RESEARCH

Analysis of SARS-CoV-2 RNA-dependent RNA polymerase as a potential therapeutic drug target using a computational approach

Syed Ovais Aftab[1,2†], Muhammad Zubair Ghouri[1,2*†], Muhammad Umer Masood[1], Zeshan Haider[1], Zulqurnain Khan[3†], Aftab Ahmad[2,4*] and Nayla Munawar[5]

Abstract

Background: The Severe acute respiratory syndrome-related coronavirus 2 (SARS-CoV-2) outbreak originating in Wuhan, China, has raised global health concerns and the pandemic has now been reported on all inhabited continents. Hitherto, no antiviral drug is available to combat this viral outbreak.

Methods: Keeping in mind the urgency of the situation, the current study was designed to devise new strategies for drug discovery and/or repositioning against SARS-CoV-2. In the current study, RNA-dependent RNA polymerase (RdRp), which regulates viral replication, is proposed as a potential therapeutic target to inhibit viral infection.

Results: Evolutionary studies of whole-genome sequences of SARS-CoV-2 represent high similarity (> 90%) with other SARS viruses. Targeting the RdRp active sites, ASP760 and ASP761, by antiviral drugs could be a potential therapeutic option for inhibition of coronavirus RdRp, and thus viral replication. Target-based virtual screening and molecular docking results show that the antiviral Galidesivir and its structurally similar compounds have shown promise against SARS-CoV-2.

Conclusions: The anti-polymerase drugs predicted here—CID123624208 and CID11687749—may be considered for in vitro and in vivo clinical trials.

Keywords: RdRp, SARS-CoV-2, Phylogenetic tree, Homology modeling, Molecular Docking, Active site

Background

Severe acute respiratory syndrome coronavirus (SARS-CoV) is a positive-sense single-stranded RNA virus from the genus Betacoronavirus, commonly known to infect bats, humans, and other mammals [1–4]. On January 30th, 2020, the Director-General of the World Health Organization (WHO) declared that the outbreak of novel coronavirus (2019-nCoV) constitutes a Public Health Emergency of International Concern (PHEIC). By April 10th, 2020, the current pandemic caused by the 2019-nCoV has reached nearly all the world's countries and has consisted of more than 1.5 million confirmed cases with more than 92,000 deaths [5]. To date, two SARS strains have been reported to cause epidemics: (1) SARS-CoV, identified in 2002–2004, and (2) SARS-CoV-2, also known as the novel coronavirus that emerged as a potential threat in late 2019 [6]. Both these strains evolved from a common Betacoronavirus ancestor; however, it is believed that SARS-CoV-2 first infected humans from a bat host during interspecies viral transmission. In support, it has been reported in China and other countries that bats are the primary reservoirs of SARS-CoV-2 [6–8]. Coronaviruses are a

---

*Correspondence: zubair.ghouri@uaf.edu.pk; aftab.ahmad@uaf.edu.pk
†Muhammad Zubair Ghouri, Syed Ovais Aftab, Muhammad Umer Masood, and Zulqurnain Khan contributed equally to this work
[2] Center for Advanced Studies in Agriculture and Food Security (CAS-AFS), University of Agriculture, Faisalabad, Pakistan
Full list of author information is available at the end of the article

Aftab *et al. J Transl Med*   (2020) 18:275

Page 2 of 15

large family of viruses reported to cause illnesses ranging from the common cold to severe diseases such as Middle East respiratory syndrome (MERS) and severe acute respiratory syndrome (SARS). SARS-CoV-2 is one of the seven coronaviruses known to cause infection in *Homo sapiens*, which also includes: 229E (HCoV-229E), NL63 (HCoV-NL63), OC43 (HCoV-OC43), HKU1 (HCoV-HKU1), MERS-CoV, and the original SARS-CoV [9–12].

The novel coronavirus is the most relevant virus of the family Coronaviridae in terms of research currently, as there remains no approved antiviral drug or vaccine against it [13]. Recently, on the basis of genomic resemblance with previously reported SARS-CoV, the International Committee on Taxonomy of Viruses (ICTV) coronavirus study group named this virus "SARS-CoV-2" [13]. It has been confirmed that SARS-CoV-2 can spread with human-to-human transmission via respiratory droplets (e.g. through coughing or sneezing) or even by contact with contaminated surfaces [14, 15]. A coronavirus epidemic was previously predicted by the WHO soon after the Ebola virus outbreak in 2016 [16, 17]; and this prediction came to fruition in the Wuhan city seafood market with the coronavirus epidemic of 2019–2020 [18–20]. Therefore, scientists are attempting to use preexisting antiviral drugs to control the virus upsurge, however, these drugs have thus far had inappreciable effects on SARS-CoV-2 [21, 22]. Efficacy of such antiviral drugs may be compromised due to changes induced by single nucleotide polymorphisms (SNPs), thereby resulting in amino acid shifts which ultimately modify functional viral protein(s). This could be the case for SARS-CoV-2, in that viral proteins are actively acquiring mutations due to SNPs and thus escape from being targeted by antiviral drugs [13, 23, 24].

Genome organization of all coronaviruses are similar and contain 5′ and 3′ untranslated regions (UTR's) for characteristic genes coding for ORF1ab, spike, envelope, membrane, and nucleocapsid proteins [24]. ORF1ab is of particular importance, as it occupies two-thirds of the CoV genome and encodes a replicase polyprotein from ORF1a and ORF1b. In addition, a slippery sequence (UUUAAAC) is present at the junction between ORF1a and ORF1b in all coronaviruses, with translation commencing at the end of slippery sequence via a − 1 RNA-mediated ribosome frame shift [25–27]. Papain-like protease (PLpro) and 3C-like protease (3CLpro) are proteins encoded by ORF1ab and cleave the replicase polyprotein into 15–16 non-structural proteins (nsps) at consensus cleavage sites. Some of these nsps encode proteins with essential functions, such as PLpro (nsp3), 3CLpro (nsp5) and RdRP (nsp12) which plays an important role in viral replication, whereas helicase (nsp13) has been recognized to unwind duplex oligonucleotides in an NTP-dependent manner [26–28].

Most RNA viruses—except for retroviruses—require an RdRp for replication and transcription of the viral genome, making it essential for their survival [29, 30]. The RdRp protein ranges from 240 to 450 kD and consists of a catalytic core with a clear resemblance to the human right hand with differentiated palm, fingers and thumb domains [30, 31]. RdRp is considered to be a conserved protein within RNA viruses, and thus could be used as an attractive candidate for understanding their biology in terms of nucleic acid synthesis and development of antiviral drugs [32–39]. RdRp plays a crucial role in the viral life cycle, and as the active site of the RdRp is the most conserved and accessible region, targeting this region for inhibition of viral replication may be an effective therapeutic approach. Scientists worldwide have proposed the use of preexisting drugs against the novel coronavirus; however, the efficacy of these drugs is somewhat limited. Others reported repurposing the use of existing antiviral agents in order to reduce time and cost compared to de novo drug discovery [40]. Ribavirin, lopinavir–ritonavir, corticosteroids, and interferon are just a few of the antiviral drugs that have been tested for use against SARS-CoV-2 [41, 42].

Soon after the COVID-19 outbreak, China initiated several in vitro studies and clinical trials to discover effective antiviral drugs. Favipiravir is the first antiviral drug that has been approved by National Medical Products Administration of China to use against SARS-CoV-2 [43]. A number of drugs—such as Sofosbuvir, Ribavirin, Galidesivir, Remdesivir, etc.—are in clinical trials for different viruses (e.g. Hepatitis C, Zika virus, Dengue and SARS-CoV-2) on the basis of anti-RdRp activity [44–51]. The core objective of the current study was to evaluate the potential of Galidesivir, similar to structurally similar antiviral compounds, to inhibit the viral replication protein RdRp. An evolutionary-based study was performed to understand the relationships of SARS-CoV-2 genome with other SARS isolates. Homology modeling for predicting SARS-CoV-2 RdRp structure was performed with a reference template of SARS-CoV RdRp. Among the eight antiviral agents examined (Ribavirin, penciclovir, nitazoxanide, nafamostat, chloroquine, Galidesivir, favipiravir, and interferon), the best suited inhibitors were screened through in silico analysis that can further be used for preclinical trials to halt viral replication after prior testing. Our results are promising and may be considered for both in vitro and in vivo clinical trials for inhibition of COVID-19.

Aftab *et al. J Transl Med*   (2020) 18:275

Page 3 of 15

Materials and methods

Sequence retrieval, alignment and evolutionary relationship of viral species

The SARS-CoV-2 whole-genome sequence was retrieved from NCBI (National Center for Biotechnology Information) under Accession # NC_045512.2. In addition, a total of 94 whole-genome viral sequences belonging only to class Nidovirales, order Coronaviradae, family Coronavirinae, and three genera (Alphacoronavirus, Betacoronavirus, and Gammacoronavirus) were downloaded from the GenBank database for sequence similarity comparison with our experimental sequence of SARS-CoV-2. The sequences were submitted to MEGA v7 in FASTA format, aligned using CLUSTALW alignment tool and maximum likelihood estimation (MLE) tree was constructed against all viral genome sequences for tracing the evolutionary record with SARS-CoV-2 using a bootstrap value of 1000.

Protein sequence retrieval and estimation of physicochemical properties

Sequence of RdRp protein for SARS-CoV-2 (YP_009725307.1) was retrieved from the NCBI protein database and contains 932 amino acids (AA). In addition, physicochemical properties—such as molecular weight, amino acid composition, atomic composition, extinction coefficient, estimated half-life, instability index, aliphatic index and grand average of hydropathicity (GRAVY)—were analyzed through ProtParam tool of ExPASy for evaluation of the primary structure of RdRp protein.

