

EVOLUTIONARY RELATIONSHIP AND REPURPOSING OF SARS INHIBITORS AGAINST SURFACE GLYCOPROTEIN OF SARS-COV-2

PRIYANKA RAJBHAR^a, DIKSHANT SINGH^b, RUKHSAR FATIMA^c, RUCHI YADAV^{d*}

Amity Institute of Biotechnology, Amity University Uttar Pradesh, Lucknow Campus, Lucknow 226028

Email: ryadav@lko.amity.edu

Received: 25 Mar 2021, Revised and Accepted: 19 Jun 2021

ABSTRACT

Objective: Coronaviruses are a group of similar viruses which cause fatal infection and responsible for affecting the upper respiratory tract in many organisms. Throughout the time these viruses have been found to affect human life by causing major pandemics like SARS, MERS and COVID-19 due to their high rate of mutation and zoonotic transmission. Repurposing of a drug could be a solution for this challenge, as many previously available drugs hold great potential to act as a drug molecule. Interfering this interaction could be a potent mechanism to stop the viral infection and propagation.

Methods: In the current study we have predicted the evolutionary relationship of nCoV using three viral proteins Nucleocapsid phosphoprotein, membrane glycoprotein and Envelop protein with accession number YP_009724397, YP_009724393 and YP_009724392 respectively. Phylogenetic tree was constructed and evaluated using the bootstrap method. Homology modeling and docking studies has been done to identify the interaction and binding affinity of SARS drugs.

Results: Phylogenetic tree shows that Nucleocapsid phosphoprotein is originated from Hypsugo Bat Coronavirus, Membrane glycoprotein is originated from MERS Corona Virus and Envelop proteins have originated from Ferret coronavirus. From the docking result we concluded that Precose (glide score-8.372) shows that it has stable and strong interaction with Spike glycoprotein.

Conclusion: Precose which is commonly known as Acarbose can act as a potential inhibitor for the spike glycoprotein. This paper described and highlighted the importance of repurposing of the previously available drug to act as a potent inhibitor in the newly discovered or novel diseases.

Keywords: COVID 19, SARS, MERS, Phylogenetic analysis, Spike glycoprotein, Homology modelling, Docking

© 2021 The Authors. Published by Innovare Academic Sciences Pvt Ltd. This is an open access article under the CC BY license (<https://creativecommons.org/licenses/by/4.0/>)
DOI: <https://dx.doi.org/10.22159/ijap.2021v13i5.42652>. Journal homepage: <https://innovareacademics.in/journals/index.php/ijap>

INTRODUCTION

Recently a severe respiratory disease was reported in Wuhan, china which is caused by a novel, or new corona virus (nCoV). These Corona Viruses causes illness and diseases ranging from common cold, fever, breathing difficulties [1]. Severe infection of CoV can cause pneumonia, kidney failure, severe acute respiratory syndromes and can even lead to death. These corona viruses (CoVs) are the largest group of viruses that belongs to the family of Coronaviridae, it also involves MERS-CoV (middle east respiratory syndrome), SARS-COV (severe acute respiratory syndrome) and nCoV (novel corona virus) [2]. These viruses can easily be transmitted between people and animals like MERS-CoV was transmitted from dromedary camels to humans, likewise, SARS-CoV was transmitted from civet cats to humans and nCoV was transmitted from bat to humans.

Currently, the sequence of nCoV has been deposited in NCBI database with accession no NC_045512, it is a single-stranded RNA with the genome size of 29903 base pair. The genome of this virus ranges in between 26-32 Kb, which is largest amongst the RNA Viruses [3]. One of the most prominent features of these coronaviruses is the club-shaped spike projections present on their surface, and these spikes projections are the defining feature of the novel corona viruses [4]. These coronaviruses contain four main structural proteins, and these are spike(s), Nucleocapsid (N), Membrane(M), and Envelop (E) protein. All these proteins are encoded within the 3' end of the viral genome. Among this structural proteins, Nucleocapsid protein is the largest, which has the length of 419 amino acid and envelop protein is the smallest which has the length of 75 amino acid [5].

The genome of these viruses is packed inside helical shaped capsid, which is formed by the nucleocapsid protein(n) and then further it is covered by an envelope. Nucleocapsid is majorly involved in processes which is related to the viral genome, although it is also involved with other aspects of the replication cycle of the corona

viruses and also with the cellular response of host to the viral infection [6]. Membrane protein is the abundant structural protein, as the shape of viral envelope is only defined by this protein. It is also called as central organizer of the coronavirus assembly, as it interacts with the other major structural proteins of the corona virus.

