# Druggability for COVID19 – *In silico* discovery of Potential Drug Compounds against Nucleocapsid (N) Protein of SARS-CoV-2

Manisha Ray<sup>1</sup>, Saurav Sarkar<sup>1</sup>, Surya Narayan Rath<sup>2</sup>\*

<sup>1</sup>All India Institute of Medical Sciences, Bhubaneswar, Odisha, 751019 <sup>2</sup>Department of Bioinformatics, Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, 751003

## **Corresponding Author**

Mr. Surya Narayan Rath, Assistant Professor Department of Bioinformatics, Odisha University of Agriculture and Technology, Bhubaneswar, Odisha, India Email: <u>snrbioinfo@gmail.com</u> Mob: +91-9937727461

#### Abstract

#### Background:

The coronavirus disease 2019 (COVID-19) was caused havoc throughout the world by creating widespread mortality and morbidity. The presence of RNA binding domain in the nucleocapsid (N) protein of SARS-CoV-2 is a potential drug target, serving multiple critical functions during the viral life cycle, especially the viral replication. The unavailability of vaccines and proper antiviral drugs encourages the researchers to identify some potential antiviral drug compounds to be used against N protein of SARS-CoV-2 for this current scenario. While vaccine development might take some time, the identification of a drug compound might decrease the widespread deaths and suffering.

**Method:** This study was analyzed the phylogenetic relationship of N protein sequence divergence with other 49 CoV species and also identified the conserved regions according to protein families through conserved domain search. Along with it, good structural binding affinities of some natural/synthetic phytocompounds/ drugs against N protein were also found using the molecular docking approaches.

**Result:** The analyzed antiviral properties, predicted binding affinities and the presence of higher numbers of Hydrogen bonds of selected compounds represent the drug-ability of these compounds. Among them, the established antiviral drug Glycyrrhizic acid and the phytochemical Theaflavin can be considered as putative drug compound against target protein of SARS-CoV-2 as they showed all the properties of a potential drug.

**Conclusion:** The findings of this study might lead to the development of a drug for the disease and helpful to reduce the risk of deadly infections in host cell due to SARS-CoV-2.

Keywords: SARS-CoV-2, COVID-19, Nucleocapsid Protein, molecular docking

#### Introduction

The outbreak of novel coronavirus infection has drastically affected the lives of the human population worldwide. This infection started as respiratory illness/pneumonia of unknown origin in Wuhan city of China at the end of the year 2019. The organism identified and termed as novel on 7<sup>th</sup> January 2020. The World Health Organization (WHO) declared it as a public health emergency of international concern as the disease spread to other regions of the world <sup>1</sup>. The official name of this infection was made as coronavirus disease 2019 (COVID-19) on 11<sup>th</sup> February 2020. The epidemic was declared a pandemic officially by WHO on 11th March 2020. The novel coronavirus is also termed, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2)<sup>1</sup>. SARS-CoV-2 infection mainly causes pneumonia, upper and lower respiratory tract infection with fever and cough as significant clinical symptoms. But some other symptoms include shortness of breath, muscle pain, confusion, headache, sore throat and acute respiratory distress syndrome, leading to respiratory or multi-organ failure including renal and neurological diseases<sup>2-3</sup>.

Corona viruses (CoVs) are a group of large enveloped viruses with positive sense, singlestranded RNA genomes. Previously identified CoVs in human disease are the alpha CoVs (hCoV-NL63, hCoV-229E) and the beta CoVs (hCoV-0C43), severe acute respiratory syndrome (SARS-CoV), and the Middle East respiratory syndrome (MERS-CoV)<sup>4</sup>. However, among these emerging, highly pathogenic human CoVs, SARS-CoV, MERS-CoV and the newly emerged SARS-CoV-2 infection can result in life-threatening disease conditions and the potential to cause pandemic <sup>2</sup>.

The outcome of SARS-CoV-2 sequencing, (NCBI Reference Sequence: NC\_045512.2) has proposed about the significant sequence level identity of SARS-CoV-2 with SARS-CoV (79%) rather than MERS-CoV (50%). Besides, the higher levels of transmissibility and pandemic risk of COVID-19 at an early stage has been reported in many studies<sup>1</sup>. In the available literatures, the size of the SARS-CoV-2 (NCBI Reference Sequence: NC\_045512.2) genome is 30KB. The genomic virion consists of four major protein regions including matrix (M) protein, an envelope (E) protein, spike (S) protein and a nucleocapsid (N) protein within the viral envelope<sup>5-6</sup>. The functional architectures of each of these viral proteins have accurately characterized. S protein primarily binds to the host cell receptor and form attachment with the host body. Alternatively, M and E proteins are involved in the formation of the viral envelope<sup>6</sup>.

Similarly, SARS-CoV-2 protein N is a multifunctional RNA binding protein, necessary for viral RNA transcription, replication and/or assembly of virus<sup>6</sup>. Interestingly, a unique N-terminal RNA binding domain of SARS-CoV-2 N protein has identified as a novel antiviral drug target site<sup>7</sup>. The viral N protein packages the genome into long, flexible and helical RNP complexes, called nucleocapsids which protect the SARS-CoV-2 virion structure<sup>5</sup>. Additionally, N protein has a significant contribution towards timely replication and reliable transmission of SARS-CoV-2 during its life cycle. Therefore N protein (PDB ID: 6VYO) can be considered as a novel drug target of SARS-CoV-2.

The SARS-CoV-2 infection has created a dangerous pandemic situation due to its quick transmission and deadly nature. It has affected both the health and economy of human population across the globe tremendously. Many ongoing pieces of research are trying to develop vaccines to control this situation, but all are in various phases of trials. Thus, the present study has focused on *in silico* discovery of potent leads from several antiviral drugs and compounds of plant origin against SARS-CoV-2 infection. The present study would throw lights on the discovery of antiviral drug against SARS-CoV-2.