Homology modeling and protein phylogenetic analysis

RdRp protein AA from SARS-CoV and SARS-CoV-2 were aligned with the help of Clustal X 2.1 software using default parameters, and the sequences were further refined by GeneDoc to find the similarity between them. This similarity helps to select a template for modeling and further analysis. RdRp proteins of SARS-CoV [32] and novel SARS-CoV-2 were compared for their amino acid residues. Based on similarity in amino acid residues of both RdRps, a reference template of SARS-CoV RdRp (PDB ID: 6NUR; kindly provided by the Ward lab) was used for further analysis [52]. Homology modeling was performed with the MODELLER v9.1 program using 6NUR as reference and ~100 structures were predicted [53]. The best possible structure was selected based on highest discrete optimized protein energy (DOPE) score. In addition, this best suited structure was placed in a Ramachandran plot for evaluation of chemical properties, bonds, and angles of RdRp. Moreover, the RdRp protein sequence of SARS-CoV-2 was subjected to pBLAST to infer its similarity with other RdRp protein sequences; 94 total sequences were selected that exhibited > 90%

similarity with our input RdRp sequence. Likewise, these 94 validated sequences were uploaded in MEGA v7 and further aligned through the CLUSTALW alignment tool. Aligned sequences were used to construct a MLE tree for understanding the phylogenies of all the validated protein sequences along with SARS-CoV-2 RdRp.

Ligand selection and virtual screening

For docking, the following ten antiviral drugs were screened: Ribavirin, Remdesivir, Sofosbuvir, Penciclovir, Nitazoxanide, Nafamostat, Chloroquine, Galidesivir, Favipiravir, and Interferon. Docking was performed using molecular operating environment (MOE) software. Based on the phylogenetic similarity of the RdRp proteins of SARS-CoV and SARS-CoV-2, the catalytic domain was selected for inhibitor docking of inhibitors against the SARS-CoV-2 RdRP. Antiviral drugs were docked against the following catalytic sites: GLY616, TRP617, ASP618, TYR619, LEU758, SER759, ASP760, ASP761, ALA762, LYS798, TYS799, TRP800, GLU811, PHE812, CIS813, and SER814. The best antiviral drug was selected with distinct S-score, root-mean-square deviation (RMSD), and energy binding score for the catalytic domain of RdRp SARS-CoV-2. The pre-eminent ligand was screened against the PubChem database for detection of drug-like compounds, and the Pfizer rule was applied to evaluate drug-like properties. Different parameters for drug evaluation were considered, such as: molecular weight (MW) < 500, LogP < 5, H-bond doner(s) < 10, and H-bonding accepter(s) < 5. Finally, compounds fulfilling the above-mentioned requirements were used for molecular docking into a new database.

Molecular docking

All the retrieved compounds were docked using selected catalytic site of the three-dimensional structure of RdRp protein. Removal of water, 3D protonation and energy minimization were carried out with the MOE software using the following parameters: force field, MMFF94X, gradient 0.05, and current geometry [54, 55]. The binding pocket of the RdRp—GLY616, TRP617, ASP618, TYR619, LEU758, SER759, ASP760, ASP761, ALA762, LYS798, TYS799, TRP800, GLU811, PHE812, CIS813, and/or SER814—was found using the MOE SiteFinder algorithm, and most optimal docked compounds were selected based on higher S-score and lower RMSD with a reference inhibitor (PubChem CID: 10445549). Docking simulations were performed for Galidesivir, a potential non-covalent inhibitor of SARS-CoV-2 RdRp, which binds strongly to the active site of RdRp. It also targets the ASP761, ALA762, LYS798, and SER814 residues with stout hydrogen bonding and a maximum docking score. For the discovery of novel drug-like compounds, we

Aftab *et al. J Transl Med*   (2020) 18:275

Page 4 of 15
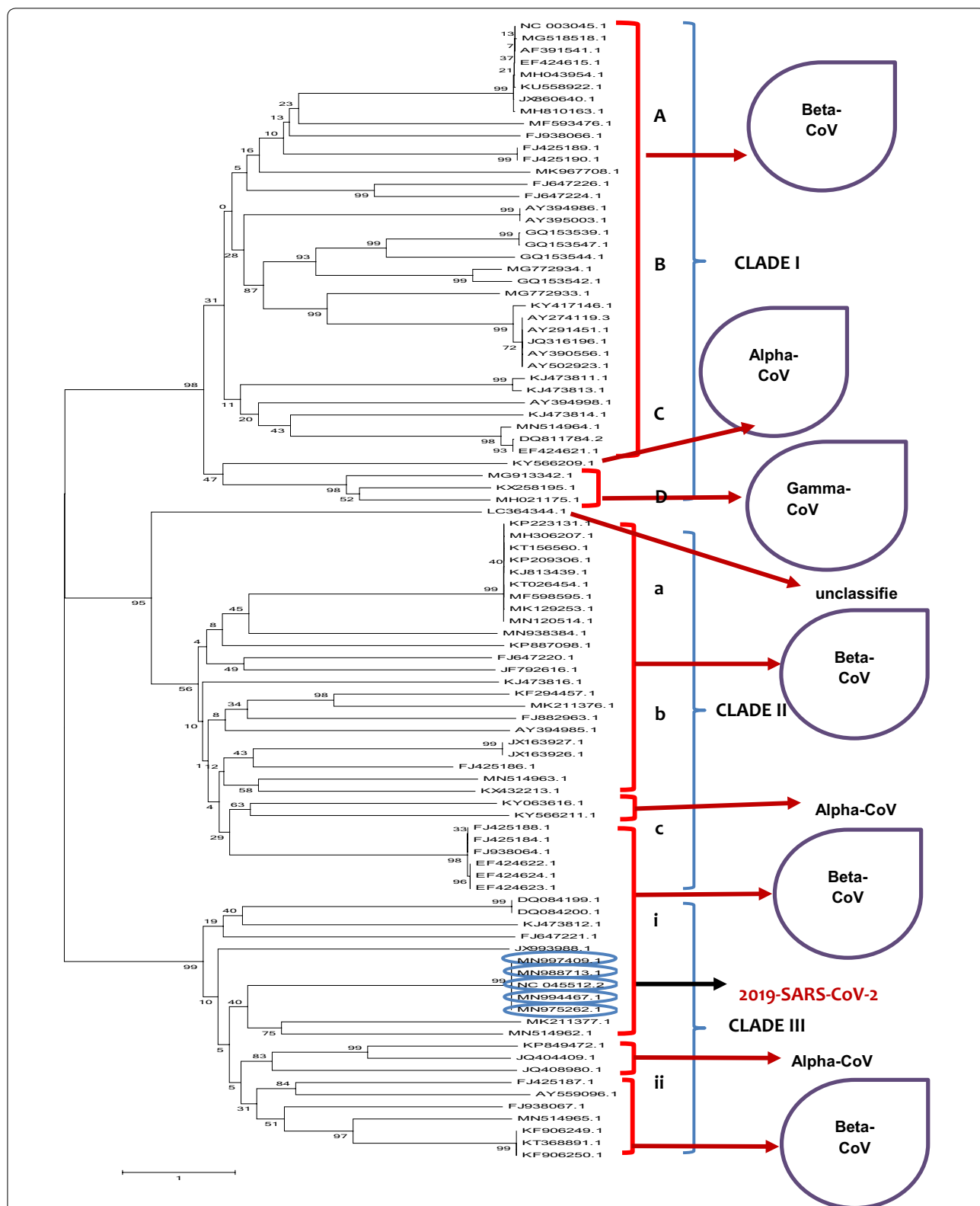
Fig. 1 Dendrogram representing the phylogeny of viruses with SARS-CoV-2. Phylogenetic tree is divided into three clades, and all clades are further divided into sub-groups. 2019-SARS-CoV-2 from clade III is closely related to its relatives from Hong Kong and USA (circled in blue). For accessions, origin and other relevant data, see Additional file 1

Aftab *et al. J Transl Med*   (2020) 18:275

Page 5 of 15

Table 1 Physicochemical properties of RdRP protein (YP_009725307.1)

Physiochemical properties	RdRP COVID-19	Amino acid composition	No.	Percent composition (%)
Molecular weight	106,660.24	Ala (A)	64	6.90
No. of amino acids	932	Arg (R)	43	4.60
Theoretical pl	6.14	Asn (N)	56	6.00
Instability index	28.32	Asp (D)	75	8.00
No. of negatively charged (Asp + Glu)	106	Cys (C)	29	3.10
No. of positively charged (Arg + Lys	94	Gln (Q)	28	3.00
aliphatic index	78.43	Glu (E)	31	3.30
Grand average of hydropathicity	−0.224	Gly (G)	45	4.80
Estimated half-life (mammalian reticulocytes, in vitro)	1.9 h	His (H)	27	2.90
Atomic composition		Ile (I)	33	3.50
C	4792	Leu (L)	83	8.90
H	7265	Lys (K)	51	5.50
N	1259	Met (M)	25	2.70
O	1401	Phe (F)	57	6.10
S	54	Pro (P)	30	3.20
Formula	$C_{4792}H_{7265}N_{1259}O_{1401}S_{54}$	Ser (S)	53	5.70
Total number of atoms	14,771	Thr (T)	61	6.50
		Trp (W)	9	1.00
		Tyr (Y)	58	6.20
		Val (V)	74	7.90
		Pyl (O)	0	0.00
		Sec (U)	0	0.00

screened compounds structurally similar to Galidesivir that may target SARS-CoV-2 RdRp as an antiviral agent. The best hits—CID123624208 and CID11687749—were selected from the docking analysis.

AdmetSAR profiling and toxicity validation
Virtually-selected drug-like compounds were used in admetSAR online tool (Immd.ecust.edu.cn/admetsar2) to evaluate their level of toxicity. This program predicts toxic effects—such as carcinogenicity, mutagenicity, irritant effect, reproductivity and drug-like physical and chemical properties of compounds—which can help for selection of compounds that can be used as safe antiviral agents on humans.

