In silico analysis of RNA-dependent RNA polymerase of the SARS-CoV-2 and potentiality of the pre-existing drugs

Samyabrata Mukhoty

The University of Burdwan

Madhab Kumar Sen

Czech University of Life Sciences Prague

Subhajit Ghosh

The University of Burdwan

Himangsu Kundu

The University of Burdwan

Suvankar Das

Calcutta University

Sunil Kanti Mondal (✉ skmondal@biotech.buruniv.ac.in)

The University of Burdwan <https://orcid.org/0000-0001-8537-2564>

Research Article

Keywords: SARS-CoV-2, RdRp, mutation, evolution, drug, molecular docking

Posted Date: August 19th, 2020

DOI: <https://doi.org/10.21203/rs.3.rs-61163/v1>

License: © ⓘ This work is licensed under a Creative Commons Attribution 4.0 International License.

[Read Full License](#)

Abstract

Background: The alarming increase in the number of SARS-CoV-2 cases worldwide, urgently demands far-reaching effective strategies to win the battle against emerging as well as re-emerging diseases. Many research laboratories are conducting clinical trials with different drugs, among which some became quite interestingly effective against this pandemic. Our aim is to investigate the potentiality of the pre-existing drugs and get a clear understanding of their effects on RNA-dependent RNA polymerase (RdRp) of the SARS-CoV-2.

Methods: Multiple sequence alignment (MSA) along with molecular phylogeny analysis were performed using homologous sequences to identify the mutations within RdRp and evolutionary relationship. Based on the published literatures, we have chosen eight drug molecules like Ribavirin, Tenofovir, Sofosbuvir, IDX-184, YAK, Setrobuvir, Remdesivir and Galidesivir. Series of molecular docking studies between template RNA and RdRp of SARS-CoV-2 has been performed in absence or presence of those drugs and cofactors nsp7 and nsp8.

Results: From MSA 13 exclusive mutations identified within RdRp of SARS-CoV-2 and phylogeny reveals its close relation with Bat coronavirus RaTG13. The interaction affinities and interacting residues as obtained from systematic molecular docking study led to a conclusion that the chosen drugs can prove to be effective against this pandemic partially.

Conclusion: Therefore, designing of inhibitors which will specifically act against nsp12 and lower the binding affinity of nsp12 towards RNA template is necessary. In this article we additionally focus on the nsp7-8 hexadecameric complex and had suggested a list of amino acid residues of nsp12 and nsp7-8 complex to develop far-reaching effective drugs.

1. Introduction

Since its outbreak in December 2019, in Wuhan, Hubei Province in China, Severe Acute Respiratory Syndrome Coronavirus 2 (SARS-CoV-2) (formally known as 2019 novel coronavirus or 2019-nCoV), had infected more than seventy lakh people worldwide, as of 18th July 2020, leading to more than 6.7 lakh deaths across the globe (Source: <https://www.worldometers.info/coronavirus/>). Although recently the number of cases had reduced in China, an exponential increase in the number of cases has been detected in countries like the United States of America, Spain, Italy, France, the United Kingdom, Germany, India, South Korea, Iran etc.. As of the present scenario, the worst affected country is the United States of America, followed by Spain (Source: <https://www.worldometers.info/coronavirus/>). In many countries, these numbers are disesteemed, mainly due to the lack of scrutiny and testing facilities. The most common existing symptoms of SARS-CoV-2 include fevers, chills and shortness of breathing and the most affected organ is the lungs. Recently, The Centers for Disease Control and Prevention had introduced new symptoms, which can be linked with this viral infection. These include muscle pain, sore

throat, loss of taste or smell and headaches, influenza like symptoms, often leads to the development of Acute Respiratory disorders (ARDS) or even multiple organ failure [1–6].

SARS-CoV-2 belongs to a group of positive-sense ssRNA enveloped viruses (60nm - 140nm diameter) known as Coronavirus. These viruses earned their name due to their characteristic presence of crown-like projections on their surfaces [7–9]. Till date, only four genera of these viruses have been identified, namely α , β , γ and δ . Among the human coronaviruses, HCoV-229E and NL63 fall under α category and MERS-CoV, SARS-CoV, HCoV-OC43 and HCoV-HKU1 fall under β category [10–13]. Scientists predict 14 functional ORFs within the SARS-CoV genome, which include the replicase and protease genes, the spike (S), envelope (E), membrane (M), and nucleocapsid (N) genes (order of appearance: 5' to 3') [10, 14]. The replicase gene (ORF1a) and protease gene (ORF1b) are known to encode polyprotein1a (pp1a) and polyprotein1ab (pp1ab), which are further processed by Papain-like protease (PLpro) and Chymotrypsin-like protease (3CLpro) to yield sixteen individual nonstructural proteins (nsp). The pp1ab is thought to be translated through a frameshifting mechanism [10, 15]. RACE-PCR and northern blotting techniques had identified the leader sequence and the transcription regulatory sequence at the 5'UTR and highly conserved s2m motif at the 3'UTRs [16]. Trimeric spike glycoprotein (or S protein) plays the most important role in coronavirus infection, by neutralizing antibodies, binding to the receptors, and mediating the membrane fusion, which finally leads to the entry of the viral particles. The S proteins of the coronaviruses remain as trimers in their pre-fusion and post-fusion states and hence can be categorized under typical class-I viral fusion proteins. To activate their fusion potentials, a protease cleavage is required. In the cases of SARS-CoV and MERS-CoV, scientists had proposed a two-step protease cleavage model, which includes priming cleavage between S1 and S2 and activating cleavage on S2' site. Coronavirus may enter the target cells through the plasma membrane or via endocytosis, which depends on the availability of the types of proteases within the cell. Several proteases like human airway trypsin-like protease (HAT), transmembrane protease serine protease-2 and 4 (TMPRSS-2 and TMPRSS-4), trypsin, cathepsins etc. are known to cleave the S proteins of Coronaviruses. The N-terminal or the C-terminal domain can serve as the receptor-binding domain, depending on the type of virus. Most coronaviruses use their C-terminal domain to bind to the receptors [17–18]. It is also being reported that SARS-CoV-2 uses human angiotensin-converting enzyme 2 (hACE2) as the receptor, to enter the human bodies [19]. The replication of viral RNA from an RNA template is catalyzed by a special class of enzymes known as RNA-dependent RNA polymerase (RdRp or RNA replicase). Besides playing an important role in the early stages of evolution, these classes of enzymes forms an integral part in the life cycle of RNA-containing viruses with no DNA stage. Although RdRps share structural similarities with the DNA-dependent DNA polymerases and reverse transcriptase, their error rate in the process of transcription is much higher than the others. These higher error rates lead to genomic variations within the viral populations. But, through the RNA recombination process, viruses can repair these mutations and genetic rearrangements may help them to acquire new genes, which may provide a selective advantage to the viral populations [20]. In general, the virus with segmented genomes prefers reassortment of RNA. This is completely contrasting in cases of RNA viruses. RNA viruses undergo recombination between two non-segmented RNA genomes, which involves polymerase jumping during the synthesis of RNA. This RNA