The new coronavirus identified in Wuhan, the capital of China's Hubei province in December 2019 showed similar symptoms as SARS-CoV and MERS-CoV and people infected with this virus suffered a severe inflammatory response. The World health organization has named the new coronavirus as 2019-nCoV which later got changed into SARS-CoV-2 or COVID 19. COVID 19 is classified as zoonotic viral disease similar to SARS-CoV and MERS-CoV which means that the patient who were infected acquired these viruses directly from animals. And is mainly transmitted through air and infects the respiratory and gastrointestinal tract of mammals and birds [7].

The name of the coronavirus comes from its resemblance to solar crown or corona-like appearance [Almeida JD, Berry DM, 1968]. The viruses are enveloped non-segmented positive-sense RNA viruses, 27-32kb in size. The virus belongs to the family *Coronaviridae* and the order *Nidovirale* and divided into four Genera i.e. alpha, beta, gamma and delta. The virus responsible for COVID-19 belongs to beta coronavirus just like as SARS and MERS [8]. The spike protein is mainly consisting of S1, S2 and S2' subunit which play important role in viral infection [9].

S1: It mainly help in the attachment of viral particle to the host cell receptor (ACE 2). Binding further led the viral particle into the endosome of host cell and induce conformational changes in spike glycoprotein.

S2: It act as a class I viral fusion protein and mediate the fusion of the virus and cell membrane. During the process of fusion, the coiled region of protein begins to form a trimer of a hairpin structure, and

start arranging the fusion peptide in close proximities to the C-terminal region of the ectodomain. And due to the formation and positioning of this structure, subsequent fusion of viral and target cell membranes take place.

ACE2: ACE2 is a type 1 integral membrane protein mainly expressed in endothelium, lungs, kidney, and heart. The extracellular domain of ACE2 enzyme contain a single catalytic metallopeptidase unit which is responsible for converting of Angiotensin 2 to Angiotensin 1-7 and thus play a crucial role in the Renin-Angiotensin system (RAS). Apart from these ACE2 is also associated with integrin function [10].

The virus enters our body through contact with an infected person or due to direct contact with the viral particles and they mainly attack the respiratory system specifically speaking the alveoli. The alveoli consist of two types of cell i.e. pneumocytes I and pneumocytes II, the virus infect the later one because of the presence of ACE2 receptor, which has been found to have a higher affinity for spike protein [11].

The coronavirus has (+) sense ssRNA, and it can enter in the cell by several methods and can proliferate i.e. either through direct translation or through replication and uses the host machinery

proteins in order to continue. Some studies have shown that SARS-CoV 2 secretes at least three virulence factor that is responsible for the production of the new viral particle and suppressing the immune response [12]. Thus, the spike protein act as a main key for entering in the cell and helps in the viral attachment. It also helps in the fusion and allow infection to begin. So, the structural study of spike protein is an important aspect to understand the molecular mechanism of viral infection and could be very important for creating vaccines and for therapeutic drug discovery.

MATERIALS AND METHODS

Phylogenetic tree construction

Protein sequence of Nucleocapsid Phosphoprotein, Membrane Glycoprotein and Envelop Protein [13] was retrieved from GenBank database (<https://www.ncbi.nlm.nih.gov/genbank/>) with the accession number YP_009724397, YP_009724393 and YP_009724392 respectively detail of these sequences is shown in table 1. These sequences were selected to construct phylogenetic tree to understand their evolutionary relationship with other organisms.

Table 1: List of proteins along with accession number that were used to study evolution of CoV-19

S. No.	Protein name	Accession number	Length
1.	Nucleocapsid Phosphoprotien	YP_009724397	419 aa
2.	Membrane Glycoprotien	YP_009724393	222 aa
3.	Envelop Protein	YP_009724392	75 aa

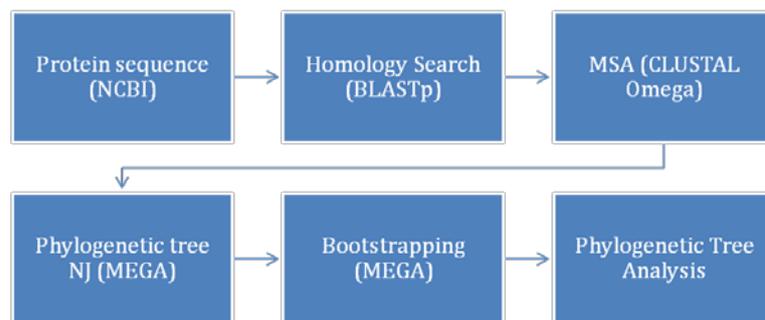


Fig. 1: Methodology used for the construction of phylogenetic tree for all three protein sequences as shown in table 1

Homologous sequence for all protein sequence Nucleocapsid Phosphoprotein, Membrane Glycoprotein and Envelop Protein was obtained using BLASTp tool (<https://blast.ncbi.nlm.nih.gov/Blast.cgi?PAGE=Proteins>) Multiple Sequence Alignment (MSA) was done using Clustal Omega tool (<https://www.ebi.ac.uk/Tools/msa/clustalo/>). MSA result was further used for the construction of phylogenetic tree by Neighbor-Joining [14] method using MEGA Tool (<https://www.megasoftware.net/>). The methodology adopted for this research work, tools and software's was mentioned in fig. 1.

