00000	0.04800 00000	1.00000 00000	0.29400 00000	0.01400 00000	0.24600 00000	1.00000 00000	0.08800 00000	0.12000 00000	0.22400 00000	0.05000 00000
MT270102.1	0.28000 00000	1.00000 00000	0.31400 00000	1.00000 00000	0.00600 00000	0.31600 00000	0.31200 00000	0.01000 00000	0.36000 00000	0.00800 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.23400 00000	0.41000 00000	0.10200 00000
MT457401.1	0.11800 00000	0.02000 00000	0.01400 00000	0.03000 00000	0.13000 00000	0.08400 00000	0.14200 00000	1.00000 00000	0.28400 00000	1.00000 00000	0.00000 00000	0.02400 00000	0.01000 00000	0.04400 00000	0.08200 00000	0.19800 00000	0.07600 00000
MT457393.1	1.00000 00000	0.28200 00000	1.00000 00000	1.00000 00000	0.25800 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.04600 00000	0.40200 00000	0.33800 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.28800 00000
MT940481.1	0.06800 00000	0.00000 00000	0.00400 00000	0.00800 00000	0.07800 00000	0.02000 00000	0.05000 00000	1.00000 00000	0.13400 00000	0.32600 00000	0.00000 00000	0.00200 00000	0.00000 00000	0.00000 00000	0.00800 00000	0.08600 00000	0.00600 00000
MN985325.1	0.27400 00000	1.00000 00000	0.25800 00000	1.00000 00000	0.01600 00000	0.40200 00000	0.33800 00000	0.02800 00000	0.19200 00000	0.04400 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.28200 00000	0.38000 00000	0.25000 00000
MT810761.1	0.08600 00000	1.00000 00000	0.05600 00000	0.21200 00000	0.00200 00000	0.05800 00000	0.10200 00000	0.21000 00000	0.33600 00000	0.01800 00000	0.04800 00000	1.00000 00000	0.02000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000
MT810789.1	0.16600 00000	0.01200 00000	0.02200 00000	0.06200 00000	0.16400 00000	0.08000 00000	0.04600 00000	1.00000 00000	0.28800 00000	0.39000 00000	0.00200 00000	0.04400 00000	0.00000 00000	0.01800 00000	0.09400 00000	0.26800 00000	0.10600 00000
MT126808.1	1.00000 00000	0.32400 00000	1.00000 00000	1.00000 00000	0.07600 00000	1.00000 00000	1.00000 00000	0.03200 00000	0.23000 00000	0.05400 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.31000 00000	0.16200 00000	0.30200 00000	0.13000 00000
MT827074.1	0.08400 00000	1.00000 00000	0.10200 00000	0.33600 00000	0.00200 00000	0.11000 00000	0.19400 00000	0.02400 00000	0.37200 00000	0.00800 00000	0.03400 00000	1.00000 00000	0.07600 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000
MT470219.1	0.32800 00000	0.14600 00000	0.09800 00000	0.23200 00000	0.13200 00000	0.38600 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.17200 00000	0.00400 00000	0.19600 00000	0.04800 00000	0.12600 00000	1.00000 00000	1.00000 00000	1.00000 00000
MT256924.2	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.16200 00000	1.00000 00000	1.00000 00000	0.18600 00000	1.00000 00000	0.21800 00000	0.28200 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000
MT263074.1	0.12000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.00200 00000	0.18200 00000	0.06400 00000	0.02800 00000	1.00000 00000	0.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.25800 00000	0.36600 00000	0.12800 00000
MW023935.1	1.00000 00000	0.02400 00000	0.24400 00000	0.12200 00000	0.38400 00000	0.20600 00000	0.18400 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.00800 00000	0.07600 00000	0.00400 00000	0.03800 00000	0.09600 00000	0.39600 00000	0.04000 00000

MT470102.1	MW041156.1	MT873892.1	MT682732.1	MT066156.1	MT582499.1	MT270102.1	MT457401.1	MT457393.1	MT940481.1	MN985325.1	MT810761.1	MT810789.1	MT126808.1	MT827074.1	MT470219.1	MT256924.2	MT263074.1	MW023935.1
1.00000 00000																		
1.00000 00000	1.00000 00000																	
0.01000 00000	1.00000 00000	0.03200 00000																
0.16400 00000	1.00000 00000	0.16600 00000	1.00000 00000															
0.11800 00000	0.23200 00000	0.11800 00000	1.00000 00000	0.08200 00000														
0.22400 00000	1.00000 00000	0.22200 00000	1.00000 00000	1.00000 00000	0.33000 00000													
0.08800 00000	0.12800 00000	0.18800 00000	0.00600 00000	0.00400 00000	0.04200 00000	0.00600 00000												
1.00000 00000	1.00000 00000	1.00000 00000	0.15200 00000	0.24800 00000	1.00000 00000	0.29000 00000	0.17800 00000											
0.00800 00000	0.06000 00000	0.00000 00000	0.00000 00000	0.00200 00000	0.00200 00000	0.00000 00000	1.00000 00000	0.15200 00000										
0.30400 00000	1.00000 00000	0.26000 00000	1.00000 00000	1.00000 00000	0.18000 00000	1.00000 00000	0.00600 00000	0.27200 00000	0.00000 00000									
1.00000 00000	1.00000 00000	1.00000 00000	0.00000 00000	0.27800 00000	0.05600 00000	0.08400 00000	0.23000 00000	0.31200 00000	0.02200 00000	1.00000 00000								
0.09200 00000	0.37600 00000	0.26600 00000	0.00000 00000	0.02200 00000	0.03400 00000	0.00000 00000	1.00000 00000	0.13200 00000	1.00000 00000	0.02800 00000	0.00200 00000							
0.16600 00000	1.00000 00000	0.22400 00000	1.00000 00000	1.00000 00000	0.31000 00000	1.00000 00000	0.01000 00000	0.29200 00000	0.00000 00000	1.00000 00000	0.18200 00000	0.04400 00000						
1.00000 00000	1.00000 00000	0.43800 00000	0.03200 00000	0.43000 00000	0.09000 00000	1.00000 00000	0.01200 00000	0.27600 00000	0.00000 00000	1.00000 00000	1.00000 00000	0.00400 00000	0.33000 00000					
1.00000 00000	1.00000 00000	1.00000 00000	0.03000 00000	0.13200 00000	0.08000 00000	0.07000 00000	0.23600 00000	1.00000 00000	0.01200 00000	0.14000 00000	1.00000 00000	1.00000 00000	0.17000 00000	0.12800 00000				
1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.15200 00000	1.00000 00000	0.02400 00000	1.00000 00000	1.00000 00000	0.23200 00000	1.00000 00000	1.00000 00000	1.00000 00000			
0.23000 00000	1.00000 00000	0.25000 00000	1.00000 00000	1.00000 00000	1.00000 00000	1.00000 00000	0.01000 00000	0.31200 00000	0.00000 00000	1.00000 00000	0.14600 00000	0.00600 00000	1.00000 00000	1.00000 00000	0.14000 00000	1.00000 00000		
0.10000 00000	0.37400 00000	0.34000 00000	0.00600 00000	0.04800 00000	0.40200 00000	0.04000 00000	1.00000 00000	1.00000 00000	0.32600 00000	0.07000 00000	0.04200 00000	1.00000 00000	0.13000 00000	0.04000 00000	0.31000 00000	0.29000 00000	0.00600 00000	

Description: The probability of rejecting the null hypothesis that sequences have evolved with the same pattern of substitution, as judged from the extent of differences in base composition biases between sequences (Disparity Index test. A Monte Carlo test (500 replicates) was used to estimate the P-values, which are shown below the diagonal. P-values smaller than 0.05 are considered significant (marked with yellow highlights). The estimates of the disparity index per site are shown for each sequence pair above the diagonal. This analysis involved 36 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Noncoding. There were a total of 29,904 positions in the final dataset.

formation signify the stable binding of remdesivir with Netherlands as compared to China, USA and Bangladesh.

The free energies of binding of remdesivir with the RBDs of S proteins from China, USA, Netherlands and Bangladesh were calculated in prime MMGBSA (Supplementary Table 2). Interestingly the remdesivir drug displayed highest binding (dG) with Netherland's RBD of S protein −54.17 kcal/mol followed by China −17.24 kcal/mol, USA −6.42 kcal/mol. By contrast remdesivir became unstable during binding and comes out from the binding pocket of RBD of S protein of Bangladesh. Therefore, the binding energy in MMGBSA displayed positive interaction energy (dG = 21.31 kcal/mol) which is not feasible for ligand binding (Supplementary Table 2). All other parameters are listed in the table which gave additional contribution to the binding energies of target and the drug ligand molecule and conformational stability.

Post dynamics structural analysis also revealed the outcomes of MMGBSA calculation. It was observed that there are series of

Table 3
Maximum likelihood estimate of substitution matrix.

Nucleotide bases	A	T/U	C	G
A	–	*3.83*	*2.19*	**7.47**
T/U	*3.56*	–	**20.85**	*2.34*
C	*3.56*	**36.48**	–	*2.34*
G	**11.38**	*3.83*	*2.19*	–

Description: Each entry is the probability of substitution (*r*) from one base (row) to another base (column). Substitution pattern and rates were estimated under the Tamura-Nei (1993) model. Rates of different transitional substitutions are shown in **bold** and those of transversionsal substitutions are shown in *italics*. Relative values of instantaneous *r* should be considered when evaluating them. For simplicity, sum of *r*values is made equal to 100, The nucleotide frequencies are A = 29.90%, T/U = 32.12%, C = 18.36%, and G = 19.62%. For estimating ML values, a tree topology was automatically computed. The maximum Log likelihood for this computation was −41,753.003. This analysis involved 36 nucleotide sequences. Codon positions included were 1st + 2nd + 3rd + Non-coding. There were a total of 29,904 positions in the final dataset.

modifications of ligand accommodation pocket, alteration of side chain secondary structures. The RBD of S protein of China displayed positional alteration of the binding cavity loop (L1) (Fig. 11A and B). While the H1 and H2 helices were absent after 100 ns of simulation (Fig. 11B). In addition, S1 and S2 sheets conformed into loops L4 and L5 respectively (Fig. 11A and B). Moreover, the drug molecule oriented itself into more favorable position for higher binding (arrow). In case of Netherlands, the drug molecule moved from outward to deeper inside the cavity (Arrow, Fig. 11C and D). This structural variation corroborated the significant outcome of MMGBSA as presented in Table 1. Moreover, alteration of L1 loops, helix H1 to loop L2 and change in shape to S1 were also observed (Fig. 11C and D). Observations also revealed that USA RBD Remdesivir bound complex displayed significant changes in the domains such as deviation in the L1 loop structure significantly changes the binding cavity (Fig. 11E and F). Sheets S1 and S2 relaxed into loops L2 and L3 after 100 ns of simulation (Fig. 11E and F). The L1 loop modification made a saddle shape cavity gave better accommodation to the drug molecule (Fig. 11F). On the other hand, remdesivir with RBD of S protein of Bangladesh showed significant structural variations where shape of L1 was altered and L2 became S2 (Fig. 11G and H). These structural variations led to shape of unfavorable pocket for ligand accommodation and thereby the drug molecule moved out from the binding cavity (arrow, Fig. 11H). Thus, the result corroborated the MMGBSA analysis of least binding in case of Bangladesh derived S protein.

3. Discussion

The recent emergence of COVID-19 has become a severe humanitarian disaster. The pathogenic SARS-CoV-2 has been spreading rapidly across the globe since the first days and has become a huge threat for almost all the countries around the world (Chakraborty and Maity, 2020). Even some countries are experiencing spike in COVID-19 infection rate for several times, which is the most alarming. Although more than 5.46 billion doses of COVID-19 vaccines have been administered worldwide, several mutations in SARS-CoV-2 variants have become a great barrier to combat this pandemic. This present study mainly focused on phylogenetic and molecular evolutionary analysis of SARS-CoV-2 among different Asian, European and American countries, with investigating the variations among spike (S) proteins to pursue the novel vaccine or drug candidate. The destructive SARS-CoV-2 uses receptor binding domain (RBD) of S protein to bind with human ACE2 receptor and then priming and cleavage at S1/S2 site by the protease enzymes for viral and host cell membrane fusion (Walls et al., 2020; Tortorici and Veesler, 2019; Wang et al., 2020). As a consequence, virus enters into the host cell and causes pathogenesis. Therefore, along with evolutionary details, here we studied some mutations in RBD regions of S proteins of SARS-CoV-2 from the most prevalent countries to understand the ultimate effects of these mutations.

