Structural Chemistry (2022) 33:1789–1797
https://doi.org/10.1007/s11224-022-02019-6

ORIGINAL RESEARCH

Identification of SARS-CoV-2 inhibitors through phylogenetics and drug repurposing

Anamika Mishra[1] · **Viswajit Mulpuru**[1] · **Nidhi Mishra**[1]

Received: 12 April 2022 / Accepted: 12 July 2022 / Published online: 26 July 2022

Abstract

The novel coronavirus that has affected the whole world is declared a pandemic by the World Health Organization. Since the emergence of this virus, researchers worldwide have searched for potential antivirals against it. Being an RNA virus, it shows a high rate of mutability and variability in its genome. In the present study, all the reported SARS-CoV-2 genomes isolated from diverse regions of the world available in the GISAID database have been considered for phylogenetic analysis. The strain identified at the root is subjected to phylogenetic analysis with genomes of other known human viruses obtained from NCBI for identifying the nearest viral neighbor. Furthermore, the phylogenetic relationship between various human viruses was used to repurpose the known antiviral drugs towards coronavirus using in silico docking approach. The phylogeny reveals the link of the COVID virus with adenovirus. The known drugs against adenovirus are considered in the present study for drug repurposing through molecular docking analysis. The reference inhibitors of the respective targets were also considered in the docking study. The protein targets, namely protease, endoribonuclease, methyltransferase, phosphatase, and spike protein, are considered for screening with the known drug of adenovirus. Ribavirin, known to treat adenoviral infection, shows the best docking score, suggesting its use as a repurposed drug to treat SARS-CoV-2. Furthermore, the potency of the ribavirin drug is analyzed using molecular dynamics studies.

Keywords COVID-19 · Docking · Drug repurposing · Molecular dynamics · Phylogenetics · SARS-CoV-2 · Virtual screening

Introduction

COVID-19 was declared a global pandemic by the World Health Organization in March 2020. It is caused by a novel Betacoronavirus officially known as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) by the International Committee on Taxonomy of Viruses (ICTV) [1, 2]. Coronaviruses that belong to the order Nidovirales, family Coronaviridae, genus Betacoronavirus, and subgenus Sarbecovirus, are single-stranded RNA viruses and are known to infect various organisms, including Aves and mammals. SARS-CoV-2 belongs to the clad of coronavirus along with severe acute respiratory syndrome coronavirus (SARS-CoV), which was reported in 2002 in Guangdong Province of China. SARS-CoV was known to spread to 26 countries, causing 774 deaths. The Middle East respiratory syndrome coronavirus (MERS-CoV) is another virus in the same family that was first reported in 2012 in Saudi Arabia. MERS-CoV is known to spread to 27 countries, causing 858 deaths [2]. Being an RNA virus, coronaviruses evolve rapidly and are more prone to frequent mutations. The clinical manifestation of the coronavirus includes pneumonia, fever, dry cough, headache, dyspnea, and diarrhea. The severe symptoms include metabolic acidosis, septic shock, and bleeding [2].

COVID-19 has shown a critical effect on the world, resulting in over 6 million deaths with 514 million confirmed cases as of May 2022. Alongside many deaths, the pandemic has also resulted in the loss of livelihoods, which has a rippling effect on the global economy. The WHO epidemiological report shows that as of December 2021, there are five different SARS-CoV-2 strains of interest, namely alpha, beta, gamma, delta, and omicron, first reported in the UK, South Africa, Brazil, India, and South Africa respectively. The current fatality rate estimate of

✉ Nidhi Mishra
nidhimishra@iiita.ac.in

[1] Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj, India

COVID-19 is 2.2%, affected by many factors, including age, underlying health conditions, and severity of illness. By the second week of May 2022, it is reported that there was a 14% increase in the number of new weekly cases in the region of the Americas and an increase of 12% in the African region. The weekly death rates of the African region were also reported to show a rise of 84%.

Since the emergence of SARS-CoV-2, researchers have been interested in the phylogenetic analysis of the virus. Phylogenetic analysis is carried out for the taxonomical classification, establishing the ancestry and the relation of a particular organism with other organisms, along with the evolutionary history of an organism. The DNA or protein sequences are used for studying phylogeny and are depicted by constructing the phylogenetic tree using mathematical graphs representing the evolutionary relationships among different organisms [3]. The phylogenetic study of SARS-CoV-2 shows its relation with bat-derived viruses (bat-SL-CoVZC45 and bat-SL-CoVZXC21) and is known to offer more similarity to these viruses than to other human infection-causing viruses [4]. As stated, SARS-CoV-2 is an RNA virus with a high rate of mutation, resulting in various virulent strains differing in their genome from region to region. In the present study, firstly, the phylogeny of the human viruses from the Coronaviridae family and the SARS-CoV-2 variants was established. Furthermore, the phylogenetic relationships among SARS-CoV-2 from different areas of the world, along with their level of variations, have been established, followed by the establishment of the phylogeny of SARS-CoV-2 with other human viruses. The search for the phylogenetic relation between SARS-CoV-2 and other human viruses would help in drug repurposing by screening for their effectiveness against SARS-CoV-2, thus decreasing the time and cost required for identifying drugs against the virus. The genomic data of SARS-CoV-2 and other human viruses were used to construct a phylogenetic tree. The genomic data of SARS-CoV-2, along with the variants, were obtained from GISAID [5], a database containing the genomic data of SARS-CoV-2 deposited by researchers and clinicians worldwide [6]. The genomic data of other human viruses were downloaded from the virus repository of the NCBI genome browser. The genome of the origin node of the SARS-CoV-2 was then compared with the genome of other human viruses. We have used two distance-based methods, an unweighted pair group method with arithmetic mean (UPGMA) and neighbor-joining (NJ), to construct a phylogenetic tree.