Results

Evolutionary relationship of SARS-CoV-2 with other viral isolates
A phylogenetic tree infers relative evolutionary distance with respect to other species. Tracing evolutionary history among all possible viral relatives using a genome sequence of interest can elucidate the adaptive behavior of species. In this study, we identified interspecies divergence among all the 94 validated viral genomic sequences

belonging to the family Coronavirinae retrieved from the NCBI database. Most of the species clustered together, reflecting prior taxonomic identification, with a few exceptions. Overall, the phylogenetic tree was divided into three different clades (Fig. 1). The accessions, origin and taxonomic classification of all sequences are given in Additional file 1. Most of the accessions in clade I belong to the Betacoronaviruses, indicating they all are closely related to each other, while only one and two accessions belong to alpha-CoV and gamma-CoV, respectively, thereby representing some distance with respect to the Betacoronavirus members. Clade II also included most of the beta-CoV members, with only two viral species clustered in a separate clade belong to alpha-CoV. Most importantly, clade III—which included the experimental SARS-CoV-2 isolate originating from the Wuhan seafood market—was found in close relation with all the other isolates of SARS-CoV-2 reported from Hong Kong and the USA (MN975262.1, MN997409.1, MN994467.1, MN988713.1; Fig. 1).

Estimation of physiochemical properties
The feasibility of protein structure depends upon three-dimensional conformation based on amino acid
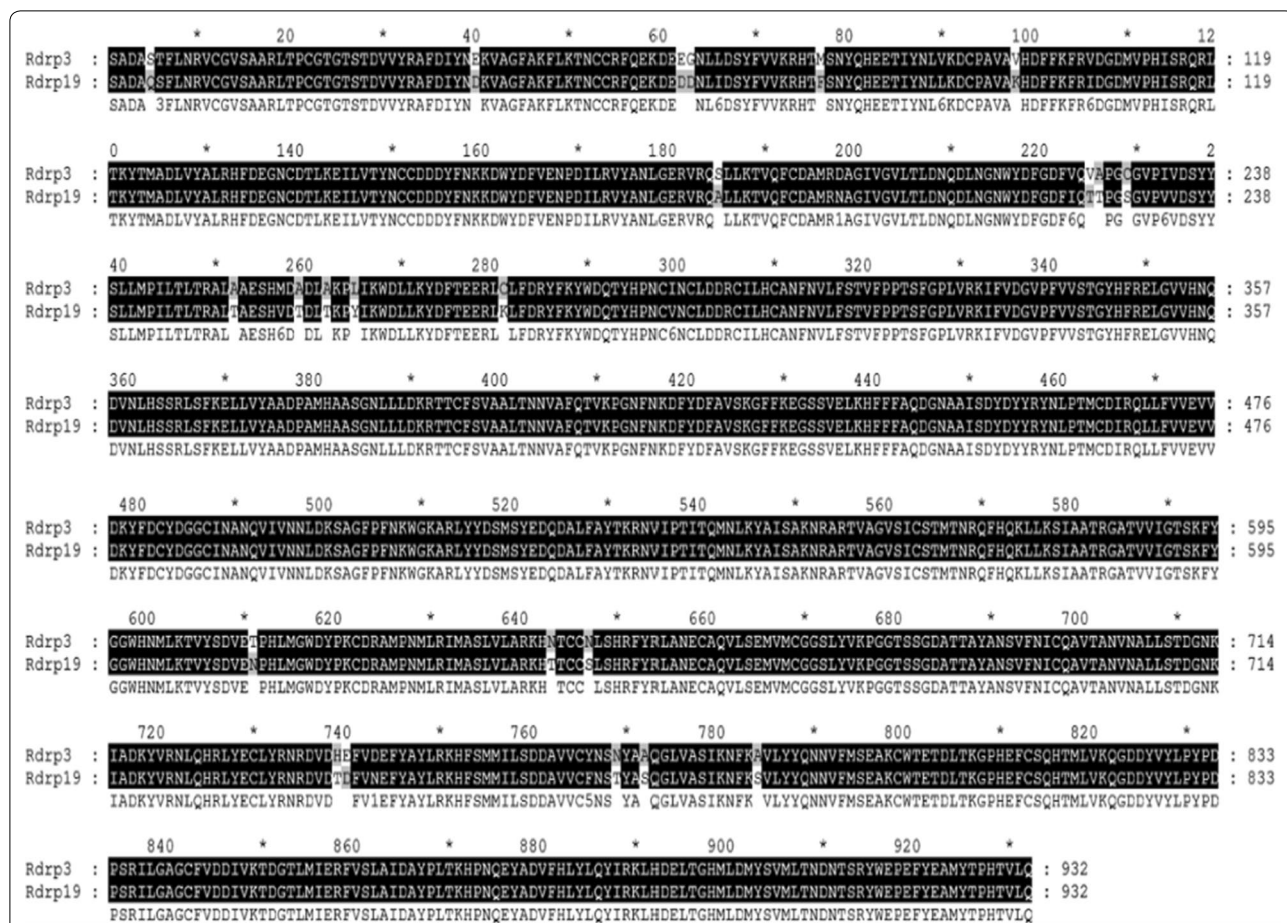
Fig. 2 RdRp Sequence alignment of SARS-CoV and SARS-CoV-2. RdRp sequences of SARS-CoV and SARS-CoV-2 were aligned for analysis of divergence, the dark region represents the conserved regions and lighter regions highlight changes in amino acids in specific sites. Only 2.46% divergence was observed between the two RdRp protein sequences
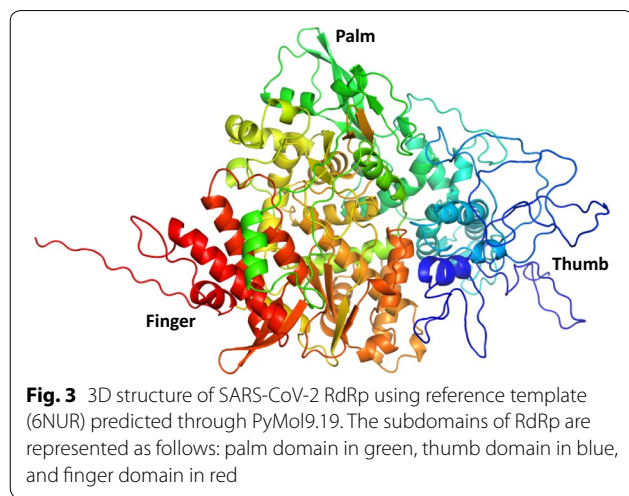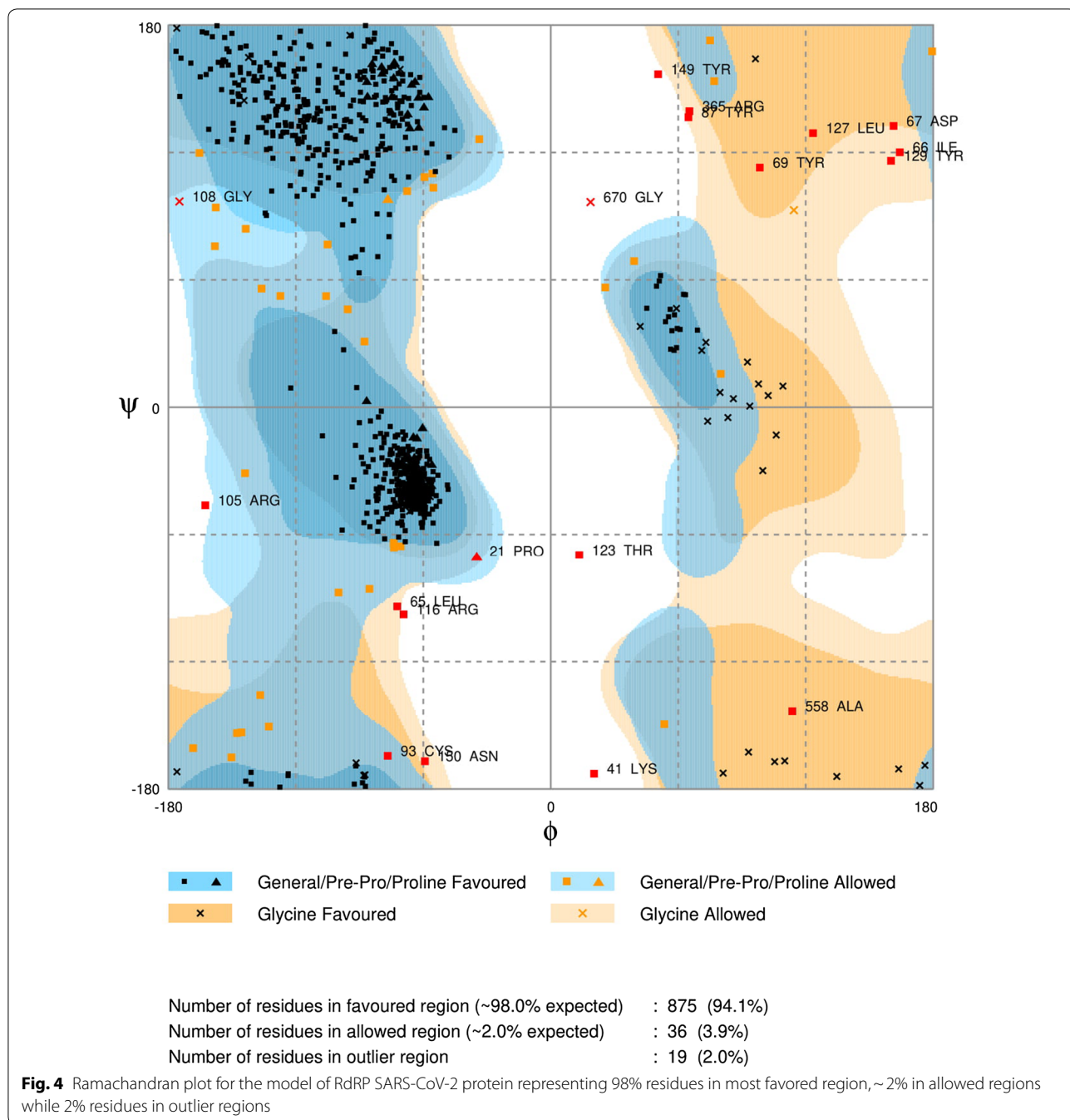

Fig. 3 3D structure of SARS-CoV-2 RdRp using reference template (6NUR) predicted through PyMol9.19. The subdomains of RdRp are represented as follows: palm domain in green, thumb domain in blue, and finger domain in red

sequence. Protein sequences were assembled based on their physicochemical properties. ProtParam (ExPASy) results for physicochemical properties revealed that the RdRp polypeptide contains 932 AA with a molecular weight of 106,660.24 Da, GRAVY score of −0.224, an instability index of 28.32, and is capable of making hydrogen bonds, categorizing it as a stable protein (Table 1).