#### **Materials and Methods**

## Sequence retrieval and construction of phylogenetic tree

Nucleocapsid protein sequences of total 49 corona virus species and/or strains including SARS-CoV-2 were retrieved in FASTA format from National Centre for Biotechnology Information (NCBI) web server (https://www.ncbi.nlm.nih.gov/) on 30<sup>th</sup> March 2020. Two N proteins of Ebola and H1N1 virus were included within study to study evolutionary divergence across species. Further, total 51 N protein sequences were aligned using MUSCLE algorithm of Molecular Evolutionary Genetics Analysis 7 (MEGA 7) package<sup>8</sup>. The resulted alignment was used to generate phylogenetic tree using Neighbour Joining (NJ) method of MEGA 7 for 1000 bootstrap replicates.

## **Conserved domain search**

Functional domains of SARS-CoV-2 N protein (YP\_009724397.2) were identified using NCBI conserved domain database (CDD) (https://www.ncbi.nlm.nih.gov/Structure/cdd/wrpsb.cgi) search. The CDD is a collection of domain models which imports information from Pfam, SMART, COG, and NCBI to provide a more accurate assessment of neighbor relationships between protein sequences<sup>9</sup>.

## **Prediction of structural element**

The secondary structure of SARS-CoV-2 N protein was predicted from its complete amino acid sequence (Accession: YP\_009724397.2) using PSIPRED 4.0 algorithm<sup>10</sup>. Similarly, protein disorder portion and membrane helix region was predicted by using DISOPRED3<sup>10</sup> and MEMSAT-SVM algorithm<sup>10</sup> of PSIPRED web server (http://bioinf.cs.ucl.ac.uk/psipred/).

## **Retrieval and preparation of 3D structure**

Available N-terminal domain structure (PDB ID: 6VYO) of SARS-CoV-2 N protein was retrieved from Protein Data Bank (PDB) (https://www.rcsb.org/). Initially, hydrogen atoms were added to protein structure after removal of all water and other hetero molecules. Further, energy minimization was performed using Discovery Studio 3.5 suite to obtain a properly optimized structure of target protein.

## **Drug binding cavity prediction**

In absence of knowledge on exact drug binding site, probable binding cavity within SARS-CoV-2 N protein was predicted using metaPocket 2.0 (https://projects.biotec.tudresden.de/metapocket/). MetaPocket tool identifies cavities on protein surface for drug binding site prediction using multiple computational approaches<sup>11</sup> such as PASS11, LIGSITE, Fpocket, SURFNET, GHECOM, and ConCavity.

#### Selection of ligand molecules

Different natural compounds of plant origin reported with antiviral, anti-inflammation, anti-influenza, anti-HIV, anti-hepatic properties were shortlisted from different literatures. In addition, few FDA approved, and investigational antiviral drugs were also selected from Drug Bank (https://www.drugbank.ca/) database for further investigation.

#### Ligand structure retrieval and correction

Three-dimensional structures of natural ligands were retrieved from PubChem (https://pubchem.ncbi.nlm.nih.gov/) database in SDF format and converted into PDB format using Discovery Studio 3.5 suite. Similarly, PDB structures of antiviral drugs were collected from the Drug Bank (https://www.drugbank.ca/). Further, structure optimization and protonation state of all ligands were achieved using Discovery Studio 3.5 suite.

## **Molecular docking**

Molecular docking was performed between all selected ligands (phytochemicals and antiviral drugs) and the drug target (N protein, PDB ID: 6VYO) separately in order to identify the most efficient inhibitor against SARS-CoV-2. AutoDock 4.2 (http://autodock.scripps.edu/) and Auto-Dock Tools 4 tool<sup>12</sup> were used to perform molecular docking study. The N-terminal RNA binding domain of SARS-CoV-2 N protein was observed as a homo tetramer structure; therefore, only chain A of the available crystal structure was employed for docking analysis. Prior to docking, Kollman charges and polar hydrogen atoms were added to the target structure. Both ligand and receptor structures were prepared using ADT tool and converted to pdbqt format before docking. A virtual grid box was set around the drug binding cavity of the target structure with size of 74, 78, 74 Å in x, y, z direction in spacing of 0.375 Å. Semi flexible docking was performed by maintaining target structure as rigid and allowing flexibility to ligand molecules within the drug-binding pocket<sup>13</sup>. Lamarckian genetic algorithm (LGA) was used with 25000000 energy evaluation steps for each dock run. Auto dock generated ten conformers based on free binding energy for each protein-ligand complex. The most energetically favorable (lowest energy) binding complex was considered for analysis. Further analysis and presentation of atomic interaction between docked complexes were performed using PyMol molecular graphics tool (www.pymol.org).

#### Results

## Molecular phylogeny ascertained sequential divergence of SARS-CoV-2 N protein

Total 49 N proteins different CoV species, including SARS-CoV-2 (**Table 1**) were retrieved to construct the phylogenetic tree. Again, protein sequences of two distance homologues of SARS-CoV-2 such as Ebola (Accession: SCD11531.1) and H1N1 (Accession: YP\_009118629.1) virus were included within the tree in order to establish sequential divergence pattern across species. The phylogenetic tree was constructed using Neighbour Joining (NJ) method<sup>14</sup> with tree evaluation step for 1000 bootstrap replicates. The resulted rooted tree (Fig.1) clustered into two major clades. Total 49 species were diversified within both of the clades (clade-I: 26; clade-II: 23). The target N protein sequence of SARS-CoV-2 (Accession: YP\_009724397.2) was grouped with SARS-CoV (Severe acute respiratory syndrome-related virus) (Accession: NP\_828858.1) sequence within clade-I with branch frequency of 100% which pointed out regarding their significant evolutionary closeness. One separate clade was formed within the tree with branch frequency of 61% among the two outgroups (Ebola and H1N1) which clearly revealed their divergence from all other 49 sequences.