Developing a new drug is always a challenging task that requires both time and extensive effort, so using a previously known drug against a particular disease can be used to treat other diseases by drug repurposing. The various advantages of drug repurposing in search of treatment for a specific disease are fewer chances of failure from a safety point of view, reduction in time, and cost-effectiveness [7].

The known drugs of the closely related virus to the SARS-CoV-2, identified through phylogenetic analysis, are used for docking against five common targets of SARS-CoV-2, main protease (PDB ID: 6LU7), NSP15, which is an endoribonuclease (PDB ID: 6VWW), spike protein (PDB ID: 6LZG), ADP-ribose phosphatase of NSP3 (PDB ID: 6VXS), and methyltransferase-stimulatory factor complex of NSP16 and NSP 10 (PDB ID: 6W61). The first protein used for docking analysis was the main protease (PDB ID: 6LU7), the most common drug target in SARS-CoV-2. The processing of polyproteins translated by the viral genome is carried out by this enzyme. The viral replication can be obstructed by impeding the activity of this enzyme [8]. Another protein is NSP15, an endoribonuclease (PDB ID: 6VWW); it cleaves RNA at the 3′ uridylate position, forming a 2′-3′ phosphodiester product. The activity of NSP 15 is to prevent the immune sensing system of the host from detecting the virus by targeting and degrading the polyuridine sequence of the virus. Manganese ($Mn2+$) is needed for the activity of NSP15 [9]. The spike protein of SARS-CoV-2 (PDB ID: 6LZG) is a glycoprotein that helps mediate the entry of the virus to the host cell. The protein interacts with human angiotensin-converting enzyme 2 (ACE2) present on the cell membrane of the host cell; this interaction is responsible for the transmission of the virus in the host cell [9, 10]. The next protein is ADP-ribose phosphatase of NSP3 (PDB ID: 6VXS) from SARS-CoV-2; NSP3 is a papain-like proteinase. It is the largest protein having several conserved domains, including the transmembrane domain encoded by a coronavirus. NSP3 has the protease activity responsible for cleaving the site between NSP2 and NSP3. Apart from this, protease activity mediates the release of NSP1, NSP2, and NSP3 from the N-terminal region of polyprotein 1a and 1ab from coronaviruses. The release of these proteins is essential for viral activity, and inhibition of NSP3 can help to fight against SARS-CoV-2 [9]. Methyltransferase-stimulatory factor complex of NSP16 and NSP 10 (PDB ID: 6W61) of SARS-CoV-2 that show methyl transferring activity, i.e., viral RNA capping, is also used as a protein target for docking analysis. mRNA in eukaryotes and most viruses are capped at the 5′ end, helping in RNA splicing, transportation of mRNA, maintaining stability, and initiation of translation. Host or viral mRNA without capping is prone to rapid degradation. The viruses require capping to protect themselves from the host's innate immune responses. It also helps in viral replication by enhancing viral translation to escape from host RNA sensors. NSP 10 acts as a stimulatory factor for NSP 16, a 2′-O-methyltransferase (2′-O-MTase); it stabilizes the SAM binding pocket and extends the substrate RNA binding groove of NSP16 by preventing their interaction 2′-O-MTase activity that can be inhibited [11].

Drug repurposing could serve efficiently in finding the treatment for SARS-CoV-2 infection, and the drugs used to treat the diseases caused by the nearest phylogenetic neighbor of SARS-CoV-2 can be considered for targeting SARS-CoV-2 infection. In the present work, the phylogeny of SARS-CoV-2 has been analyzed, and the known drug of its nearest neighbor is considered for drug repurposing using various computational techniques such as sequence alignment, virtual screening, and molecular dynamics simulations.

Material and methods

Retrieval and filtering of the coronavirus genome sequences

The coronavirus genome sequences reported in the GISAID, a primary source for influenza and novel coronavirus genomic data, were analyzed in this study. About 30,800 sequences were downloaded from the EpiCoV repository of the GISAID in the FASTA format. To filter out similar sequences from the downloaded sequences, a single genome was picked randomly from each region, namely Africa, Asia, Central America, Europe, North America, Oceania, and South America. The complete set of downloaded sequences was subject to similarity matching with the seven sequences picked randomly from the stated regions. The latest complete and high coverage sequences from all the available variants of SARS-CoV-2 were also downloaded from GISAID along with the genomic sequences of all viruses from the Coronaviridae family with humans as a host from NCBI. To visualize the evolutionary history, the downloaded sequences were subjected to sequence alignment to construct a phylogenetic tree using the neighbor-joining method. Locality sensitive hashing (LSH) technique [12] was used to calculate the similarities between the studied genomes using the sequences picked randomly from each region as a reference. Using the similarity score generated by the LSH, all the sequences that showed less than 70% similarity towards reference genomes were subjected to further study. Three thousand one hundred fifty-one genomic isolates were identified using this approach.