## Modeling and structure prediction

Results of comparison of amino acid residues demonstrated that the RdRps of SARS-CoV and SARS-CoV-2 are mostly conserved. Overall, divergence was observed in only 23 out of 932 AA (Fig. 2). According to our results, 2.46% divergence was observed, thus we selected the SARS-CoV RdRp (6NUR) protein structure as a reference template for the purpose of homology modeling. The protein sequence of SARS-CoV-2 RdRp (YP_009725307.1) was retrieved from NCBI and submitted to a protein databank for PSI-BLAST (https://www.rcsb.org/), which depicts 96% similarity with the

Fig. 4 Ramachandran plot for the model of RdRP SARS-CoV-2 protein representing 98% residues in most favored region, ~ 2% in allowed regions while 2% residues in outlier regions

6NUR protein [52]. Using the 6NUR protein structure as a template, 100 models were generated with MODEL-LER and the SARS-CoV 3D structure was selected with a DOPE score of -108,832.65625 and GA score 1 (Fig. 3). RAMPAGE Ramachandran plots indicate that most of the amino acids are within favorable regions (875 AA), making up 94.1% of the expected residues. There was no need for refinement analysis, as the Ramachandran plot shows 94.1% in most favored regions. Moreover, outlier

AA represented on the plot shows the desired AA which are to be targeted by our ligand reside in favored regions (Fig. 4).

The SARS-CoV-2 RdRp protein sequence was queried using pBLAST and 94 validated sequences were downloaded from the NCBI database for evaluating its phylogenetic relationship with other viral RdRp proteins. An MLE tree constructed for all validated proteins revealed that most of the proteins shared close association with

Aftab *et al. J Transl Med*   (2020) 18:275

Page 8 of 15
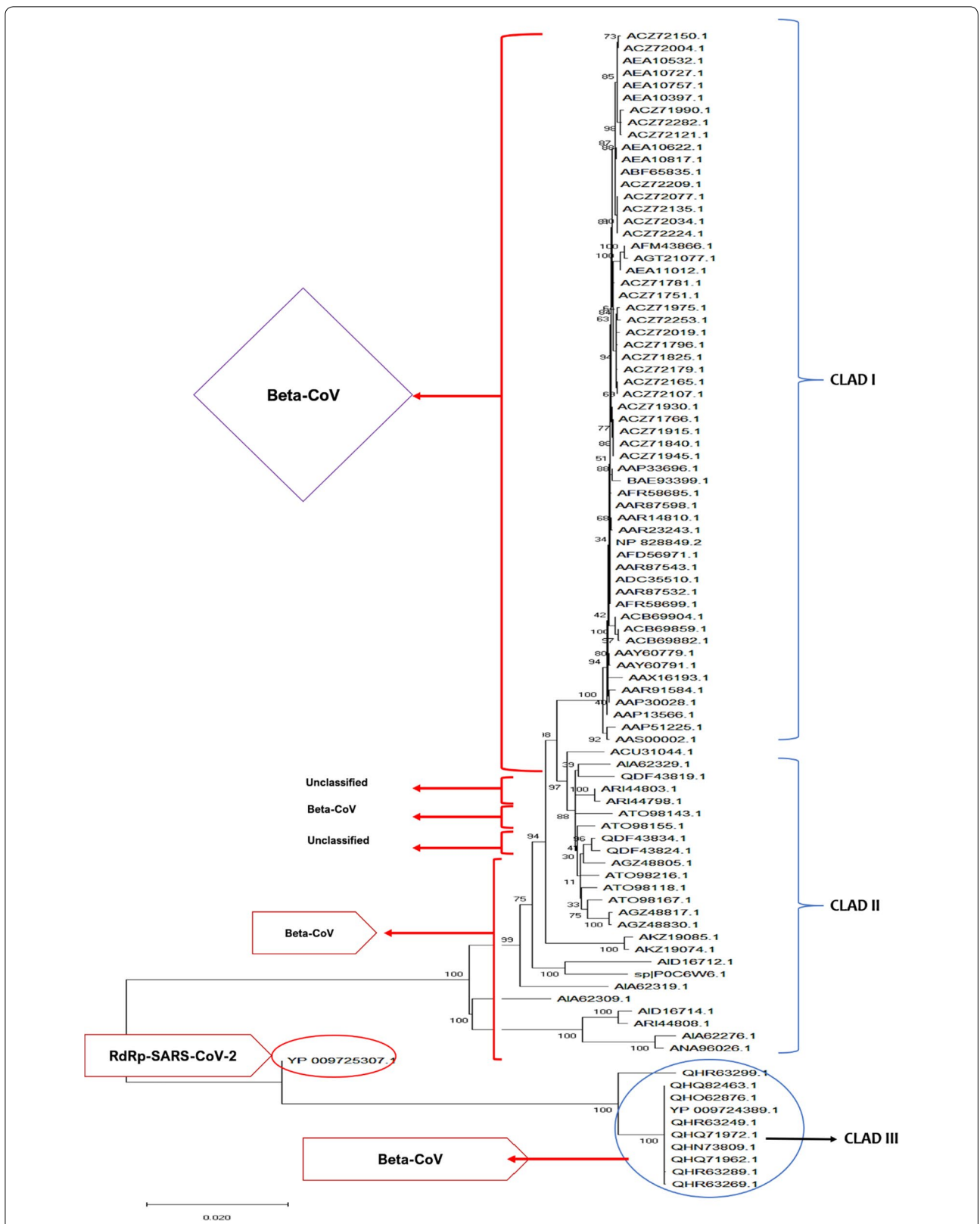
Fig. 5 Dendrogram representing evolutionary relationship of RdRp proteins of SARS-CoV-2 with other RdRp proteins. SARS-CoV-2 accession (red circle) is closely associated with other SARS RdRp proteins (blue circle). For accessions, origin and other relevant data, see Additional file 1: Data 2

Aftab *et al. J Transl Med*   (2020) 18:275

Page 9 of 15

Table 2 Docking scores of selected ligands against RdRp SARS-CoV-2

Sr. no	Inhibitor	S-score	RMSD
1.	Remdesivir	− 14.06038	2.357701
2.	Galidesivir	− 12.5338	1.0217
3.	Ribavirin	− 11.8884	0.9554
4.	Sofosbuvir	− 11.132	2.08
5.	Penciclovir	− 10.9772	0.9461
6.	Chloroquine	− 10.3804	1.6397
7.	Nitazoxanide	− 9.7636	2.4603
8.	Nafamostat	− 9.76403	2.4605
9.	Interferon	− 9.62582	1.7336
10.	Favipiravir	− 9.36643	1.2231

each other. Overall, the RdRp protein tree was divided into three clades: I, II, and III. Most of the viral RdRps in clade I belong to beta-CoV; the same is true for clades II and III with most accessions belonging to beta-CoV, while a few remain unclassified and require more research for proper designation. Accession QHR63299.1 was also unclassified but shows close association with other SARS RdRp relatives (Fig. 5).