Phylogenetic tree build for all three protein sequences was verified using bootstrap method using MEGA tool [15]. This method gives the bootstrapping values that signifies the relationship among different species and clades information [16]. The bootstrap values were represented on the edges of the phylogenetic tree it is calculated out of 100 replicates.

Structure prediction by homology modelling

The complete genome sequence of SARS-CoV-2 was published in the NCBI database (www.ncbi.nlm.nih.gov) with Accession no. MN908947 under the title-Severe acute respiratory syndrome coronavirus 2 Wuhan-Hu-1. Form this database the sequence of surface glycoprotein was retrieved with Accession no. QHD43416.

Sequence alignment of surface glycoprotein for Homology modeling was done using protein BLAST and the highest aligned sequence was selected to as a template. On the basis of the template, the protein 3D structure was predicted using the Schrodinger software suite version10.4.018 (Schrodinger 2011) [17]. Modelled protein structure was verified by Ramachandran plot analysis using PROCHECK software which was further used for binding site prediction, grid generation and Glide docking [18].

Selection of potential drug compound as ligand

The selection of ligands was done by going through various research papers in the PubChem database (<https://pubchem.ncbi.nlm.nih.gov/>) and the final list was created consisting of six compounds as shown in table 2. These compounds were prepared for docking, using the ligand preparation method of the Schrodinger software suite and docking was done using Glide dock method as implemented in Schrodinger software suite.

Binding site prediction and docking

Binding site of the modelled structure was predicted using the sitemap tool of Schrodinger software suit and the predicted binding site were then used for grid generation. The docking was done using Glide dock tool of Schrodinger software suit. A Ligand-protein interaction map was studied to identify binding properties and efficiency of the selected ligands against these proteins.

Table 2: Ligand used against COVID 19 protein which can act as potential inhibitor

S. No.	Ligand name	Pubchem ID	Molecular formula	Molecular Wt.
1	Precose	444254	C ₂₅ H ₄₃ NO ₁₈	645.6 g/mol
2	N-(1-Naphthyl)-2-(phenylthio)ethanethioamide	3246501	C ₁₈ H ₁₅ NS ₂	309.5 g/mol
3	6MP-Arabinoside	3034423	C ₁₀ H ₁₂ N ₄ O ₄ S	284.29 g/mol
4	(2S)-2-(4-Methoxy-3,5-dimethylphenyl)-5-methyl-2-(3-pyrimidin-5-ylphenyl)-1,3-dihydroimidazol-4-amine	60202318	C ₂₃ H ₂₅ N ₅ O	387.5 g/mol
5	Verapamil	2520	C ₂₇ H ₃₈ N ₂ O ₄	454.6 g/mol
6	2-(Phenethylthio)acetic acid	292540	C ₁₀ H ₁₂ O ₂ S	196.27 g/mol

RESULTS AND DISCUSSION

Phylogenetic analysis

BLAST And MSA Results of Nucleocapsid Phosphoprotein (fig. 2), Membrane Glycoprotein (fig. 3), Envelop Protein (fig. 4). BLAST result obtained from BLASTp tool was visualized in MEGA tool

for the extensive analysis and identification of conserved and variation regions. Coloured regions show the similarity between the homologous sequence and mutations that is insertions and deletions are represented in the form of gaps [19]. In some positions of the alignment, substitutions have also been represented.