Our multiple sequence alignment and phylogenetic analysis results revealed that retrieved SARS-CoV-2 genomes were greatly conserved and highly similar to reference genome of Wuhan NC_045512.2 that is supported by several recent reports (Sheikh et al., 2020; Kumar et al., 2020). Results also indicated mutations in all the genomes of SARS-CoV-2. A study confirmed that, the original epicenter of SARS-CoV-2 outbreak was Chinese province (Li et al., 2020b). However, human SARS coronavirus was clearly derivative of viruses from wild animals of southern China wet markets which is confirmed from multiple lineages of SARS. Some results also indicated that the probable hosts for SARS-coronaviruses were bats (Li et al., 2020b; Vijaykrishna et al., 2007). The phylogenetic tree constructed by ML method with highest log likelihood (−41,753.00) (Fig. 1) showed that SARS-CoV-2 genomes were highly conserved and also supported that transmission of COVID-19 initiated from Wuhan of China (Zhou et al., 2020). The bootstrap consensus tree constructed by ME method recapitulated the evolutionary history (Fig. 2).

Molecular evolutionary statistics is essential for evaluating and reconstructing the models of evolutionary antiquities between SARS-CoV-2 genomes in order to comparative analysis of SARS-CoV-2 genomic sequences (Kumar et al., 2020; Benvenuto et al., 2020). Evaluation of the best ML DNA/protein models for nucleotide substitutions revealed that the TN93 + G model had the lowest BIC scores among the 24 models, and thus it can be considered as the best nucleotide substitution pattern model (Table 1). The Tamura-Nei model (TN93 + G) suggested the differences that included both transitions as well as transversions. It also reveals the type of transitions (i.e. purine-purine and pyrimidine-pyrimidine) and signifies that transversional substitution rates are equal (Kumar, 1996). Mutations that were observed in multiple sequence alignment also supported these results. Our TN93 + G model indicates that during transmission among people of selected countries, transitions and transversions occurred at multiple degrees in the SARS-CoV-2 genomes. However, a recent study reported that among 24 and 88 competitive models, HKY and GTR + G model is the best nucleotide substitution pattern model (Li et al., 2020b).

The test for homogeneity of substitution patterns among the genome sequences indicated that the sequences evolved with the same pattern of nucleotide substitution (probability of rejecting the null hypothesis), as considered from the extent of distinctions in base composition biases between genomes (Table 2). It also provided the noticeable evolutionary link in genomes between the phylogenetic precisions. It has been clarified that homologous recombination can happen in the receptor binding domain of S-protein of SARS-CoV-2 that may govern the transmission of the disease in various species (Ji et al., 2020). Our results also revealed that there was a minor discrepancy among the genomes that was showed by the MSA, phylogenetic tree as well as evolutionary analysis of ME method. However, new divergences among the genomes were observed due to the adaptation of virus into the human, and mutations emerging from human to human transfer aided to comprehend the COVID-19 transmission dynamics (Sheikh et al., 2020). Besides, the ML substitution matrix was calculated to investigate molecular evolution of SARS-CoV-2 genomes (Table 3). Results signified that there were both transitional and transversional substitutions, where transitional substitution rate was greater. From the result, it can be assured that SARS-CoV-2 of selected regions experienced differences in the rate of evolution as well as in transmission. Estimation of Transition/Transversion bias (R) (3.00) represented the DNA evolution pattern and sequence distances among the genomes. This calculation predicted that COVID-19 evolution occurred from Wuhan and declared the presence of mutations among the genomes (Benvenuto et al., 2020). However, SARS-CoV-2 has maximum genomic similarity with Bat coronavirus (BCoV) that is determined by examination of sequences and assessment in combination with relative synonymous codon usage (RSCU) bias between numerous animal species, and also show the analogous codon usage bias with snake (Ji et al.,

Table 4
Codon-based test of neutrality for analysis between sequences.

	MT416725.2	MT415321.1	MT320891.2	MT994286.1	MT847220.1	MT476385.1	MT630432.1	MT755886.1	MT240479.1	MW031800.1	NC_045512.2	MN975262.1	MT890462.1	MT635445.1	MT956918.1	MT956914.1	MT320538.2
MT416725.2																	
MT415321.1	0.08789																
MT3208 91.2	0.83240	0.52148															
MT9942 86.1	0.45580	0.99422	0.81391														
MT8472 20.1	0.00198	0.03413	0.62821	0.31824													
MT4763 85.1	0.01574	0.22347	0.94272	0.62836	0.01574												
MT6304 32.1	0.68280	0.85335	0.52650	0.91459	0.39410	0.92621											
MT7558 86.1	0.29680	0.92569	0.54999	0.95822	0.16586	0.50320	0.84128										
MT2404 79.1	0.61781	0.18123	0.32616	0.99408	0.81345	0.44086	0.21674	0.22187									
MW031 800.1	0.89575	0.57113	0.39280	0.71148	0.81500	0.59992	0.24652	0.60221	0.16223								
NC_045 512.2	0.29669	0.92579	0.46471	0.95578	0.16578	0.50307	0.84142	0.89506	0.24652	0.60234							
MN9752 62.1	0.06326	0.29108	0.81373	0.36790	0.03203	0.12146	0.53226	0.45889	0.75139	0.83294	0.02718						
MT8904 62.1	0.04774	0.54183	0.71337	0.83225	0.01574	0.15987	0.50274	0.79284	0.29305	0.34343	0.79270	0.22319					
MT6354 45.1	0.08789	1.00000	0.52154	0.99431	0.03405	0.22313	0.85345	0.92631	0.18128	0.57174	0.92641	0.29066	0.54133				
MT9569 18.1	0.05477	0.31931	0.63337	0.88700	0.02122	0.14075	0.96401	0.86862	0.23652	0.70318	0.86850	0.21149	0.35101	0.31931			
MT9569 14.1	0.02116	0.08581	0.87307	0.66594	0.00821	0.05484	0.63663	0.52471	0.37606	0.97892	0.52459	0.10718	0.14065	0.08581	0.15987		
MT3205 38.2	0.03406	0.15987	0.75158	0.77301	0.01318	0.08794	0.79187	0.68294	0.30154	0.84109	0.68281	0.15139	0.22324	0.15987	0.08581	0.02718	
MT4701 02.1	0.05476	0.31931	0.63349	0.88683	0.02121	0.14072	0.96389	0.86852	0.23656	0.70329	0.86840	0.21142	0.35094	0.31931	0.15987	0.04774	0.31931
MW041 156.1	0.94299	0.41834	0.13543	0.55016	0.72678	0.82736	0.44113	0.45913	0.07613	0.32720	0.34369	0.95565	0.60314	0.41842	0.52143	0.75177	0.63348
MT8738 92.1	0.05477	0.31931	0.63344	0.88691	0.02118	0.14055	0.96392	0.86800	0.23654	0.70382	0.86788	0.21122	0.35064	0.31931	0.15987	0.04774	0.08581
MT6827 32.1	0.68268	0.85330	0.52653	0.91463	0.39415	0.92616	0.34367	0.84121	0.21670	0.24651	0.84136	0.53234	0.50273	0.85345	0.96407	0.63669	0.79193
MT0661 56.1	0.63926	0.60067	0.24661	0.75161	0.43763	0.88550	0.60323	0.63340	0.13541	0.44113	0.50280	0.52391	0.84155	0.60072	0.74629	0.95538	0.89615
MT5824 99.1	0.83268	0.52134	0.30153	0.64700	0.62848	0.94247	0.52635	0.54981	0.17493	0.39265	0.46462	0.81395	0.71317	0.52137	0.63321	0.87285	0.75138
MT2701 02.1	0.84119	0.15982	0.52144	0.99415	0.88569	0.57160	0.24411	0.24651	0.18125	0.17965	0.92626	0.29081	0.33952	0.15982	0.24138	0.46476	0.34368
MT4574 01.1	0.72658	0.75123	0.43002	0.73676	0.45596	0.83272	0.61826	0.45875	0.23111	0.54220	0.74859	0.73939	0.94244	0.95546	0.87238	0.88693	0.99414
MT4573 93.1	0.63912	0.60044	0.40825	0.74362	0.43804	0.88595	0.60332	0.63307	0.16574	0.44084	0.63320	0.77326	0.84118	0.86755	0.74599	0.95578	0.89593
MT9404 81.1	0.22304	0.86783	0.64718	0.93758	0.12127	0.38922	0.97949	0.95599	0.27652	0.71323	0.95583	0.36794	0.63692	0.86782	0.68268	0.39387	0.52411
MN9853 25.1	0.12129	0.52442	0.89595	0.56319	0.06337	0.22320	0.75771	0.68272	0.52128	0.94222	0.08581	0.04774	0.38943	0.52406	0.39435	0.21129	0.29101
MT8107 61.1	0.05484	0.31931	0.41803	0.87208	0.02120	0.14066	0.22333	0.86834	0.23640	0.70347	0.92596	0.29109	0.35084	0.31931	0.15987	0.15987	0.08581
MT8107 89.1	0.22341	0.86831	0.45872	0.85234	0.12152	0.38972	0.50282	0.95662	0.27621	0.71261	0.89533	0.45888	0.63749	0.86831	0.68319	0.68291	0.52460
MT1268 08.1	0.94314	0.41820	0.13541	0.55006	0.72714	0.82706	0.44105	0.45894	0.07612	0.32703	0.34363	0.95591	0.60291	0.41836	0.52125	0.75151	0.63326
MT8270 74.1	0.08787	1.00000	0.52153	0.99431	0.03404	0.22311	0.85335	0.92636	0.18128	0.57178	0.92646	0.29062	0.54129	1.00000	0.31931	0.08581	0.15987
MT4702 19.1	0.03406	0.15987	0.75158	0.77301	0.01319	0.08803	0.79187	0.68335	0.30154	0.84061	0.68323	0.15148	0.22342	0.15987	0.08581	0.02718	0.04774
MT2569 24.2	0.35610	0.95607	0.63317	0.88733	0.22732	0.53281	0.94248	0.99370	0.37581	0.71389	0.86801	0.68271	0.75801	0.95591	0.81427	0.56316	0.68238
MT2630 74.1	0.68268	0.85345	0.52653	0.91463	0.39400	0.92630	0.34367	0.84142	0.21676	0.24656	0.84156	0.53210	0.50277	0.85345	0.96389	0.63650	0.79174
MW023 935.1	0.89576	0.57174	0.32726	0.62517	0.81397	0.60049	0.46467	0.60303	0.16231	0.15987	0.60316	0.83203	0.34366	0.57174	0.70394	0.97994	0.84152

2020; Ceraolo and Giorgi, 2020).

