



## OPEN ACCESS

## EDITED BY

Nimisha Ghosh,  
Siksha O Anusandhan University, India

## REVIEWED BY

Mohsen Norouzinia,  
Shahid Beheshti University of Medical  
Sciences, Iran  
Xuedong An,  
China Academy of Chinese Medical Sciences,  
China

## \*CORRESPONDENCE

Ravins Dohare,  
✉ ravinsdohare@gmail.com  
Kapil Dev,  
✉ kdev@jmi.ac.in

RECEIVED 11 September 2023

ACCEPTED 08 January 2024

PUBLISHED 02 February 2024

## CITATION

Tasneem A, Sultan A, Singh P, Bairagya HR,  
Almasoudi HH, Alhazmi AYM, Binshaya AS,  
Hakami MA, Alotaibi BS, Abdulaziz Eisa A,  
Alolaiqy ASI, Hasan MR, Dev K and Dohare R  
(2024), Identification of potential therapeutic  
targets for COVID-19 through a structural-  
based similarity approach between SARS-CoV-  
2 and its human host proteins.  
*Front. Genet.* 15:1292280.  
doi: 10.3389/fgene.2024.1292280

## COPYRIGHT

© 2024 Tasneem, Sultan, Singh, Bairagya,  
Almasoudi, Alhazmi, Binshaya, Hakami, Alotaibi,  
Abdulaziz Eisa, Alolaiqy, Hasan, Dev and  
Dohare. This is an open-access article  
distributed under the terms of the [Creative  
Commons Attribution License \(CC BY\)](https://creativecommons.org/licenses/by/4.0/). The use,  
distribution or reproduction in other forums is  
permitted, provided the original author(s) and  
the copyright owner(s) are credited and that the  
original publication in this journal is cited, in  
accordance with accepted academic practice.  
No use, distribution or reproduction is  
permitted which does not comply with these  
terms.

# Identification of potential therapeutic targets for COVID-19 through a structural-based similarity approach between SARS-CoV-2 and its human host proteins

Alvea Tasneem<sup>1</sup>, Armiya Sultan<sup>2</sup>, Prithvi Singh<sup>1</sup>,  
Hriday R. Bairagya<sup>3</sup>, Hassan Hussain Almasoudi<sup>4</sup>,  
Abdulfattah Yahya M. Alhazmi<sup>5</sup>, Abdulkarim S. Binshaya<sup>6</sup>,  
Mohammed Ageeli Hakami<sup>7</sup>, Bader S. Alotaibi<sup>7</sup>,  
Alaa Abdulaziz Eisa<sup>8</sup>, Abdulaziz Saleh I. Alolaiqy<sup>9</sup>,  
Mohammad Raghiful Hasan<sup>7</sup>, Kapil Dev<sup>2\*</sup> and Ravins Dohare<sup>1\*</sup>

<sup>1</sup>Centre for Interdisciplinary Research in Basic Sciences, Jamia Millia Islamia, New Delhi, India,

<sup>2</sup>Department of Biotechnology, Faculty of Natural Sciences, Jamia Millia Islamia, New Delhi, India,

<sup>3</sup>Department of Bioinformatics, Maulana Abul Kalam Azad University of Technology, Haringhata, West Bengal, India, <sup>4</sup>Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Najran University, Najran, Saudi Arabia, <sup>5</sup>Clinical Pharmacy Department, Umm Al-Qura University, Makkah, Saudi Arabia, <sup>6</sup>Department of Medical Laboratory Sciences, College of Applied Medical Sciences, Prince Sattam Bin Abdulaziz University, Alkharj, Saudi Arabia, <sup>7</sup>Department of Clinical Laboratory Sciences, College of Applied Medical Sciences, Al-Quwayyah, Shaqra University, Riyadh, Saudi Arabia, <sup>8</sup>Department of Medical Laboratory Technology, College of Applied Medical Sciences, Taibah University, Medina, Saudi Arabia, <sup>9</sup>Forensic Medical Services Center, Ministry of Health, Qassim, Saudi Arabia

**Background:** The COVID-19 pandemic caused by SARS-CoV-2 has led to millions of deaths worldwide, and vaccination efficacy has been decreasing with each lineage, necessitating the need for alternative antiviral therapies. Predicting host–virus protein–protein interactions (HV-PPIs) is essential for identifying potential host-targeting drug targets against SARS-CoV-2 infection.

**Objective:** This study aims to identify therapeutic target proteins in humans that could act as virus–host-targeting drug targets against SARS-CoV-2 and study their interaction against antiviral inhibitors.

**Methods:** A structure-based similarity approach was used to predict human proteins similar to SARS-CoV-2 (“hCoV-2”), followed by identifying PPIs between hCoV-2 and its target human proteins. Overlapping genes were identified between the protein-coding genes of the target and COVID-19-infected patient’s mRNA expression data. Pathway and Gene Ontology (GO) term analyses, the construction of PPI networks, and the detection of hub gene

**Abbreviations:** COVID-19, coronavirus disease 2019; h18, rRNA helix 18; hCoV-2, human proteins structurally similar to SARS-CoV-2; HIV, human immunodeficiency virus; HPs, human proteins; HV-PPIs, host–virus protein–protein interactions; NSP, non-structural protein; ORF, open reading frame; RPS2, ribosomal protein subunit 2; RPS3, ribosomal protein subunit 3; SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; T\_hCoV-2, targets of hCoV-2; TIG-DEGs, targets and infected group DEGs; and ZIKV, Zika virus.

modules were performed. Structure-based virtual screening with antiviral compounds was performed to identify potential hits against target gene-encoded protein.

**Results:** This study predicted 19,051 unique target human proteins that interact with hCoV-2, and compared to the microarray dataset, 1,120 target and infected group differentially expressed genes (TIG-DEGs) were identified. The significant pathway and GO enrichment analyses revealed the involvement of these genes in several biological processes and molecular functions. PPI network analysis identified a significant hub gene with maximum neighboring partners. Virtual screening analysis identified three potential antiviral compounds against the target gene-encoded protein.

**Conclusion:** This study provides potential targets for host-targeting drug development against SARS-CoV-2 infection, and further experimental validation of the target protein is required for pharmaceutical intervention.

#### KEYWORDS

COVID-19, virus–host target proteins, structural-based similarity, PPI network, SARS-CoV-2

## 1 Introduction

At present, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has spread rapidly, infecting ~769 million people globally, including ~6 million reported deaths as of 16 August 2023 (<https://covid19.who.int/>). Treatment options for COVID-19 disease include oral medications and intravenous therapy for mild and severe patients, while vaccines are primarily administered as a preventive measure (Ali et al., 2020; Morse et al., 2020; Martín-Sánchez et al., 2023). The use of vaccines initially curbed the SARS-CoV-2 pandemic. However, with the emergence of the Omicron variant (Fan et al., 2022), the efficacy of vaccination against the infection has decreased with each viral lineage. Antiviral drug candidates have provided effective treatment against viral infections, yet their use may be limited by the rapid evolutionary rate of viruses and the possibility of resistance emerging (Dai et al., 2020; Dolgin, 2021; Burki, 2022). Therefore, it is essential to continue the development of alternative antiviral therapies to combat the ongoing threat of emerging viral infections. Host-directed therapy (HDT) has gained momentum in the past decade (Tripathi et al., 2021). Host-targeting drugs with a low evolutionary divergence compared to viruses may avoid treatment failure. Similarities in motif sequence or protein structures between the virus and host proteins can provide clues for predicting virus–human protein–protein interactions (PPIs) (Mariano and Wuchty, 2017; Liu X. et al., 2021). Thus, investigating these PPIs between a virus and its human host could contribute to discovering the viruses' potential human host target proteins (Liu X. et al., 2021).