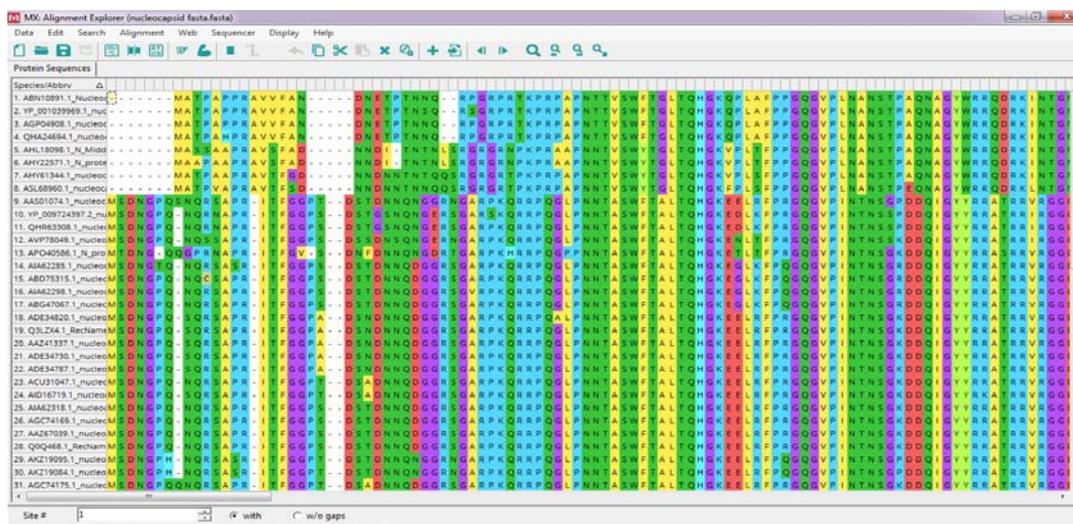


Fig. 3: Multiple sequence alignment of nucleocapsid phosphoprotein

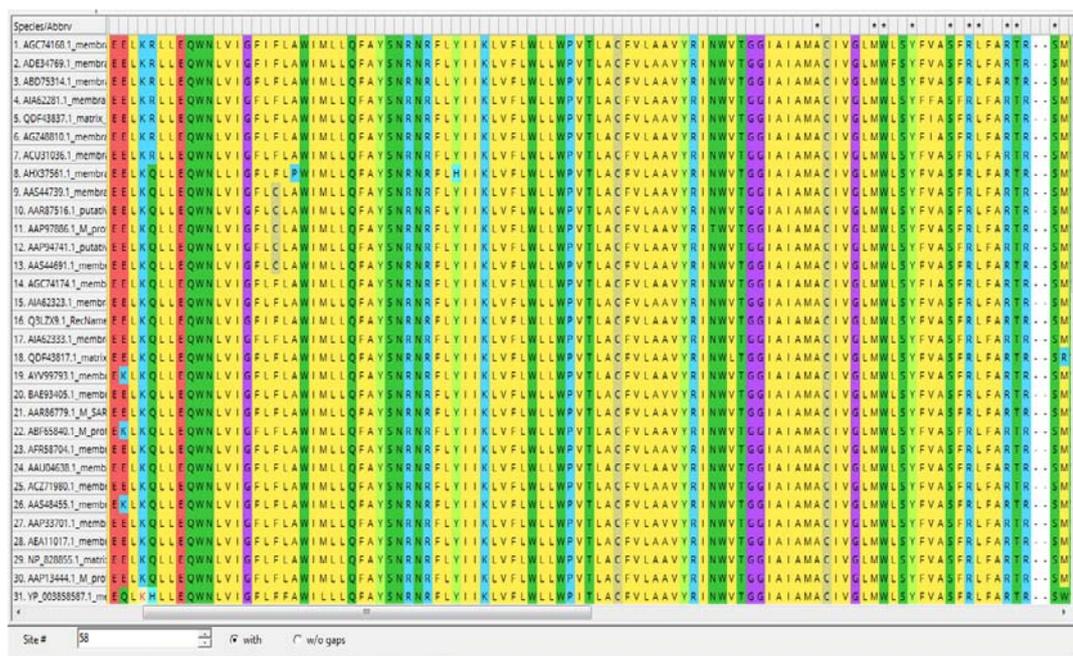


Fig. 4: Multiple sequence alignment of membrane glycoprotein

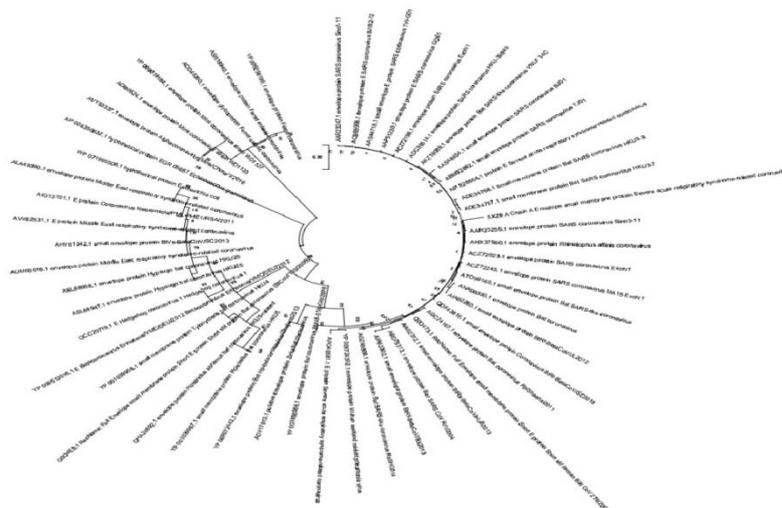


Fig. 8: Phylogenetic tree of envelop protein

BLAST and MSA result show the similarity among all three protein and results of the phylogenetic tree shows that Nucleocapsid phosphoprotein is originated from Hypsugo Bat Corona virus, Membrane glycoprotein is originated from MERS Corona Virus and Envelop proteins have originated from Ferret corona virus.

Homology modeling of spike glycoprotein

Homology modeling is a method which uses template to predict the structure of query sequence and to build 3D model based on homology. The template identified for this procedure was 6acc_C (Spike glycoprotein of SARS-CoV) which is homologous with the SARS-CoV-2 spike glycoprotein with identity scores of 75% (fig. 8).

Predicted 3D structure of spike glycoprotein (SARS-CoV-2) is shown in (fig. 9). Helices and sheet's secondary structure can be seen in the modelled protein.

Structure verification of the modelled protein was done by Ramachandran plot analysis using ProCheck software which showed 86.0% of residues in favoured region, 11.8% Number of residues in allowed region and 2.2% residues in outlier region (fig. 10). The result show that modelled structure of surface glycoprotein of SARS 2-CoV was accurately modelled and can be further used for binding site prediction and docking studies.

Binding site prediction and grid generation

Docking between ligands and modelled spike glycoprotein structure was done using the slide dock method. The first step in the process of docking was the identification of the binding site; it was done using Sitemap tool of Schrodinger software (fig. 11A).

Binding site mainly shows the site score size, D score, volume, phobic-philic nature and residue position for each site predicted and the best binding site with the highest score was selected. The next important step in docking was grid generation; it basically defines the binding positions on the target protein. The site with highest site score predicted using site map tool is used for grid generation and the grid are made using the grid generation program of glide dock as given in Schrodinger software. Grid map of modelled spike glycoprotein is shown in (fig. 11B).

The last main step was the docking which is done between the ligand and the modelled protein which mainly uses glide dock method as a tool.

Glide docking result

All the six-ligand were docked against the predicted active site (grid) of the protein in order to identify the best ligand which could act as an inhibitor for the selected protein. Table 3 shows the docking result of all six ligands with their respective glide score.