#### Functional domain identified for SARS-CoV-2 N protein

The complete sequence of SARS-CoV-2 N protein (Accession: YP\_009724397.2) comprises of 419 amino acids. All functional domain regions within the N protein sequence of SARS-CoV-2 were identified from its conserved pattern among the members of beta corona virus nucleocapsid protein family. The conserved domains were observed within the aligned region of SARS-CoV-2 N protein from 14-368 amino acids (**Fig.2A**) with the members of the superfamily (pfam00937) (**Fig.2B**). The CD search identified one. N-terminal (50-175 amino acids) and one C-terminal (258-359 amino acids) functional domain (**Fig.2C**) with good bit score (424.07) and lowest e-value (7.05e-148). The nucleocapsid N terminal domain (NTD) of SARS-CoV-2 was showed significant similarities with the conserved domain of family cd21554 whereas the C terminal domain (CTD) found conserved within the family members of cd21595 (**Fig.2D**).

## Structural elements of SARS-CoV-2 N protein

In the absence of full-length structure, the secondary structural elements of SARS-CoV-2 N protein were predicted from its primary sequence using PSIPRED web server. Secondary structural elements such as 2 long, 8 medium, 2 short helical regions and 2 medium, 9 short  $\beta$ -sheets were predicted within the complete sequence of SARS-CoV-2 N protein (**Fig.3**). Most of

the NTD (50-175) regions were predicted as  $\beta$ -sheets and coils. On the contrary, structural elements such as helices,  $\beta$ -sheets and coils were observed within CTD (258-359) regions (**Fig.3**). Further, highly disordered regions of SARS-CoV-2 N protein were observed above the cut off score (0.5) from amino acid positions 1-50, 180-250, and 350-419 (**Fig.4A**). However, significant disorder portions were absent within the both NTD (50-175) and CTD (258-359) regions (**Fig.4A**). According to MEMSAT-SVM algorithm, the sub-cellular localization of SARS-CoV-2 nucleocapsid NTD was found as cytoplasmic, whereas a small C-terminal transmembrane region was noticed from 302-317 amino acids (**Fig.4B**).

## Structure preparation and active site identification of N protein NTD

Homology search using BLASTP algorithm revealed the structure of N-terminal RNA binding domain occupied 30% region of SARS-CoV-2 N protein (Accession: YP\_009724397.2) sequence with 100% identity. Therefore, the three-dimensional structure of SARS-CoV-2 N protein was retrieved and processed for structural correction and optimization. The possible drug-binding cavity of SARS-CoV-2 N protein was predicted in the absence of literary evidence. Algorithm of metaPocket was generated top three hits after clustering the results of PASS11, LIGSITE, Fpocket, SURFNET, GHECOM, and ConCavity. Out of these three, the large active pocket was considered a possible drug-binding cavity (**Fig.5**).

#### Structure preparation natural/synthetic ligands against SARS-CoV-2 N protein

As of literature, a total of eight natural compounds of plant origin and three synthetic compounds (**Table 2**) were identified with antiviral properties, therefore, prepared to dock against SARS-CoV-2 N protein. Again, seven antiviral drugs (**Table 3**) were also included within the study to discover potent inhibitor against N protein of SARS-CoV-2. Finally, 3D structures of a total of eighteen ligands were extracted from online databases (PubChem/Drug Bank) and prepared for docking study.

#### Molecular docking identified efficient ligand against SARS-CoV-2 N protein

Molecular docking is an efficient technique to identify the binding affinity of a drug compound against a drug target<sup>15, 16</sup>. Therefore, all possible inhibitors were docked separately against SARS-CoV-2 N protein to discover effective ligand and important atomic interaction between protein-ligand complexes within the drug-binding cavity. The resulted in free binding energy, and the inhibition constant of each binding complex was reported in **table 4**. According to docking energy score and inhibition constant (KI), total eight antiviral compounds such as

Glycyrrhizic acid (-12.61 kcal/mol; KI: 573.72 pm), Theaflavin (-10.35 kcal/mol; KI: 26.03 nM), Diosgenin (-10.06 kcal/mol; KI: 42.53 nM), U18666A (-9.08 kcal/mol; KI: 219.38 nM), Ethyl brevifolincarboxylate (-9.07 kcal/mol; KI: 226.42 nM), Quercitrin (-9.04 kcal/mol; KI: 238.18 nM), Curcumin (-8.68 kcal/mol; KI: 434.59 nM), and Ladanein (-8.19 kcal/mol; KI: 988.63 nM) showed good binding efficiency than rest of the compounds (Table 4). Presence of an ample number of polar interactions has a significant contribution towards the stability of a specific ligand within the binding site of drug target. Therefore, h-bond interaction between the drug target and ligands were inspected. Interestingly, good binding affinity and strong h-bond interaction within distance  $\leq 3.5$  Å from binding cavity were identified in case of ten suitable compounds such as Glycyrrhizic acid (-12.61 kcal/mol; h-bond: 16 nos), Theaflavin (-10.35 kcal/mol; h-bond: 11 nos), Ethyl brevifolincarboxylate (-9.07 kcal/mol; h-bond: 6 nos), Quercitrin (-9.04 kcal/mol; h-bond: 11 nos), Curcumin (-8.68 kcal/mol; h-bond: 5 nos), Ladanein (-8.19 kcal/mol; h-bond: 8 nos), Apigenin (-7.98 kcal/mol; h-bond: 6 nos), Tenofovir (-6.92 kcal/mol; h-bond: 9 nos), Resveratrol (-6.91 kcal/mol; h-bond: 5 nos), Ribavirin (-6.41 kcal/mol; h-bond: 12 nos), indicated about their efficacy to block the important site within the RNA binding domain of SARS-CoV-2 N protein (Table 4, Table 5, Fig. 6A-6J). To its support, few amino acid residues such as PHE 66, PRO 67, ARG 68, GLY 69, GLN 70, TYR 123, TRP 132, and ALA 134 were found commonly interacting with all of these ligands within the binding cavity of SARS-CoV-2 N protein. However, presence of h-bond interaction with quite good binding energy and inhibition constant values were also noticed in case of rest seven antiviral compounds such as Diosgenin (-10.06 kcal/mol; KI: 42.53 nM; h-bond: 3 nos), U18666A (-9.08 kcal/mol; KI: 219.38 nM; h-bond: 2 nos), Berberine (-7.87 kcal/mol; KI: 1.69 uM; h-bond: 2 nos), Emodin (-7.82 kcal/mol; KI: 1.86 uM; h-bond: 6 nos), Quercetin (-7.47 kcal/mol; KI: 3.33 uM; h-bond: 8 nos), Hydroxychloroquine (-7.35 kcal/mol; KI: 4.07 uM; h-bond: 2 nos), Chloroquine (-6.86 kcal/mol; KI: 9.34 uM; h-bond: 1 nos) inbound form with SARS-CoV-2 N protein (Table 4, Fig. 7A-6G). Overall docking study confirmed the binding potential of the discussed phytochemicals and drugs, against drug target, Nucleocapsid protein of SARS-CoV-2.