Although finding intraspecies PPIs is common, predicting PPIs among different species is rare; therefore, this study may allow researchers to understand not only how a pathogenic protein interacts with its host on a molecular level but also how such interactions function in a more extensive cellular network (Lee et al., 2008). In this study, to elucidate the host–virus protein–protein interactions (HV-PPIs), a structure-based interaction approach was implemented to identify PPIs among

human proteins (HPs) with known interactions that are structurally similar to viral proteins (Franzosa and Xia, 2011; Keskin et al., 2016; Rajasekharan et al., 2016). The primary limitation of using a structure-based interaction approach is the inadequate availability of three-dimensional structures and protein interaction of either pathogens or their host (Mariano and Wuchty, 2017). However, several studies have used a structural similarity approach to identify human–pathogen PPIs. Davis et al. were the first to demonstrate the use of 3D structural homology for predicting interspecies PPIs. They scanned the genomes of the host and pathogen for proteins similar to known putative interactions, using structural information, and filtered the remaining interactions based on the biological context for various human pathogens (Davis et al., 2007).

Similarly, Doolittle and Gomez (2011) predicted 502 interactions between HIV and human proteins by using the structure similarity between nine HIV-1 proteins and human proteins with known interactions, functional data from RNAi studies, and cellular component annotation from the Gene Ontology database (Doolittle and Gomez, 2011). Sagar et al. (2021) used a structural similarity approach to identify host–pathogen interactions between 10 Zika virus (ZIKV) proteins and their two distinct hosts, human and *D. melanogaster* (Sagar et al., 2021). Based on previous studies, it has been proposed that if there are pairs of structure-similar virus–host proteins with known interacting host proteins, the target host proteins may also interact with the viral proteins (Sagar et al., 2021). Furthermore, these target host proteins could either enhance the function of the virus or have an inhibitory effect. Hence, exploring the common pathways, molecular functions, biological processes, cellular compartments, and other properties of the host proteins interacting with the virus can help inhibit infectious disease. These virus–host-targeting proteins may act as a potential drug target for the design of antiviral inhibitors.

The present study used a structure-based similarity approach to predict human host proteins similar to SARS-CoV-2 (“hCoV-2”), followed by the identification of PPIs between hCoV-2 and its

target human proteins from different protein interaction databases (referred to as “T\_hCoV-2”). Additionally, overlapping genes were identified between the protein-coding genes of T\_hCoV-2 and COVID-19-infected patient’s messenger RNA (mRNA) expression data obtained from the NCBI-GEO. The overlapped genes between target and infected groups (TIG-DEGs) were further processed to identify differentially expressed genes (DEGs). Pathway and Gene Ontology (GO) enrichment analyses, construction of PPI networks, and detection of hub gene modules for TIG-DEGs were performed. Last, structure-based virtual screening with an antiviral compound database was performed to identify potential hits against the target gene-encoded protein. The vision of the present study was to identify target proteins in humans that may act as host-targeting drug targets against SARS-CoV-2 infection. Further experimental validation of the target protein is required for pharmaceutical intervention against SARS-CoV-2.

## 2 Materials and methods

### 2.1 Viral protein structure collection

The experimentally determined three-dimensional structures of SARS-CoV-2 [envelope protein (PDB ID: 7K3G), main protease (PDB ID: 5R7Y), spike glycoproteins (PDB ID: 6VSB, 6VXX), HR2 domain (PDB ID: 6LVN, 6M1V), NSP2 (PDB ID: 7MSW, 7MSX), NSP3 (PDB ID: 6W02, 6W6Y, 6W9C), NSP7 (PDB ID: 6WTC), NSP8 (6M5I), NSP9 (PDB ID: 6W9Q, 6W4B, 6WCI), NSP12 (PDB ID: 6XEZ, 6M71), NSP13 (PDB ID: 5RL6), NSP14 (PDB ID: 7DIY), NSP15 (PDB ID: 6VWW, 7K0R), NSP16 (PDB ID: 6W4H), ORF3a (PDB ID: 6XDC), ORF7a (PDB ID: 6W37), and ORF8 (PDB ID: 7JTL)] were obtained from the RCSB Database (<https://www.rcsb.org/>) (Rose et al., 2017). The remaining coronavirus structures [membrane protein (M), NSP4, and ORF6] that are not available in the Protein Data Bank (PDB) were modeled using Contact-guided Iterative Threading ASSEMBLY Refinement (C-I-TASSER) (<https://zhanggroup.org/C-I-TASSER/>), an extended version of I-TASSER for high-accuracy protein structure and function predictions (Zheng et al., 2021). The modeled structures of M protein, NSP4, and ORF6 were verified using the Structural Analysis Verification Server (SAVES) program (<https://saves.mbi.ucla.edu/>) and protein structure analysis (ProSA)-web program (<https://prosa.services.came.sbg.ac.at/prosa.php?pdbCode=1zyb>) (Wiederstein and Sippl, 2007). The structural quality (PROCHECK) (Laskowski et al., 1993), non-bonded interactions (ERRAT) (Colovos and Yeates, 1993), and energy profile or Z-score (ProSA using molecular mechanics force field) were thoroughly checked and calibrated. The Z-scores obtained for the M protein, NSP4, and ORF6 were close to the related native conformations regarding their residue length, as determined by X-ray and NMR structures available in the RCSB database. Additionally, nearly 83%–88% of residues in these respective modeled structures were found to occupy favorable allowed regions in the Ramachandran plot (Hoofst et al., 1997). It indicates that the modeled structures are high quality and likely biologically relevant.

### 2.2 Identification of structurally similar proteins among SARS-CoV-2 and its host

The structurally similar human proteins of SARS-CoV-2 were obtained by submitting the PDB IDs of viral protein structures (either known or C-I-TASSER-generated) to the Dali webserver (<http://ekhidna2.biocenter.helsinki.fi/dali/>) (Holm, 2022). In this study, only the *Homo sapiens* proteins with structural similarity score (Z-score)  $\geq 2$  were considered structurally similar proteins of the respective SARS-CoV-2 protein and are referred to as “hCoV-2” proteins. The PDB codes of hCoV-2 obtained from Dali were mapped to their corresponding UniProt ID and gene names by the UniProt Retrieve/ID mapping tool (<https://www.uniprot.org/id-mapping>) (Pundir et al., 2016).

### 2.3 Refinement of predictions

The Dali database may contain multiple structures of the same proteins in the PDB, which could result in repetitive interaction predictions. This issue can arise with certain SARS-CoV-2 proteins with multiple PDB structures, leading to repeating similar predictions. Hence, duplicate PDB IDs were removed before converting them to UniProt accession ID and gene names for target prediction. In addition, multiple hCoV-2 proteins (human proteins structurally similar to SARS-CoV-2) can have common cellular partners. Among these, unique pairs of interactions between human UniProt accession and SARS-CoV-2 proteins were considered.

### 2.4 Prediction of SARS-CoV-2–host protein interactions

Human endogenous protein interactors of SARS-CoV-2 were identified by obtaining the target protein interactors of hCoV-2 during various cellular processes. These cellular partners of hCoV-2 were obtained from HPRD (<https://www.hprd.org/>), BioGRID (<https://thebiogrid.org/>), HIPPIE (<http://cbdm-01.zdv.uni-mainz.de/~mschaefer/hippie/>), and MINT (<https://mint.bio.uniroma2.it/>) databases and are referred to as T\_hCoV-2 (target protein interactors of hCoV-2) (Baolin et al., 2007; Licata et al., 2012; Alanis-Lobato et al., 2017; Oughtred et al., 2019). These datasets are from literature-curated interactions established through *in vitro* and/or *in vivo* methods among human proteins. These cellular proteins, known to interact with hCoV-2 proteins, are presumed to interact with SARS-CoV-2 proteins due to their structural similarities with hCoV-2 proteins.