Fig. 8: Template identification for surface glycoprotein of 2019-nCoV

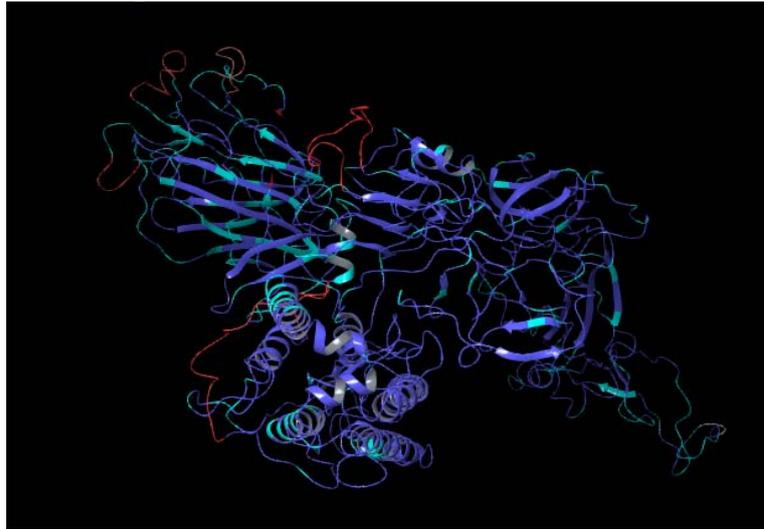


Fig. 9: Modeled structure of surface glycoprotein of 2019-nCoV

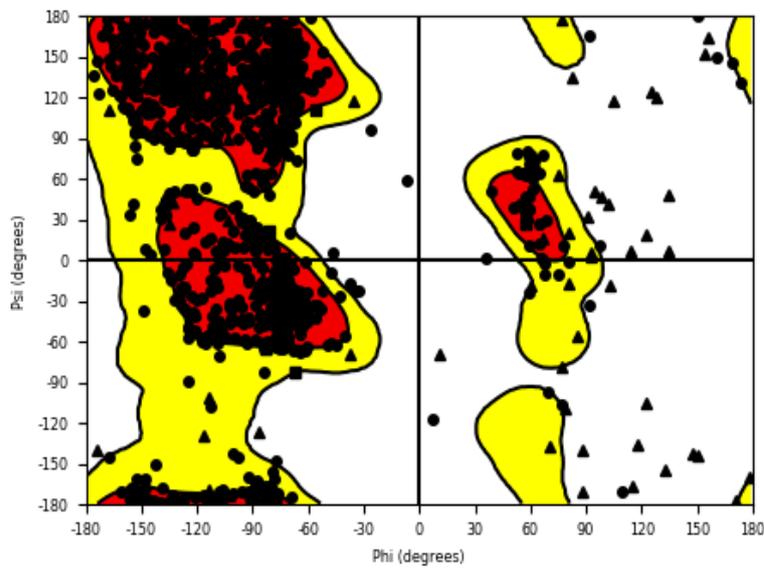


Fig. 10: Structure verification using ramachandran plot analysis

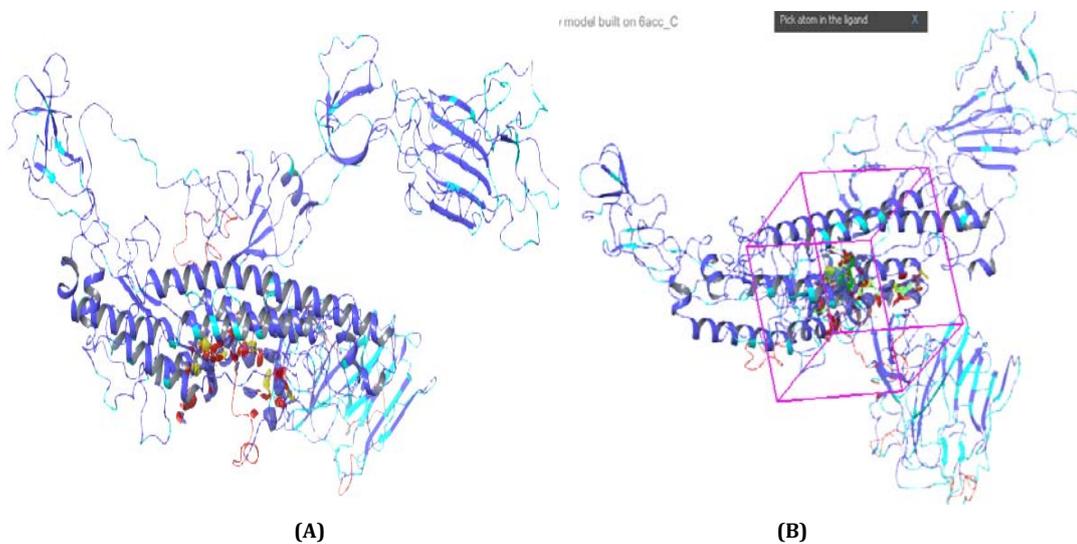


Fig. 11: (A) Binding site prediction (B) Grid generation of the spike glycoprotein

Table 3: Docking result of ligands against COVID 19 protein

S. No.	Ligand name	Glide score
1.	Precose	-8.372
2.	N-(1-Naphthyl)-2-(phenylthio)ethanethioamide	-5.695
3.	6MP-Arabinoside	-5.383
4.	(2S)-2-(4-Methoxy-3,5-dimethylphenyl)-5-methyl-2-(3-pyrimidin-5-ylphenyl)-1,3-dihydroimidazol-4-amine	-4.605
5.	Verapamil	-4.119
6.	2-(Phenethylthio)acetic acid	-3.740

From the docking result we concluded that the highest glide score was of the ligand with PubChem Id 444254 and is commonly known as Precose (fig. 12) having glide score or docking score of -8.372 which shows that it has stable and strong interaction with Spike glycoprotein.

Precose is also commonly known as Acarbose and it's a pseudotetrasaccharide which is mainly responsible for the inhibition of the alpha-glucosidase and alpha-amylase with anti-hyperglycemic activity. Structure and molecular analysis of compound (Precose) shows that it has exact mass of 645.2480g/mol and 14-hydrogen bond donors and 19-hydrogen bond acceptors are present which help in the interaction [20].

Protein-ligand interaction map of Precose was studied in order to find out the type of bond formed and amino-acid involved in the binding and is shown in fig. 13 which shows that compound makes hydrogen bond with PHE 823, VAL826, PRO 863, ASP 867 and HIS 1058 amino acid residue of the active site of spike glycoprotein of SARS-CoV-2 which potentially suggested that it could act as an inhibitor for the spike glycoprotein and further interfere the binding between the spike protein and host receptor.