recombination mechanism can be considered similar to the generation of defective interfering RNA (a very common phenomenon among RNA viruses) except the fact that the RNA recombination mechanism has been found in cases of limited viral species [21]. Only a few members belonging to picornavirus and coronavirus have been reported to undergo RNA recombination [22, 23]. Several plant viruses, like Brome Mosaic Virus and Cowpea Chlorotic Mottle Virus, has also been reported to undertake RNA recombination, but only in some rare situations [24, 25, 21]. The viral replication and transcription processes of CoVs are mainly facilitated by a set of non-structural proteins (products of cleavage from viral polyproteins) and the RNA-dependent RNA polymerase (RdRp). Non-structural proteins 7 and 8 (nsp7 and nsp8) act as cofactors for the COVID-19-RdRp (also known as nsp12) and play an important role in the replication and transcription cycle of these viruses [26]. In brief, the structure of the SARS-CoV-2 RdRp consists of a polymerase domain (RdRp domain) and a unique N-terminal domain that forms architecture similar to the nidovirus RdRp-associated nucleotidyltransferase (NiRAN) [27, 28]. In addition to this confirmation, the cryo-EM map revealed that an interface domain connects the polymerase domain and NiRAN domain. A further detail on the structure of the COVID-19 viral RNA-dependent RNA polymerase is provided in the article published by Gao et. al, 2020 [28]. Several research papers had been published recently, which provides a clear idea about the molecular pathogenesis of the SARS-CoV-2 [10, 29–31].

The alarming increase in the number of SARS-CoV-2 cases worldwide, urgently demands far-reaching effective strategies, which include the development of effective vaccines and drugs, to win the battle against emerging as well as re-emerging diseases. The aim of our study is to investigate the potentiality of the pre-existing drugs and get a clear understanding of their effects on RNA-dependent RNA polymerase of the SARS-CoV-2.

2. Materials And Methods

2.1 Retrieval of the sequences

Database from Zhang Lab (freely accessible from <https://zhanglab.ccmb.med.umich.edu/COVID-19/>) was searched to identify the amino acid sequence of RNA-directed RNA polymerase (RdRp) of SARS-CoV-2. Thereafter, its homologous sequences were obtained from the BLASTP 2.2.32 (developed and maintained by National Center for Biotechnology Information; freely accessible from <https://blast.ncbi.nlm.nih.gov/Blast.cgi>) search with the already retrieved RdRp protein sequence as query [32]. In total, ORF1ab polyprotein from SARS-CoV-2 and 28 different organisms (selection criteria: e-value < 0.01, in BLASTP search) were collected and saved in FASTA format until further processing.

2.3 Multiple Sequence Alignment and Phylogenetic Analysis

Identification of the corresponding homologous regions among a bunch of input sequences provide useful pieces of information related to the biological relationships among the sequences of interest.

Analysis of protein sequence alignments has been considered as an important step in bioinformatics-based researches, including biomedical research. The sequence alignment profile of our sequences of interest was obtained from Clustal Omega tool (freely accessible from <https://www.ebi.ac.uk/Tools/msa/clustalo/>) and JalView option was also used alongwith to get a better and well-defined representation of the sequence logos [33, 34]. In addition to the sequence alignment, MEGAX (Molecular Evolutionary Genetic Analysis X) program was used to construct informative maximum-likelihood phylogenetic trees (bootstrap value: 1000; method: Jones-Taylor-Thornton (JTT) model) [35, 36].