Ligand selection and virtual screening

All the selected ligands from a previous literature survey were docked with the RdRp protein of SARS-CoV-2. The docking score, RMSD, and binding energy of ligands with receptor RdRp are shown in Table 2. The results demonstrated that all the ligands exhibit interaction with the pocket of RdRp, while, Remdesivir has the highest S-score and lowest binding energy with the active pocket of RdRp of SARS-CoV-2. However, in virtual screening, Remdesivir and Sofosbuvir do not provided us with novel compounds against SARS-CoV-2 RdRp. Therefore, we have selected Galidesivir with S-score (− 12.5338) and binding energy (− 13.922) that provided us with some
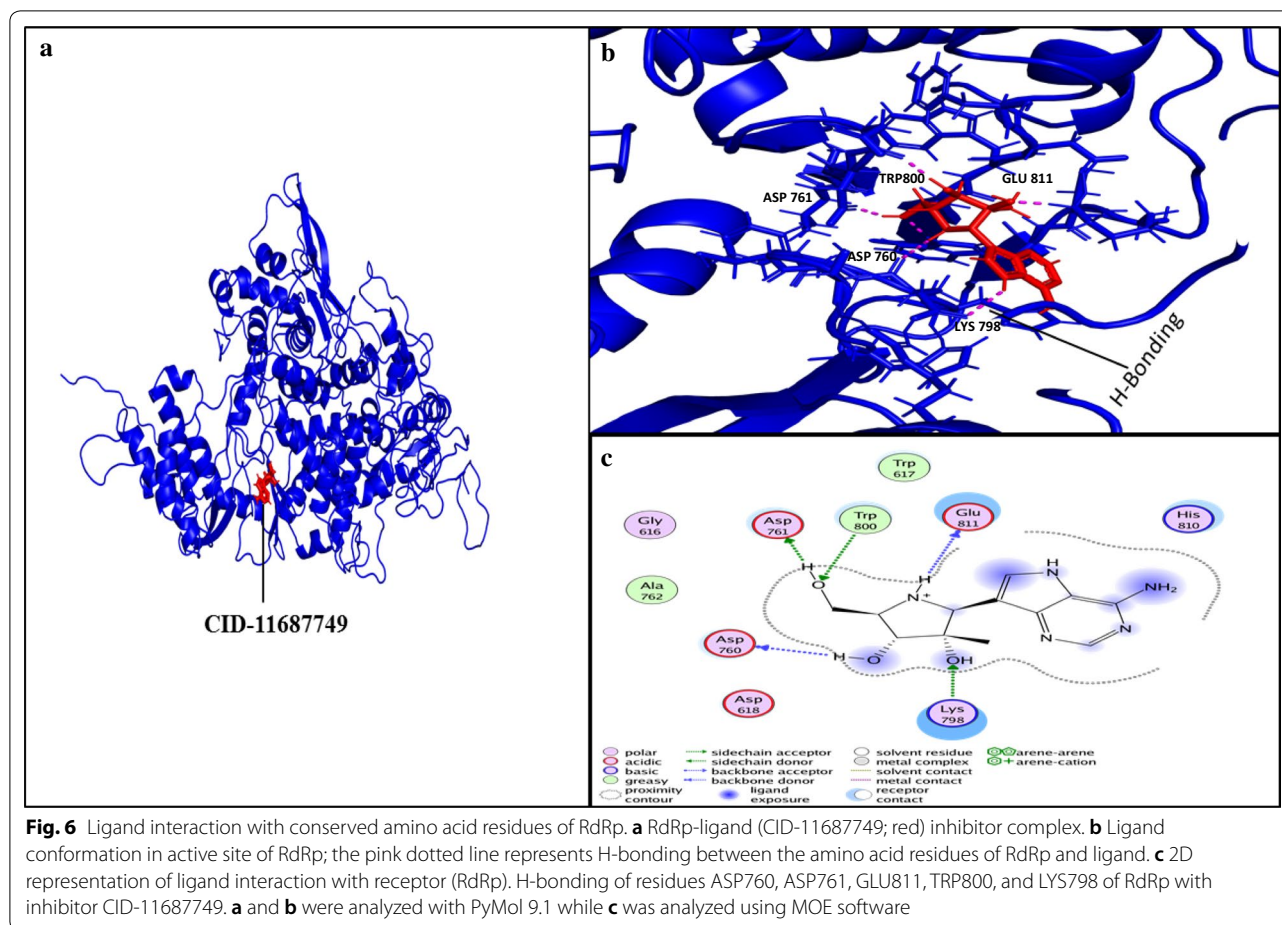
novel compounds in virtual screening (Table 3). Compounds with > 95% structural similarity to Galidesivir were selected for ligand-based virtual screening from the PubChem database. A total of 1061 compounds were retrieved and Pfizer's rule of five was applied to all the compounds, out of which 677 were aligned after minimizing energy in a new database for docking with the protein of interest RdRp.

Docking simulations

Docking simulations demonstrated that the ligand Galidesivir strongly binds within active sites containing the following residues: ASP760, ASP761, GLY616, TRP617, ASP618, TYR619, PRO620, LYS621, CYS622, LEU758, SER759, ALA762, ALA797, LYS798, CYS799, TRP800, HIS810, GLU811, PHE812, CYS813, SER814, and GLN815. Of these, the sites ASP761, ALA762, LYS798, and SER814 strongly bind Galidesivir, while the other AA found in close vicinity were involved in making the interface of ligand and receptor complex (Fig. 6). CID123624208 and CID11687749 interactions with RdRP are presented in Figs. 6 and 7. We suggest two non-toxic drug-like compounds predicted to bind ASP761, ASP760, SER814 and LYS798 sites (Figs. 6, 7) through stout hydrogen bonding. ADMET profiling for drug-like compounds demonstrated positive results for blood–brain barrier (BBB), human intestinal absorption (HIA), renal organic cation transporter (ROCT) for CID11687749 and CID123624208 (Table 3), whereas $CaCO_2$ permeability results were found to be negative for both the compounds. Ames toxicity testing results revealed that the compounds were non-toxic and non-carcinogenic (Table 3, Fig. 8). CID123624208 and CID11687749 were the best suggested drug-like compounds analyzed through in silico analysis and may act as strong inhibitors against SARS-CoV-2. We suggest that

Table 3 Docking score, binding energy, and Lipinski's rule scan results for selected compounds

PubChem ID	S-score	RMSD	Binding energy kcal/mol	Lipinski's drug-likeness score
CID123624208	− 12.2589	0.4971	− 20.812	MW = 309.282, LogP = − 1.247, H-bond donors = 6, H-bond acceptors = 4, and tPSA = 168.820
CID11687749	− 12.7751	0.9957	− 19.601	MW = 279.30, LogP = − 1.247, H-bond donor = 6, H-bond acceptor = 4, tpSA = 140.31

Fig. 6 Ligand interaction with conserved amino acid residues of RdRp. **a** RdRp-ligand (CID-11687749; red) inhibitor complex. **b** Ligand conformation in active site of RdRp; the pink dotted line represents H-bonding between the amino acid residues of RdRp and ligand. **c** 2D representation of ligand interaction with receptor (RdRp). H-bonding of residues ASP760, ASP761, GLU811, TRP800, and LYS798 of RdRp with inhibitor CID-11687749. **a** and **b** were analyzed with PyMol 9.1 while **c** was analyzed using MOE software
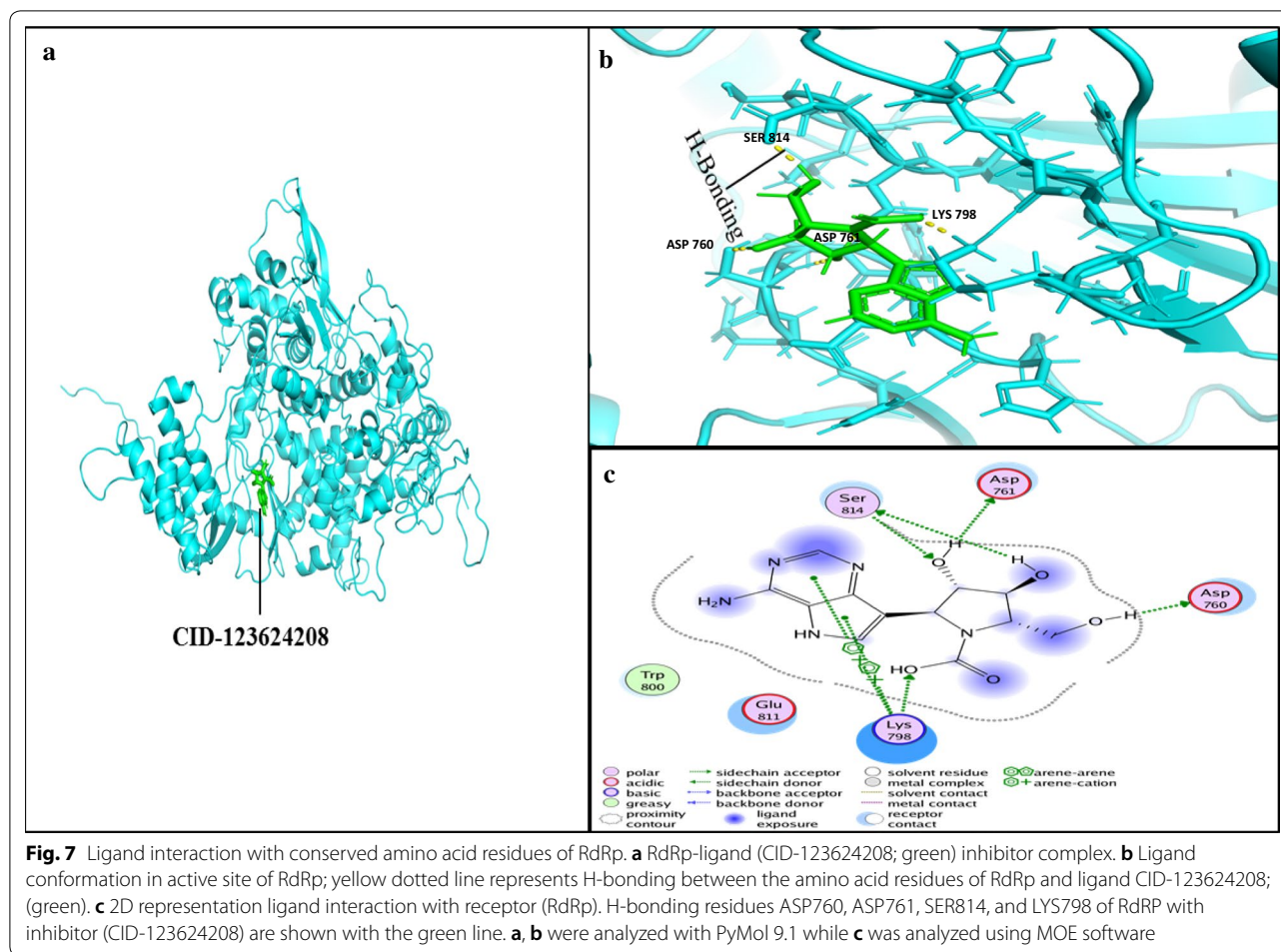
these novel compounds could be used for preclinical trials as antiviral agents against SARS-CoV-2.

Discussion

Sequencing of novel viruses such as SARS-CoV-2 helps to identify target regions for intervention. Scientists took less than 60 days to trace the origin of SARS-CoV-2, however, this virus may show similarity to other viruses based on nucleotide sequences, such as MERS, which could help to find attractive therapeutic options [53, 54, 56, 57]. It has been reported that nucleotide sequence-based analysis of SARS-CoV-2 and SARS-CoV (Bats origin) represent a close relation with each other but sufficiently divergent due to single nucleotide mutations [13, 21, 23, 56, 57]. In our study, 94 validated sequences were used for construction of a phylogenetic tree to trace out similarity between all the sequences from alpha-, beta-, and gamma coronaviruses. Most of the viral species examined showed resemblance to each other and clustered together. SARS-CoV-2 whole-genome sequences reported from Hong Kong and USA (MN975262.1, MN997409.1, MN994467.1, MN988713.1) revealed close relation with each other as they originated from same clade and are considered as sister species (Fig. 1). Our results are consistent with the results of Wu et al. [58], who demonstrated that SARS-CoV-2 and bat-origin SARS-CoV clustered together, revealing similarity in their genomes (Fig. 1).