### 2.5 Comparison with the pre-processed dataset and differentially expressed gene analysis

The target protein interactors of hCoV-2, referred to as T\_hCoV-2, of each of the SARS-CoV-2 proteins, were merged into one dataset corresponding to their HUGO Gene Nomenclature Committee (HGNC) symbols (<https://www.genenames.org/>) by

eliminating the duplicate gene IDs to avoid redundancy. Pre-processed data corresponding to COVID-19 mRNA expression profiling were obtained from the NCBI-GEO (<https://www.ncbi.nlm.nih.gov/geo/>) (Jha et al., 2022). The unique target genes of T<sub>h</sub>CoV-2 are then compared with the COVID-19 patient gene sample dataset from which only the overlapping genes between targets and infected datasets were considered for further DEG analyses.

To identify DEGs between T<sub>h</sub>CoV-2 and COVID-19 samples, a two-sample statistical *t*-test was used, followed by obtaining their  $\log_2$  (fold change) and Benjamin–Hochberg (BH) *p* – value through the limma R package (Ritchie et al., 2015). Genes with BH *p* < 0.05 and  $|\log_2$  (FC)| > 2 were considered DEGs. DEGs with  $\log_2$  (FC) > 2 and  $\log_2$  (FC) < –2 were categorized as up- and downregulated, respectively. The DEGs that overlapped between targets and infected groups were referred to as TIG-DEGs.

## 2.6 Pathway, Gene Ontology terms, and PPI network construction of TIG-DEGs

Pathway and GO term enrichment data for TIG-DEGs were compiled using various libraries, *i.e.*, Kyoto Encyclopedia of genes and genomes (KEGG) and GO-biological process (BP), GO-molecular function (MF), and GO-cellular compartment (CC) terms available within the Enrichr database (<https://maayanlab.cloud/Enrichr/>) (Kuleshov et al., 2016). All the pathways and GO terms corresponding to *p* < 0.01 were determined as statistically significant.

The genes corresponding to the enrichment analysis were then merged into one dataset (eliminating duplicate gene names) and subjected to PPI network construction using the STRING v11.5 web-based tool (<https://string-db.org/>) (Szklarczyk et al., 2021). The construction of this PPI network was based on the highest confidence score >0.9, and it was visualized using Cytoscape v3.9.1 (Shannon et al., 2003). The topological properties of the PPI network were analyzed using the NetworkAnalyzer plugin in Cytoscape. Furthermore, the Molecular Complex Detection (MCODE) plugin was used to identify highly correlated gene clusters/modules within the PPI network. The parameters set in MCODE for cluster detection were as follows: “degree cutoff = 2,” “node score cutoff = 0.2,” “k-score = 2,” “max. depth = 100,” and “cut style = haircut.” The top-scoring genes in the PPI cluster were considered the hub genes.

## 2.7 Target identification and preparation for molecular docking

The RPS3 gene has the most significant interaction partners of the top-scoring PPI cluster genes. Ribosomal protein subunit 3 (RPS3) is a protein-coding gene involved in the viral mRNA translation pathway. Since the viruses exploited the ribosomal proteins and the ribosomal biogenic processes to facilitate their replication, the RPs have been considered effective targets for developing antiviral agents (Dong et al., 2021). The three-dimensional electron microscopy structure of the human ribosomal subunit, PDB ID: 6ZLW (resolution: 2.60 Å), bounded to the SARS-CoV-2 NSP1 protein, was downloaded from the

Protein Data Bank (<https://www.rcsb.org/>) (Rose et al., 2017). The two  $\alpha$ -helices of NSP1 (residues: 154–179) of SARS-CoV-2 directly interact with RPS3, RPS2, and the phosphate backbone of rRNA helix 18 (h18) of 40S ribosomal protein. This interaction with RPS3, RPS2, and h18 rigidly anchors NSP1, obstructing the mRNA entry channel. Therefore, for this study, we took a subcomplex comprising RPS3, RPS2, and h18 as the targets for the molecular docking studies.

## 2.8 Virtual screening and ligand structure preparation

An *in silico* high-throughput virtual screening of known therapeutic COVID-19 drugs from the PubChem library was obtained (<https://pubchem.ncbi.nlm.nih.gov/#query=covid-19>) (Kim et al., 2019). The total search results were filtered thrice for final docking studies. Following are the filters for the selection of final screened compounds: filter 1, database filters (molecular weights should be from 100 to 500 g/mol, rotatable bond count from 0 to 7, h-bond acceptor count from 0 to 10, the polar area from 0 to 150 Å<sup>2</sup>, and XlogP from –6.3 to 5); filter 2, calculation of physicochemical properties (no Lipinski’s rule violations, no lead likeness violations, and high GI absorption) using the SwissADME web server (<http://www.swissadme.ch/>) (Daina et al., 2017); and filter 3, calculation of pharmacokinetics and toxicological properties (cLogP, solubility, TPSA, drug-likeness, drug score, mutagenic, tumorigenic, irritant, and reproductive effective) using the Osiris Property Explorer (Sander, 2001).

The 3D structures of the final screened compounds were generated using the OpenBabel program (O Boyle et al., 2011). The 2D SDF files were used as input files for OpenBabel for converting them to 3D structures, followed by energy minimization in MMFF94 force field (Halgren, 1996). Finally, all the structures of ligands were converted to the PDBQT format with appropriate rotatable bonds.

## 2.9 Molecular docking

The PDBQT file of the subcomplex receptor was prepared using Kollman united atom charges for molecular docking study using AutoDockTools v 1.5.6 (Singh and Kollman, 1984; Morris et al., 2009). As the subcomplex receptor contained two proteins and one nucleic acid, the tentative binding site for the inhibitor is unknown. Hence, a blind docking study using AutoDock Vina (v1.1.2) was performed (Trott and Olson, 2009). Grid point spacing was set at 1 Å, grid point 56 was taken in each direction of the grid box, and the concerned box was centered at the subcomplex receptor. The ligands bind to two probable sites of the subcomplex receptor—one at the rRNA h18 (ligand-binding site-I) and another at the RPS3 protein (II). The best three conformations of each ligand at two binding locations were identified after cluster analysis according to their lowest binding energy. The receptor–ligand hydrogen bond interactions were analyzed using the Swiss PDB viewer (Guex and Peitsch, 1997) and visualized using PyMOL v 2.5 (the PyMOL Molecular Graphics System, Version 1.2r3pre, Schrödinger, LLC).

TABLE 1 List of 20 SARS-CoV-2 proteins selected for identifying structurally similar human host proteins.

| S. no. | SARS-CoV-2 protein                          | Chain | Method | Sequence length | Resolution (Å) |
|--------|---|-------|--------|-----------------|----------------|
| 1      | <b>Envelope protein</b>                     |       |        |                 |                |
|        | 7K3G  | A     | NMR    | 31              | NA*            |
| 2      | <b>NSP2</b>                                 |       |        |                 |                |
|        | 7MSW  | A     | EM     | 638             | 3.76           |
|        | 7MSX  | A     | EM     | 638             | 3.15           |
| 3      | <b>Papain-like protease (NSP3)</b>          |       |        |                 |                |
|        | 6W02  | A     | X-ray  | 170             | 1.5            |
|        | 6W6Y  | A     | X-ray  | 170             | 1.45           |
|        | 6W9C  | A     | X-ray  | 317             | 2.7            |
| 4      | <b>Main protease (NSP5)</b>                 |       |        |                 |                |
|        | 6LU7  | A     | X-ray  | 306             | 2.16           |
| 5      | <b>NSP7</b>                                 |       |        |                 |                |
|        | 6WTC-A                                      | A     | X-ray  | 86              | 1.85           |
| 6      | <b>NSP8</b>                                 |       |        |                 |                |
|        | 6M5I-A                                      | A     | X-ray  | 198             | 2.5            |
| 7      | <b>NSP9</b>                                 |       |        |                 |                |
|        | 6W9Q-A                                      | A     | X-ray  | 133             | 2.05           |
|        | 6W4B-A                                      | A     | X-ray  | 117             | 2.95           |
|        | 6WC1-A                                      | A     | X-ray  | 116             | 2.4            |
| 8      | <b>RNA-dependent RNA polymerase (NSP12)</b> |       |        |                 |                |
|        | 6XEZ-A                                      | A     | EM     | 932             | 3.5            |
|        | 6M71-A                                      | A     | EM     | 942             | 2.9            |
| 9      | <b>Helicase (NSP13)</b>                     |       |        |                 |                |
|        | 5RL6  | A     | X-ray  | 601             | 1.92           |
| 10     | <b>NSP14</b>                                |       |        |                 |                |
|        | 7DIY  | B     | X-ray  | 294             | 2.69           |
| 11     | <b>NSP15</b>                                |       |        |                 |                |
|        | 6VWW  | A     | X-ray  | 370             | 2.2            |
|        | 7K0R  | A     | EM     | 362             | 3.3            |
| 12     | <b>NSP16</b>                                |       |        |                 |                |
|        | 6W4H  | A     | X-ray  | 301             | 1.8            |
| 13     | <b>ORF3a</b>                                |       |        |                 |                |
|        | 6XDC  | A     | EM     | 284             | 2.9            |
| 14     | <b>ORF7a</b>                                |       |        |                 |                |
|        | 6W37  | A     | X-ray  | 67              | 2.9            |
| 15     | <b>ORF8</b>                                 |       |        |                 |                |
|        | 7JTL  | A     | X-ray  | 107             | 2.04           |