2.4 Molecular Docking studies

Based on the previously published literatures, eight drug molecules were chosen for this study, which include Ribavirin (CID: 37542), Tenofovir (CID: 464205), Sofosbuvir (CID: 45375808), IDX-184 (CID: 135565589), YAK (CID: 11492563), Setrobuvir (CID: 135565932), Remdesivir (CID: 121304016) and Galidesivir (CID: 10445549) [37, 38]. These drug molecules were chosen based on their potentials to affect the polymerase activity of RdRp through inhibition/blocking of the catalytic domain which is responsible for the binding with the RNA. The related informations of the chosen drugs were retrieved from PubChem (<https://pubchem.ncbi.nlm.nih.gov/>). Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) repository of SWISS-MODEL (freely accessible from: <https://swissmodel.expasy.org/repository/species/2697049>) was searched to obtain the structures of RNA-directed RNA polymerase (<https://swissmodel.expasy.org/interactive/JDUya4/models/02>) and nsp7 / nsp8 hetero-oligomeric complex (<https://swissmodel.expasy.org/interactive/JhWqVj/models/01>). In case of RNA-directed RNA polymerase structure, among the two models, the best model was chosen based on the sequence similarity with its template (SMTL ID: 6m71.1). Molecular docking studies were conducted through HDock web-server, which is freely accessible from <http://hdock.phys.hust.edu.cn/> [39]. The receptor-ligand contacts were identified through UCSF Chimera (<https://www.cgl.ucsf.edu/chimera/>) and online tool Protein-Ligand Interaction Profiler (<https://projects.biotec.tu-dresden.de/plip-web/plip/index>) [40, 41].

3. Results

3.1 Phylogenetic analysis

Viral genomes can be differentiated from other replicons based on their mechanisms of survival strategies. A viral genome encodes all the necessary informations to maintain and fulfill its infectious stages within its host and they have tripartite survival strategy. Viruses also differ from other cellular organisms in terms of their disability to self-maintain itself and also to self-replicate itself [42]. In the 21st century, with the changes in globalization, pathogens had also evolved among themselves to cope up with the new environments. Moreover, the evolution of host immune responses induces selection pressures, which results in unbalanced survival abilities, even among the concurrent strains. These had

increased the scope of a new branch of science, called “phylodynamics”. Phylodynamics deals with those informations where the phylogenetic properties and the epidemic dynamics of viruses gets interconnected [43, 44].

We had conducted phylogenetic and sequence alignment analysis, to investigate the relationships between the taxa. The organisms involved in the formation of each sub-clusters within a cluster are likely to share homology among them, whereas, the differences in the branch length among the taxa within a particular cluster or among the taxa of different clusters indicate the differences in the mutation acquired during the course of evolution. The more the branch length, the greater will be the accumulation of the number of mutations acquired by the organisms. In all the cases, we found a very close relation between Bat coronavirus RaTG13 and SARS-CoV-2. The similar relationship has been found by various researchers, which leads to a solid conclusion that SARS-CoV-2 is genetically almost similar to RaTG13 (isolated from bat in Yunnan in 2013) [45, 46]. From the phylogenetic tree based on entire length of ORF1ab polyprotein, we observed that SARS-CoV-2 is homologous to Bat coronavirus RaTG13. Moreover, the branch length for Bat coronavirus RaTG13 is slightly higher than SARS-CoV-2. But, in cases of other phylogenetic trees, the branch length pattern for Bat coronavirus RaTG13 and SARS-CoV-2, was found to be in reverse order (Figure 1 (A-C)). This concludes that the number of acquired mutations is slightly more in Bat coronavirus RaTG13 than in SARS-CoV-2 along the length of ORF1ab polyprotein.

Multiple sequence analysis result showed differences in amino acids occurred at 12 positions within the sequence of ORF1ab (corresponds to RdRp). At site 4495 (corresponding to 90th amino acid of RdRp), valine is the most frequently used amino acid. But in some cases, valine has been replaced by Leucine (Bat_coronavirus_RaTG13/1-7095 and SARS-CoV-2) and Isoleucine in some cases. The substitution in both the cases is within the same amino acid group, i.e., neutral and non-polar. The site 4497 (corresponds to 92nd amino acid residue of RdRp) is almost occupied by Glutamate (negatively charged, polar & hydrophilic), except for Bat_SARS-like_coronavirus/1-7092, where the glutamate is replaced by glycine (non-polar, aliphatic). Moreover, in cases of Bat_coronavirus_RaTG13/1-7095 and SARS-CoV-2, glutamate is replaced by aspartate. Both glutamate and aspartate belong to the same group. In some cases, we have also found the substitution of glutamate by asparagine, a neutral, non-polar amino acid. At sites 4560 (which corresponds to 155th amino acid of RdRp) and 4576 (which corresponds to 171th amino acid of RdRp), the site is mostly occupied by aspartic acid (a negatively charged, hydrophilic amino acid) and isoleucine (non-polar) respectively, with some exceptions. The site 4594 (corresponds to 184th amino acid of RdRp), is occupied by glutamine (polar, uncharged), with some exceptions in some Bat_SARS_coronaviruses, where it is replaced by a charged and polar amino acid residue arginine. Sites 4564 (corresponds to 249th amino acid of RdRp) and 4671 (corresponds to 266th amino acid of RdRp) are mostly occupied by isoleucine (non-polar, aliphatic), with some exceptions. Sites 4698 (corresponds to 293rd amino acid of RdRp), 5016 (corresponds to 611th amino acid of RdRp), 5042 (corresponds to 637th amino acid of RdRp), 5048 (corresponds to 643rd amino acid of RdRp) and 5212 (corresponds to 804th amino acid of RdRp) are mostly occupied by Threonine, Threonine, Valine, Serine and Lysine, respectively with some exceptions. The detailed analysis of these mutated sites is shown in Figure 1(D).