Physicochemical properties of the RdRp protein used in our study revealed satisfactory results in terms of molecular weight, GRAVY score, and instability index capable of making strong hydrogen bonds, and are similar to previous data reported by Dwivedi et al. [59]. Homology modeling results within the current study are consistent with those reported by Xu et al. [32] and Sarwar et al. [60]. However, the latter group only predicted 10 models, instead of the ~ 100 structures here. In addition, of these possible structures we selected the most

Aftab *et al. J Transl Med*   (2020) 18:275

Page 11 of 15

Fig. 7 Ligand interaction with conserved amino acid residues of RdRp. **a** RdRp-ligand (CID-123624208; green) inhibitor complex. **b** Ligand conformation in active site of RdRp; yellow dotted line represents H-bonding between the amino acid residues of RdRp and ligand CID-123624208; (green). **c** 2D representation ligand interaction with receptor (RdRp). H-bonding residues ASP760, ASP761, SER814, and LYS798 of RdRP with inhibitor (CID-123624208) are shown with the green line. **a**, **b** were analyzed with PyMol 9.1 while **c** was analyzed using MOE software

suitable among them to predict the 3D structure of the RdRp protein of SARS-CoV. Moreover, in our study, the RdRp protein sequence was searched for similarity with other viral RdRps. Our results demonstrated that all the SARS-CoV-2 RdRp proteins relatively clustered together, which included isolates from Hong Kong, USA and Wuhan (Fig. 5). It is suggested that the close association between all these RdRps is due to the similarity in their genome organization. Therefore, more research is needed to explore the differentiation between these viral RdRps.

Docking simulations of results of our study are in line with Xu et al. [32] regarding their strong interaction with the following active sites: ASP760, ASP761, GLY616, TRP617, ASP618, TYR619, PRO620, LYS621, CYS622, LEU758, SER759, ALA762, ALA797, LYS798, CYS799, TRP800, HIS810, GLU811, PHE812, CYS813, SER814, GLN815 (Figs. 6, 9). This group also reported that ASP760 and ASP761 are responsible for composing

the RdRp catalytic domain, while SER814 is involved in positioning of the priming nucleotide, and our results are consistent with the previous studies [60–64]. Similarly, the LYS798 residue was found to be involved in stabilizing the core structure of the RdRp domain, the finding is in line with other studies [63–65]. In a recent study, Gao et al. reported that nucleoside analogs such as Remdesivir and Sofosbuvir strongly binds with ASP760, ASP761 and ASP618 residues to inhibit RdRp of SARS-CoV-2 [66]. Our study is similar to Gao et al. reporting nucleoside analogs can form hydrogen bonds with active site of RdRp.

Our study indicates that the novel predicted drug-like compounds CID123624208 and CID11687749 have a strong affinity with the residues of the RdRp catalytic domain (Figs. 6, 7). Strong S-score, binding energy, and RMSD values suggest that these compounds could be used as potential inhibitors against the RdRp of

Aftab *et al. J Transl Med*   (2020) 18:275

Page 12 of 15
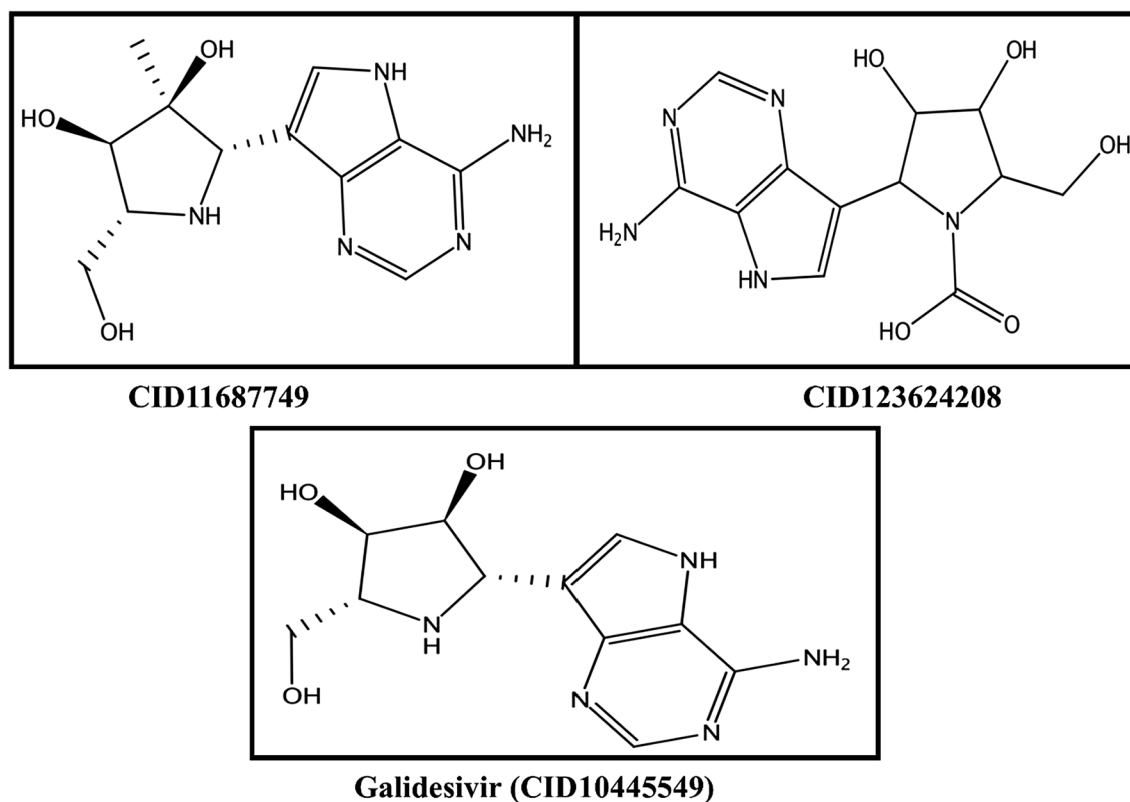
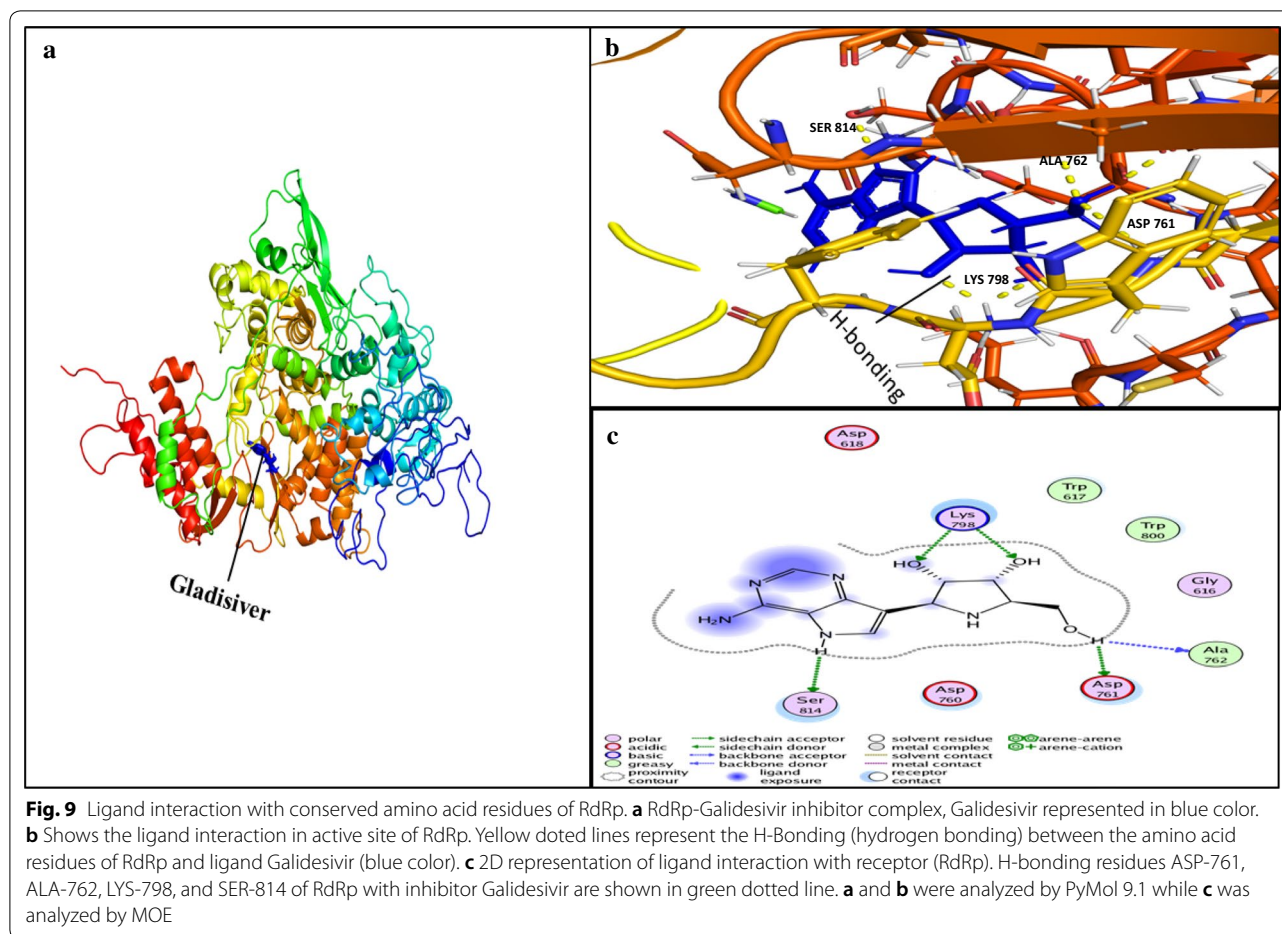**Fig. 8** 2D structure of selected inhibitor with respective PubChem IDs

SARS-CoV-2. Numerous other viral inhibitors have also been reported which are in clinical trials including Remdesivir, Sofosbuvir Galidesivir etc., for which we have already tested with virtual screening [67]. However, some of them are stated in Table 2 with their S-score and RMSD values. Thus, in summary, Galidesivir and the two drugs-like compounds CID123624208 and CID11687749 screened in the present study could more likely have potential as therapeutic drugs targeting SARS-CoV-2.