There are two main methods of natural selection: purifying selection and Darwinian or positive selection, where the former functions to remove the deleterious mutations from the sequences and the later assist to develop the genotypes conferring capability on the sequences (Hughes, 1999). The Darwinian selection might maintain a polymorphism. Therefore, the number of synonymous nucleotide substitutions per site (dS) outstrips the total number of non synonymous nucleotide substitutions per site (dN) in several protein-coding genes, that proves the eradication of the vital segment of non-synonymous mutations by purifying selection (Nei and Gojobori, 1986; Lai et al., 2020). Commonly, for adaptive evolution, a significant surplus of dN over the dS substitution is used as a sign (Benvenuto et al., 2020; Ji et al., 2020; Yang and Bielawski, 2000). To understand the emergence and spread of COVID-19, analysis of the rate of amino acid-changing [non-synonymous (dN)] and silent [synonymous (dS)] nucleotide changes or

MT470102.1	MW041156.1	MT873892.1	MT682732.1	MT066156.1	MT582499.1	MT270102.1	MT457401.1	MT457393.1	MT940481.1	MN985325.1	MT810761.1	MT810789.1	MT126808.1	MT827074.1	MT470219.1	MT256924.2	MT263074.1
0.52152																	
0.15987	0.52143																
0.96395	0.44114	0.96401															
0.74640	0.18018	0.74630	0.60329														
0.63332	0.23657	0.63326	0.52635	0.32617													
0.24139	0.41830	0.24139	0.24412	0.60058	0.52130												
0.87253	0.43001	0.87287	0.71448	0.48444	0.65289	0.22205											
0.74610	0.33886	0.74632	0.60326	0.45950	0.64726	0.18130	0.52143										
0.68257	0.55012	0.92632	0.97960	0.75156	0.64690	0.32616	0.74910	0.75166									
0.39425	0.74622	0.39413	0.75780	0.86772	0.89577	0.52425	0.93774	0.99404	0.56324								
0.15987	0.41806	0.15987	0.96438	0.60031	0.52094	0.24133	0.81425	0.68292	0.68315	0.52461							
0.68308	0.45877	0.68313	0.97895	0.63293	0.54931	0.32593	0.88750	0.81446	0.81475	0.68294	0.92595						
0.52134	0.09847	0.52135	0.44099	0.71121	0.23653	0.41824	0.42977	0.33869	0.55001	0.74602	0.41800	0.45868					
0.31931	0.41846	0.31931	0.85351	0.60077	0.52134	0.15982	0.63375	0.60083	0.86783	0.52401	0.31931	0.86831	0.41840				
0.08581	0.63348	0.31931	0.79193	0.89615	0.75138	0.34368	0.99397	0.89579	0.86778	0.29101	0.08581	0.52460	0.63326	0.15987			
0.81412	0.52132	0.81419	0.94248	0.74611	0.63307	0.95614	0.93761	0.88701	0.88745	0.92613	0.95661	0.99310	0.52123	0.95607	0.68238		
0.96377	0.44124	0.96389	0.34370	0.60335	0.52635	0.24412	0.61840	0.60344	0.97960	0.75755	0.96438	0.97895	0.44115	0.85350	0.79174	0.94248	
0.70405	0.32731	0.70400	0.24656	0.44120	0.39267	0.17969	0.71460	0.71353	0.71326	0.94278	0.70353	0.71268	0.32723	0.57174	0.84152	0.71408	0.24656

Description: The probability of rejecting the null hypothesis of strict-neutrality (dN = dS) (below diagonal) is shown. Values of P less than 0.05 are considered significant at the 5% level and are highlighted. The test statistic (dN - dS) is shown above the diagonal. dS and dN are the numbers of synonymous and non-synonymous substitutions per site, respectively. The variance of the difference was computed using the analytical method. Analyses were conducted using the Nei-Gojobori method . This analysis involved 36 nucleotide sequences. All ambiguous positions were removed for each sequence pair (pairwise deletion option). There were a total of 9215 positions in the final dataset.

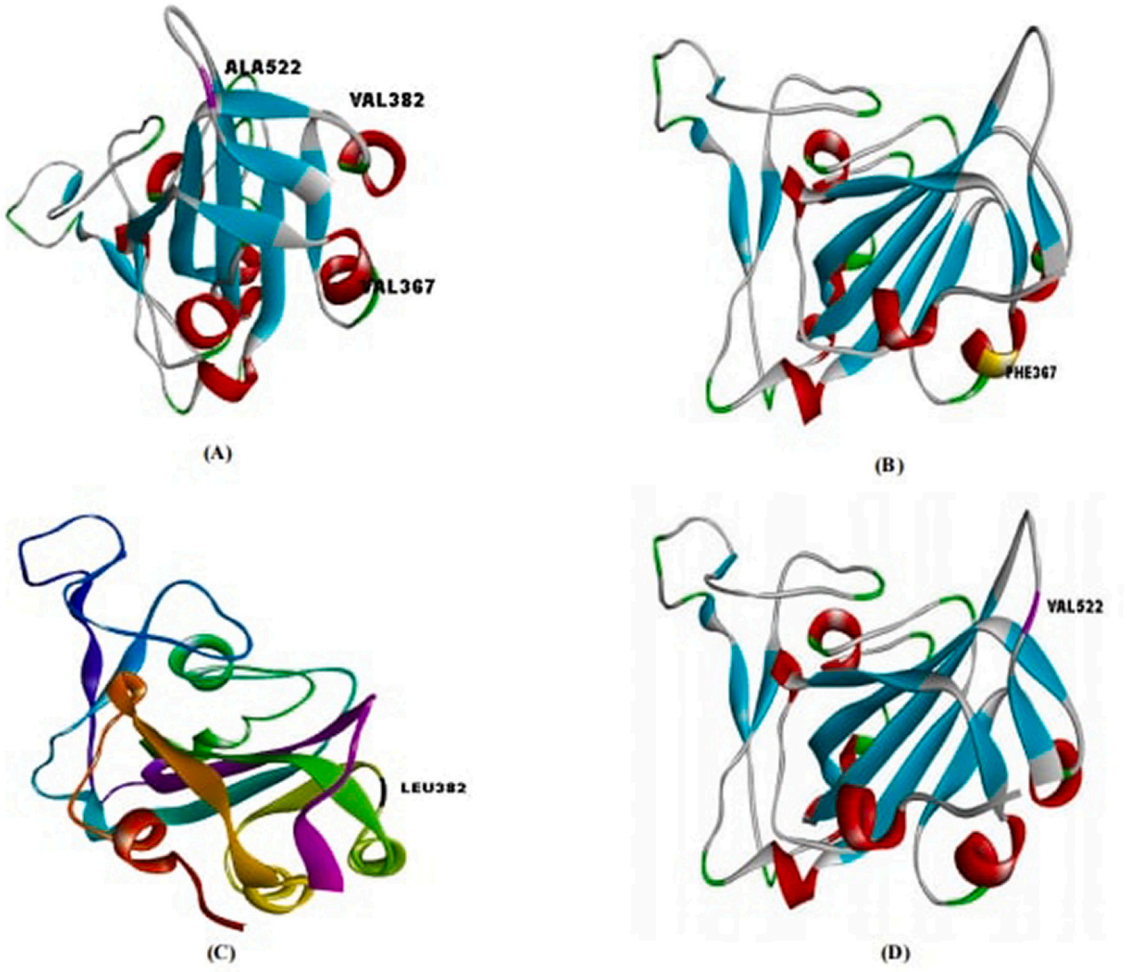

Fig. 3. Comparison of mutated and non-mutated models of RBD. (**A**) Non-mutated regions in the representative model of China (NC_045512.2: YP_009724390.1); (**B**) Val/ Phe367 mutation in the model of Netherland (MT457401.1: QJS39627.1); (**C**) Val/ Leu 382 mutation in the model ofBangladesh (MT847220.1:QMU94884.1); (**D**) Ala/ Val522 mutation in the model of the USA (MT940481.1: QNL36022.1).

substitutions in protein-coding sequences of SARS-CoV-2 can provide the substantial evidence. Particularly, the proportions of dN and dS fixation are used to limit the level of discerning pressure of proteins. For this reason, codon-based test of neutrality was conducted using the SARS-CoV-2 sequences (Table 4). 14fold proliferation was seen in the dN/dS proportion from the lineage leading to SARS-CoV-2, which implied that the majority of dN mutations currently segregating into humans have unfavorable effect on viral suitability (Wang et al., 2021). Our results ensured that SARS-CoV-2 experienced subsequent effective transition from the co-exiting genome of SARS and also confirmed that considerable positive selection happened in the viral genomes of the selected countries at the time of transmission. Tajima's Neutrality Test was performed to identify the selection and nucleotide polymorphisms in the genomes of SARS-CoV-2 (Supplementary Table 1). The result represented that there were 108 segregating sites, 0.000335 nucleotide diversity, and − 2.293243 was the value of Tajima test statistic. This allele frequencies based result clarified that paired selections with unexpected inhabitant's contraction occurred during the transmission of COVID-19 in Asian, European and American countries. Thus, growing association between maximum likelihood tree and genetic distances among the genomic sequences was signified, which marked evolutionary proportions as well as evolutionary models. However, according to a study, <0.001% substitutions per site was the genetic distance between different SARS-CoV-2 genomes.

We found 3 mutations within the RBD region in the spike (S) protein among the 36 SARS-CoV-2 viruses. The homology modeling was conducted by using SARS-CoV-2 spike receptor-binding domain bound with ACE2 (PDB ID: 6M0J) as template in order to explore the effects of these mutations in binding with remdesivir. The model of reference sequence (China; NC_045512.2: YP_009724390.1) was considered as the representative of all S proteins which don't have mutations in RBD regions. The 3 mutations, Val/ Phe367, Val/ Leu 382 and Ala/ Val522, were discovered in Netherland (MT457401.1: QJS39627.1), Bangladesh (MT847220.1:QMU94884.1) and the USA (MT940481.1: QNL36022.1), respectively (Fig. 3). Before molecular docking, it's crucial to assess the quality of the homology models. For this purpose, ramachandran plot, local and global quality, and overall quality factor of the models were estimated. All the models showed more than 90% residues in ramachandran favored regions indicating the satisfactory quality of the models (Figs. 4, 5, 6, 7). Besides, Qmean Z scores were − 1.66, −1.69, −1.66 and − 1.67 for the models of YP_009724390.1 (China), QJS39627.1 (Netherland), QMU94884.1 (Bangladesh) and QNL36022.1 (USA), respectively, while *Z*-scores around zero and more than −4.0 signifies the models as of good quality (Benkert et al., 2009) (Figs. 4, 5, 6, 7). Likewise, Overall quality factors of these models were more than 90 each which indicated the superior quality of these models (Figs. 4, 5, 6, 7).

Molecular docking study revealed that mutations in the RBD regions of S proteins affected the binding interaction of remdesivir and RBD regions of S proteins. The reference model of China had the good binding