Furthermore, out of the twenty-five mutations observed along the length of RdRp, seventeen of these mutations are observed outside the polymerase domain of RdRp and fall within the NiRAN domain. For all of these seventeen mutations, SARS-CoV2 and Bat coronavirus RaTG13 shows occurrence of identical amino acids at all the equivalent positions except for position 4603 and 4654 of ORF1ab polyprotein, which corresponds to position 198 and 249 of RdRp respectively and these two organisms shows striking amino acid changes from all of the other taxa. In the position of 198 of RdRp, a change from Asp-Asn is observed from Bat coronavirus RaTG13 to that of the SARS-CoV2. Also, in the position of 249, another amino acid change from Ile to Arg is observed from Bat coronavirus RaTG13 to SARS-CoV2 (Figure 1(B-D)). Moreover, all of the mutable positions within the polymerase domain of RdRp exhibits the presence of identical amino acids at equivalent positions for these two organisms. Also, by comparing all of the twenty-five positions between SARS-CoV2 to other SARS-CoVs (SARS coronavirus Shanghai QXC2, SARS coronavirus TJF, SARS coronavirus HKU and SARS coronavirus Urbani) it is observed that, SARS-CoV2 shows amino acids similarities with the SARS counterparts more frequently within the conserved NiRAN domain at the positions of Asp4497(corresponds to position 92 of RdRp), Asp4560(corresponds to position 155 of RdRp), Ile4576(corresponds to position 171 of RdRp), Gln4589(corresponds to position 184 of RdRp), Ile4671(corresponds to position 266 of RdRp), Thr4698(corresponds to position 293 of RdRp). On the other hand, the occurrence of the identical amino acids in the equivalent portions between SARS-CoV2 and the SARS counterparts within the variable polymerase domain is two out of all of the eight amino acid substitutions that is being observed within the polymerase domain. They are, Val5042 (corresponds to position 637 of RdRp) and Lys5212(corresponds to position 807 of RdRp).

The variability of the amino acid characters within the polymerase domain between SARS-CoV2 and the SARS counterparts may be the reason for SARS-CoV2 RdRp efficiency.

3.2 Molecular docking studies

Based on our analysis of several already published reports, we had chosen the aim of our study. In order to fulfill the aim, we had conducted multiple molecular docking studies with SARS-CoV-2-nsp12. The SARS-CoV-2-nsp12 polymerase activity is activated by nsp7 and nsp8. To maintain similar situation as *in vivo*, we had conducted molecular docking to investigate the RdRp activity of SARS-CoV-2-nsp12 present in complexed state with nsp7-8 hexadecamer. Our result showed that the RNA template binds to the nsp7-8 complex with more affinity rather than binding to SARS-CoV-2-nsp12 within the SARS-CoV-2-nsp7-8-12 complex. Previous study reports showed that nsp7 takes part in polymerase activity and nsp8 also possesses non-canonical RdRp activity [47]. These leads to a conclusion that nsp7 and nsp8 within this complex must carry an RNA binding domain. The NTP entrance channel within the nsp12 (formed by the hydrophilic residues like Lys545, Arg553, Arg555 of Motif F), facilitates the entry of the incoming NTPs [48]. After the initial binding of the template or parental RNA with nsp7, the RNA is expected to mediate its entry into the active site of nsp12 polymerase domain (formed by Motif A & Motif C) and synthesis of new RNA strand takes place [28]. Moreover, nsp7-8 complex has been reported to interact with nsp12, at the sites of Thr409, Lys411, Trp509, Gly510, Gly897, Met899 of nsp12, which fall within the polymerase domain of the later. Therefore, it may also be concluded that nsp7-8 complex also occupies the nsp12

polymerase domain, as a result of which the viral polymerase domain becomes unavailable to many drugs. The binding affinity between only nsp12 to RNA is slightly greater (-328.84 with RMSD value of 133.19) than the binding affinity for nsp7-8-12 complex with the same RNA template (-317.09 with RMSD value of 83.68), but lesser RMSD value indicates that nsp7-8-12 complex has better confirmation to bind with the template RNA.

Subsequent molecular docking studies were conducted to investigate interactions between nsp12 and the chosen drugs (that might act as potential inhibitor of nsp12). These studies were carried out to examine the alterations of binding between nsp12 and RNA template by introducing the drug to nsp12 prior to its RNA binding. Our results showed that all the drugs eventually bind to the residues of nsp12 that fall under the polymerase domain of nsp12. The drug IDX184 forms hydrogen bonds with the residues Thr817, Leu819, Tyr831 and His872 of nsp12 polymerase domain (as indicated by Chimera and PLIP) and hydrophobic interactions with the residues Lys807, Tyr831 and pie stacking with residues Tyr831 as indicated by PLIP. Ribavirin forms multiple hydrogen bonds with Tyr530 and Val535 of nsp12 polymerase domain as indicated by Chimera and PLIP but no pie stacking or salt bridge interactions (although numerous hydrogen bonds are also found outside the polymerase domain of nsp12) and Sofosbuvir forms hydrogen bonds with Arg914 and Tyr915 of nsp12 polymerase domain as indicated by Chimera and PLIP. Also, additional hydrogen bonds are found between Sofosbuvir with Tyr915 and Glu919 of nsp12 polymerase domain as indicated by PLIP. Sofosbuvir also forms multiple hydrophobic interactions with the residues Tyr595, Tyr903 and Tyr915, pie stacking interactions with Tyr595 and salt bridge interactions with Arg914. Exceptions are for Galidesevir (hydrogen bonds are formed between the drug molecule and nsp12 with the residues Asn52, Arg116, Lys121, Tyr217 & Asp218 of nsp12 as indicated by PLIP, besides, the pie-stacking interactions are observed between Galidesevir and nsp12 with the residue Asp217 of nsp12 as indicated by PLIP and no salt bridge interaction is found in this case) and Tenofovir (hydrogen bonds are formed with residues Thr120, Thr123 of nsp12 as indicated by both Chimera and PLIP, and Cys53, Cys54 of nsp12 as indicated by Chimera). We had also got indications from Chimera software results that upon introduction of these chosen drugs (except for Setrobuvir) to nsp12, even though the binding affinity values between nsp12 and RNA template get lowered to some negligible extent, but still the template can form multiple hydrogen bonds with different amino acid sites within the nsp12 polymerase domain (Lys500, Ser501, Arg569, Lys577, Tyr689, Tyr689 and Tyr903). In case of Setrobuvir, the binding affinity between nsp12 and template RNA also get lowered (to a negligible extent), but unlike others in this case, the template RNA forms hydrogen bonds with those amino acid sites of nsp12, which are mostly positioned outside its polymerase domain (Tyr38, Asp40, Lys41, Thr76, Ser78, Asn79, Gly220 and Asp221). Molecular docking result also shows that when Sofosbuvir binds to nsp12, it increases the binding affinity of nsp12 towards RNA template. Moreover, the template RNA can still establish multiple contacts with the polymerase domain of nsp12. This can conclude preliminarily that Sofosbuvir can no longer be treated as a potential inhibitor of nsp12.