Sequence alignment and phylogenetic analysis of the coronavirus genome

To perform the phylogenetic analysis on the obtained sequences, the genomic sequence alignment was performed using the MEGAX software and Clustal Omega [13]. Furthermore, the aligned genomic sequences were subjected to Biopython to generate the distance matrix and the phylogenetic tree. The tree has been generated from the distance matrix using the UPGMA method. As it is known that UPGMA is inherently rooted at the deepest point of the tree, the genome corresponding to the deepest point is selected for further analysis. The phylogeny of all the viruses infecting the humans from Coronaviridae family, along with various variants of SARS-CoV-2, was generated using the NJ method.

Retrieval of all the known human virus genomes

The NCBI genome browser was used to obtain the genomes of various viruses that infect humans. The virus repository of the NCBI genome browser was filtered for viruses with humans as a host, and the complete genomes of these viruses were downloaded using the FTP service provided by NCBI. A total of 558 viral genome FASTA files were downloaded using this method for further study.

Comparison of human viral genomes with that of coronavirus using phylogenetic techniques

For comparison of the human viral genomes with that of coronavirus, all the viral genomes obtained in this study were aligned with the genome of the coronavirus strain identified at the root node during the phylogenetic analysis of various coronavirus strains. The viral sequences were aligned using the MEGAX software, and the phylogenetic tree of these aligned genomes was constructed using Biopython employing the neighbor-joining method.

Identification of the most similar human virus for drug repurposing

The phylogenetic tree constructed using the neighbor-joining method is further analyzed to identify the viral genome that is most similar to the coronavirus strain studied in this work, which has been used for further analysis. Based on the distances, it has been identified that adenovirus is the nearest neighbor to the coronavirus. Furthermore, various drugs that help in the treatment of adenoviral infections are identified through literature search [14], and these drugs were subjected to docking studies to identify their potency towards coronavirus. ETE toolkit was used to visualize all the phylogenetic trees [15].

Docking analysis

For identification and repurposing of anti-adenovirus drugs towards coronavirus, the identified drugs from the literature search [14] as shown in Table 1 were subjected to docking studied using Glide, an efficient docking tool from Schrodinger. The

Table 1 List of drugs used as ligands for docking analysis

S. no.	Name of the ligands
1	Cidofovir [(S)-HPMPC; (S)-1-(3-hydroxy-2-phosphonomethoxypropyl)cytosine; VISTID]
2	Brincyclovir (BCV; hexadecyloxypropyl-cidofovir; CMX001)
3	(S)-HPMPA [(S)-9-(3-hydroxy-2-phosphonomethoxypropyl)adenine]
4	USC-187 (alkyl tyrosinamide-ester prodrug of HPMPA)
5	(S)-HPMPO-DAPy [2,4-diamino-6-[3-hydroxy-2-(phosphonomethoxy)-propoxy]pyrimidine]
6	(S)-2242 [2-amino-7-(1,3-dihydroxy-2-propoxymethyl)purine]
7	Ganciclovir (GCV)
8	Zalcitabine (2′3′-dideoxycytidine, ddC)
9	Alovudine
10	Trifluridine (3FT) and vidarabine (Vira-A)
11	Ribavirin (1-ß-D-ribofuranosyl-1,2,4-triazole-3-carboxamide)

MG-101 (PubChem ID: 443118), NSC95397 (PubChem ID: 262093, ZINC000000082673, ZINC000038661771, and ZINC000023398144), and ellagic acid (PubChem ID: 5281855) were considered reference inhibitors for main protease, NSP 15 an endoribonuclease, ADP-ribose phosphatase of NSP3, methyltransferase complex of NSP16 and NSP 10, and spike protein respectively [16–20]. The binding affinity, glide score, and ligand interaction diagrams are analyzed to identify the potential drugs that could be repurposed against coronavirus. The three-dimensional structures of the drugs were optimized using LigPrep, and the tautomer of the drugs used as ligands was generated using Epik. The affinity grid maps of the binding site were generated using the available ligand within the crystallized structure, where available. The Sitemap module of the Schrodinger suite was used to identify the active sites where the known binding sites were unavailable. The coronavirus proteins considered in this study are spike protein, methyltransferase, protease, endoribonuclease, and phosphatase.