Use of FDA approved anti-polymerase drugs has been the objective of virologists in treating the patients of new viral infection caused by SARS-CoV-2. The choice of FDA approved drugs was a wise decision in this current emergency arises due to COVID-19 pandemic; since they were previously tested before approval by the FDA. The present study is aimed at finding the best alternative drugs to inhibit the RdRp of SARS-CoV-2. Galidesivir and its suggested compounds pose no toxicity and are safe to consume proved in our in silico analysis. The only limitation of the study is to confirm

the results of this study using in vivo and in vitro analysis. In the future, further research is direly needed to validate the results of this in silico evaluation for preclinical trials of these potent drug-like compounds CID123624208 and CID11687749.

Conclusion

The rapidly growing coronavirus cases urge the development of new therapeutics and vaccines against SARS-CoV-2. In the previous studies so far, RdRp-CoV (nsp12) is the most frequently used and suggested potential target to inhibit viral replication. In this study, we tested some important anti polymerase drugs, whether they are currently in clinical trials or in the market to stop viral infection on an emergency basis. However, Galidesivir and its drug-like compounds CID123624208 and CID11687749 have shown an effective attachment to the priming site of viral RdRp, which could lead to replication failure. Moreover, for further safety, health concerns and vaccine development, the suggested compounds of our in silico

Aftab *et al. J Transl Med*   (2020) 18:275

Page 13 of 15

Fig. 9 Ligand interaction with conserved amino acid residues of RdRp. **a** RdRp-Galidesivir inhibitor complex, Galidesivir represented in blue color. **b** Shows the ligand interaction in active site of RdRp. Yellow doted lines represent the H-Bonding (hydrogen bonding) between the amino acid residues of RdRp and ligand Galidesivir (blue color). **c** 2D representation of ligand interaction with receptor (RdRp). H-bonding residues ASP-761, ALA-762, LYS-798, and SER-814 of RdRp with inhibitor Galidesivir are shown in green dotted line. **a** and **b** were analyzed by PyMol 9.1 while **c** was analyzed by MOE

assessment need further in vitro analysis for future confirmation which will lead towards preclinical trials.

Supplementary information

Supplementary information accompanies this paper at https://doi.org/10.1186/s12967-020-02439-0.

Additional file 1. Origin and Accession # of recently emerged SARS coronaviruses and viral RdRps.

Abbreviations
AA: amino acids; BBB: blood–brain barrier; DOPE: discrete optimized protein energy; HIA: human intestinal absorption; ICTV: International Committee on Taxonomy of Viruses; MERS: middle east respiratory syndrome; MLE: maximum likelihood estimation; MOE: molecular operating environment; MW: molecular weight; NCBI: National Center for Biotechnology Information; PHEIC: Public Health Emergency of International Concern; RMSD: root-mean-square deviation; ROCT: renal organic cation transporter; SARS: severe acute respiratory syndrome; WHO: World Health Organization.

Authors' contributions
Authors MUM, and SOA and MZG designed the study. Authors ZH and MZG performed molecular docking and complied the manuscript. MUM has prepared Additional file 1: Tables. Final manuscript proofreading and errors were resolved by authors MZG, ZH, and SOA. ZK reviewed the manuscript for further improvements and prepared final draft. AH and NM improved the manuscript for English and grammatical errors. MZG and SOA has revised the manuscript according to reviewers comments. AH reviewed the revised manuscript after prior update.

Availability of data and materials
All the supported data is included in the manuscript and where applicable hyperlinks are provided. Some of the data has been placed in additional information submitted with this manuscript.

Ethics approval and consent to participate
Not applicable.

Consent for publication
All the authors have read and approved the manuscript in all respects for publication.

Aftab *et al. J Transl Med*   (2020) 18:275

Page 14 of 15

Author details
[1] Center of Agricultural Biochemistry and Biotechnology (CABB), University of Agriculture, Faisalabad, Pakistan. [2] Center for Advanced Studies in Agriculture and Food Security (CAS-AFS), University of Agriculture, Faisalabad, Pakistan. [3] Institute of Plant Breeding and Biotechnology, MNS University of Agriculture, Multan, Pakistan. [4] Department of Biochemistry, University of Agriculture, Faisalabad, Pakistan. [5] Department of Chemistry, United Arab Emirates University, Al-Ain, UAE.

References
1. Yang L, Wu Z, Ren X, Yang F, He G, Zhang J, Dong J, Sun L, Zhu Y, Du J, Zhang S. Novel SARS-like betacoronaviruses in bats, China, 2011. Emerg Infect Dis. 2013;19:989.
2. Wong S, Lau S, Woo P, Yuen KY. Bats as a continuing source of emerging infections in humans. Rev Med Virol. 2007;17:67–91.
3. Hu B, Ge X, Wang LF, Shi Z. Bat origin of human coronaviruses. Virol J. 2015;12:221.
4. Li W, Shi Z, Yu M, Ren W, Smith C, Epstein JH, Wang H, Crameri G, Hu Z, Zhang H, Zhang J. Bats are natural reservoirs of SARS-like coronaviruses. Science. 2005;310:676–9.
5. WHO. Coronavirus Disease 2019 (COVID-19) Situation Report-81. 2020. https://www.who.int/docs/default-source/coronaviruse/situation-reports/20200410-sitrep-81-covid-19.pdf?sfvrsn=ca96eb84_2. Accessed 15 Apr 2020.
6. Lau SK, Li KS, Huang Y, Shek CT, Tse H, Wang M, Choi GK, Xu H, Lam CS, Guo R, Chan KH. Ecoepidemiology and complete genome comparison of different strains of severe acute respiratory syndrome-related Rhinolophus bat coronavirus in China reveal bats as a reservoir for acute, self-limiting infection that allows recombination events. J Virol. 2010;84:2808–19.
7. Ahlquist P, Noueiry AO, Lee WM, Kushner DB, Dye BT. Host factors in positive-strand RNA virus genome replication. J Virol. 2003;77:8181–6.
8. Ge XY, Li JL, Yang XL, Chmura AA, Zhu G, Epstein JH, Mazet JK, Hu B, Zhang W, Peng C, Zhang YJ. Isolation and characterization of a bat SARS-like coronavirus that uses the ACE2 receptor. Nature. 2013;503:535–8.
9. Cascella M, Rajnik M, Cuomo A, et al. Features, evaluation and treatment coronavirus (COVID-19) [Updated 2020 Mar 8]. In: StatPearls [Internet]. Treasure Island (FL): StatPearls Publishing; 2020. https://www.ncbi.nlm.nih.gov/books/NBK554776/.
10. Yin Y, Wunderink RG. MERS, SARS and other coronaviruses as causes of pneumonia. Respirology. 2018;23:130–7.
11. Chang L, Yan Y, Wang L. Coronavirus disease 2019: coronaviruses and blood safety. Transfus Med Rev. 2020;34(2):75–80. https://doi.org/10.1016/j.tmrv.2020.02.003.
12. Ashour HM, Elkhatib WF, Rahman M, Elshabrawy HA. Insights into the recent 2019 novel coronavirus (SARS-CoV-2) in light of Past human coronavirus outbreaks. Pathogens. 2020;9:186.
13. Nguyen TM, Zhang Y, Pandolfi PP. Virus against virus: a potential treatment for 2019-nCov (SARS-CoV-2) and other RNA viruses. Cell Res. 2020;30:189–90.
14. Peng X, Xu X, Li Y, et al. Transmission routes of 2019-nCoV and controls in dental practice. Int J Oral Sci. 2020;12:9. https://doi.org/10.1038/s41368-020-0075-9.
15. Lu CW, Liu XF, Jia ZF. 2019-nCoV transmission through the ocular surface must not be ignored. Lancet. 2020;395:e39.
16. Kieny MP. After Ebola, a Blueprint Emerges to Jump-Start R&D. 2015. https://blogs.scientificamerican.com/guest-blog/after-ebola-a-blueprint-emerges-to-jump-start-r-d/.
17. Boulos MN, Geraghty EM. Geographical tracking and mapping of coronavirus disease COVID-19/severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) epidemic and associated events around the world: how 21st century GIS technologies are supporting the global fight against outbreaks and epidemics.
18. Chan JF, Yuan S, Kok KH, To KK, Chu H, Yang J, Xing F, Liu J, Yip CC, Poon RW, Tsoi HW. A familial cluster of pneumonia associated with the 2019 novel coronavirus indicating person-to-person transmission: a study of a family cluster. Lancet. 2020.
19. Stoermer M. Homology models of the papain-like protease PLpro from coronavirus 2019-nCoV. 2020. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.11799705.v1.
20. Jung SM, Kinoshita R, Thompson RN, Hayashi K, Linton NM, Yang Y, Akhmetzhanov AR, Nishiura H. Epidemiological identification of a novel infectious disease in real time: Analysis of the atypical pneumonia outbreak in Wuhan, China, 2019-20. medRxiv. 2020.
21. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, Shi Z, Hu Z, Zhong W, Xiao G. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020:1-3.
22. Jin YH, Cai L, Cheng ZS, Cheng H, Deng T, Fan YP, Fang C, Huang D, Huang LQ, Huang Q, Han Y. A rapid advice guideline for the diagnosis and treatment of 2019 novel coronavirus (2019-nCoV) infected pneumonia (standard version). Military Medical Research. 2020;7:4.
23. Kruse RL. Therapeutic strategies in an outbreak scenario to treat the novel coronavirus originating in Wuhan, China. F1000Research. 2020;9:72.
24. Lai MM, Baric RS, Makino S, Keck JG, Egbert J, Leibowitz JL, Stohlman SA. Recombination between nonsegmented RNA genomes of murine coronaviruses. J Virol. 1985;56:449–56.
25. Woo PC, Wang M, Lau SK, Xu H, Poon RW, Guo R, Wong BH, Gao K, Tsoi HW, Huang Y, Li KS. Comparative analysis of twelve genomes of three novel group 2c and group 2d coronaviruses reveals unique group and subgroup features. J Virol. 2007;81:1574–85.
26. Woo PC, Huang Y, Lau SK, Tsoi HW, Yuen KY. *In silico* analysis of ORF1ab in coronavirus HKU1 genome reveals a unique putative cleavage site of coronavirus HKU1 3C-like protease. Microbiol Immunol. 2005;49:899–908.
27. Woo PC, Huang Y, Lau SK, Yuen KY. Coronavirus genomics and bioinformatics analysis. Viruses. 2010;2:1804–20.
28. Jia Z, Yan L, Ren Z, Wu L, Wang J, Guo J, Zheng L, Ming Z, Zhang L, Lou Z, Rao Z. Delicate structural coordination of the Severe Acute Respiratory Syndrome coronavirus Nsp13 upon ATP hydrolysis. Nucleic Acids Res. 2019;47:6538–50.
29. Chatterjee S. Key Steps in the early evolution of life from the origin of protein synthesis to modern cellular life. Preprints. 2020. p. 2020020155.
30. Wu J, Liu W, Gong P. A structural overview of RNA-dependent RNA polymerases from the Flaviviridae family. Int J Mol Sci. 2015;16:12943–57.
31. Smertina E, Urakova N, Strive T, Frese M. Calicivirus RNA-dependent RNA polymerases: evolution, structure, protein dynamics and function. Front Microbiol. 2019;10:1280.
32. Xu X, Liu Y, Weiss S, Arnold E, Sarafianos SG, Ding J. Molecular model of SARS coronavirus polymerase: implications for biochemical functions and drug design. Nucleic Acids Res. 2003;31:7117–30.
33. Hansen JL, Long AM, Schultz SC. Structure of the RNA-dependent RNA polymerase of poliovirus. Structure. 1997;5:1109–22.
34. Ago H, Adachi T, Yoshida A, Yamamoto M, Habuka N, Yatsunami K, Miyano M. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Structure. 1999;7:1417–26.
35. Bressanelli S, Tomei L, Roussel A, Incitti I, Vitale RL, Mathieu M, De Francesco R, Rey FA. Crystal structure of the RNA-dependent RNA polymerase of hepatitis C virus. Proc Natl Acad Sci. 1999;96:13034–9.
36. Lesburg CA, Cable MB, Ferrari E, Hong Z, Mannarino AF, Weber PC. Crystal structure of the RNA-dependent RNA polymerase from hepatitis C virus reveals a fully encircled active site. Nat Struct Biol. 1999;6:937–43.
37. Kinsella E, Martin SG, Grolla A, Czub M, Feldmann H, Flick R. Sequence determination of the Crimean-Congo hemorrhagic fever virus L segment. Virology. 2004;321:23–8.
38. Liang B, Li Z, Jenni S, Rahmeh AA, Morin BM, Grant T, Grigorieff N, Harrison SC, Whelan SP. Structure of the L protein of vesicular stomatitis virus from electron cryomicroscopy. Cell. 2015;162:314–27.
39. Gerlach P, Malet H, Cusack S, Reguera J. Structural insights into bunyavirus replication and its regulation by the vRNA promoter. Cell. 2015;161:1267–79.
40. Jia H, Gong P. A structure-function diversity survey of the RNA-dependent RNA polymerases from the positive-strand RNA viruses. Front Microbiol. 2019;10:1945.