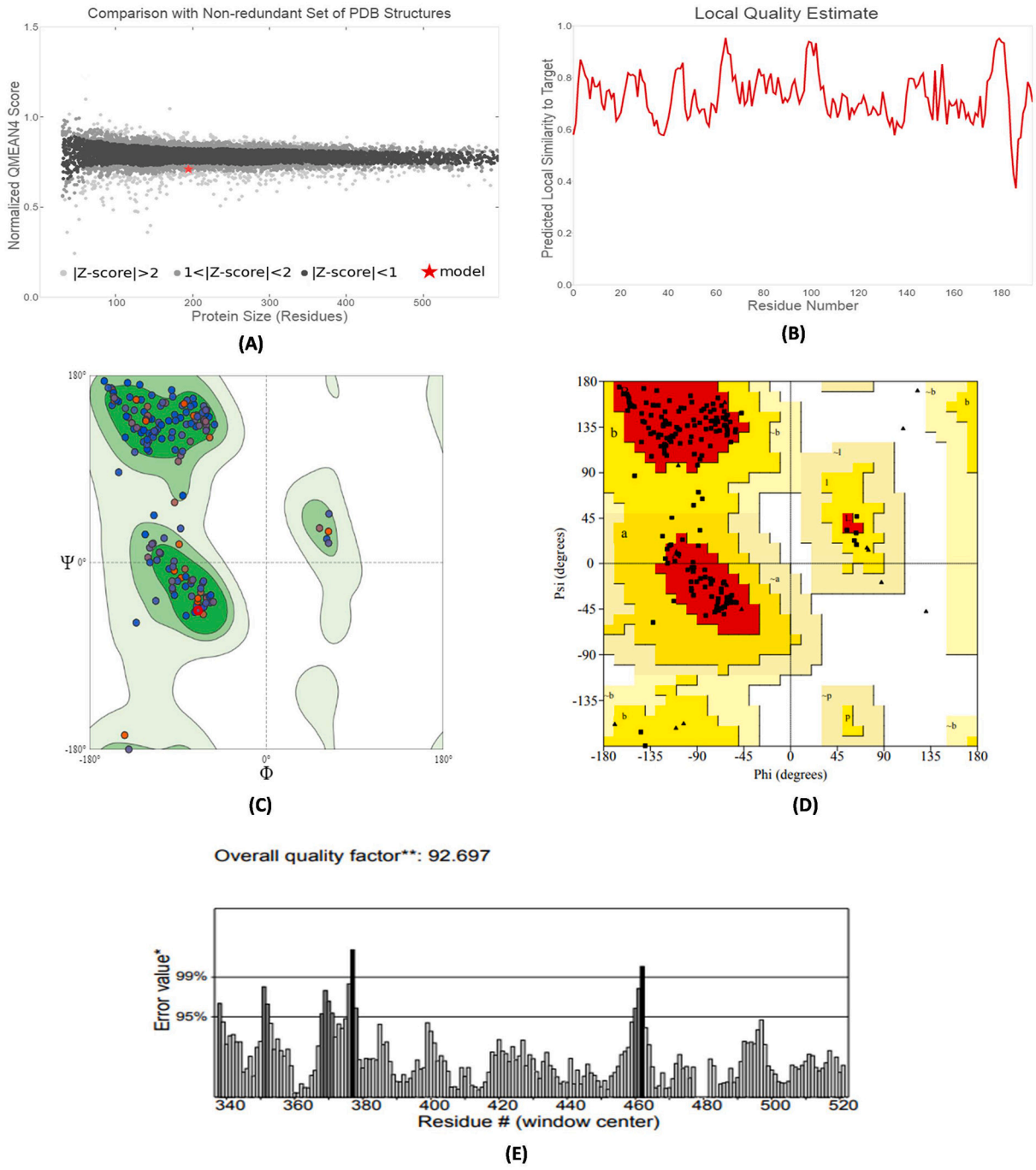

Fig. 4. Quality assessment of the RBD model of China (NC_045512.2: YP_009724390.1). (**A**)Qmean score, (**B**) local quality estimate, (**C**) molprobity ramachandran plot, (**D**) ramachandran plot analysed via procheck, and (**E**) overall quality factor.

energy while model of Netherland displayed considerably high binding as compared to China, USA and Bangladesh. In a previously studied report investigated considerable binding of remdesivir with S protein with different strategies used in molecular docking (Eweas et al., 2021). Although we have reported lesser binding energies as compared to the previous report (Eweas et al., 2021) but we drew a conclusion of binding pattern with the mutants where remdesivir highly effective against S protein of Netherland, which portrayed the novelty of our study. MD simulation studies of the reference and mutants conjugated with remdesivir deciphered the relative motion and properties of proteins. Except RBD of S protein of Bangladesh, all the mutants viz. and reference China S protein displayed much stable conformation complexed with remdesivir. MMGBSA outcomes exhibited the more realistic behavior of the remdesivir, whereas RBD of S protein of Netherland exhibited a high binding conformation of the remdesivir perhaps the first ever report. But in contrast, model of Bangladesh became incompetent in holding the drug for 100 ns, whereas China and USA proteins were able to make a moderate to low binding, respectively. Therefore, it could be suggested that apart from China and Netherlands, remdesivir is not suitable for targeting S proteins of the USA and Bangladesh. The results also suggested that mutations are causing the drugs to bind less effectively as like the mutations of Val/ Leu 382 (Bangladesh) and Ala/ Val522 (the USA). Finally, it can be said that conformational changes in the RBD regions occurred due to the mutations that resulted in the changing of binding sites, and thus binding affinity of drugs with these mutated regions are being changed extensively. Therefore, on designing drugs or

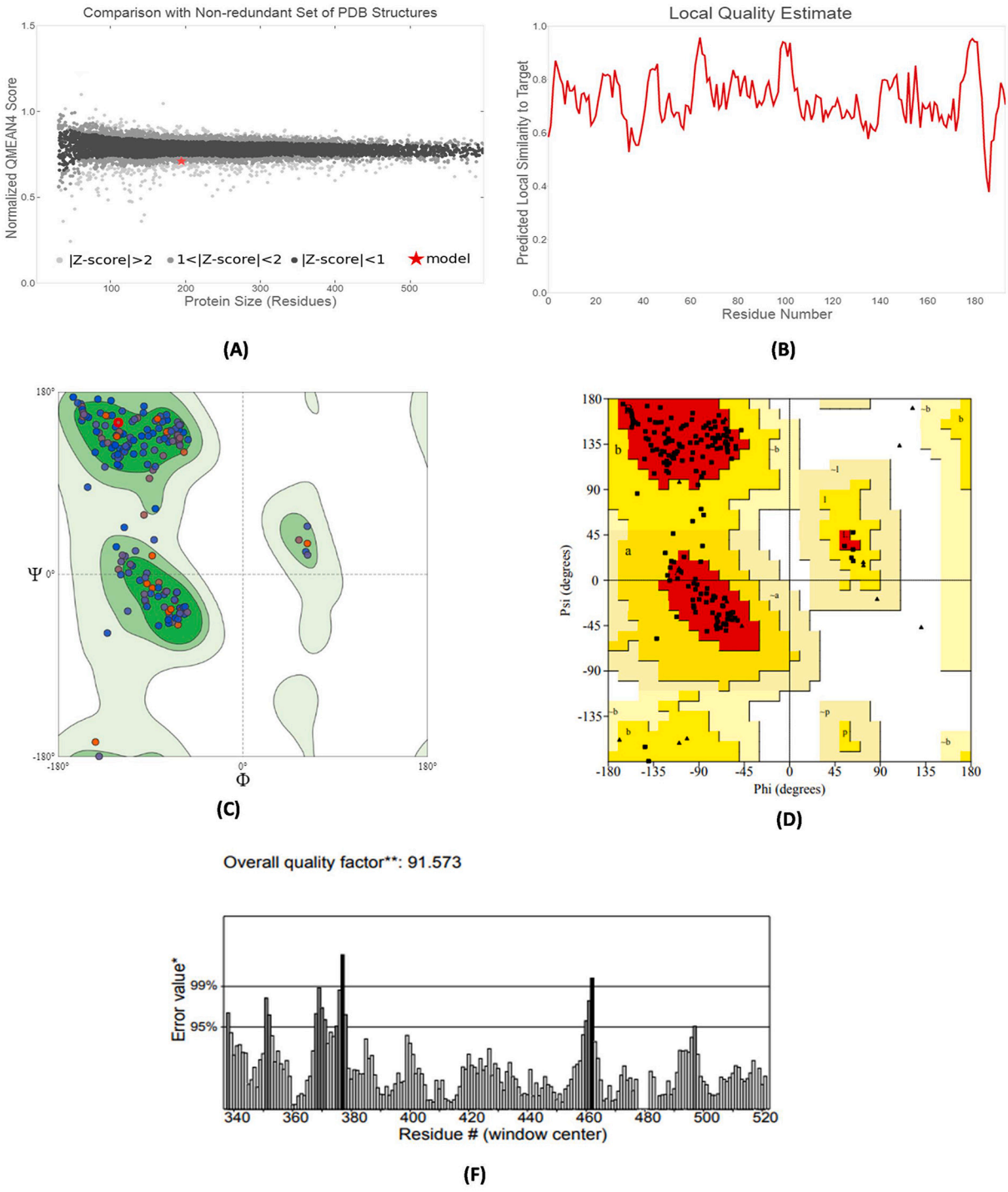

Fig. 5. Quality assessmentof the RBD model of Netherland (MT457401.1: QJS39627.1). (**A**)Qmean score, (**B**) local quality estimate, (**C**) molprobity ramachandran plot, (**D**) ramachandran plot analysed via procheck, and (**E**) overall quality factor.

vaccines against SARS-CoV-2, these mutations in RBD domain may cause serious effects.

4. Conclusions

The study revealed high similarity among sequences and also presence of mutations in all the sequences based on the results of sequence alignment and phylogenetic analysis of SARS-CoV-2 genomes. Two different types of mutation (purine-purine and pyrimidine-pyrimidine) were present in these sequences, which was indicated by the substitution model TN93 + G. This evolutionary model ensured that several degrees of transitions and transversions occurred in the genomes retrieved from Asian, European and American countries. Our mutational analysis also suggested that during transmission of COVID-19 among people, considerable positive selection took place. Furthermore, we found that mutations in RBD regions have significant effects on drug discovery and might have crucial roles in binding with the ACE2 receptor. These results indirectly indicate how the mutations of different variants, such as Alpha, Beta, Gamma, Delta, etc., can obstruct the success of drugs or vaccines. Therefore, we concluded that several mutations (whether these are in Alpha, Beta, Gamma, Delta, etc., or in other variants) should be considered during drug or vaccine development. We

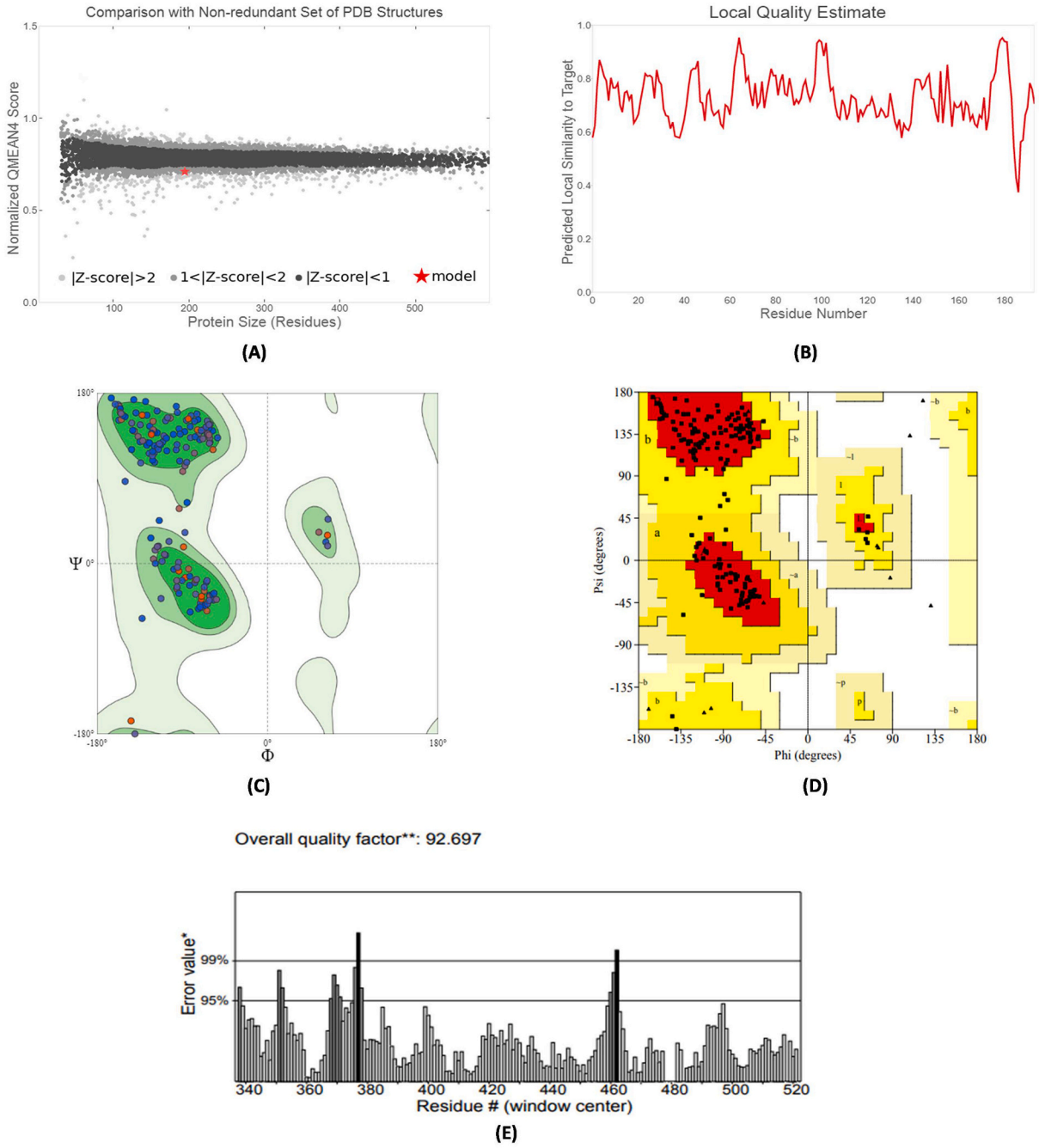

Fig. 6. Quality assessmentof the RBD model of Bangladesh (MT847220.1:QMU94884.1). (**A**)Qmean score, (**B**) local quality estimate, (**C**) molprobity ramachandran plot, (**D**) ramachandran plot analysed via procheck, and (**E**) overall quality factor.

believe that our study provides a view of how mutations in RBD domain can affect the drug efficacy.