Molecular dynamics simulation

The target protein with the best docking score was subjected to MDS to validate the stability of protein–ligand interaction. MDS was performed for 50 ns on GROMACS (version 2018.2) [21] using the CHARMM36 all-atom force field [22]. SwissParam web server [23] was used to generate ligand topology as the force field used for simulation (CHARMM36) lacked the force field parameters for ligands. GROMACS compatible protein and ligand files were generated with in-house ad hoc scripts before the protein, and ligand topologies were combined, solvated, minimized, and equilibrated. The protein–ligand complex was solvated with a TIP3P explicit model of water molecules. The covalent bonds of the particles were using a linear constraint solver (LINCS) algorithm. Later, the system was neutralized using Cl− and Na + ions. The energy minimization of the system was performed using the steepest descent algorithm until the maximum force is less than 10.0 kJ/mol evaluated using the particle mesh Ewald electrostatic interactions. Furthermore, the NVT and NPT conserved ensembles were generated and equilibrated at 1 bar pressure and 300 K temperature using the Berendsen thermostat and Berendsen pressure coupling algorithm. The MD simulations were performed using the leapfrog algorithm with an integration time step of 2 fs for 50 ns at constant temperature and pressure of 300 K and 1 bar, respectively.

Trajectory analysis

The molecular dynamic trajectories were analyzed using GROMACS analysis utilities to derive and conclude results. The stability of the protein complex was determined by calculating the root mean square deviation (RMSD) between the initial structure and the simulated structure. At the same time, the calculation of RMSF was carried out to determine the rigidity of the secondary structure. The analysis toolkit was also used to generate the PDB files of the protein–ligand complex to analyze the protein–ligand interactions at various simulation periods. The strength of binding between protein and ligand was established by analyzing the change in interaction energies between protein and ligand throughout the simulation time. The stability of the complex was further validated by analyzing modifications in secondary structure per residue versus time during the simulation period using the DSSP algorithm [24]. The timeline analysis of the evolution of the secondary structure was plotted using Ghostscript. The UCSF Chimera software [25] was used for structural alignment and visualization. MATLAB was used to plot graphs.

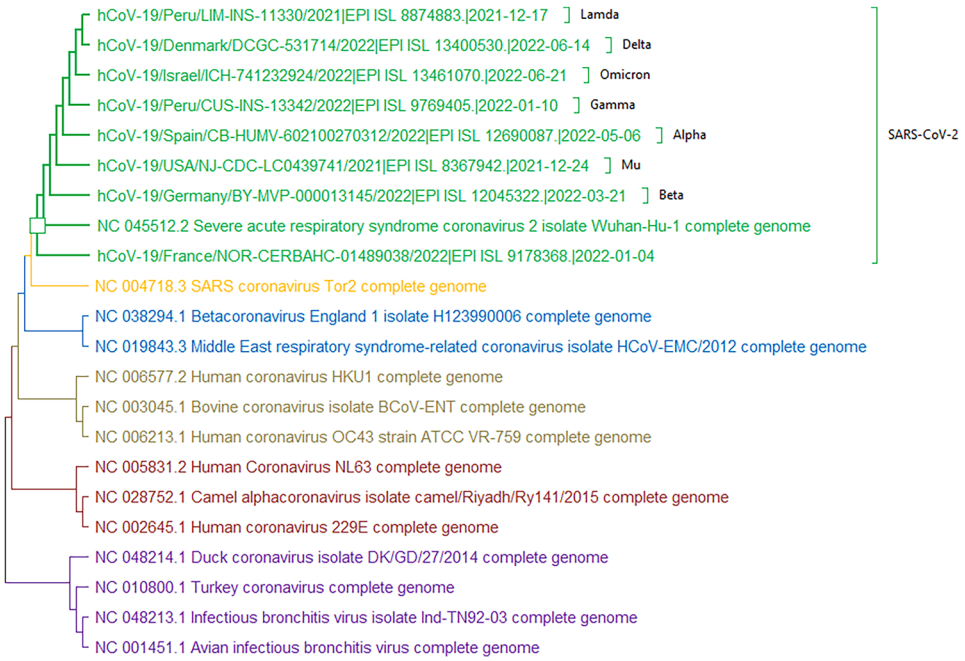

Fig. 1 The phylogenetic tree representing the evolutionary history of the coronavirus

Results and discussion

Phylogenetic analysis of coronavirus genome

Phylogenetic studies play an essential role in forecasting an organism's evolutionary history [26]. Analyzing the phylogeny of coronavirus would provide information on its evolutionary history and the most closely related neighbor. Firstly, the phylogenetic tree of the genomic sequences of the viruses from the Coronaviridae family, along with various variants of the SARS-CoV-2, is constructed (Fig. 1). The phylogenetic tree shows that coronavirus shows a high rate of mutation with respect to time. It also shows that the first coronavirus strain that evolved to infect humans is from the avian population.

Phylogenetic analysis of a large number of SARS-CoV-2 strains from different regions of the world revealed that there had been a variety of variations in the genomes isolated in the different areas. To filter out one genome from these large numbers of strains, the corresponding genome nearest to the root node as calculated through phylogenetic analysis using the UPGMA method is selected for further study. The analysis of the phylogenetic tree reveals that the strain isolated from China (Asia) is calculated to be nearest to the root node of all the strains used in this study. The table showing the five least distant strains excluding animals from the root as calculated by the UPGMA method is shown in Table 2. The phylogenetic tree constructed using the identified strains is shown in the figure (Online Resource 1). From the phylogenetic tree, it can be deduced that the isolates from the Europe region show large variations in their genome.