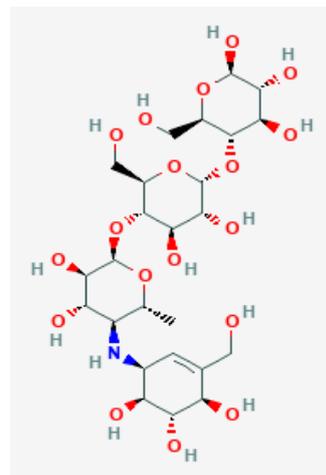
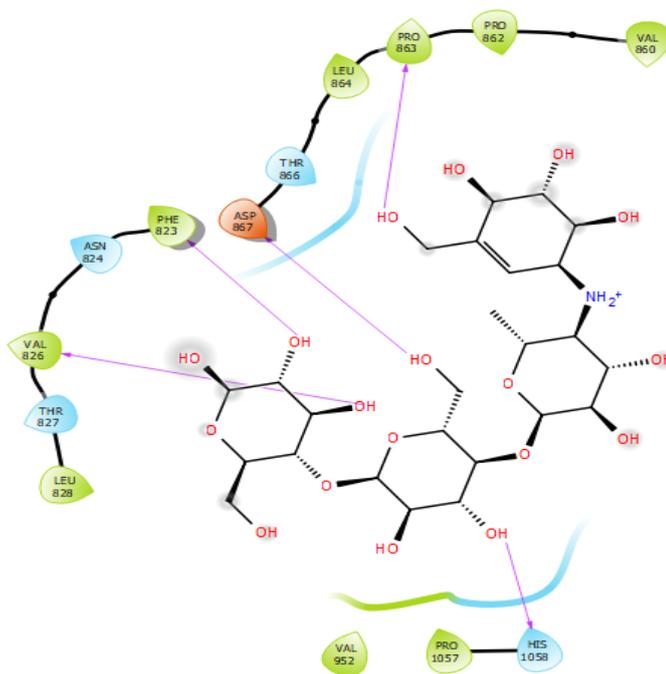
Fig. 12: Precose (C₂₅H₄₃NO₁₈) chemical structure

Fig. 13: Interaction of surface glycoprotein of 2019-nCoV with precose ligand. It shows four hydrogen bond interaction with the surface glycoprotein, with glide score of -8.372

Except, Precose other five ligands which were used out of the total six has docking score in between -5.69 to -3.74 which could be considered less stable as well as weak in interaction with the modeled protein structure.

CONCLUSION

Evolutionary biology plays a very important role to understand the evolution of specific genes, proteins, or species. Current work

focuses on the evolutionary relationship of three important protein viz. Nucleocapsid Phosphoprotein, Membrane Glycoprotein and Envelop Protein of SARS-CoV-2 that is responsible for COVID-19 disease. Protein sequence of these proteins were used for homologs identification and prediction of evolutionary relationships with other viruses. Extensive analysis of MSA and Phylogenetic tree shows that SARS-CoV-2 have similarity with the other family of viruses. Result shows that Nucleocapsid phosphoprotein has origin

from Hypsugo Bat Corona virus, Membrane glycoprotein originated from MERS Corona Virus and Envelop proteins have evolutionary relationship with Ferret coronavirus. This study can be used for drug repurposing of MERS Virus for SARS-CoV-2 and to understand the mutations and gene conservation through evolution of this SARS-CoV-2 virus.

Viruses are sub-microscopic agents that mainly replicate when get inside a host and causes numerous infectious diseases in all life-forms. Throughout the time viruses has caused many deadly diseases influenza, chickenpox, AIDS, Ebola, SARS, MERS etc. Due to this virulence property the infection become much more severe and fatal in most case. Coronavirus has diverse class of viruses which is responsible of many diseases in the past like SARS, MERS and recently SARS-2 which is responsible for taking many lives and has caused major economic loss to many countries. Thus, developing a potential drug or vaccine is main task in order to stop the infection as well as to safe many lives.

Structural properties of Surface Glycoprotein of SARS 2-CoV was studied and protein sequence of surface glycoprotein was retrieved from NCBI database. Homology modeling was done and further binding site and grid generation was done for docking studies. Docking result mainly helps in finding the insight of bond formation, ligand efficiency, binding affinity and stability of protein-ligand interaction. The result mainly showed Precose which is commonly known as Acarbose can act as a potential inhibitor for the spike glycoprotein, while the protein-ligand interaction map also showed the important amino acid with their position. This paper described and highlighted the importance of repurposing of the previously available drug to act as potent inhibitor in newly discovered or novel diseases.