Keeping in mind about the *in vivo* situation, we had also conducted another series of molecular docking experiment with our drugs of interest to investigate that whether these drugs confer the RNA binding ability of nsp7-8-12 complex upon binding with the complex. Our results clearly indicated that whenever

the nsp7-8-12 forms complex, the polymerase domain of nsp12 is protected by the cofactor and as a result, the drugs are unable to get the access of nsp12 polymerase domain. Additionally, the template RNA is still able to bind with nsp7 and nsp8 chain within the nsp7-8-12 complex through the formation of hydrogen bonds with the residues of Lys27, Arg21, Glu23 of nsp7 and Arg75, Gln73 of nsp8 as indicated by Chimera. PLIP also showed that template RNA forms hydrogen bonds with Arg21, Ser26 of nsp7 and Gln73, Arg75, Arg80 of nsp8 within the nsp7-8-12 complex. None of these drugs inhibits the binding affinity of template RNA to nsp7 and nsp8 chain within the nsp7-8-12 complex. Meanwhile, when sofosbuvir is docked against nsp7-8-12 complex and then RNA is introduced for docking against nsp7-8-12 to Sofosbuvir complex, it was observed that Sofosbuvir increased the binding affinity of nsp7-8-12 complex for RNA template, which is indicating that this drug can no longer be treated as the inhibitor of nsp12.

Lastly, we have performed a final series of docking studies with nsp12 complexed with each of these eight drugs individually, which is then allowed to interact with nsp7-8 and ultimately docked against template RNA. We have found that the RNA template still can be able to bind with nsp7 and nsp8 chains of nsp7-8-12 complex, although the binding sites may vary for each of the selected drugs. Besides these, one contact is also been made by the template RNA at Ala406 site of polymerase domain of nsp12, upon introducing Remdisivir with nsp12. Also, template RNA is able to bind with the polymerase domain of nsp12 through the formation of hydrogen bonds with the residues of Thr402, Asn403, Gln408 and Gly670 upon the introduction of Remdisivir to nsp12, which is then complexed with nsp7-8 prior to the RNA binding, which brings about the polymerization by nsp12. Also, upon the introduction of Sofosbuvir to nsp12 before the formation of nsp7-8-12 complex, it is observed that, RNA template forms hydrogen bonds with nsp12 complexed with Sofosbuvir at the residues Asp499, Lys500, Ser501, Tyr521, Arg569, Lys577, Asp684, Tyr689, Ser814, Arg836 and Tyr903 of nsp12 polymerase domain, as indicated by Chimera and PLIP and salt bridge interactions are also observed between RNA template and nsp12 at the site of Arg513, Lys593 and Arg836 of nsp12. As a result, these drugs do not seem to alter the binding affinity of template RNA upon binding with nsp12 when the latter forms complex with its cofactor nsp7-8 complex. The results of the docking experiments are shown in the figures 2-4 and supplementary table 1a-f.

4. Discussion

4.1 Phylogenetic analysis:

From the phylogeny analysis on the basis of complete length of ORF1ab or only the mutable sites, it is clear that SARS-Cov-2 is evolutionary close to Bat coronavirus RaTG13. The branch length of the phylogenetic tree based on full length of ORF1ab depicts that the total number of acquired mutations in Bat coronavirus RaTG13 is little bit higher than SARS-CoV2 and the occurrence of identical amino acids in the equivalent mutable positions along the entire length of ORF1ab between Bat coronavirus RaTG13 and SARS-CoV2 reflects evolutionary relatedness among these two taxa.

Upon considering the twenty five selected mutable positions along the length of ORF1ab corresponding to the viral RdRp, we have found that the total number of mutations acquired is highest in SARS-CoV2, although the evolutionary relationship between Bat coronavirus RaTG13 and SARS-CoV2 remains the same and out of these mutations observed along the length of RdRp, and amino acid changes within the NiRAN domain(1-397 amino acid) of RdRp shows that all the equivalent positions are occupied by identical amino acids between SARS-CoV2 and Bat coronavirus RaTG13 with only two exceptions within ORF1ab polyprotein corresponding to positions 198 and 249 of RdRp respectively. A change from Asp-Asn in the position of 198 of RdRp in SARS-CoV2 will have the potential to form hydrogen bonds with it's ligand, while the change from Ile-Arg in the position of 249 of SARS-CoV2 RdRp, will have the potential of forming salt bridge interactions more efficiently when it binds to any ligand. Besides, it is also observed that the equivalent positions along the length of ORF1ab corresponding to the polymerase domain of RdRp shows the presence of the identical amino acids between SARS-CoV2 and the other SARS counterparts in only two positions, at the positions of Val637 and Lys807 of RdRp respectively. Furthermore, the changes from Asp-Asn residue in SARS-CoV2 from that of the SARS counterparts at the position of 5148 of ORF1ab(corresponding to the position 743 of RdRp) reflects that Asn residue will more efficiently form hydrogen bonds upon the interaction of SARS-CoV2 RdRp with any ligand. These notable amino acid changes is the main reason of the evolutionary divergence among SARS-CoV2 and the SARS counterparts and the greater number of acquired mutations in SARS-CoV2 RdRp, particularly in the polymerase domain may bring about the more efficient RNA polymerase activity in SARS-CoV2 RdRp.