Comparison of human viral genomes with coronavirus

The phylogenetic analysis of the SARS-CoV-2 strain identified in the previous step and the human viral genomes downloaded from the NCBI virus repository revealed that the adenovirus is the nearest neighbor to the coronavirus. The table showing the five least distant viruses concerning coronavirus as calculated by the neighbor-joining method

Table 2 Distances of the top five nearest strains from the root node along with isolates from bat and pangolin

Region	Distance from root Node
Bat (China)	0.409678
China (Asia)	0.409678
Wuhan (Asia)	0.409678
Pangolin (China)	0.409678
Colombia (SouthAmerica)	0.432357
Wales (Europe)	0.432357
Italy (Europe)	0.432357

Table 3 Distances of the top five nearest strains from the root node along with isolates from bat and pangolin

Accession no	Name	Relative distance
Studied strain	Coronavirus	0
AY597011.2	Human coronavirus	−0.006943
KX868297.2	Human adenovirus	0.015946
AY567487.2	Human coronavirus	−0.016879
FJ217161.1	Bundibugyo ebolavirus	−0.052827
AB543336.1	Human parainfluenza virus	−0.061091

can be seen in Table 3. The phylogenetic tree reveals the similarity between various human viruses and the coronavirus, as shown in Fig. 2. It is evident from the table that SARS-CoV-2 shows the nearest similarity with adenovirus; further as identified from the phylogenetic analysis, the antiviral drugs that help against adenoviral infections [14] are studied for repurposing them towards SARS-CoV-2 target proteins.

Docking analysis

Docking is one of the most important computational tools commonly used to study the interaction of the small, potent drug molecules against targets and is known to have significant importance in drug discovery [27]. Docking analysis was carried out by Maestro Schrödinger in 2019 to find the strength of the interaction between the known drugs of

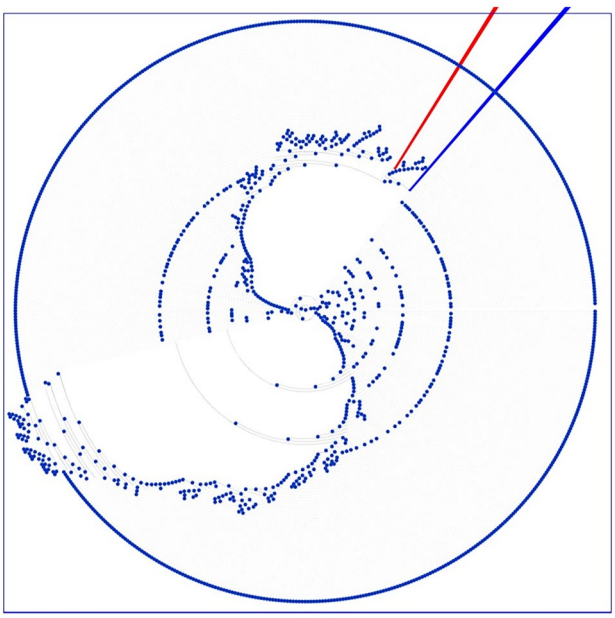

Fig. 2 The phylogenetic tree is generated by the NJ method. The blue and red pointers point to the studied coronavirus strain and adenovirus, respectively

Table 4 Top five nearest neighbors to coronavirus as calculated by the NJ method along with their accession numbers

Protein	Best-docked ligand along with reference inhibitors	Docking score of best-docked ligand and reference inhibitors	Glide emodel	Interacting residues of top-scored repurposed drug with the respective target as shown in Fig. 3
Main protease (PDB ID: 6LU7)	Ribavirin	−5.923	−45.803	THR292, PRO293, PHE294, ILE249, HIE246, VAL202, GLU240, ILE200, GLN110, GLY109, PRO108, GLN107
	MG-101	−6.556	−67.893	
NSP 15 an endoribonuclease (PDB ID: 6VWW)	NSC95397	−4.100	−77.062	ASN6899, LEU6898, ASP6897, SER6896, GLY6911, ASP6912, CYS6913, VAL6916, PHE6868, GLY6869, PHE6947, ASP6931, TYR6930, TYR6930, MET6929
	Alovudine	−6.654	−46.099	
ADP-ribose phosphatase of NSP3 (PDB ID: 6VXS)	Alovudine	−6.8	−46.099	GLY46, GLY47, GLY48, VAL49, ALA50, ASN40, ALA39, ALA38, PRO125, LEU126, LEU127, ALA129, GLY130, ILE131, PHE132
	ZINC000000082673	−9.348	−50.175	
Methyltransferase complex of NSP16 and NSP 10 (PDB ID: 6W61)	ZINC000038661771	−4.091	−61.577	GLU69, LYS71, LYS90, THR167, THR196, SER198, ARG199, ASN200 LEU252, ASP273, SER274, THR275, VAL276, LYS277, TYR279, VAL295, ILE296, ASP297
	ZINC000023398144	−2.353	−72.946	
	Ribavirin	−9.451		
Spike protein (6LZG)	Ellagic acid	−5.103	−85.041	TYR365, SER366, LEU368, TYR369, PHE377, CYX379, VAL382, SER383, PRO384, THR385, LEU387, ASN388, ASP389, LEU390, PHE392, PHE515, LEU434, CYX432, CYX525
	Ribavirin	−4.96	−11.348	