5. Methods

5.1. Retrieval of SARS-CoV-2 genome sequences

The nucleotide sequences of SARS-CoV-2 were retrieved from the GenBank (https://www.ncbi.nlm.nih.gov/genbank/). We collected 36 genome sequences from 18 different countries (India, Iran, Bangladesh, Saudi Arabia, Pakistan, China, Russia, Spain, France, UK, Italy, Germany, Netherland, USA, Mexico, Brazil, Colombia, and Peru) of Asian, European, and American, and we selected two genome sequences randomly from each country (Supplementary file 1).

5.2. Bioinformatics analysis

A pairwise sequence alignment tool called Basic Local Alignment Search Tool (BLAST) was used for the identification of the sequences (Altschul et al., 1990). Additionally, CLUSTAL-O (Sievers et al., 2011), MUSCLE (Edgar, 2004), and MAFFT (Katoh et al., 2019) were used for performing multiple sequence alignment. For further analysis, FASTA file was generated.

Phylogenetic analysis was carried out by MEGA-X (v10.1.8), and in this case, we used the maximum likelihood (ML) algorithm (Tamura and Nei, 1993; Tamura, 2000; Hiendleder, 1998). A minimum evolutionary (ME) tree was constructed to conclude the evolution of SARS-CoV-2 (Rzhetsky and Nei, 1992). The computed evolutionary distances were in the units of the number of base substitutions per site. Using the Close-

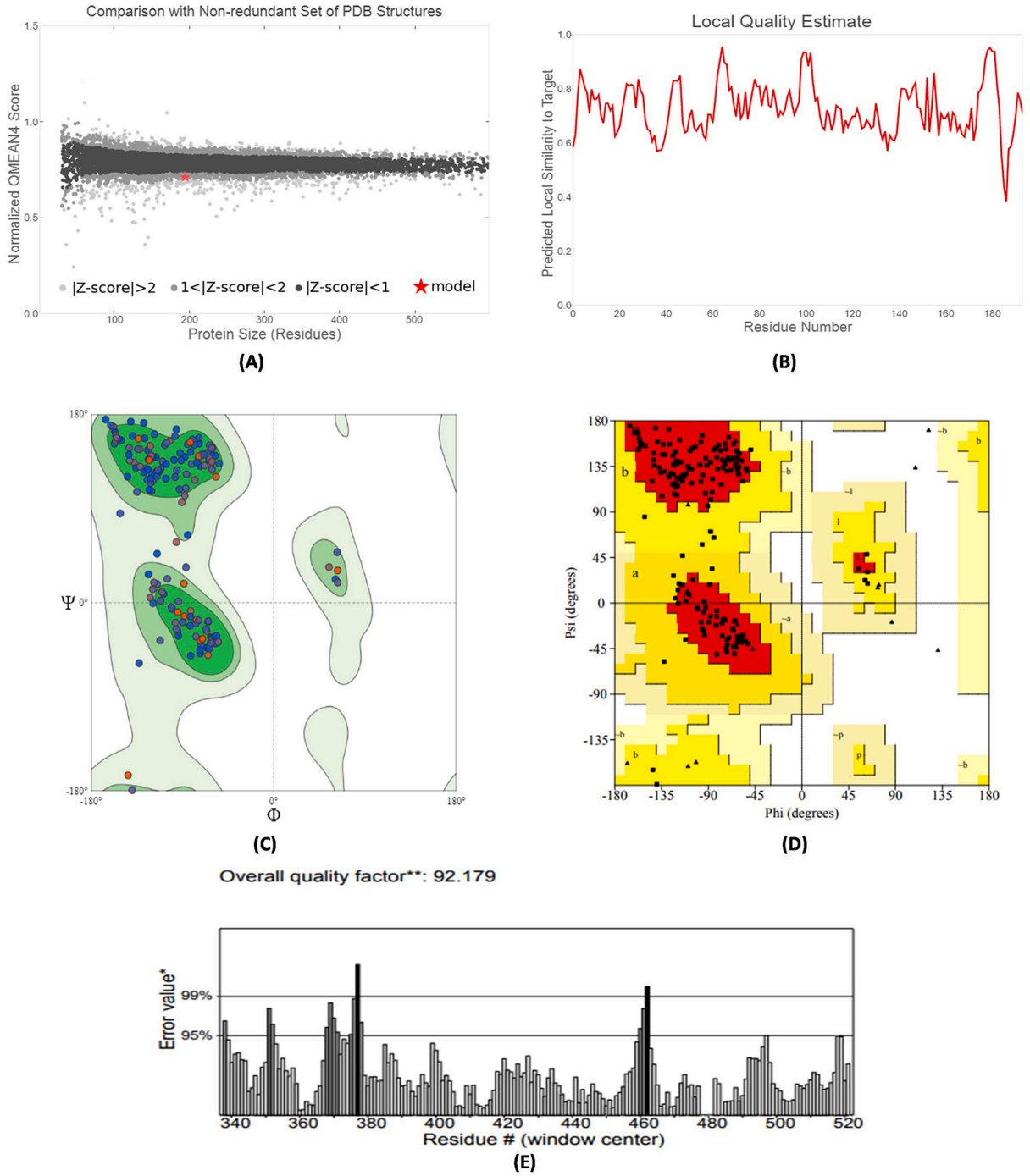

Fig. 7. Quality assessmentof the RBD model of the USA (MT940481.1: QNL36022.1). (**A**)Qmean score, (**B**) local quality estimate, (**C**) molprobity ramachandran plot, (**D**) ramachandran plot analysed via procheck, and (**E**) overall quality factor.

Neighbor-Interchange (CNI) algorithm (Thomas, 2001) at a search level of 1, this evolutionary tree was searched. The initial tree was generated by using the Neighbor-joining algorithm (Saitou and Nei, 1987). In addition, validation of the phylogenetic trees was completed employing the investigation on 100 bootstrapped input datasets (Felsenstein, 1985). At the end, for the visualization of that trees, the Interactive Tree of Life (iTOL; EMBL, Heidelberg, Germany) was used (Letunic and Bork, 2019).

Moreover, the ML DNA/protein models were found by analyzing nucleotide models substitution patterns (Thomas, 2001). Models with the minimum Bayesian information criterion (BIC) scores were considered to describe the superior substitution pattern. For all the models, the AICc value (Akaike Information Criterion, corrected), the Maximum Likelihood value (lnL), and the number of parameters (including branch lengths) were also evaluated. Transition/transversion bias (r) values were also calculated for all the models. These parameters were also followed by the calculation of rates of base substitutions (r) and nucleotide frequencies (f) for each nucleotide pair (Thomas, 2001). The disparity index test of pattern heterogeneity was used to explore the homogeneity of substitution patterns between genome sequences (Kumar and Gadagkar, 2001). *P*-values were calculated by the Monte Carlo test (500 replicates).

The ML substitution matrix with substitution pattern and rates were calculated under the Tamura-Nei (1993) model (Tamura and Nei, 1993).

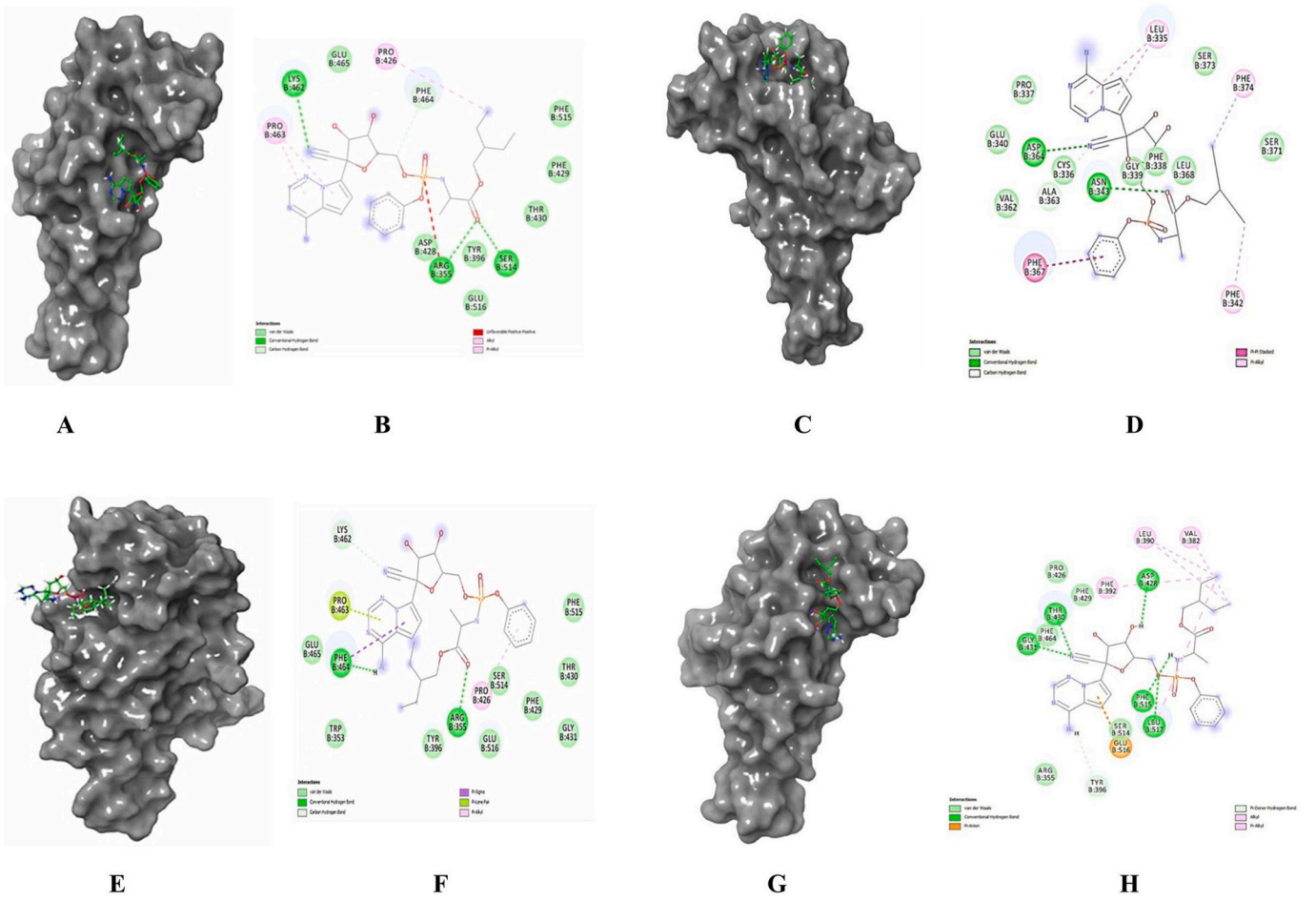

Fig. 8. Molecular docking best poses of remdesivir with RBDs of S proteins of (**A**) China, (**C**) USA, (**E**) Bangladesh and (**G**) Netherlands. 2D plots of RBDs of S proteins (**B**) China, (**D**) USA, (**F**) Bangladesh and (**H**) Netherlands, their active residues interacting with the Remdesivir.

Tree topology was calculated automatically for computing ML values, and in evaluating relative values of instantaneous r were considered. The Kimura (2-parameter model) was used for the ML estimation of Transition/Transversion bias. Tree topology was automatically calculated for ML values (Kimura, 1980). Nei-Gojobori method was used for Codon-based test of neutrality to analyze among sequences (Nei and Gojobori, 1986). Analysis of the probability of rejecting the null hypothesis of strict-neutrality (dN = dS) was also performed.