#### Discussion

The SARS-CoV-2 or COVID19 pandemic has created an alarming situation due to severe infection and death rate worldwide. Researchers all over the world are in search to identify novel drug/vaccine target as well as the development of drug/vaccine to combat the disease. In support of the present scenario, the current study has tried to conduct some critical analyses on important drug target, i.e. Nucleocapsid (N) protein of SARS-CoV-2. The present research also focuses on *in silico* discovery of potent natural/synthetic compounds against the virus.

The phylogenetic study among different coronavirus species community identified the close relation and less diversification between N proteins of SARS-CoV and SARS-CoV-2, which indicates the high similarities between those species. The protein family sequence similarity search or the conserved domain search points out the versatility of SARS-CoV-2 N protein, which is predicted by the conserved amino acid regions from different members CoV superfamilies such as SARS-CoV, Murine CoV (Murine Hepatitis Virus) and Alpha Cov-1 species (Feline infectious peritonitis virus).

It has also been reported that N protein has a vital role in the survival and growth of SARS-CoV-2. Thus authors focused on the discovery of potential natural or synthetic compounds to block its regular mechanism. Primary sequence analysis resulted in two crucial functional domain regions both in N and C terminals of SARS-CoV-2. Interestingly, the NTD comprises RNA binding site, which signifies its importance towards a viral cellular mechanism. To its support, the available crystal structure of NTD SARS-CoV-2 N protein was retrieved and utilized in further study. The SARS-CoV-2 N protein had no binding site information including drug binding sites till the end of March 2020, which influences the researchers to predict the drug-binding pocket in RNA binding domain of N protein. But recently, Kang et al. reported about the crystal structure and showed the drug-binding pocket (including the amino acids Tyr 110/124, Tyr 112/126, Ala56, Gly68, Thr55, Ser 67, Arg89, Tyr102) of N protein with PDB ID 6M3M<sup>17</sup> whereas this present study predicted the binding domain in SARS-CoV-2 N protein (PDB ID: 6VYO) with amino acids positioned from 64-71, 84, 123-124, 131-140. This study represents the maximum similarities between the crystal structure binding pocket and the presently identified drug-binding pocket in N protein, which should be considered while deciding a drug for trial in the treatment of the disease.

Today, the death report of COVID19 from different corner of the globe is drastically increasing due to the absence of an effective antiviral drug. To overcome this situation, eighteen compounds, including natural compounds of plant origin and antiviral drugs, were docked into the drug-binding cavity of N protein to identify potential ligands against SARS-CoV-2. This study has been able to find the binding efficiency of a few phytochemicals (Theaflavin, curcumin, ladanein), and a few drug compounds (glycyrrhizic acid, ethyl brevifolin caboxylate, and quercitrin) against N protein of the virus. This might serve as information about their potential to be a treatment option for SARS-CoV-2. The antiviral effects of phytochemicals such as Theaflavin, curcumin, and ladanein, against many pathogenic viruses, have already been well studied and reported. Theaflavin is known to prevent from influenza virus by inhibiting its replication<sup>18</sup>.

Similarly, Curcumin has anti viral properties against H1N1 Influenza and FIPV<sup>19</sup>. Again, the inhibitory effect of ladanein against hepatitis C virus infection<sup>20</sup> is also well studied. Thus, these compounds may be useful as an anti-infective agent against COVID19. Antiviral drugs such as Glycyrrhizic acid, Ethyl brevifolincarboxylate, and Quercitrin have inhibitory effect against<sup>21, 22</sup> hepatitis B and C virus. But, Glycyrrhizic acid and quercetin are associated with severe side effects such as hypokalemia, oedema, rhabdomyolysis or myoglobinuria, mitochondrial toxicity and mutagenicity<sup>23-24</sup>. However, according to the resulted binding affinities and the presence of H-bonds glycyrrhizic acid and Theaflavin can be considered as suitable drug compounds against SARS-CoV-2 N protein. In regards to toxicity associated with glycyrrhizic acid, the use of natural compound, i.e. Theaflavin may be more effective against COVID19. Other than the mentioned natural/synthetic compounds, few others such as Diosgenin<sup>20</sup>, U18666A<sup>25</sup>, Apigenin (*Ocimum sanctum*)<sup>26</sup>, Resveratrol (*Vitis labrusca*)<sup>27</sup>, Berberine (*Berberis vulgaris*)<sup>28</sup>, Emodin (*Radix et Rhizoma Rhei, Radix Polygoni Multiflori*)<sup>29</sup>, Tenofovir (*Phyllanthus niruri*)<sup>22</sup> has shown stable binding interaction with SARS-CoV-2 N protein. Hence they may also be studied for further validation.

## Conclusion

The COVID19 outbreak has caused havoc throughout the world, changing the course of human lives. Researchers are trying to design a vaccine against SARS-CoV2 but that might take some time. This study attempts to find a drug for treating the disease condition, which will help to save human lives and mitigate the sufferings of millions of people infected by the virus worldwide. Some antivirals phytocompounds and synthetic drugs have been analyzed in this *in silico* study, which would target the N protein, responsible for replication of SARS-CoV-2 in the host body. Of all the compounds in this study, glycyrrhizic acid and Theaflavin can be used as the antiviral drug, as they showed a higher binding affinity with the target protein. They might be effective to inhibit the viral effects and prevent the infections in the host cell, serving as "The Treatment" of the disease.