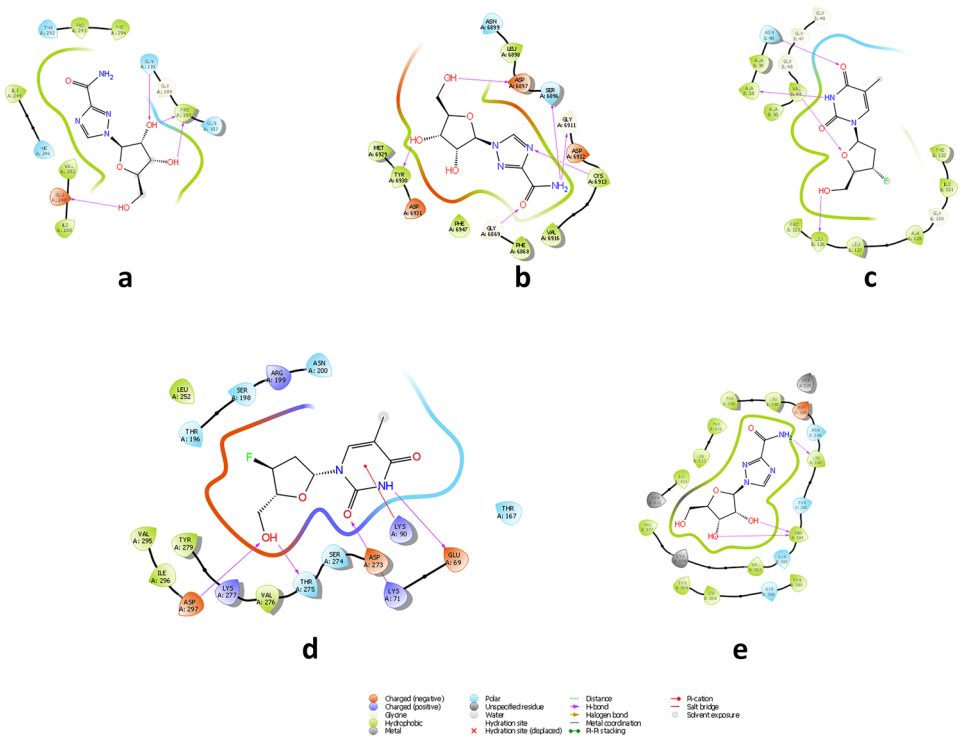

Fig. 3 The interaction diagrams correspond to each protein's top-scoring ligand. **A** Interactions between protease and ribavirin. **B** Interactions between methyltransferase and ribavirin. **C** Interactions between phosphatase and alovudine. **D** Interactions between endoribonuclease and alovudine. **E** Interactions between spike protein and ribavirin

adenovirus infection along with known inhibitors against the structures of target proteins, namely spike protein, methyltransferase, protease, endoribonuclease, and phosphatase of SARS-CoV-2. The docking score of top-scored drugs is mentioned in Table 4. As evident from the docking score, ribavirin, an antiviral agent used against adenovirus, shows promising binding affinity with methyltransferase with a docking score of − 9.451 kcal/mol and can be considered for repurposing against coronavirus. The interaction diagrams corresponding to the top-scoring ligand of each protein are shown in Fig. 3.

Molecular dynamics simulation

As stated, it was subjected to molecular dynamic analysis to study the ligand's inhibitory activity further. The RMSD plot of the protein–ligand complex reveals the stability of the compound. The RMSD plot is shown in Fig. 4. To further evaluate the stability of protein, the RMSF and the change in interaction energies are analyzed. The RMSF plot, as seen in Fig. 5, reveals that the changes in protein structure throughout the simulation were within the threshold concerning the ligand. As stated, the strength of nonbonded interaction between methyltransferase and ribavirin was quantified by calculating interaction energy between them as represented in Fig. 6; the graph shows a stable interaction confirming the stable interaction between methyltransferase and ribavirin.