(Continued on following page)



TABLE 1 (Continued) List of 20 SARS-CoV-2 proteins selected for identifying structurally similar human host proteins.

| S. no.  | SARS-CoV-2 protein      | Chain              | Method                 | Sequence length            | Resolution (Å) |
|---|-------------------------|--------------------|------------------------|----------------------------|----------------|
| 16  | <b>S2 subunit</b>       |                    |                        |                            |                |
|   | 6LVN                    | A                  | X-ray                  | 36                         | 2.47           |
|   | 6M1V                    | A                  | X-ray                  | 119                        | 1.5            |
| 17  | <b>Spike protein</b>    |                    |                        |                            |                |
|   | 6VSB                    | A                  | EM                     | 1288                       | 3.46           |
| Modeled Structures using C-I-TASSER (predicted) |                         |                    |                        |                            |                |
|   |                         | <i>Sequence ID</i> | <i>Sequence Length</i> | <i>Estimate TM-score**</i> |                |
| 18  | <b>Membrane protein</b> | QHD43419           | 222                    | 0.59                       |                |
| 19  | <b>NSP4</b>             | QHD43415.4         | 500                    | 0.61                       |                |
| 20  | <b>ORF6</b>             | QHD43420           | 61                     | 0.50                       |                |

### 3 Result

#### 3.1 Identification of structurally similar human proteins of SARS-CoV-2 and hCoV-2 host PPIs

We used the Dali Lite webserver to identify human host proteins that exhibit structural similarity to the proteins of SARS-CoV-2. Twenty SARS-CoV-2 proteins were selected for this study (17 proteins from known sources, i.e., available via the RCSB PDB, and 3 proteins predicted or modeled using the C-I-TASSER program) (Table 1). The viral protein structures were then submitted to the Dali webserver to compare 3D structural coordinates of virus and human host proteins. To make the study more comprehensive and avoid redundancy, all the structurally similar host proteins for each viral protein were merged into one dataset, and duplicates were removed. For example, for NSP3 of SARS-CoV-2, the human proteins structurally similar to these three PDB IDs: 6W02, 6W6Y, and 6W9C were merged into one column, and duplicate human PDB IDs were then removed. A similar protocol was followed for the other viral proteins, with more than one PDB ID obtained for this study. It resulted in the identification of 6,116 human proteins (hCoV-2) similar to 20 SARS-CoV-2 proteins (Table 2 and Supplementary Table S1).

Afterward, we identified all possible interacting partners of hCoV-2 proteins of the 20 proteins of SARS-CoV-2 from four different databases (HPRD, MINT, HIPPIE, and BioGRID). The target protein interactors of hCoV-2 from each database corresponding to 165,051 genes were merged into one unique dataset without duplicates. A total of 19,051 unique interactors were identified for these 20 SARS-CoV-2 proteins. These unique interactors are termed T\_hCoV-2 proteins (Table 2 and Supplementary Table S2).

#### 3.2 Comparison with the pre-processed dataset and differentially expressed gene analysis

The pre-processed SARS-CoV-2 mRNA expression profile (GSE164805) datasets comprised five controls, and 10 COVID-19

(five mild and five severe) samples were compared with the T\_hCoV-2 genes. The overlapping genes between targets and mRNA datasets amounted to 8,336 genes (Figure 1A). These genes were then differentially expressed corresponding to  $p$ -value  $< 0.05$  and  $|\log_2(\text{fold change})| > 2$ . The DEGs overlapping between targets and infection severity groups were considered TIG-DEGs, which amount to 1,120 unique genes. Amongst these genes, 663 and 457 were up- and downregulated, respectively (Figure 1B). The most highly up- and downregulated DEGs were HBD [ $\log_2(FC) = 6.00$ ] and TEX101 [ $\log_2(FC) = -10.34$ ], respectively (Figure 1C).

#### 3.3 Pathway and Gene Ontology enrichment analyses

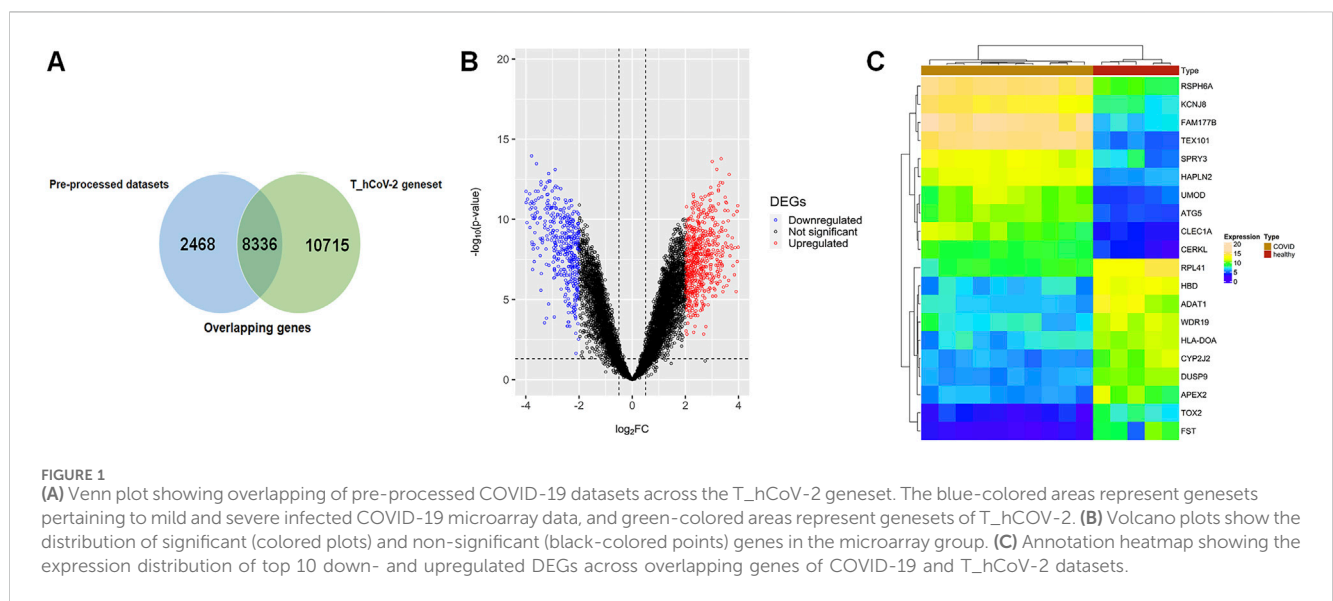
All the 1,120 TIG-DEGs participated in the Gene Ontology and pathway enrichment analyses. All the significant pathways and GO-BP, GO-MF, and GO-CC terms ( $p$ -value  $< 0.01$ ) are shown in Supplementary Table S3. The most significant pathways and GO-BP, GO-MF, and GO-CC terms were coronavirus disease ( $p$ -value =  $8.49 \times 10^{-08}$ ), cellular macromolecule biosynthetic process ( $p$ -value =  $1.88 \times 10^{-08}$ ), RNA binding ( $p$ -value =  $5.64 \times 10^{-05}$ ), and ribosome ( $p$ -value =  $1.18 \times 10^{-06}$ ) (Supplementary Table S3). For a more comprehensive representation, only the top 10 most significant pathways and GO terms are shown in Figure 2.