## Acknowledgments

We are thankful to Dr. Pawan Kumar Agrawal, Vice chancellor, Odisha University of Agriculture and Technology, Bhubaneswar for his moral support and valuable suggestion.

## **Conflict of interest**

The authors declare no competing interest.

#### References

- Adhikari SP, Meng S, Wu YJ, et al. Epidemiology, causes, clinical manifestation and diagnosis, prevention and control of coronavirus disease (COVID-19) during the early outbreak period: a scoping review. Infectious Diseases of Poverty 2020; 9(1): 29.
- Jiang S, Hillyer C, Du L. Neutralizing Antibodies against SARS-CoV-2 and Other Human Coronaviruses. Trends in Immunology 2020; 41(5): 355-359.
- 3. Ou X, Liu Y, Lei X, et al. Characterization of spike glycoprotein of SARS-CoV-2 on virus entry and its immune cross-reactivity with SARS-CoV. Nature Communications 2020; 11(1): 1620.
- Wang Q, Zhang Y, Wu L, Niu S, et al. Structural and Functional Basis of SARS-CoV-2 Entry by Using Human ACE2. Cell 2020; 181:1–11.
- 5. McBride R, van Zyl M, Fielding BC. The coronavirus nucleocapsid is a multifunctional protein. Viruses 2014; 6(8): 2991–3018.
- Yang P, Wang X. COVID-19: a new challenge for human beings. Cellular and Molecular immunology 2020; 17:555–557.
- Lin SM, Lin SC, Hsu JN et al. Structure-Based Stabilization of Non-native Protein– Protein Interactions of Coronavirus Nucleocapsid Proteins in Antiviral Drug Design. J. Med Chem 2020; 63(6): 3131-3141.
- Kumar S, Stecher G, Tamura K. MEGA7: Molecular Evolutionary Genetics Analysis Version 7.0 for Bigger Datasets. Molecular Biology and Evolution 2016; 33(7): 1870– 1874.
- Marchler-Bauer A, Anderson J B, Derbyshire MK, DeWeese et al. CDD: a conserved domain database for interactive domain family analysis. Nucleic Acids Research 2007; 35: D237–D240.
- Buchan DWA, Jones DT. The PSIPRED Protein Analysis Workbench: 20 years on. Nucleic Acids Research 2019; doi:10.1093/nar/gkz297.
- Huang B. MetaPocket: A Meta Approach to Improve Protein Ligand Binding Site Prediction. OMICS: A Journal of Integrative Biology 2009; 13(4): 325–330.

- 12. Rizvi SM, Shakil S, Haneef M. A simple click by click protocol to perform docking: AutoDock 4.2 made easy for non-bioinformaticians. EXCLI J 2013; 12:831-57.
- 13. Fuhrmann J, Rurainski A, Lenhof HP, Neumann D. A New Lamarckian Genetic Algorithm for Flexible Ligand-Receptor Docking. J Compput Chem 2010; 31(9):1911-8.
- 14. Bogusz M, Whelan S. Phylogenetic tree estimation with and without alignment: New distance methods and Benchmarking. System Biology 2017; 66(2):218-231.
- 15. Sahoo M, Jena L, Rath SN, Kumar S. Identification of Suitable Natural Inhibitor against Influenza A (H1N1) Neuraminidase Protein by Molecular Docking. Genomics Inform 2016, 14(3):96-103.
- 16. Jagadeb M, Rath SN, Sonawane A. In silico discovery of potential drug molecules to improve the treatment of isoniazid resistant Mycobacterium tuberculosis. J Biomol Struct Dyn 2018; doi:10.1080/07391102.2018.1515116.
- Kang S, Yang M, Hong Z, et al. Crystal structure of SARS-CoV-2 nucleocapsid protein RNA binding domain reveals potential unique drug targeting sites. Acta Pharmaceutica Sinica B 2020; S2211-3835(20)30550-5.
- Sahoo M, Jena L, Rath SN, Kumar S. Identification of Suitable Natural Inhibitor against Influenza A (H1N1) Neuraminidase Protein by Molecular Docking. Genomics Inform 2016; 14(3):96-103.
- Zorofchian Moghadamtousi S, Abdul Kadir H, Hassandarvish P, Tajik H, Abubakar S, Zandi K. A Review on Antibacterial, Antiviral, and Antifungal Activity of Curcumin. BioMed Research International 2014; 1–12.
- Ashfaq UA, Idrees S. Medicinal plants against hepatitis C virus. World J Gastroenterol. 2014; 20(11):2941-7.
- 21. Chen F, Chan KH, Jiang Y et al. In vitro susceptibility of 10 clinical isolates of SARS coronavirus to selected antiviral compounds. J Clin Virol 2004; 31(1):69-75.
- 22. Mohan M, James P, Valsalan R, Nazeem PA. Molecular docking studies of phytochemicals from Phyllanthus niruri against Hepatitis B DNA Polymerase. Bioinformation 2015; 11(9):426-31.
- John C. Glycyrrhizic Acid Toxicity Caused by Consumption of Licorice Candy Cigars. CJEM 2009; 11(1):94-6.

- 24. Chen R, Lin J, Hong J, Han D et al. Potential Toxicity of Quercetin: The Repression of Mitochondrial Copy Number *via* Decreased *POLG* Expression and Excessive *TFAM* Expression in Irradiated Murine Bone Marrow. Toxicol Rep 2014; 1:450-458.
- 25. Doki T, Tarusawa T, Hohdatsu T, Takano T. In Vivo Antiviral Effects of U18666A Against Type I Feline Infectious Peritonitis Virus. Pathogens 2020; 9(1): 67.
- Alhazmi MI. Molecular docking of selected phytocompounds with H1N1 Proteins. Bioinformation 2015; 11(4):196-202.
- 27. Rafe T, Shawon PA, Salem L et al. Preventive role of Resveratrol against inflammatory cytokines and related diseases. Curr Pharm Des 2019; 25(12):1345-1371.
- Kaliyaperumal S, Periyasamy K, Balakrishnan U, Palanivel P, Egbuna C. Antiviral phytocompounds for drug development. Phytochemicals as Lead Compounds for New Drug Discovery 2020; 239–244.
- 29. Ho TY, Wu SL, Chen JC, Li CC, Hsiang CY. Emodin blocks the SARS coronavirus spike protein and angiotensin-converting enzyme 2 interaction. Antiviral Res 2007; 74(2):92-101.