The analytical method was used to estimate the variance of the difference (Nei and Gojobori, 1986). Furthermore, Tajima's neutrality test was done to calculate the nucleotide mutation hypothesis by DNA polymorphism (Nei and Tajima, 1981) and calculation of D value was done to distinguish the genome sequences evolving from random and non-random processes (Thomas, 2001).

5.3. Homology modeling

Crystal structure of SARS-CoV-2 spike receptor-binding domain bound with ACE2 (PDB: 6M0J) was extracted from Protein Data Bank (https://www.rcsb.org/) (Rose et al., 2017) to perform modeling of RBD regions (amino acid residues 333–526) (Lan et al., 2020) of Spike (S) protein of SARS-CoV-2. For homology modeling, sequences of S-protein were submitted to Swiss-Model (Biasini et al., 2014). It was a template-based modeling and the SARS-CoV-2 spike receptor-binding domain bound with ACE2 (PDB: 6M0J) was used as a template. Swiss-Model is an automatics web-based molecular modeling tool and creates models more efficiently. As a result, building and analysis of protein homology models are easy at various levels of complexity.

5.4. Quality assessment of the models

MolProbity (Williams et al., 2018) and QMEAN server (Benkert et al., 2009) were used for validation of the constructed model. In additional, ERRAT (MacArthur et al., 1994) and Procheck (Laskowski et al., 1993) were also used for validation. To evaluate amino acid distributions in protein structure, the SAVES v5.0 (https://servicesn.mbi.ucla.edu/SAVES/) tool was utilized.

5.5. Molecular docking of remdisivir against receptor binding domains (RBDs) of S- proteins

Molecular Docking is an important part of scrutinizing the specific drug target against pathogens like viruses and it also simplifies the drug screening process (Meng et al., 2011). This process provides the value of binding affinity between ligand and receptor complex that would help to find the superior one (Huang and Zou, 2010; López-Vallejo et al., 2011). In this study, molecular interaction studies were carried out using the modeled receptor binding domains (RBDs) of S proteins of SARS-CoV-2 against remdisivir in Autodock vs 4.2.6 (Morris et al., 1996). The ligand molecule remdesivir was fetched from pubchem public domain database (https://pubchem.ncbi.nlm.nih.gov/compound/Remdesivir) in 3D SDF format and subsequently converted to PDB in Biovia Discovery Studio visualizer (https://discover.3ds.com/discovery-studio-visualizer). The 3D coordinates of the ligand was checked, hydrogens were added and energy was minimized to get the using 1000 steps conjugate gradient and 1000 steps of steepest descent algorithm in UCSF Chimera vs 1.14. After energy minimization of the ligand, autodock steps were performed

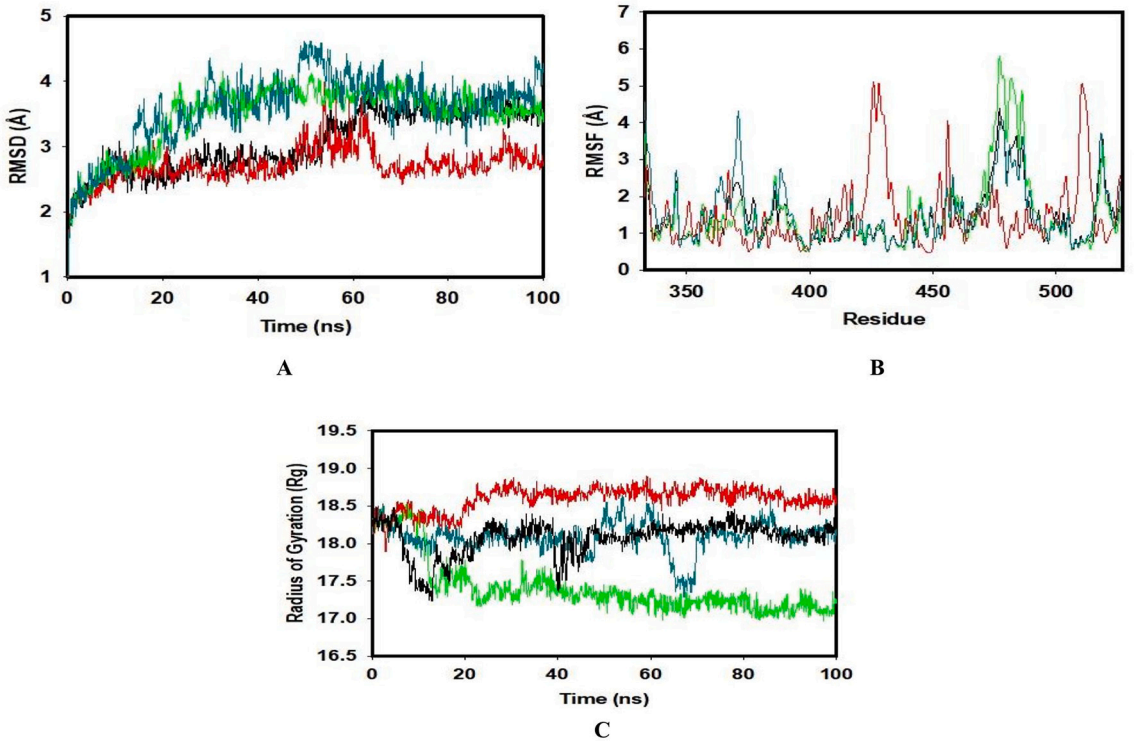

Fig. 9. MD simulation of 100 ns displaying (**A**) RMSD plots Cα-backbone of Netherlands (---), China (—), USA (---) and Bangladesh (---). (**B**) RMSF plots of Cα-backbone of Netherlands(—), China (---), USA (---) and Bangladesh (---). (**C**) Radius of gyration plots of Cα-backbone of Netherlands (---), China (---), USA (—) and Bangladesh (---).

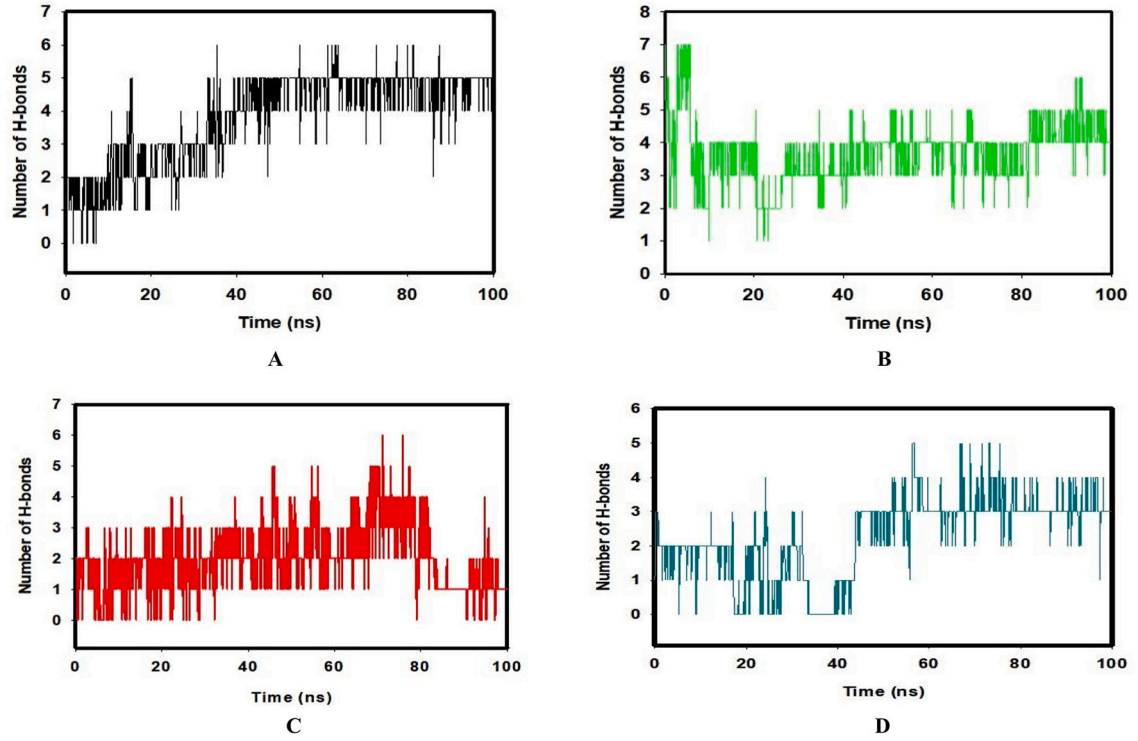

Fig. 10. Formation of average number of Hydrogen bonds (H-bonds) throughout the 100 ns of simulation between RBDs of S proteins of (**A**) Netherlands, (**B**) China, (**C**) Bangladesh and (**D**) USA, and Remdesivir.

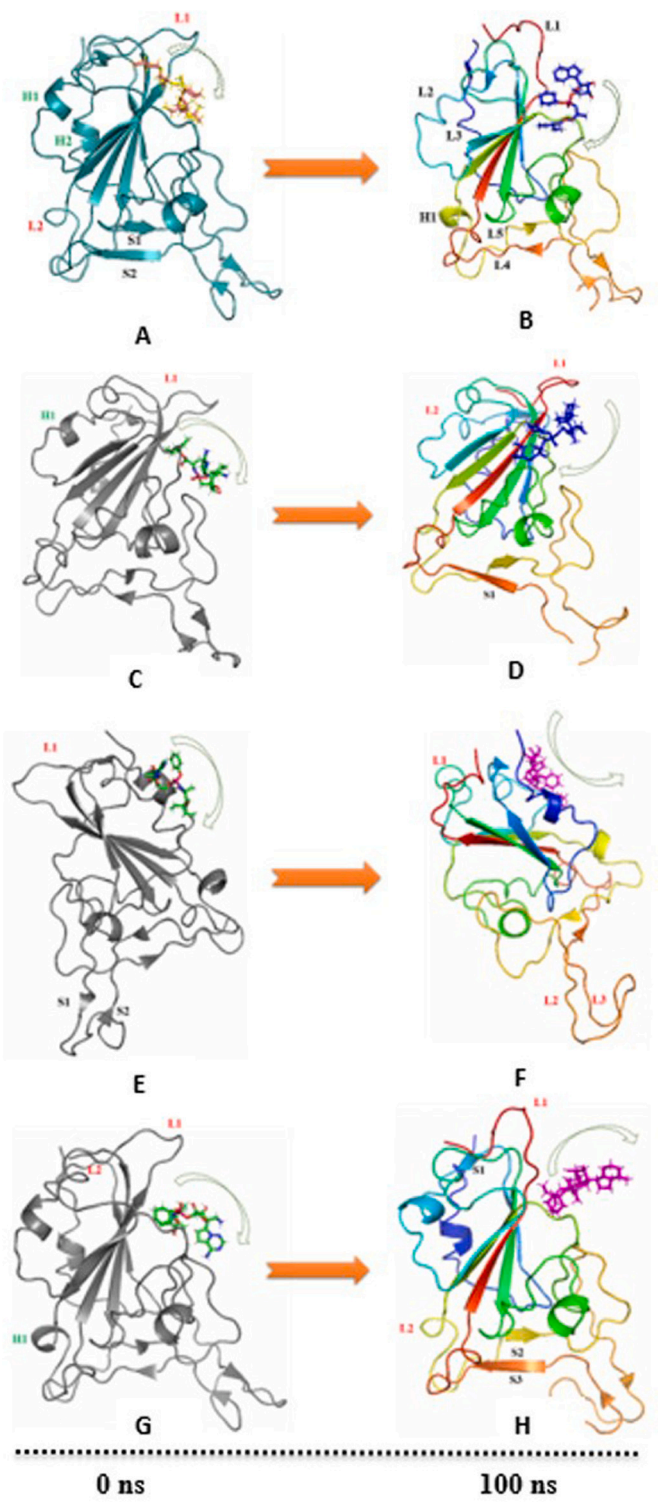

Fig. 11. Structural analysis of pre and post dynamic final trajectory to compare the structural variation of RBDs of S proteins occurred during simulation (**A**) China, (**B**) Netherland, (**C**) USA and (**D**) Bangladesh, in a bound complex with remdesivir. Arrow indicates the movement of the remdesivir at the binding cavity of S proteins.

as described elsewhere (Ghosh et al., 2019). The simulated grid box size of China, the USA, Netherlands and Bangladesh derived models S proteins with the ligand for docking was generated 46.49 × 16.639 × 6.359 with a spacing of 1 Å, respectively. The grid box was saved in grid parameter file (gpf) format. Lamarckian genetic algorithm (LGA) was used for binding of protein and ligand approximation The best docking result with lowest binding energy which was in the docking log file (dlg) format was picked and was converted to Protein data bank (pdb) format by using PyMOL 2.4 software (https://pymol.org/2/). The bonding interactions of docking complexes of RBDs of S proteins with the ligand remdesivir were analysed using BIOVIA Discovery Studio visualizer.