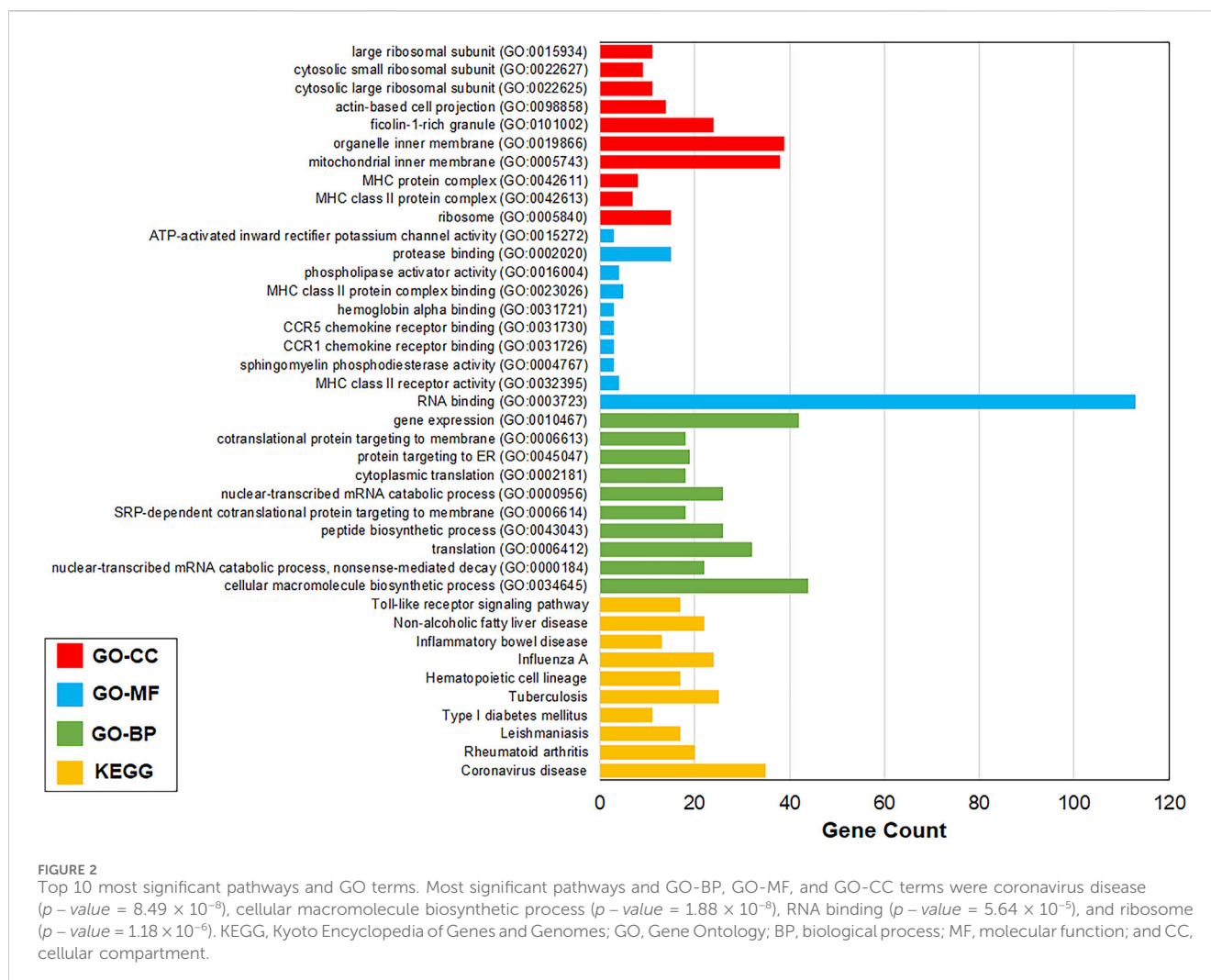
#### 3.4 PPI network analysis of TIG-DEGs

A total of 606 genes comprising all the pathways and GO terms ( $p$ -value  $< 0.01$ ) were inputted into the STRING database. The PPI network constructed using Cytoscape software included 302 nodes, and 688 edges are shown in Figure 3A. A total of 302 genes out of 1,120 TIG-DEGs participated in the PPI network at the selected highest confidence level, i.e.,  $> 0.9$ . Moreover, Molecular Complex Detection (MCODE) revealed that the most highly significant cluster with a *top score* = 18 was selected

**TABLE 2** Number of hCoV-2 structurally similar proteins (known and predicted) to the SARS-CoV-2 proteins and its unique host interactors known as T\_hCoV-2 proteins.

| S. no. | SARS-CoV-2 protein                   | Structure type | Number of hCoV-2 similar proteins | Targeted T_hCoV-2 proteins |
|--------|--------------------------------------|----------------|-----------------------------------|----------------------------|
| 1      | Envelope protein                     | Known          | 86                                | 5863                       |
| 2      | NSP2                                 | Known          | 371                               | 1158                       |
| 3      | Papain-like protease (NSP3)          | Known          | 83                                | 342                        |
| 4      | Main protease (NSP5)                 | Known          | 67                                | 204                        |
| 5      | NSP7                                 | Known          | 393                               | 473                        |
| 6      | NSP8                                 | Known          | 576                               | 513                        |
| 7      | NSP9                                 | Known          | 76                                | 90                         |
| 8      | RNA-dependent RNA polymerase (NSP12) | Known          | 06                                | 04                         |
| 9      | Helicase (NSP13)                     | Known          | 466                               | 257                        |
| 10     | NSP14                                | Known          | 17                                | 10                         |
| 11     | NSP15                                | Known          | 30                                | 33                         |
| 12     | NSP16                                | Known          | 60                                | 34                         |
| 13     | ORF3a                                | Known          | 266                               | 67                         |
| 14     | ORF7a                                | Known          | 678                               | 825                        |
| 15     | ORF8                                 | Known          | 628                               | 41                         |
| 16     | S2 subunit                           | Known          | 995                               | 558                        |
| 17     | Spike protein                        | Known          | 23                                | 12                         |
| 18     | Membrane protein                     | Predicted      | 532                               | 8167                       |
| 19     | NSP4                                 | Predicted      | 08                                | 15                         |
| 20     | ORF6                                 | Predicted      | 755                               | 385                        |
|        | <b>Total</b>                         |                | <b>6,116</b>                      | <b>19,051</b>              |





amongst all the 17 identified clusters, consisting 19 nodes and 171 edges (Figure 3B). Essential centrality measures such as node degree, betweenness, closeness, average shortest path length, clustering coefficient, radiality, and topological coefficient of the PPI network are provided in Supplementary Table S4. Based on the centrality measures, the significant hub gene RPS3 (ribosomal protein subunit 3, degree = 25), with the highest number of interacting partners, was identified.