4.2 Molecular docking studies:

Series of molecular docking experiment clearly indicates that whenever the nsp7-8-12 forms complex, the polymerase domain of nsp12 is protected by the cofactor and as a result, the drugs are unable to get the access of nsp12 polymerase domain. Additionally, the template RNA is still able to bind with nsp7 and nsp8 chain within the nsp7-8-12 complex through the formation of hydrogen bonds with the residues of Lys27, Arg21, Glu23 of nsp7 and Arg75, Gln73 of nsp8 as indicated by Chimera (Figure 2-4). PLIP also showed that template RNA forms hydrogen bonds with Arg21, Ser26 of nsp7 and Gln73, Arg75, Arg80 of nsp8 within the nsp7-8-12 complex (Supplementary Table 1a-1f). None of these drugs inhibits the binding affinity of template RNA to nsp7 and nsp8 chain with in the nsp7-8-12 complex. Meanwhile, when sofosbuvir is docked against nsp7-8-12 complex and then RNA is introduced for docking against nsp12-7-8 to Sofosbuvir complex, it was observed that Sofosbuvir increased the binding affinity of nsp12-7-8 complex for RNA template, which is indicating that this drug can no longer be treated as the inhibitor of nsp12.

5. Conclusion

From the above systematic studies it's clear that RdRp of SARS-CoV-2 is highly mutable in comparison to its homologous sequences and evolutionary close to Bat coronavirus RaTG13, which might also be the study of interest for drug design. Furthermore, we would like to conclude that the drugs those are already been treated as the potential inhibitors of nsp12 do not lower the binding affinity of RNA for nsp12 to a

significant level. These drugs seem to inhibit the nsp12 partially, that is also when the nsp12 does not form any complex with its cofactors. Therefore, designing of inhibitors which will specifically act against nsp12 and that will have the ability to lower the binding affinity of RNA template for nsp12 is necessary. Besides, specific drug molecules have to be designed against nsp7–8 hexadecameric complex with the ability of binding to the residues of nsp7–8 complex, which in turn binds with nsp12. We have found that Asn69 of nsp7 interacts with Gly897 of nsp12, Cys72 of nsp7 interacts with Gly510 of nsp12, Glu73 of nsp7 interacts with the Arg513 of nsp12. Also, Arg96, Arg111, Asn136 and Met174 of nsp8 interacts with the Trp509, Leu900, Lys411 and Thr409 of nsp12 respectively during the formation of nsp7–8–12 complex. So, by blocking these sites of nsp7–8 complex by designing of specific drug molecules which will interact with the sites of nsp7–8 mentioned above will result in the alteration of binding affinity of nsp7–8 complex to nsp12. Also by designing specific drugs, which will eventually bind with the sites of nsp7–8–12 complex that is being found responsible for binding with RNA (Gln69 of nsp8, Trp29 of nsp7 and Arg21 of nsp7) may also provide fruitful result in blocking of the polymerase activity of nsp7–8–12 complex.

Declarations

Compliance with Ethical Standards: This article does not contain any studies with animals performed by any of the authors.

Acknowledgements

The authors would like to acknowledge the combined efforts of all the scientists in the worldwide, who are working day and night to discover a long-term solution against this pandemic. M.K.S. acknowledges Czech University of Life Sciences, Prague for providing the necessary infrastructural support. S.K.M. acknowledges The University of Burdwan for its necessary support. Authors also acknowledge Mr. Abhishek Tewari and Renuka Das for their opinion.

Author Contributions

All the authors combinedly designed, performed and analyzed the experiments and the results. Madhab Kumar Sen wrote the initial draft of the manuscript. Sunil Kanti Mondal, Samybrata Mukhoty, Subhajit Ghosh, Himangsu Kundu and Suvankar Das proof-read the draft the manuscript and finalized it.

References

1. Bryson-Cahn C, Duchin J, Makarewicz VA, Kay M, Rietberg K, Napolitano N, Lynch, JB. A Novel Approach for a Novel Pathogen: using a home assessment team to evaluate patients for 2019 novel coronavirus (SARS-CoV-2). *Clinical Infectious Diseases* 2020; doi:10.1093/cid/ciaa256