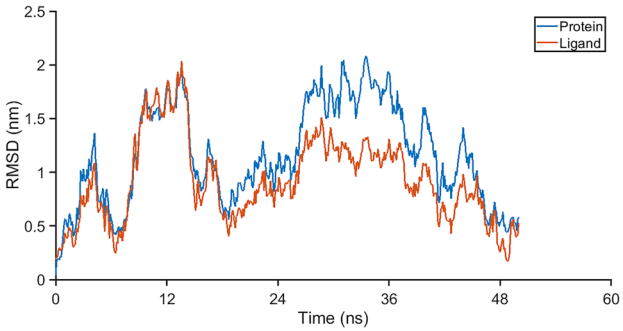

Fig. 4 RMSD vs. time graph of protein and ligand

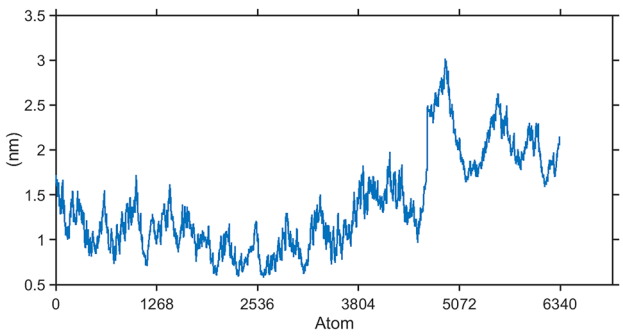

Fig. 5 RMSF plot of protein

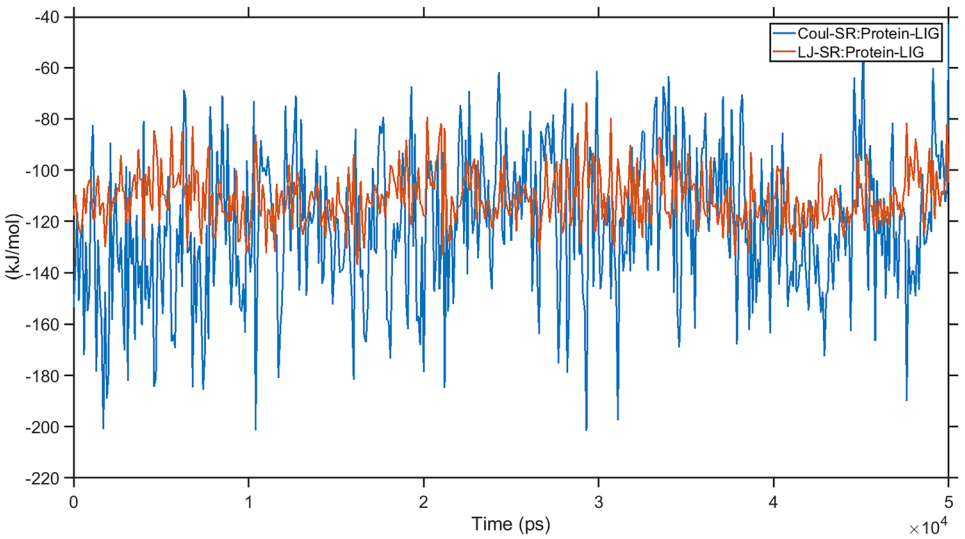

Fig. 6 Nonbonded interaction energy graph between protein and ligand

When considering diverse genomes from coronavirus, it is seen that the root of the phylogenetic tree lies at the genome isolated from the bat. Further phylogenetic analysis of various human virus genomes shows that the coronavirus is most closely related to adenovirus. The virtual screening of adenoviral drugs against coronavirus suggests that ribavirin can be repurposed against different target proteins of coronavirus.

Conclusions

The emergence of SARS-CoV-2 as a pandemic in the world demands the search for an efficient potential drug for the treatment of COVID-19. This study attempted to identify potential antiviral drugs for the SARS-CoV-2 using a bioinformatics approach. The phylogenetic analysis of the SARS-CoV-2 strains obtained from GISAID suggests that the samples identified from the European region showed large diversity in their genome as compared to other regions. The phylogenetic analysis also reveals that the coronavirus strains infecting the humans emerged from the avian population. The SARS-CoV-2 strain isolated at Wuhan, China can be identified as the strain corresponding to the root node as calculated by the UPGMA method. Furthermore, the phylogenetic analysis of various human viral genomes obtained from the NCBI genome browser confirms the close relation between SARS-CoV-2 and adenovirus. Considering adenovirus as the nearest neighbor to the coronavirus, as suggested by the phylogenetic analysis, the literature search for drugs against adenovirus reveals diverse small molecules. For drug repurposing studies, the molecular docking analysis of the identified small molecules towards various protein structures corresponding to coronavirus suggests that ribavirin, a known antiviral drug against adenovirus infection, can be repurposed against coronavirus. The molecular dynamic simulation studies further confirm the potency of ribavirin towards coronavirus and could help find the treatment for COVID-19.

Supplementary Information The online version contains supplementary material available at https://doi.org/10.1007/s11224-022-02019-6.

Acknowledgements The research work was carried out in the laboratory of Chemistry and Nanotechnology, Department of Applied Sciences, Indian Institute of Information Technology Allahabad, Prayagraj, India. The molecular dynamic simulation results reported in this work were performed on the Central Computing Facility of IIITA, Prayagraj, India.

Author contribution A. M. and N. M. designed the study, A. M. took part in the docking studies, and A. M and V. M. took part in the phylogenetic analysis and molecular dynamics simulation studies. A. M and V. M wrote the main manuscript and prepared the figures. N. M. reviewed the manuscript. All authors read and approved the final manuscript.

Availability of data and material All data generated or analyzed during this study are included in this published article.