### 3.5 Virtual screening of ligands

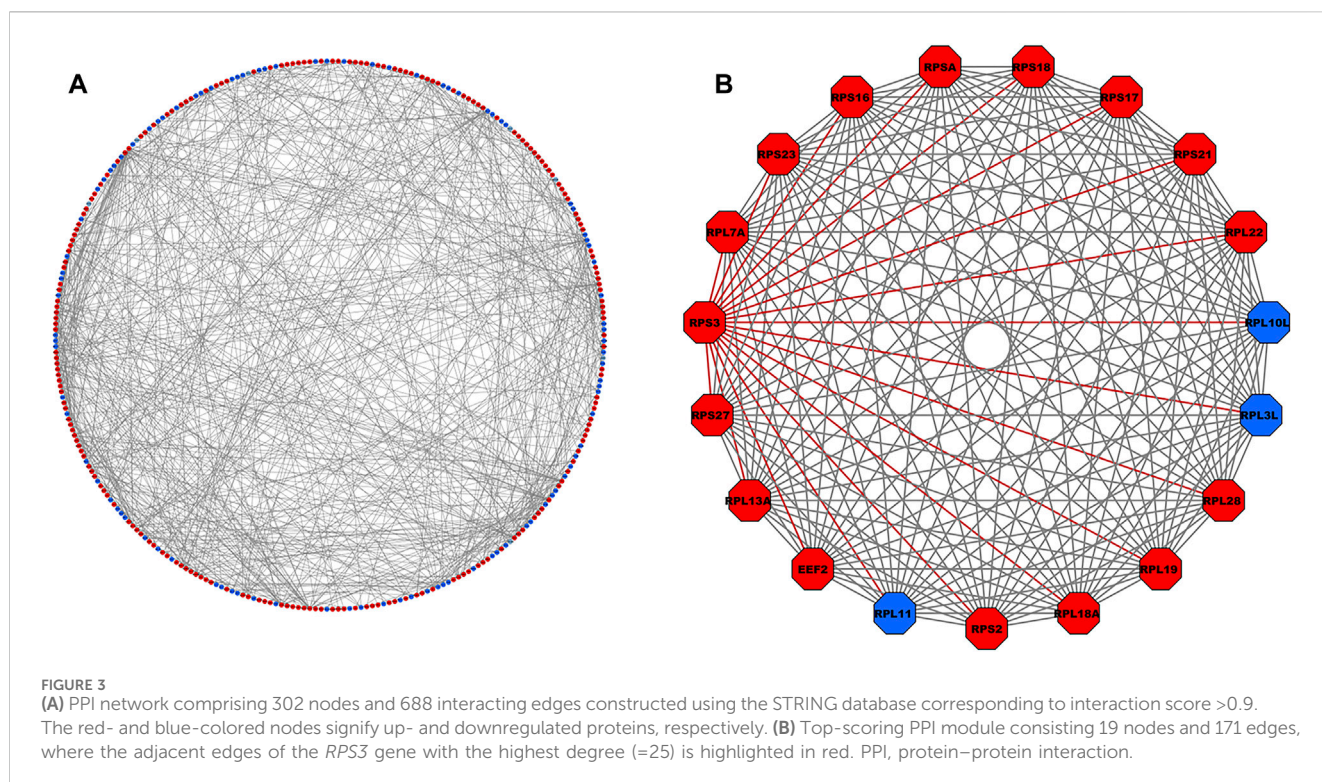
The PubChem library was searched with the keyword “COVID-19,” resulting in 1,709 compounds. The compounds were filtered out through three filtering processes (Martín-Sánchez et al., 2023): database filters (Ali et al., 2020), pharmacokinetic properties, and (Morse et al., 2020) toxicological properties. Through database filtering, only 886 candidates were selected that have a molecular weight between 100 and 500 g/mol, rotatable bond count between 0 and 7, h-bond acceptor count between 0 and 10, polar area between 0 and 150 Å<sup>2</sup>, and XlogP between -6.3 and 5. Government organizations, research and development, journal

publishers, and NIH initiatives were selected as the data source category. The SwissADME webserver was accessed to calculate pharmacokinetic parameters. After the pharmacokinetic filtration process, 42 compounds that show zero violations of Lipinski’s rule, zero lead likeness violations, and high gastrointestinal (GI) were selected. To calculate toxicological parameters, the OSIRIS Property Explorer was used to predict the toxicity risks and drug scores of the 42 compounds. The program evaluated the risks of different side effects, such as mutagenic, irritant, tumorigenic, reproductive effects, and drug-related properties. Furthermore, the overall drug score was calculated by summing up various parameters such as cLogP, solubility (logS), molecular weight, drug-likeness, and toxicity risk. The final filtration process resulted in 28 compounds showing no toxicity risks and drug score values between 0.40 and 1.00, making it a good library for further docking studies (Table 3).

### 3.6 Molecular docking

In this study, blind docking was performed since the subcomplex receptor of ribosomal protein has no ligand-





binding site (Figures 4A, B). The rotatable bonds of the 28 ligands were kept flexible, while the subcomplex macromolecule was adopted as a rigid structure. The analysis of the molecular docking results shows that most of the ligands screened bind into two active sites: (I) the center of RPS3 protein and (II) the head of rRNA helix 18 (Figure 4C). The top three binding poses with the best binding energies at each ligand-binding site are given in Table 4. The best binding energies of compound baicalein, vadadustat, and estetrol are observed at the head of the rRNA helix 18 site with binding energies  $-8.24$ ,  $-8.01$ , and  $-7.98$  kcal/mol, respectively. Similarly, baicalein, estetrol, and 3-[[[(3-methyl-1,2,4-oxadiazol-5-yl)methyl]carbamoyl]benzoic acid show the best binding energies at the center of RPS3 protein with binding energies  $-7.54$ ,  $-7.33$ , and  $-7.24$  kcal/mol, respectively. Baicalein has been shown to have the highest binding pose for both ligand-binding sites. All the compounds form hydrogen bonds with the rRNA helix18 site that contains the following nucleotides: U607, C608, U609, G610, G611, G613, U630, A629, and U631, whereas the RPS3-binding site contains the following residues: Ala30, Gly33, Arg54, Arg94, Ala100, Gln101, Glu103, and Ser104 (Table 4). The interactions of baicalein, estetrol, and vadadustat (rRNA h18) are shown in Figure 5A, whereas the interactions of baicalein, estetrol, and 3-[[[(3-methyl-1,2,4-oxadiazol-5-yl)methyl]carbamoyl]benzoic acid (RPS3 protein) are shown in Figure 5B. In Figure 5A, (i) baicalein binds in the region of rRNA h18 through h-bond interactions with U607, G609, G610, U630, and U631; (ii) estetrol shows h-bond interactions with U607, U609, G611, G613, and A629; and (iii) vadadustat forms h-bond interactions with U608, G610, G611, and U630. Meanwhile, in Figure 5B, (iv) baicalein binds in the region of RPS3 protein

through h-bond interactions with residues Ala30, Gly33, Arg54, Arg94, and Ala100; (v) estetrol forms h-bond interactions with residues Arg94, Ala100, and Gln101; and (vi) 3-[[[(3-methyl-1,2,4-oxadiazol-5-yl)methyl]carbamoyl]benzoic acid forms h-bond with residues Arg54, Arg94, Gln101, Glu103, and Ser104.

## 4 Discussion

COVID-19 has highlighted several therapeutic gaps in healthcare. One of the most significant gaps is the lack of effective antiviral drugs to treat the disease. Although some drugs have been repurposed to treat COVID-19, such as remdesivir, their efficacy is limited, and they are not a cure for the disease. Another gap is the lack of effective treatments for severe cases of COVID-19. While some patients recover from the disease with supportive care, there are still many who die from the disease. There is an urgent need for more effective treatments to reduce the mortality rate of COVID-19.