2. Contini C, Di Nuzzo M, Barp N, Bonazza A, De Giorgio R, Tognon M, Rubino S. The novel zoonotic COVID-19 pandemic: An expected global health concern. *The Journal of Infection in Developing Countries* 2020; doi:10.3855/jidc.12671
3. Lauer SA, Grantz KH, Bi Q, Jones FK, Zheng Q, Meredith HR, Lessler J. The incubation period of coronavirus disease 2019 (COVID-19) from publicly reported confirmed cases: estimation and application. *Annals of internal medicine* 2020; doi:10.7326/M20-0504
4. Parvin F, Islam S, Urmay Z, Ahmed S. THE SYMPTOMS, CONTAGIOUS PROCESS, PREVENTION AND POST TREATMENT OF COVID-19. *European Journal of Physiotherapy and Rehabilitation Studies* 2020; doi: 10.5281/zenodo.3779252
5. Zu ZY, Jiang MD, Xu PP, Chen W, Ni QQ, Lu GM, Zhang LJ. Coronavirus disease 2019 (COVID-19): a perspective from China. *Radiology* 2020; doi:10.1148/radiol.2020200490
6. Sharma A, Tiwari S, Deb M.K, MARTY J.L. Severe Acute Respiratory Syndrome Coronavirus-2 SARS-CoV-2) : A global pandemic and treatments strategies. *International Journal of Antimicrobial Agents* 2020; doi: <https://doi.org/10.1016/j.ijantimicag.2020.106054>.
7. Ou, X, Liu Y, Lei X, Li P, Mi D, Ren L, Xiang Z. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. *Nature communications* 2020; doi:10.1038/s41467-020-15562-9
8. Kadam S, Sukhramani G, Bishnoi P, Pable A, Barvkar, V. Molecular and Structural Insights into COVID-19 Pandemic. *Preprints* 2020; doi:10.20944/preprints202005.0519.v1
9. Tsiambas E, Papanikolaou V, Chrysovergis A, Mastronikolis N, Ragos V, Kavantzias N, Kyrodimos E. Coronavirus in Hematologic Malignancies: Targeting Molecules Beyond the Angiotensin-Converting Enzyme 2 (ACE2) Wall in COVID-19. *Pathology & Oncology Research* 2020; doi:10.1007/s12253-020-00810-6
10. Li X, Geng M, Peng Y, Meng L, Lu S. Molecular immune pathogenesis and diagnosis of COVID-19. *Journal of Pharmaceutical Analysis* 2020; doi:10.1016/j.jpha.2020.03.001
11. Ullah MA, Araf Y, Sarkar B, Moin AT, Reshad RAI, Rahman, MH. Pathogenesis, Diagnosis and Possible Therapeutic Options for COVID-19. *Preprints* 2020; doi: 10.20944/preprints202004.0372.v1.
12. Ye ZW, Yuan S, Yuen KS, Fung SY, Chan CP, Jin DY. Zoonotic origins of human coronaviruses. *International journal of biological sciences* 2020; doi:10.7150/ijbs.45472
13. Perlman S, Netland J. Coronaviruses post-SARS: update on replication and pathogenesis. *Nature reviews microbiology* 2009; doi:10.1038/nrmicro2147
14. Walls AC, Park YJ, Tortorici MA, Wall A, McGuire AT, Veesler D. Structure, function, and antigenicity of the SARS-CoV-2 spike glycoprotein. *Cell* 2020; doi:10.1016/j.cell.2020.02.058
15. Fehr AR, Perlman S. Coronaviruses: an overview of their replication and pathogenesis. In *Coronaviruses*; Humana Press, New York, NY 2015;. pp1-23.
16. Robertson MP, Igel H, Baertsch R, Haussler D, Ares Jr M, Scott WG. The structure of a rigorously conserved RNA element within the SARS virus genome. *PLoS Bio* 2004; doi:10.1371/journal.pbio.0030005

17. Coutard B, Valle C, de Lamballerie X, Canard B, Seidah NG, Decroly E. The spike glycoprotein of the new coronavirus 2019-nCoV contains a furin-like cleavage site absent in CoV of the same clade. *Antiviral research*2020; doi:10.1016/j.antiviral.2020.104742
18. Yan R, Zhang Y, L, Y, Xia L, Guo Y, Zhou Q. Structural basis for the recognition of SARS-CoV-2 by full-length human ACE2. *Science*2020; doi:10.1126/science.abb2762
19. Zhou P, Yang XL, Wang XG, Hu B, Zhang L, Zhang W, Chen HD. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature*2020; doi:10.1038/s41586-020-2012-7
20. Domingo E. Molecular basis of genetic variation of viruses: error-prone replication. *Virus as Populations* 2020; doi:10.1016/B978-0-12-816331-3.00002-7
21. Lai MMC. Genetic recombination in RNA viruses. In *Genetic Diversity of RNA Viruses*pp21-32;Springer, Berlin, Heidelberg 1992; doi:https://doi.org/10.1007/978-3-642-77011-1_2
22. Ledinko N. Genetic recombination with poliovirus type 1: studies of crosses between a normal horse serum-resistant mutant and several guanidine-resistant mutants of the same *Virology*1963; doi:10.1016/0042-6822(63)90145-4
23. Lai MM, Baric RS, Makino S, Keck JG, Egbert J, Leibowitz JL, Stohlman S.A. Recombination between non-segmented RNA genomes of murine coronaviruses. *Journal of Virology* 1985;56(2):449-456.
24. Bujarski JJ, Kaesberg P. Genetic recombination between RNA components of a multipartite plant virus. *Nature*, 1986; doi:10.1038/321528a0
25. Allison R, Thompson C, Ahlquist P. Regeneration of a functional RNA virus genome by recombination between deletion mutants and requirement for cowpea chlorotic mottle virus 3a and coat genes for systemic infection. *Proceedings of the National Academy of Sciences*1990; doi:10.1073/pnas.87.5.1820
26. Abdul Kadhim AH, Hadi NR, Abdul hussein M. Preprocessing of the Candidate Antiviral Drugs against COVID-19 in Models of SARS cov2 Targets. *Prensa Med Argent* 2020; 106, 2.
27. Lehmann KC, Gulyaeva A, Zevenhoven-Dobbe JC, Janssen GM, Ruben M, Overkleeft HS, Sidorov, IA. Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses. *Nucleic acids research*2015; doi:10.1093/nar/gkv838
28. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Ge J. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. *Science* 2020; doi:10.1126/science.abb7498
29. Huang C, Wang Y, Li X, Ren L, Zhao J, Hu Y, Cheng Z. Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. *The lancet*2020; doi:10.1016/S0140-6736(20)30183-5
30. Lin L, Lu L, Cao W, Li T. Hypothesis for potential pathogenesis of SARS-CoV-2 infection—a review of immune changes in patients with viral pneumonia. *Emerging microbes & infections* 2020; doi:10.1080/22221751.2020.1746199
31. Zhang H, Penninger JM, Li Y, Zhong N, Slutsky AS. Angiotensin-converting enzyme 2 (ACE2) as a SARS-CoV-2 receptor: molecular mechanisms and potential therapeutic target. *Intensive care medicine* 2020; doi:10.1007/s00134-020-05985-9

32. Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman, DJ. Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. *Nucleic acids research* 1997; doi:10.1093/nar/25.17.3389
33. Clamp M, Cuff J, Searle SM, Barton GJ. The jalview java alignment editor. *Bioinformatics* 2004; doi:10.1093/bioinformatics/btg430
34. Sievers F, Higgins DG. Clustal omega. *Current protocols in bioinformatics* 2014; doi:10.1002/0471250953.bi0313s48
35. Hall BG. Building phylogenetic trees from molecular data with MEGA. *Molecular biology and evolution* 2013; doi:10.1093/molbev/mst012
36. Kumar S, Stecher G, Li M, Knyaz C, Tamura, K. MEGA X: molecular evolutionary genetics analysis across computing platforms. *Molecular biology and evolution* 2018; doi:10.1093/molbev/msy096
37. Elfiky AA. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase RdRp): A molecular docking study. *Life sciences* 2020; doi:10.1016/j.lfs.2020.117592
38. Elfiky AA. SARS-CoV-2 RNA dependent RNA polymerase RdRp) targeting: An in-silico perspective. *Journal of Biomolecular Structure and Dynamics* 2020; doi:10.1080/07391102.2020.1761882
39. Yan Y, Zhang D, Zhou P, Li B, Huang, SY. HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic acids research* 2017;45(W1):W365-W373.
40. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE. UCSF Chimera—a visualization system for exploratory research and analysis. *Journal of computational chemistry* 2004; doi:10.1002/jcc.20084
41. Salentin S, Schreiber S, Haupt VJ, Adasme MF, Schroeder, M. PLIP: fully automated protein–ligand interaction profiler. *Nucleic acids research* 2015; doi:10.1093/nar/gkv315
42. Ludmir EB, Enquist LW. Viral genomes are part of the phylogenetic tree of life. *Nature Reviews Microbiology* 2009; doi:10.1038/nrmicro2108-c4
43. Rife BD, Mavian C, Chen X, Ciccozzi M, Salemi M, Min J, Prosperi MC. Phylodynamic applications in 21st century global infectious disease research. *Global Health Research and Policy* 2017; doi:10.1186/s41256-017-0034-y
44. Chen X. Understanding the development and perception of global health for more effective student education. *The Yale journal of biology and medicine* 2014;87(3): 231.
45. Kovarik A, Matyasek R. Mutation patterns of human SARS-COV-2 and bat RaTG13 coronaviruses genomes are strongly biased towards C > U indicating rapid evolution in their hosts. *Molecular Genetics* 2020; doi: [21203/rs.3.rs-21377/v1](https://doi.org/10.1038/s41525-020-01377-1)
46. Tang X, Wu C, Li X, Song Y, Yao X, Wu X, Cui J. On the origin and continuing evolution of SARS-CoV-2. *National Science Review* 2020; doi:10.1093/nsr/nwaa036
47. TeVelthuis AJ, van den Worm SH, Snijder EJ. The SARS-coronavirus nsp7+ nsp8 complex is a unique multimeric RNA polymerase capable of both de novo initiation and primer extension. *Nucleic acids*

48. Barakat K, Ahmed M, Tabana Y, Ha M. A Deep Dive into the SARS-Cov-2 Polymerase Assembly: Identifying Novel Allosteric Sites and Analyzing the Hydrogen Bond Networks and Correlated Dynamics. *bioRxiv* 2020; doi: 10.1101/2020.06.02.130849.

Figures

Figure 1

Phylogenetic tree analysis. Informative maximum-likelihood phylogenetic trees with 1000 bootstrap values considering (A) entire ORF1ab sequences, (B) 12 most mutable positions within RdRp, (C) observed 25 mutations (including 12 most mutable positions/sites: 3,4,6,7,8, 13,15, 17, 18, 19, 20, 25 within JalView) within RdRp of SARS-CoV-2 and (D) JalView representation of multiple sequence alignment of ORF1ab for the 25 selective mutable positions within RdRp.

Figure 2

Molecular interaction (Visualization through UCSF Chimera) between (A) nsp12 and RNA and (B) nsp7-8-12 complex and RNA.

Figure 3

Molecular docking between nsp7-8-12 complex and RNA, considering that nsp12 first interact to drug in the successive interactions.

Figure 4

Binding energy values obtained from HDock for different possible interactions among nsp12, nsp7-8 hexa-decameric complex, drug and RNA. More negative values indicate more efficient interaction.

Supplementary Files

This is a list of supplementary files associated with this preprint. Click to download.

- [Supplementaryfiles.docx](#)