41. Zhou Y, Hou Y, Shen J, et al. Network-based drug repurposing for novel coronavirus 2019-nCoV/SARS-CoV-2. Cell Discov. 2020;14:6.
42. Yu WC, Hui DSC, Chan-Yeung M. Antiviral agents and corticosteroids in the treatment of severe acute respiratory syndrome (SARS). Thorax. 2004;59:643–5.
43. Elfiky AA. Anti-HCV, nucleotide inhibitors, repurposing against COVID-19. Life Sci. 2020;248:117477.
44. Elfiky AA, Mahdy SM, Elshemey WM. Quantitative structure-activity relationship and molecular docking revealed a potency of anti-hepatitis C virus drugs against human corona viruses. J Med Virol. 2017;89:1040–7.
45. Elfiky AA. Zika viral polymerase inhibition using anti-HCV drugs both in market and under clinical trials. J Med Virol. 2016;88:2044–51.
46. Elfiky AA, Ismail A. Molecular dynamics and docking reveal the potency of novel GTP derivatives against RNA dependent RNA polymerase of genotype 4a HCV. Life Sci. 2019;238:116958.
47. Ezat AA, Elfiky AA, Elshemey WM, Saleh NA. Novel inhibitors against wildtype and mutated HCV NS3 serine protease: an in silico study. Virus Dis. 2019;30:207–13.
48. Yang PL, Gao M, Lin K, Liu Q, Villareal VA. Anti-HCV drugs in the pipeline. Curr Opin Virol. 2011;1:607–16.
49. Markland W, McQuaid T, Jain J, Kwong J. Broad-spectrum antiviral activity of the IMP dehydrogenase inhibitor VX-497: a comparison with Ribavirin and demonstration of antiviral additivity with alpha interferon. Antimicrob Agents Chemother. 2000;44:859–66.
50. Wang M, Cao R, Zhang L, Yang X, Liu J, Xu M, et al. Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. Cell Res. 2020;1:3.
51. McQuaid T, Savini C, Seyedkazemi S. Sofosbuvir, a significant paradigm change in HCV treatment. J Clin Transl Hepatol. 2015;3:27–35.
52. Kirchdoerfer RN, Ward AB. Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors. Nat Commun. 2019;10:1–9.
53. Qamar MT, Alqahtani SM, Alamri MA, Chen LL. Structural basis of SARS-CoV-2 3CLpro and anti-COVID-19 drug discovery from medicinal plants. J Pharm Anal. 2020. https://doi.org/10.1016/j.jpha.2020.03.009.
54. ul Qamar MT, Maryam A, Muneer I, Xing F, Ashfaq UA, Khan FA, Anwar F, Geesi MH, Khalid RR, Rauf SA, Siddiqi AR. Computational screening of medicinal plant phytochemicals to discover potent pan-serotype inhibitors against dengue virus. Sci Rep. 2019;9(1):1–6.
55. Qamar MU, Kiran S, Ashfaq UA, Javed MR, Anwar F, Ali MA. Discovery of novel dengue NS2B/NS3 protease inhibitors using pharmacophore modeling and molecular docking based virtual screening of the zinc database. Int J Pharmacol. 2016;12:621–32.
56. Ceraolo C, Giorgi FM. Genomic variance of the 2019-nCoV coronavirus. J Med Virol. 2020;92:522.
57. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y. Genomic characterization and epidemiology of 2019 novel coronavirus: implications for virus origins and receptor binding. Lancet. 2020;395:565–74.
58. Wu A, Peng Y, Huang B, Ding X, Wang X, Niu P, Meng J, Zhu Z, Zhang Z, Wang J, Sheng J. Genome composition and divergence of the novel coronavirus (2019-nCoV) originating in China. Cell Host Microbe. 2020;27:372.
59. Dwivedi VD, Arora S, Pandey A. Computational analysis of physico-chemical properties and homology modeling of carbonic anhydrase from Cordyceps militaris. Netw Model Anal Health Inf Bioinform. 2013;2:209–12.
60. Sarwar MW, Riaz A, Nahid N, Al Qahtani A, Ahmed N, Nawaz-Ul-Rehman MS, Younus A, Mubin M. Homology modeling and docking analysis of ßC1 protein encoded by Cotton leaf curl Multan betasatellite with different plant flavonoids. Heliyon. 2019;5:e01303.
61. Tao Y, Farsetta DL, Nibert ML, Harrison SC. RNA synthesis in a cage—structural studies of reovirus polymerase λ3. Cell. 2002;111:733–45.
62. Butcher SJ, Grimes JM, Makeyev EV, Bamford DH, Stuart DI. A mechanism for initiating RNA-dependent RNA polymerization. Nature. 2001;410:235–40.
63. Ding J, Das K, Hsiou Y, Sarafianos SG, Clark AD, Jacobo-Molina A, Tantillo C, Hughes SH, Arnold E. Structure and functional implications of the polymerase active site region in a complex of HIV-1 RT with a double-stranded DNA template-primer and an antibody Fab fragment at 2.8 Å resolution. J Mol Biol. 1998;284:1095–111.
64. Jacobo-Molina A, Ding J, Nanni RG, Clark AD, Lu X, Tantillo C, Williams RL, Kamer G, Ferris AL, Clark P. Crystal structure of human immunodeficiency virus type 1 reverse transcriptase complexed with double-stranded DNA at 3.0 A resolution shows bent DNA. Proc Natl Acad Sci. 1993;90:6320–4.
65. Huang H, Chopra R, Verdine GL, Harrison SC. Structure of a covalently trapped catalytic complex of HIV-1 reverse transcriptase: implications for drug resistance. Science. 1998;282:1669–75.
66. Gao Y, Yan L, Huang Y, Liu F, Zhao Y, Cao L, Wang T, Sun Q, Ming Z, Zhang L, Ge J. Structure of the RNA-dependent RNA polymerase from COVID-19 virus. Science. 2020;368:779–82.
67. Elfiky AA. Ribavirin, Remdesivir, Sofosbuvir, Galidesivir, and Tenofovir against SARS-CoV-2 RNA dependent RNA polymerase (RdRp): a molecular docking study. Life Sci. 2020;253:117592.

Publisher's Note