The present study used bioinformatics tools to explore the structural similarities between human host proteins and SARS-CoV-2 proteins and their interacting partners. This analysis led to the identification of an impressive number of 6,116 human proteins and 19,051 unique interacting partners of SARS-CoV-2 proteins. Further analyses of the target interactors of hCoV-2 proteins, combined with the differentially expressed genes (DEGs) of COVID-19 patients, revealed a total of 1,120 TIG-DEGs that participated in Gene Ontology and pathway enrichment analyses. Remarkably, the most significant pathways and GO terms identified were related to coronavirus disease, the cellular macromolecule biosynthetic process, RNA binding, and ribosome. Subsequently,

TABLE 3 Physicochemical properties of 28 selected drugs calculated after filtering thrice database filter, pharmacokinetic filter, and toxicological filter. The following drugs show no toxicity risks (mutagenicity, tumorigenic, irritant, and reproductive effective).

| S.no. | PubChemID | Compound Name  | Chemical formula   | Osiris Property Explorer |             |              |               |             | Swiss ADME webserver |          |                 |             |
|-------|-----------|--|--|--------------------------|-------------|--------------|---------------|-------------|----------------------|----------|-----------------|-------------|
|       |           |  |  | MW                       | cLogP       | Solubility   | Drug-likeness | Drug score  | TPSA                 | Lipinski | Bioavailability | Synthetic   |
| 1     | 3365      | Fluconazole  | C <sub>13</sub> H <sub>12</sub> F <sub>2</sub> N <sub>6</sub> O                | 306.27                   | -0.11       | -2.17        | 1.99          | 0.87        | 81.65                | 0        | 0.55            | 2.45        |
| 2     | 4236      | Modafinil  | C <sub>15</sub> H <sub>15</sub> NO <sub>2</sub> S                              | 273.35                   | 0.38        | -3.84        | 0.52          | 0.69        | 79.37                | 0        | 0.55            | 2.83        |
| 3     | 4594      | Omeprazole   | C <sub>17</sub> H <sub>19</sub> N <sub>3</sub> O <sub>3</sub> S                | 345.42                   | 2.09        | -2.72        | -0.51         | 0.6         | 96.31                | 0        | 0.55            | 3.58        |
| 4     | 5329      | Sulfamethoxazole   | C <sub>10</sub> H <sub>11</sub> N <sub>3</sub> O <sub>3</sub> S                | 253.28                   | 0.44        | -3.02        | 2.77          | 0.88        | 106.6                | 0        | 0.55            | 2.73        |
| 5     | 13730     | 2'-Deoxyadenosine  | C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>3</sub>                  | 251.24                   | -0.89       | -2.72        | -5.08         | 0.47        | 119.31               | 0        | 0.55            | 3.6         |
| 6     | 27125     | <b>Estetrol</b>  | <b>C<sub>18</sub>H<sub>24</sub>O<sub>4</sub></b>                               | <b>304.38</b>            | <b>2.17</b> | <b>-3.22</b> | <b>-0.61</b>  | <b>0.58</b> | <b>80.92</b>         | <b>0</b> | <b>0.55</b>     | <b>3.99</b> |
| 7     | 35370     | Zidovudine   | C <sub>10</sub> H <sub>13</sub> N <sub>5</sub> O <sub>4</sub>                  | 267.24                   | -1.02       | -1.44        | 2.12          | 0.91        | 134.07               | 0        | 0.55            | 3.93        |
| 8     | 73115     | 1-((2S,3R,4S,5S)-3-fluoro-4-hydroxy-5-(hydroxymethyl)tetrahydrofuran-2-yl)-5-methylpyrimidine-2,4(1H,3H)-dione | C <sub>10</sub> H <sub>13</sub> FN <sub>2</sub> O <sub>5</sub>                 | 260.22                   | -1.4        | -1.3         | 2.56          | 0.93        | 104.55               | 0        | 0.55            | 3.95        |
| 9     | 446541    | Mycophenolic acid  | C <sub>17</sub> H <sub>20</sub> O <sub>6</sub>                                 | 320.34                   | 2.3         | -3.01        | -0.61         | 0.58        | 93.06                | 0        | 0.56            | 3.02        |
| 10    | 2566008   | 1-[4-(5-Chlorothiophen-2-yl)sulfonylpiperazin-1-yl]ethanone  | C <sub>10</sub> H <sub>13</sub> ClN <sub>2</sub> O <sub>3</sub> S <sub>2</sub> | 308.8                    | 1.67        | -1.28        | 5.94          | 0.93        | 94.31                | 0        | 0.55            | 3.08        |
| 11    | 3759658   | 1-[4-(Thiophene-2-sulfonyl)-piperazin-1-yl]ethanone  | C <sub>10</sub> H <sub>14</sub> N <sub>2</sub> O <sub>3</sub> S <sub>2</sub>   | 274.36                   | 0.86        | -0.52        | 6.03          | 0.95        | 94.31                | 0        | 0.55            | 2.88        |
| 12    | 3803220   | 2-(4-Acetylpiperazin-1-yl)sulfonylbenzotrile   | C <sub>13</sub> H <sub>15</sub> N <sub>3</sub> O <sub>3</sub> S                | 293.34                   | 0.7         | -1.38        | 1.42          | 0.85        | 89.86                | 0        | 0.55            | 2.62        |
| 13    | 5281605   | <b>Baicalein</b>   | <b>C<sub>15</sub>H<sub>10</sub>O<sub>5</sub></b>                               | <b>270.24</b>            | <b>2.34</b> | <b>-2.86</b> | <b>0.75</b>   | <b>0.75</b> | <b>90.9</b>          | <b>0</b> | <b>0.55</b>     | <b>3.02</b> |
| 14    | 5360515   | Naltrexone   | C <sub>20</sub> H <sub>23</sub> NO <sub>4</sub>                                | 341.4                    | 1.58        | -3.19        | 4.67          | 0.85        | 70                   | 0        | 0.55            | 4.72        |
| 15    | 6453528   | Sulodexide   | C <sub>12</sub> H <sub>17</sub> N <sub>5</sub> O <sub>4</sub>                  | 295.29                   | -0.99       | -2.44        | -0.88         | 0.6         | 114.55               | 0        | 0.55            | 4.14        |
| 16    | 9926791   | Tofacitinib  | C <sub>16</sub> H <sub>20</sub> N <sub>6</sub> O                               | 312.37                   | 1.19        | -3.59        | -2.54         | 0.46        | 88.91                | 0        | 0.55            | 3.26        |
| 17    | 15277004  | 5'-Thiothymidine   | C <sub>10</sub> H <sub>14</sub> N <sub>2</sub> O <sub>4</sub> S                | 258.29                   | -0.39       | -2.28        | -0.51         | 0.65        | 123.12               | 0        | 0.55            | 3.71        |
| 18    | 23634441  | <b>Vadadustat</b>  | <b>C<sub>14</sub>H<sub>11</sub>ClN<sub>2</sub>O<sub>4</sub></b>                | <b>306.7</b>             | <b>1.2</b>  | <b>-3.15</b> | <b>-3.53</b>  | <b>0.45</b> | <b>99.52</b>         | <b>0</b> | <b>0.56</b>     | <b>2.31</b> |
| 19    | 25126798  | Ruxolitinib  | C <sub>17</sub> H <sub>18</sub> N <sub>6</sub>                                 | 306.37                   | 1.68        | -3.92        | -6.97         | 0.41        | 83.18                | 0        | 0.55            | 3.16        |
| 20    | 25271624  | Adrafinil, (R)-  | C <sub>15</sub> H <sub>15</sub> NO <sub>3</sub> S                              | 289.35                   | 0.27        | -4.04        | 0.85          | 0.7         | 85.61                | 0        | 0.55            | 3.21        |
| 21    | 27646027  | 1-acetyl-N-(6-methoxypyridin-3-yl)piperidine-4-carboxamide   | C <sub>14</sub> H <sub>19</sub> N <sub>3</sub> O <sub>3</sub>                  | 277.32                   | 0.99        | -1.93        | 5.77          | 0.94        | 71.53                | 0        | 0.55            | 2.37        |
| 22    | 47110626  | 1-acetyl-N-(2-hydroxyphenyl)piperidine-4-carboxamide   | C <sub>14</sub> H <sub>18</sub> N <sub>2</sub> O <sub>3</sub>                  | 262.3                    | 1.36        | -1.88        | 6.08          | 0.94        | 69.64                | 0        | 0.55            | 1.75        |