5.6. MD simulation and MMGBSA calculations

MD simulations studies of the RBDs of S proteins with the ligand remdesivir were performed for 100 ns using the Desmond 2018-4from Schrödinger, LLC. The OPLS-2005 force field (Bowers et al., 2006; Chow et al., 2008; Shivakumar et al., 2010) and explicit solvent model with the SPC water molecules were used in this system (Jorgensen et al., 1996). Na^+ ions were added to neutralize the charge. 0.15 M, NaCl solutions added to the system to simulate the physiological environment. The NPT ensemble was set up by using the Nose-Hoover chain coupling scheme (Martyna et al., 1992) with temperature 300 K, relaxation time of 1.0 ps and pressure 1 bar was maintained in all the simulations. A time step of 2 fs was used. The Martyna-Tuckerman–Klein chain coupling scheme (Martyna et al., 1994) barostat method was used for pressure control with a relaxation time of 2 ps. The particle mesh Ewald method (Toukmaji and Board Jr, 1996) was used for calculating long-range electrostatic interactions and the radius for the Coulomb interactions were fixed at 9 Å. RESPA integrator was used to calculate the non-bonded forces. The root mean square deviation (RMSD) was to monitor the stability of the MD simulations.

The binding free energy (ΔG_{bind}) of the docked complexes during MD simulations of RBDs of S proteins with the ligand remdesivir was estimated using the prime molecular mechanics generalized born surface area (MM-GBSA) module at the (Schrodinger suite, LLC, New York, NY, 2017–4). The OPLS 2005 force field, VSGB solvent model, and rotamer search algorithms were used to define the binding free energy during the calculation (Wang et al., 2018). The MD trajectories frames after MD run were selected at each 10 ns interval. The following formula was used to calculate the total free energy binding:

$$\Delta G_{bind} = G_{complex} - (G_{protein} + G_{ligand})$$

Where, ΔG_{bind} = binding free energy, $G_{complex}$ = free energy of the complex, $G_{protein}$ = free energy of the target protein, and G_{ligand} = free energy of the ligand. The MMGBSA outcome trajectories were analyzed further for post dynamics structure modifications.

Supplementary data to this article can be found online at https://doi.org/10.1016/j.meegid.2021.105128.

Author contributions

Conceptualization, Md Bashir Uddin, Deog-Hwan Oh and Syed Sayeem Uddin Ahmed; Data curation, Md Bashir Uddin, Emran Hossain Sajib, Syeda Farjana Hoque and Md. Nazmul Islam Bappy; Formal analysis, Md Bashir Uddin, Emran Hossain Sajib, Syeda Farjana Hoque, Md. Nazmul Islam Bappy, Fazle Elahi, Arabinda Ghosh, Samuel Muhit and Syed Sayeem Uddin Ahmed; Funding acquisition, Deog-Hwan Oh and Syed Sayeem Uddin Ahmed; Investigation, Md Bashir Uddin, Emran Hossain Sajib, Syeda Farjana Hoque, Md. Nazmul Islam Bappy and Syed Sayeem Uddin Ahmed; Methodology, Md Bashir Uddin, Emran Hossain Sajib, Syeda Farjana Hoque, Md. Nazmul Islam Bappy, Fazle Elahi, Arabinda Ghosh, Mohammad Mahmudul Hassan, Mahmudul Hasan and Syed Sayeem Uddin Ahmed; Project administration, Md Bashir Uddin and Syed Sayeem Uddin Ahmed; Resources, Mahmudul Hasan and Syed Sayeem Uddin Ahmed; Software, Md Bashir Uddin, Emran Hossain Sajib, Md. Nazmul Islam Bappy, Arabinda Ghosh, Mahmudul Hasan, Ramachandran Chelliah, Se Jin Park and Syed Sayeem Uddin Ahmed; Supervision, Md Bashir Uddin and Syed Sayeem Uddin Ahmed; Validation, Ramachandran Chelliah, Se Jin Park, Tamnna Jahan Mony,

Deog-Hwan Oh and Syed Sayeem Uddin Ahmed; Visualization, Fazle Elahi, Samuel Muhit, Mohammad Mahmudul Hassan, Se Jin Park and Deog-Hwan Oh; Writing – original draft, Md Bashir Uddin, Emran Hossain Sajib, Syeda Farjana Hoque, Arabinda Ghosh and Syed Sayeem Uddin Ahmed; Writing – review & editing, Md Bashir Uddin, Emran Hossain Sajib, Syeda Farjana Hoque, Md. Nazmul Islam Bappy, Fazle Elahi, Arabinda Ghosh, Samuel Muhit, Mohammad Mahmudul Hassan, Mahmudul Hasan, Ramachandran Chelliah, Se Jin Park, Tamnna Jahan Mony, Deog-Hwan Oh and Syed Sayeem Uddin Ahmed.

Funding

This research supported by Korea Research Fellowship (KRF) Program (KRF Grant No: 2020H1D3A1A02081423, through the National Research Foundation of Korea (NRF) funded by the Ministry of Science and ICT, Seoul, Republic of Korea.

Declaration of Competing Interest

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

References

Altschul, S.F., Gish, W., Miller, W., Myers, E.W., Lipman, D.J., 1990. Basic local alignment search tool. J. Mol. Biol. 215, 403–410.
Benkert, P., Künzli, M., Schwede, T., 2009. QMEAN server for protein model quality estimation. Nucleic Acids Res. 37, W510–W514.
Benvenuto, D., Giovanetti, M., Ciccozzi, A., Spoto, S., Angeletti, S., Ciccozzi, M., 2020. The 2019-new coronavirus epidemic: evidence for virus evolution. J. Med. Virol. 92, 455–459.
Bertram, S., Dijkman, R., Habjan, M., Heurich, A., Gierer, S., Glowacka, I., Welsch, K., Winkler, M., Schneider, H., Hofmann-Winkler, H., et al., 2013. TMPRSS2 activates the human coronavirus 229E for cathepsin-independent host cell entry and is expressed in viral target cells in the respiratory epithelium. J. Virol. 87, 6150–6160.
Biasini, M., Bienert, S., Waterhouse, A., Arnold, K., Studer, G., Schmidt, T., Kiefer, F., Gallo Cassarino, T., Bertoni, M., Bordoli, L., Schwede, T., 2014. SWISS-MODEL: modelling protein tertiary and quaternary structure using evolutionary information. Nucleic Acids Res. 42, W252–W258.
Bosch, B.J., van der Zee, R., de Haan, C.A., Rottier, P.J., 2003. The coronavirus spike protein is a class I virus fusion protein: structural and functional characterization of the fusion core complex. J. Virol. 77, 8801–8811.
Bowers, K.J., Chow, D.E., Xu, H., Dror, R.O., Eastwood, M.P., Gregersen, B.A., Klepeis, J. L., Kolossvary, I., Moraes, M.A., Sacerdoti, F.D., 2006. Scalable algorithms for molecular dynamics simulations on commodity clusters. In: SC'06: Proceedings of the 2006 ACM/IEEE Conference on Supercomputing. IEEE, p. 43.
Burke, R.M., Midgley, C.M., Dratch, A., Fenstersheib, M., Haupt, T., Holshue, M., Ghinai, I., Jarashow, M.C., Lo, J., McPherson, T.D., et al., 2020. Active monitoring of persons exposed to patients with confirmed COVID-19 - United States, January–February 2020. MMWR Morb. Mortal. Wkly Rep. 69, 245–246.
Ceraolo, C., Giorgi, F.M., 2020. Genomic variance of the 2019-nCoV coronavirus. J. Med. Virol. 92, 522–528.
Chakraborty, I., Maity, P., 2020. COVID-19 outbreak: migration, effects on society, global environment and prevention. Sci. Total Environ. 728, 138882.
Chan, J.F., Yuan, S., Kok, K.H., To KK, Chu, H., Yang, J., Xing, F., Liu, J., Yip, C.C., Poon, R.W., et al., 2020. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet 395, 514–523.
Chen, J., 2020. Pathogenicity and transmissibility of 2019-nCoV-A quick overview and comparison with other emerging viruses. Microbes Infect. 22, 69–71.
Chen, L., Liu, W., Zhang, Q., Xu, K., Ye, G., Wu, W., Sun, Z., Liu, F., Wu, K., Zhong, B., et al., 2020. RNA based mNGS approach identifies a novel human coronavirus from two individual pneumonia cases in 2019 Wuhan outbreak. Emerg Microbes Infect. 9, 313–319.
Chow, E., Rendleman, C.A., Bowers, K.J., Dror, R.O., Hughes, D.H., Gullingsrud, J., Sacerdoti, F.D., Shaw, D.E., 2008. Desmond performance on a cluster of multicore processors. In: DE Shaw Research Technical Report DESRES/TR–2008-01.
Du, L., Kao, R.Y., Zhou, Y., He, Y., Zhao, G., Wong, C., Jiang, S., Yuen, K.Y., Jin, D.Y., Zheng, B.J., 2007. Cleavage of spike protein of SARS coronavirus by protease factor Xa is associated with viral infectivity. Biochem. Biophys. Res. Commun. 359, 174–179.
Eastman, R.T., Roth, J.S., Brimacombe, K.R., Simeonov, A., Shen, M., Patnaik, S., Hall, M.D., 2020. Remdesivir: a review of its discovery and development leading to emergency use authorization for treatment of COVID-19. ACS Cent Sci. 6, 672–683.
Edgar, R.C., 2004. MUSCLE: multiple sequence alignment with high accuracy and high throughput. Nucleic Acids Res. 32, 1792–1797.
Eweas, A.F., Alhossary, A.A., Abdel-Moneim, A.S., 2021. Molecular docking reveals ivermectin and remdesivir as potential repurposed drugs against SARS-CoV-2. Front. Microbiol. 11.
Felsenstein, J., 1985. Confidence limits on phylogenies: an approach using the bootstrap. Evolution 39, 783–791.
Ghinai, I., McPherson, T.D., Hunter, J.C., Kirking, H.L., Christiansen, D., Joshi, K., Rubin, R., Morales-Estrada, S., Black, S.R., Pacilli, M., et al., 2020. First known person-to-person transmission of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in the USA. Lancet 395, 1137–1144.
Ghosh, A., Sutradhar, S., Baishya, D., 2019. Delineating thermophilic xylanase from bacillus licheniformis DM5 towards its potential application in xylooligosaccharides production. World J. Microbiol. Biotechnol. 35, 34.
Hamner, L., Dubbel, P., Capron, I., Ross, A., Jordan, A., Lee, J., Lynn, J., Ball, A., Narwal, S., Russell, S., et al., 2020. High SARS-CoV-2 attack rate following exposure at a choir practice - Skagit County, Washington, march 2020. MMWR Morb. Mortal. Wkly Rep. 69, 606–610.
Hiendleder, S., 1998. A low rate of replacement substitutions in two major Ovis aries mitochondrial genomes. Anim. Genet. 29, 116–122.
Hoffmann, M., Kleine-Weber, H., Schroeder, S., Krüger, N., Herrler, T., Erichsen, S., Schiergens, T.S., Herrler, G., Wu, N.H., Nitsche, A., et al., 2020. SARS-CoV-2 cell entry depends on ACE2 and TMPRSS2 and is blocked by a clinically proven protease inhibitor. Cell 181, 271–280 e278.
Huang, S.Y., Zou, X., 2010. Advances and challenges in protein-ligand docking. Int. J. Mol. Sci. 11, 3016–3034.
Huang, C., Wang, Y., Li, X., Ren, L., Zhao, J., Hu, Y., Zhang, L., Fan, G., Xu, J., Gu, X., et al., 2020a. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet 395, 497–506.
Huang, Y., Yang, C., Xu, X.F., Xu, W., Liu, S.W., 2020b. Structural and functional properties of SARS-CoV-2 spike protein: potential antivirus drug development for COVID-19. Acta Pharmacol. Sin. 41, 1141–1149.
Hughes, A.L., 1999. Adaptive Evolution of Genes and Genomes. Oxford University Press, USA.
Ji, W., Wang, W., Zhao, X., Zai, J., Li, X., 2020. Cross-species transmission of the newly identified coronavirus 2019-nCoV. J. Med. Virol. 92, 433–440.
Jorgensen, W.L., Maxwell, D.S., Tirado-Rives, J., 1996. Development and testing of the OPLS all-atom force field on conformational energetics and properties of organic liquids. J. Am. Chem. Soc. 118, 11225–11236.
Katoh, K., Rozewicki, J., Yamada, K.D., 2019. MAFFT online service: multiple sequence alignment, interactive sequence choice and visualization. Brief. Bioinform. 20, 1160–1166.
Kimura, M., 1980. A simple method for estimating evolutionary rates of base substitutions through comparative studies of nucleotide sequences. J. Mol. Evol. 16, 111–120.
Kumar, S., 1996. Patterns of nucleotide substitution in mitochondrial protein coding genes of vertebrates. Genetics 143, 537–548.
Kumar, S., Gadagkar, S.R., 2001. Disparity index: a simple statistic to measure and test the homogeneity of substitution patterns between molecular sequences. Genetics 158, 1321–1327.
Kumar, S., Maurya, V.K., Prasad, A.K., Bhatt, M.L.B., Saxena, S.K., 2020. Structural, glycosylation and antigenic variation between 2019 novel coronavirus (2019-nCoV) and SARS coronavirus (SARS-CoV). Virusdisease 31, 13–21.
Lai, C.-C., Shih, T.-P., Ko, W.-C., Tang, H.-J., Hsueh, P.-R., 2020. Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) and corona virus disease-2019 (COVID-19): the epidemic and the challenges. Int. J. Antimicrob. Agents 105924.
Lan, J., Ge, J., Yu, J., Shan, S., Zhou, H., Fan, S., Zhang, Q., Shi, X., Wang, Q., Zhang, L., Wang, X., 2020. Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. Nature 581, 215–220.
Laskowski, R.A., MacArthur, M.W., Moss, D.S., Thornton, J.M., 1993. PROCHECK: a program to check the stereochemical quality of protein structures. J. Appl. Crystallogr. 26, 283–291.
Lau, Y.L., Peiris, J.S., 2005. Pathogenesis of severe acute respiratory syndrome. Curr. Opin. Immunol. 17, 404–410.
Letunic, I., Bork, P., 2019. Interactive tree of life (iTOL) v4: recent updates and new developments. Nucleic Acids Res. 47, W256–w259.
Li, Q., Guan, X., Wu, P., Wang, X., Zhou, L., Tong, Y., Ren, R., Leung, K.S.M., Lau, E.H.Y., Wong, J.Y., et al., 2020a. Early transmission dynamics in Wuhan, China, of novel coronavirus-infected pneumonia. N. Engl. J. Med. 382, 1199–1207.
Li, X., Zai, J., Zhao, Q., Nie, Q., Li, Y., Foley, B.T., Chaillon, A., 2020b. Evolutionary history, potential intermediate animal host, and cross-species analyses of SARS-CoV-2. J. Med. Virol. 92, 602–611.
Liu, J., Liao, X., Qian, S., Yuan, J., Wang, F., Liu, Y., Wang, Z., Wang, F.S., Liu, L., Zhang, Z., 2020a. Community transmission of severe acute respiratory syndrome coronavirus 2, Shenzhen, China, 2020. Emerg. Infect. Dis. 26, 1320–1323.
Liu, C., Zhou, Q., Li, Y., Garner, L.V., Watkins, S.P., Carter, L.J., Smoot, J., Gregg, A.C., Daniels, A.D., Jervey, S., Albaiu, D., 2020b. Research and development on therapeutic agents and vaccines for COVID-19 and related human coronavirus diseases. ACS Cent Sci. 6, 315–331.
López-Vallejo, F., Caulfield, T., Martínez-Mayorga, K., Giulianotti, M.A., Nefzi, A., Houghten, R.A., Medina-Franco, J.L., 2011. Integrating virtual screening and combinatorial chemistry for accelerated drug discovery. Comb. Chem. High Throughput Screen. 14, 475–487.
Lu, R., Zhao, X., Li, J., Niu, P., Yang, B., Wu, H., Wang, W., Song, H., Huang, B., Zhu, N., et al., 2020. Genomic characterisation and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet 395, 565–574.