FUNDING

No funds were provided for this research

AUTHORS CONTRIBUTIONS

Authors (a) have done the work on evolutionary relationship analysis (b and c) Docking and Interaction analysis and interpretation of the data; (d) drafting the article or revising it critically for important intellectual content and approval of the final version.

CONFLICTS OF INTERESTS

The author(s) declare(s) that there is no conflict of interest

REFERENCES

- Shang J, Ye G, Shi K, Wan Y, Luo C, Aihara H, *et al.* Structural basis of receptor recognition by SARS-CoV-2. *Nature* 2020;581:221-4.
- Sakurai A, Sasaki T, Kato S, Hayashi M, Tsuzuki SI, Ishihara T, *et al.* Natural history of asymptomatic SARS-CoV-2 infection. *New England J Med* 2020;27:885-6.
- Tillett RL, Sevinsky JR, Hartley PD, Kerwin H, Crawford N, Gorzalski A, *et al.* Genomic evidence for reinfection with SARS-CoV-2: a case study. *Lancet Infect Dis* 2021;21:52-8.
- Kim D, Lee JY, Yang JS, Kim JW, Kim VN, Chang H. The architecture of SARS-CoV-2 transcriptome. *Cell* 2020;181:914-21.
- Helms J, Kremer S, Merdji H, Clere Jehl R, Schenck M, Kummerlen C, *et al.* Neurologic features in severe SARS-CoV-2 infection. *New England J Med* 2020;382:2268-70.
- Lan J, Ge J, Yu J, Shan S, Zhou H, Fan S, *et al.* Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor. *Nature* 2020;581:215-20.
- Zhou L, Zhang M, Wang J, Gao J. Sars-Cov-2: underestimated damage to nervous system. *Travel Med Infect Dis* 2020;36:101642.
- Hu B, Guo H, Zhou P, Shi ZL. Characteristics of SARS-CoV-2 and COVID-19. *Nature Rev Microbiol* 2020;6:1-4.
- Dinnes J, Deeks JJ, Berhane S, Taylor M, Adriano A, Davenport C, *et al.* Rapid, point-of-care antigen and molecular-based tests for diagnosis of SARS-CoV-2 infection. *Cochrane Database Systematic Rev* 2021;3. <https://doi.org/10.1002/14651858.CD013705.pub2>
- Xiao F, Tang M, Zheng X, Liu Y, Li X, Shan H. Evidence for gastrointestinal infection of SARS-CoV-2. *Gastroenterology* 2020;158:1831-3.
- Helms J, Kremer S, Merdji H, Clere Jehl R, Schenck M, Kummerlen C, *et al.* Neurologic features in severe SARS-CoV-2 infection. *New England J Med* 2020;382:2268-70.
- Hosier H, Farhadian SF, Morotti RA, Deshmukh U, Lu-Culligan A, Campbell KH, *et al.* SARS-CoV-2 infection of the placenta. *J Clin Investigation* 2020;130:4947-53.
- Yurkovetskiy L, Wang X, Pascal KE, Tomkins Tinch C, Nyalile TP, Wang Y, *et al.* Structural and functional analysis of the D614G SARS-CoV-2 spike protein variant. *Cell* 2020;183:739-51.
- Heydarabadi FH, Baessi K, Bashar R, Fazeli M, Sheikholeslami F. A phylogenetic study of new rabies virus strains in different regions of Iran. *Virus Genes* 2020;56:361-8.
- Miccio Fonseca LC. MEGA 7 Tox an analysis of "Risk/Treatment needs and progress protocol" by Kang, Beltrani, Manheim, Spriggs, Nishimura, Sinclair, Stachniuk, Pate, Righthand, Prentky, and Worling. *J Child Sexual Abuse* 2020;29:351-72.
- Simon C. An evolving view of phylogenetic support. *Syst Biol* 2020. DOI:10.1093/sysbio/syaa068
- Bhachoo J, Beuming T. Investigating protein-peptide interactions using the schrodinger computational suite. *Modeling Peptide Protein Interactions*; 2017. p. 235-54.
- Reddy KK, Rathore RS, Srujana P, Burri RR, Reddy CR, Sumakanth M, *et al.* Performance evaluation of docking programs-Glide, GOLD, AutoDock and SurflexDock, using free energy perturbation reference data: a case study of fructose-1, 6-bisphosphatase-AMP analogs. *Mini Rev Med Chem* 2020;20:1179-87.
- Grazina L, Costa J, Amaral JS, Mafra I. High-resolution melting analysis as a tool for plant species authentication. In: *Crop Breeding*; 2021. p. 55-73.
- Kaur J, Famta P, Khurana N, Vyas M, Khatik GL. Pharmacotherapy of type 2 diabetes. In: *Obesity and Diabetes Springer*; 2020. p. 679-94.