Code availability Not applicable.

Declarations

Competing interests The authors declare no competing interests.

References

1. Lupia T, Scabini S, Mornese Pinna S, Di Perri G, De Rosa FG, Corcione S (2020) J Glob Antimicrob Resist 21:22–27
2. Helmy YA, Fawzy M, Elaswad A, Sobieh A, Kenney SP, Shehata AA (2020) JCM 9:1225

3. Choudhuri S, Kotewicz M (2014) Bioinformatics for beginners: genes, genomes, molecular evolution, databases, and analytical tools, Elsevier/AP, Amsterdam ; Boston
4. Lu R, Zhao X, Li J, Niu P, Yang B, Wu H, Wang W, Song H, Huang B, Zhu N, Bi Y, Ma X, Zhan F, Wang L, Hu T, Zhou H, Hu Z, Zhou W, Zhao L, Chen J, Meng Y, Wang J, Lin Y, Yuan J, Xie Z, Ma J, Liu WJ, Wang D, Xu W, Holmes EC, Gao GF, Wu G, Chen W, Shi W, Tan W (2020) Lancet 395:565–574
5. Shu Y, McCauley J (2017) Euro Surveill 22. https://doi.org/10.2807/1560-7917.ES.2017.22.13.30494
6. Forster P, Forster L, Renfrew C, Forster M (2020) Proc Natl Acad Sci USA 117:9241–9243
7. Pushpakom S, Iorio F, Eyers PA, Escott KJ, Hopper S, Wells A, Doig A, Guilliams T, Latimer J, McNamee C, Norris A, Sanseau P, Cavalla D, Pirmohamed M (2019) Nat Rev Drug Discov 18:41–58
8. Zhang L, Lin D, Sun X, Curth U, Drosten C, Sauerhering L, Becker S, Rox K, Hilgenfeld R (2020) Science 368:409–412
9. Yoshimoto FK (2020) Protein J 39:198–216
10. Walls AC, Park Y-J, Tortorici MA, Wall A, McGuire AT, Veesler D (2020) Cell 181:281-292.e6
11. Maurya AK, Mishra N (2020) J Biomol Struct Dyn 1–16
12. Indyk P, Motwani R (1998) In: Proceedings of the Thirtieth Annual ACM Symposium on Theory of Computing STOC '98. ACM Press, Dallas, Texas, United States, pp 604–613
13. Kumar S, Stecher G, Li M, Knyaz C, Tamura K (2018) Mol Biol Evol 35:1547–1549
14. Galabov AS (2019) Chemotherapy of Adenovirus Infections. In: Desheva Y (ed) Adenoviruses. IntechOpen
15. Huerta-Cepas J, Serra F, Bork P (2016) Mol Biol Evol 33:1635–1638
16. Narayanan A, Narwal M, Majowicz SA et al (2022) Identification of SARS-CoV-2 inhibitors targeting Mpro and PLpro using in-cell-protease assay. Commun Biol 5:169. https://doi.org/10.1038/s42003-022-03090-9
17. Canal B, Fujisawa R, McClure AW et al (2021) Identifying SARS-CoV-2 antiviral compounds by screening for small molecule inhibitors of nsp15 endoribonuclease. Biochem J 478:2465–2479. https://doi.org/10.1042/BCJ20210199
18. Patel DC, Hausman KR, Arba M et al (2022) Novel inhibitors to ADP ribose phosphatase of SARS-CoV-2 identified by structure-based high throughput virtual screening and molecular dynamics simulations. Comput Biol Med 140:105084. https://doi.org/10.1016/j.compbiomed.2021.105084
19. Bobrovs R, Kanepe I, Narvaiss N et al (2021) Discovery of SARS-CoV-2 Nsp14 and Nsp16 methyltransferase inhibitors by high-throughput virtual screening. Pharmaceuticals 14:1243. https://doi.org/10.3390/ph14121243
20. David AB, Diamant E, Dor E et al (2021) Identification of SARS-CoV-2 receptor binding inhibitors by in vitro screening of drug libraries. Molecules 26:3213. https://doi.org/10.3390/molecules26113213
21. Van Der Spoel D, Lindahl E, Hess B, Groenhof G, Mark AE, Berendsen HJC (2005) J Comput Chem 26:1701–1718
22. Huang J, MacKerell AD (2013) J Comput Chem 34:2135–2145
23. Zoete V, Cuendet MA, Grosdidier A, Michielin O (2011) J Comput Chem 32:2359–2368
24. Kabsch W, Sander C (1983) Biopolymers 22:2577–2637
25. Pettersen EF, Goddard TD, Huang CC, Couch GS, Greenblatt DM, Meng EC, Ferrin TE (2004) J Comput Chem 25:1605–1612
26. Wiley EO (2010) Why trees are important. Evo Edu Outreach 3:499–505. https://doi.org/10.1007/s12052-010-0279-0
27. Meng XY, Zhang HX, Mezei M, Cui M (2011) Molecular docking: a powerful approach for structure-based drug discovery. Curr Comput Aided Drug Des 7(2):146–157