Species Name	NCBI Accession	Length
Duck Corornavirus (Avian CoV)	AKF17732.1	414
Turkey Coronavirus (Avian CoV)	YP_001941174.1	409
Infectious Bronchitis Virus (Avian CoV)	NP_040838.1	409
Infectious Bronchitis Virus (Avian CoV)	AKV63212.1	409
Rat CoV Parker (Murine CoV)	YP_003029852.1	454
Murine Hepatitis Virus (Murine CoV)	AAU06361.1	454
Murine Hepatitis Virus (Murine CoV)	NP_045302.1	454
Bovine Coronavirus (Beta CoV)	NP_150083.1	448
Human Coronavirus OC43 (Beta CoV)	YP_009555245.1         448	
Middle East Respiratory Syndrome- Related Coronavirus (MERS CoV)	Middle East Respiratory Syndrome- Related Coronavirus (MERS CoV)YP_007188585.1	
Mink Coronavirus 1	YP_009019186.1	376
Feline Infectious Peritonitis Virus (Alpha Coronavirus 1)	YP_004070199.1	377
Transmissible Gastroenteritis Virus (Alpha Coronavirus 1)	NP_058428.1	382
Rousettus bat Coronavirus HKU9	YP_001039975.1	468
Pipistrellus bat Coronavirus HKU5	YP_001039969.1	427
Canada Goose Coronavirus	YP_009755908.1	414
Tylonycteris Bat Coronavirus HKU4	YP_001039960.1	423
	Species NameDuck Corornavirus (Avian CoV)Turkey Coronavirus (Avian CoV)Infectious Bronchitis Virus (Avian CoV)Infectious Bronchitis Virus (Avian CoV)Rat CoV Parker (Murine CoV)Murine Hepatitis Virus (Murine CoV)Murine Hepatitis Virus (Murine CoV)Bovine Coronavirus (Beta CoV)Human Coronavirus (Beta CoV)Human Coronavirus (OL43 (Beta CoV))Middle East Respiratory Syndrome- Related Coronavirus (MERS CoV)Mink Coronavirus 1Feline Infectious Peritonitis Virus (Alpha Coronavirus 1)Transmissible Gastroenteritis Virus (Alpha Coronavirus 1)Rousettus bat Coronavirus HKU9Pipistrellus bat Coronavirus HKU5Canada Goose Coronavirus HKU4	Species NameNCBI AccessionDuck Coronavirus (Avian CoV)AKF17732.1Turkey Coronavirus (Avian CoV)YP_001941174.1Infectious Bronchitis Virus (Avian CoV)NP_040838.1Infectious Bronchitis Virus (Avian CoV)AKV63212.1Rat CoV Parker (Murine CoV)YP_003029852.1Murine Hepatitis Virus (Murine CoV)AAU06361.1Murine Hepatitis Virus (Murine CoV)NP_045302.1Bovine Coronavirus (Beta CoV)NP_045302.1Human Coronavirus (Beta CoV)NP_150083.1Human Coronavirus OC43 (Beta CoV)YP_009555245.1Middle East Respiratory Syndrome- Related Coronavirus (MERS CoV)YP_007188585.1Feline Infectious Peritonitis Virus (Alpha Coronavirus 1)YP_004070199.1Transmissible Gastroenteritis Virus (Alpha Coronavirus 1)NP_058428.1Rousettus bat Coronavirus HKU9YP_001039975.1Pipistrellus bat Coronavirus HKU5YP_009755908.1Tylonycteris Bat Coronavirus HKU4YP_001039960.1

**Table 1:** Nucleocapsid proteins from different Corona virus species collected from NCBI are listed here

18	Severe Acute Respiratory Syndrome-	NP_828858.1	422
	Related Coronavirus (SARS-CoV)		
19	Severe Acute Respiratory Syndrome	YP_009724397.2	419
	Coronavirus-2 (SARS-CoV-2)		
20	Alpha coronavirus Bat-	YP_009755894.1	432
	CoV/P.kuhlii/Italy/3398-19/2015		
21	Miniopterus Bat Coronavirus 1	YP_001718609.1	389
22	Wencheng Sm Shrew Coronavirus	YP_009389428.1	366
23	Coronavirus AcCoV-JC34	YP_009380526.1	389
24	Lucheng Rn rat Coronavirus	YP_009336487.1	391
25	NL63-related bat Coronavirus	APD51488.1	433
26	NL63-related bat Coronavirus	YP_009328939.1	407
27	Rousettus bat Coronavirus	YP_009273009.1	443
28	Ferret Coronavirus	BAV31353.1	374
29	BtMr-AlphaCoV/SAX2011	YP_009199613.1	429
30	BtNv-AlphaCoV/SC2013	YP_009201734.1	431
31	BtRf-AlphaCoV/HuB 2013	YP_009199794.1	383
32	BtRf-AlphaCoV/YN2012	YP_009200739.1	375
33	Swine enteric Coronavirus	YP_009199247.1	382
34	Camel Alpha Coronavirus	YP_009194643.1	382
35	Beta Coronavirus HKU24	YP_009113031.1	443
36	Bat-Hp-Betacoronavirus/Zhejiang 2013	YP_009072446.1	418
37	Betacoronavirus Erinaceus/VMC/DEU/2012	YP_009513018.1	424
38	Bat Coronavirus	YP_008439206.1	425