MacArthur, M.W., Laskowski, R.A., Thornton, J.M., 1994. Knowledge-based validation of protein structure coordinates derived by X-ray crystallography and NMR spectroscopy. Curr. Opin. Struct. Biol. 4, 731–737.
Martyna, G.J., Klein, M.L., Tuckerman, M., 1992. Nosé–hoover chains: the canonical ensemble via continuous dynamics. J. Chem. Phys. 97, 2635–2643.
Martyna, G.J., Tobias, D.J., Klein, M.L., 1994. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 101, 4177–4189.
Meng, X.Y., Zhang, H.X., Mezei, M., Cui, M., 2011. Molecular docking: a powerful approach for structure-based drug discovery. Curr. Comput. Aid. Drug Des. 7, 146–157.
Morris, G.M., Goodsell, D.S., Huey, R., Olson, A.J., 1996. Distributed automated docking of flexible ligands to proteins: parallel applications of AutoDock 2.4. J. Comput. Aided Mol. Des. 10, 293–304.
Naqvi, A.A.T., Fatima, K., Mohammad, T., Fatima, U., Singh, I.K., Singh, A., Atif, S.M., Hariprasad, G., Hasan, G.M., Hassan, M.I., 2020. Insights into SARS-CoV-2 genome, structure, evolution, pathogenesis and therapies: structural genomics approach. Biochim. Biophys. Acta Mol. basis Dis. 1866, 165878.
Nei, M., Gojobori, T., 1986. Simple methods for estimating the numbers of synonymous and nonsynonymous nucleotide substitutions. Mol. Biol. Evol. 3, 418–426.
Nei, M., Tajima, F., 1981. DNA polymorphism detectable by restriction endonucleases. Genetics 97, 145–163.
Rose, P.W., Prlić, A., Altunkaya, A., Bi, C., Bradley, A.R., Christie, C.H., Costanzo, L.D., Duarte, J.M., Dutta, S., Feng, Z., et al., 2017. The RCSB protein data bank: integrative view of protein, gene and 3D structural information. Nucleic Acids Res. 45, D271–d281.
Rzhetsky, A., Nei, M., 1992. A Simple Method for Estimating and Testing Minimum-Evolution Trees.
Saitou, N., Nei, M., 1987. The neighbor-joining method: a new method for reconstructing phylogenetic trees. Mol. Biol. Evol. 4, 406–425.
Sanders, J.M., Monogue, M.L., Jodlowski, T.Z., Cutrell, J.B., 2020. Pharmacologic treatments for coronavirus disease 2019 (COVID-19): a review. Jama 323, 1824–1836.
Sheikh, J.A., Singh, J., Singh, H., Jamal, S., Khubaib, M., Kohli, S., Dobrindt, U., Rahman, S.A., Ehtesham, N.Z., Hasnain, S.E., 2020. Emerging genetic diversity among clinical isolates of SARS-CoV-2: lessons for today. Infect. Genet. Evol. 84, 104330.
Shivakumar, D., Williams, J., Wu, Y., Damm, W., Shelley, J., Sherman, W., 2010. Prediction of absolute solvation free energies using molecular dynamics free energy perturbation and the OPLS force field. J. Chem. Theory Comput. 6, 1509–1519.
Sievers, F., Wilm, A., Dineen, D., Gibson, T.J., Karplus, K., Li, W., Lopez, R., McWilliam, H., Remmert, M., Söding, J., et al., 2011. Fast, scalable generation of high-quality protein multiple sequence alignments using Clustal omega. Mol. Syst. Biol. 7, 539.
Tamura, K., 2000. On the estimation of the rate of nucleotide substitution for the control region of human mitochondrial DNA. Gene 259, 189–197.
Tamura, K., Nei, M., 1993. Estimation of the number of nucleotide substitutions in the control region of mitochondrial DNA in humans and chimpanzees. Mol. Biol. Evol. 10, 512–526.
Thomas, R.H., 2001. Molecular evolution and Phylogenetics. Heredity 86, 385.
Tortorici, M.A., Veesler, D., 2019. Structural insights into coronavirus entry. Adv. Virus Res. 105, 93–116.
Toukmaji, A.Y., Board Jr., J.A., 1996. Ewald summation techniques in perspective: a survey. Comput. Phys. Commun. 95, 73–92.
Vijaykrishna, D., Smith, G.J., Zhang, J.X., Peiris, J.S., Chen, H., Guan, Y., 2007. Evolutionary insights into the ecology of coronaviruses. J. Virol. 81, 4012–4020.
Walls, A.C., Park, Y.J., Tortorici, M.A., Wall, A., McGuire, A.T., Veesler, D., 2020. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. Cell 181, 281–292 e286.
Wang, H., Pipes, L., Nielsen, R., 2021. Synonymous mutations and the molecular evolution of SARS-Cov-2 origins. Virus evolution 7 (1), veaa098.
Wang, M., Wang, Y., Kong, D., Jiang, H., Wang, J., Cheng, M., 2018. In silico exploration of aryl sulfonamide analogs as voltage-gated sodium channel 1.7 inhibitors by using 3D-QSAR, molecular docking study, and molecular dynamics simulations. Comput. Biol. Chem. 77, 214–225.
Wang, Q., Zhang, Y., Wu, L., Niu, S., Song, C., Zhang, Z., Lu, G., Qiao, C., Hu, Y., Yuen, K. Y., et al., 2020. Structural and functional basis of SARS-CoV-2 entry by using human ACE2. Cell 181, 894–904 e899.
WHO, 2020. Advice on the Use of Masks in the Context of COVID-19: Interim Guidance, 5 June 2020. World Health Organization.
Williams, C.J., Headd, J.J., Moriarty, N.W., Prisant, M.G., Videau, L.L., Deis, L.N., Verma, V., Keedy, D.A., Hintze, B.J., Chen, V.B., et al., 2018. MolProbity: more and better reference data for improved all-atom structure validation. Protein Sci. 27, 293–315.
Wu, C., Liu, Y., Yang, Y., Zhang, P., Zhong, W., Wang, Y., Wang, Q., Xu, Y., Li, M., Li, X., et al., 2020. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm. Sin. B 10, 766–788.
Xu, Z., Shi, L., Wang, Y., Zhang, J., Huang, L., Zhang, C., Liu, S., Zhao, P., Liu, H., Zhu, L., et al., 2020. Pathological findings of COVID-19 associated with acute respiratory distress syndrome. Lancet Respir. Med. 8, 420–422.
Yan, R., Zhang, Y., Li, Y., Xia, L., Guo, Y., Zhou, Q., 2020. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. Science 367, 1444–1448.
Yang, Z., Bielawski, J.P., 2000. Statistical methods for detecting molecular adaptation. Trends Ecol. Evol. 15, 496–503.
Zhang, H., Penninger, J.M., Li, Y., Zhong, N., Slutsky, A.S., 2020. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. Intensive Care Med. 46, 586–590.
Zhou, P., Yang, X.L., Wang, X.G., Hu, B., Zhang, L., Zhang, W., Si, H.R., Zhu, Y., Li, B., Huang, C.L., et al., 2020. A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature 579, 270–273.