	CDPHE15/USA/2006		
39	Rousettus Bat CoronavirusV HKU10	YP_006908646.1	402
40	Rabbit Coronavirus HKU14	YP_005454249.1	444
41	Beluga Whale Coronavirus SW1	YP_001876448.1	379
42	Miniopterus Bat Coronavirus HKU8	YP_001718616.1	422
43	Rhinolophus Bat Coronavirus HKU2	YP_001552240.1	375
44	Scotophilus Bat Coronavirus 512	YP_001351688.1	394
45	Human Coronavirus HKU1	YP_173242.1	441
46	Human CoV NL63	YP_003771.1	377
47	Bat Coronavirus BM48-31/BGR/2008	YP_003858591.1	417
48	Human Coronavirus 229E	NP_073556.1	389
49	Porcine Epidemic Diarrhea Virus	NP_598314.1	441

SN: Serial Number

SN Natural Pub Chem Source/P. Virus **Property** Ref /Synthetic CID Name compounds H1N1 18 1 Theaflavin 135403798 Camellia Prevents sinensis influenza by (Tea plant) inhibiting replication using potentially directs virucidal effect 2 Curcumin 969516 Curcuma Antiviral FIPV, 19 activity against HIV, longa L. FIPV. Influenza (turmeric) Inhibition of HIV-1 and HIV-2 proteases Inhibition of haemagglutinati on 20 3 99474 Synthetic Effectively Hepatitis Diosgenin blocks the C Virus replication of Hepatitis C virus 4 Marrubium Ladanein 3084066 Effectively Hepatitis 20 inhibits the post C Virus peregrinum L. attachment entry step of Hepatitis C Virus Hepatitis Phyllanthus Inhibits virus 22, 20 5 Quercetin 5280343 niruri replication and B/C viral Virus nucleocapsid formation by inhibiting DNA polymerase of hepatitis **B** 5487248 Synthetic Inhibits virus Hepatitis 22 6 Ethyl brevifolincarbox **B** Virus replication and vlate viral nucleocapsid formation by inhibiting DNA polymerase of hepatitis B

 Table 2: Eleven ligand molecules (natural and synthetic) are reported along with antiviral properties.

7	Quercitrin	5280459	Phyllanthus niruri	Inhibit virus replication and viral nucleocapsid formation by inhibiting DNA polymerase of hepatitis B	Hepatitis B Virus	22
8	U18666A	9954082	Synthetic	Inhibits the proliferation of type 1 FIPV	Feline Infectiou s peritoniti s virus	25
9	Apigenin	5280443	<i>Ocimum</i> <i>sanctum</i> (Tulsi)	Prevents the early multiplication of H1N1 virus and control the viral growth	H1N1	26
10	Resveratrol	445154	Vitis labrusca	Effectively reduce the inflammatory cell production and pro- inflammatory cytokine accumulation	Inflamm atory virus	27
11	Allicin	65036	Allium sativum (garlic)	Inhibit virus penetration and proliferation (Inhibit cell proliferation, protect the heart injury, liver damage, anti inflammation)	Influenza	28

SN: Serial number; CID: Compound ID; P. name: Plant name; Ref: Reference

SN	Synthetic /Natural Drug Compound	Drug Bank ID	Status	Source/ P. name	Treatment /Property	Ref
1	Glycyrrhizic acid (Glycyrrhizin)	DB13751	Approved, Experimental	Glycyrrhiza glabra	Inhibit viral replication of SARS-CoV	21
2	Ribavirin	DB00811	Approved	Synthetic	Effective against Chronic Hepatitis C virus, SARS- CoV	PMID:18 565019 and 21
3	Tenofovir	DB14126	Experimental, Investigational	Phyllanthus niruri	Hepatitis B Virus	22
4	Berberine	DB04115	Approved, Investigational	Berberis vulgaris	Prevents the HIV-PI induced inflammation.	28
5	Emodin	DB07715	Investigational	Radix et Rhizoma Rhei, Radix Polygoni Multiflori	Blocks the S protein of SARS-CoV and ACE2 interaction	29
6	Chloroquine	DB00608	Approved, Investigational Vet approved	Synthetic	HIV, influenza A/H5N1, SARS-CoV, human coronavirus 229E	PMID: 2364870 8
7	Hydroxy chloroquine	DB01611	Approved	Synthetic	HIV, DENV	PMID:25 321315

Table 3: Seven antiviral drugs were reported along with medicinal value

SN: Serial number; DB ID: Drug Bank ID; P. name: Plant name; Ref: Reference; PMID: PubMed ID

SN	Ligands	Docking	Intermolecular	Inhibition
	(Phytochemicals/	energy scores	energy(kcal/mol)	Constant (KI)
	Drugs)	(kcal/mol)		
1	Glycyrrhizic acid	-12.61	-14.7	573.72pM
	(Glycyrrhizin)			
2	Theaflavin	-10.35	-13.63	26.03nM
3	Diosgenin	-10.06	-10.35	42.53nM
4	U18666A	-9.08	-10.87	219.38nM
5	Ethyl	-9.07	-10.86	226.42nM
	brevitolincarboxylate			
6	Quercitrin	-9.04	-12.02	238.18nM
7	Curcumin	-8.68	-11.66	434.59nM
8	Ladanein	-8.19	-9.68	988.63nM
9	Apigenin	-7.98	-9.17	1.43uM
10	Berberine	-7.87	-8.47	1.69uM
11	Emodin	-7.82	-8.71	1.86uM
12	Quercetin	-7.47	-9.26	3.33uM
13	Hydroxy chloroquine	-7.35	-10.04	4.07uM
14	Tenofovir	-6.92	-8.41	8.53uM
15	Resveratrol	-6.91	-8.4	8.63uM
16	Chloroquine	-6.86	-9.25	9.34uM
17	Ribavirin	-6.41	-8.2	19.88uM
18	Allicin	-4.69	-6.18	363.41uM

**Table 4:** Docking scores of eighteen ligands against SARS-CoV-2 N protein are reported.

SN: Serial number; pM: pico molar; nM: nano molar; uM: micro molar

SL	Phytochemicals	H-Bond	Bond	Length (A <sup>0</sup> )
	/Drugs	Residues		
1	Glycyrrhizic acid	LYS 65	NZO	2.84
	(Glycyrrnizin)	PHE 66	NO	3.18
			NO	2.93
			ОНО	2.49
		PRO 67	ОНО	3.22
		ARG 68	NEO	3.10
			NH1O	2.59
		GLY 69	NO	2.95
		GLN 70	ОНО	2.71
			ОНО	3.23
		TYR 123	ОНО	2.91
			ОНО	3.06
		GLY 124	ОНО	3.38
		TRP 132	ОНО	2.77
		ALA 134	NO	2.87
			ОНО	3.09
2	Theaflavin	PHE 66	NOH53	3.35
		GLY 69	N05	2.60
		GLN 70	ОН60	2.45
		TYR 123	ОНО9	2.96
			ОНО9	3.16

**Table 5:** Polar interaction (distance  $\leq 3.5$  Å) between selected antiviral compounds and<br/>Neucleocapsid protein of SARS-CoV-2 are reported.

		ILE 130	OH64	2.83
		TRP 132	ОН53	2.95
			N01	3.26
		ALA 134	ООН59	3.13
			NOH59	3.21
			OHO10	3.23
3	Ethyl	PHE 66	NO5	3.55
	brevitolincarboxylate	ARG 68	NEO6	3.12
			NH1O6	3.45
		GLY 69	NO8	2.89
		GLN 70	OH31	2.48
		TRP 132	OH30	2.91
			005	2.78
			N05	3.03
		ALA 134	ОН35	2.45
			NO8	3.03
			N07	2.98
4	Quercitrin	PHE 66	OH43	2.81
			OH51	2.96
			NO10	2.83
		PRO 67	OH41	2.64
		GLY 69	N04	3.00
		GLN 70	OH42	2.98
			OH41	3.04

		TYR 123	ОНО2	3.53
			OH07	2.52
		GLY 124	OH50	2.96
		ALA 134	N04	2.68
5	Curcumin	PHE 66	NO5	3.21
		GLY 69	N04	2.86
		ASN 126	OH40	2.69
		LYS 127	NO3	2.56
		ALA 134	N04	2.81
6	Ladanein	PRO 67	005	3.30
		ARG 68	NEO2	2.98
		GLY 69	NO3	2.89
			N04	3.06
		GLN 70	OH30	2.98
			005	2.44
		ALA 134	OH31	2.50
			NO3	2.75
7	Apigenin	ARG 68	NEO3	3.01
			NH1O3	3.51
		GLY 69	N01	3.25
		GLN 70	OH29	2.94
		ALA 134	N01	3.19
		THR 135	ОН30	3.00
8	Tenofovir	PRO 67	0N	3.44

		ARG 68	NEO	3.16
			NEO	3.10
			NH1O	2.86
			NH1O	2.96
		GLN 70	ON	2.48
			ON	3.17
		ALA 134	ON	3.25
			ON	2.77
9	Resveratrol	GLY 69	NO2	2.76
		GLN 70	NO2	2.94
			OO2H28	2.59
		TYR 123	OHO1	2.82
		ALA 134	NO2H28	3.03
10	Ribavirin	PRO 67	005	3.49
		GLY 69	NO5	2.93
		GLN 70	005	3.07
		TYR 123	OHN8	2.76
		TRP 132	ON9	2.66
		ALA 134	N05	3.02
			005	3.24
			ON7	2.81
			OO2H24	2.68
		GLU 136	OE2H25	2.85
		GLY 137	NO4H27	3.28

	ALA 138	NO4H27	2.98
		OH27	2.99



Graphical Abstract



**Fig.1**: Phylogenetic tree were presented among 49 nucleocapsid protein sequences of SARS-CoV and SARS-CoV-2 from different species. The number in the left side of tree denotes bootstrap frequency for each taxon. The N protein of out group (Ebola and H1N1) sequences and

the target SARS-CoV-2 protein were highlighted using red and blue outline respectively. Similarly, N protein sequence of SARS-CoV was highlighted using green outline.



**Fig.2**: Conserved functional domains of SARS-CoV-2 nucleocapsid protein. (A) Sequence alignment between SARS-CoV-2 and members of super family (pfam00937). (B) The alignment between SARS-CoV-2 and consensus sequence of pfam00937 nucleocapsid protein. The conserved amino acid patterns were highlighted using boxes. (B) All functional domain regions of SARS-CoV-2 nucleocapsid protein were presented in schematic diagram. N-NTD: Nucleocapsid protein N terminal domain; N-CTD: Nucleocapsid protein C terminal domain. (D) The sequence alignment of N-NTD (50-175) and N-CTD of SARS-CoV-2 with their respective conserved domain family.



**Fig.3**: Predicted secondary structural elements for full length N protein of SARS-CoV-2. Helix: Pink cylinder; Sheet: Yellow cylinder.



**Fig.4**: The disorder plot of SARS-CoV-2 nucleocapsid protein was deciphered. Along X-axis the amino acid residue number and along Y-axis cut off value were presented. Black colour dots were used to plot disorder value on Y-axis for the corresponding amino acids on X-axis.



**Fig.5**: (A) Cartoon representation of SARS-CoV-2 nucleocapsid protein (PDB ID: 6VYO, Chain A) structure. B-sheet: Pink colour arrows; Coil: tube. (B) Space filling representation. Active drug binding pocket was highlighted using red colour within the structure.



**Fig.6:** Polar interaction between SARS-CoV-2 N protein with natural/synthetic compounds (A) glycyrrhizin, (B) theaflavin, (C) ethyl brevifolincarboxylate, (D) quercitrin, (E) curcumin, (F) ladanein, (G) apigenin, (H) tenofovin (I) resveratrol, (J) ribavirin



**Fig.7:** Polar interaction between SARS-CoV-2 N protein with natural/synthetic compounds (A) diosgenin, (B) U18666A, (C) berberine, (D) emodin, (E) quercetin, (F) hydroxyl chloroquine, (G) chloroquine