(Continued on following page)

**TABLE 3 (Continued)** Physiochemical properties of 28 selected drugs calculated after filtering thrice database filter, pharmacokinetic filter, and toxicological filter. The following drugs show no toxicity risks (mutagenicity, tumorigenic, irritant, and reproductive effective).

| S.no. | PubChemID | Compound Name   | Chemical formula  | Osiris Property Explorer |       |            |               | Swiss ADME webserver |        |          |                 |           |
|-------|-----------|---|---|--------------------------|-------|------------|---------------|----------------------|--------|----------|-----------------|-----------|
|       |           |   |   | MW                       | cLogP | Solubility | Drug-likeness | Drug score           | TPSA   | Lipinski | Bioavailability | Synthetic |
| 23    | 47289091  | N-[3-(thiomorpholine-4-carbonyl)phenyl]acetamide                          | C <sub>13</sub> H <sub>16</sub> N <sub>2</sub> O <sub>2</sub> S | 264.34                   | 1.44  | -2.35      | 3.05          | 0.91                 | 74.71  | 0        | 0.55            | 2.01      |
| 24    | 61865966  | 3-[[[3-Methyl-1,2,4-oxadiazol-5-yl)methyl]carbamoyl]benzoic acid          | C <sub>12</sub> H <sub>11</sub> N <sub>3</sub> O <sub>4</sub>   | 261.23                   | 0.41  | -1.34      | -0.92         | 0.61                 | 105.32 | 0        | 0.56            | 2.37      |
| 25    | 135398593 | 2'-Deoxyinosine   | C <sub>10</sub> H <sub>12</sub> N <sub>4</sub> O <sub>4</sub>   | 252.23                   | -1.41 | -2.02      | 3.04          | 0.93                 | 113.26 | 0        | 0.55            | 3.51      |
| 26    | 145998218 | 1-(4-Methylpiperazin-1-yl)-2-[(1H-pyrrolo [2,3-b]pyridin-3-yl)ethan-1-one | C <sub>14</sub> H <sub>18</sub> N <sub>4</sub> O                | 258.32                   | 0.76  | -1.37      | 9.16          | 0.95                 | 52.23  | 0        | 0.55            | 2.12      |
| 27    | 146019222 | N[alpha]-acetyl-N-(3-bromoprop-2-yn-1-yl)-L-tyrosinamide                  | C <sub>14</sub> H <sub>15</sub> BrN <sub>2</sub> O <sub>3</sub> | 339.18                   | 1.08  | -2.93      | -3.14         | 0.46                 | 78.43  | 0        | 0.55            | 2.74      |
| 28    | 146037571 | N-(3-[[[(2R)-4-oxoazetidin-2-yl]oxy]phenyl]-2-(pyrimidin-5-yl)acetamide   | C <sub>15</sub> H <sub>14</sub> N <sub>4</sub> O <sub>3</sub>   | 298.3                    | 0.67  | -2.31      | 1.63          | 0.85                 | 93.21  | 0        | 0.55            | 2.95      |

Abbreviations, MW, molecular weight (g/mol); cLogP, calculated Log P; TPSA, topological polar surface area (in Å). Bold values represent the final compounds selected in the result section.

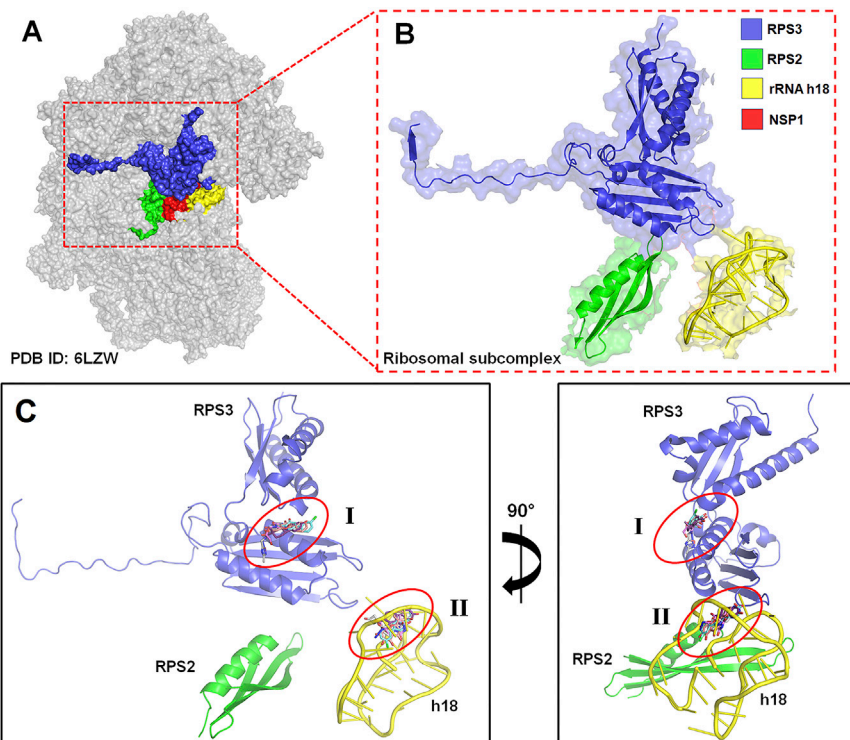
the PPI network analysis identified *RPS3* as the hub gene in a cluster of 19 nodes and 171 edges.

It should be noted that the approach of identifying host proteins structurally similar to viral proteins has been used before, as evidenced by [Gordon et al. \(2020\)](#). Nonetheless, the current study identified a much higher number of host proteins and interacting partners, due to the utilization of a different method of merging structurally similar host proteins and removing duplicates. Similarly, the approach of comparing the target interactors of hCoV-2 proteins with the DEGs of COVID-19 patients has been used before, as demonstrated by [Blanco-Melo et al. \(2020\)](#). Despite this, the present study identified a larger number of TIG-DEGs, indicating greater level of sensitivity in the methodology used.

The PPI network analysis of TIG-DEGs using Cytoscape software is a well-established method in systems biology that allows the identification of hub genes and protein complexes. The identification of *RPS3* as a hub gene in the present study is particularly noteworthy, as it has also been identified by [Prasad et al. \(2021\)](#) as a hub gene in their PPI network analysis of SARS-CoV-2 interactors. The current study and findings from these two earlier studies suggest that *RPS3* could be a potential therapeutic target for COVID-19.

Last, filtering potential drug candidates from an extensive library of compounds based on various pharmacokinetic and toxicological properties is a conventional strategy in drug discovery ([Reyaz et al., 2021](#); [Sultan et al., 2021](#); [Sultan et al., 2022](#)). This study identified three compounds—baicalein, vadadustat, and estretol—with promising properties, such as binding affinity and residual interaction with *RPS3*, for COVID-19 treatment. These compounds could be further tested *in vitro* and *in vivo* for their efficacy. Although different research groups have tried to investigate that baicalein may act as an anti-SARS-CoV-2 drug through *in vitro* analysis, its safety and efficacy in SARS-CoV-2-infected transgenic animals have not been studied yet ([Huang et al., 2020](#); [Su et al., 2020](#); [Liu H. et al., 2021](#); [Dinda et al., 2023](#)). However, [Song et al. \(2021\)](#) investigated the therapeutic effect of baicalein against SARS-CoV-2 both *in vivo* and *in vitro* ([Song et al., 2021](#)). Further experimental studies are recommended to prove that baicalein may act as an effective anti-COVID-19 molecule.

In summary, this study's bioinformatics analysis provides novel insights into the biology of SARS-CoV-2 infection and potential therapeutic targets. Although some of the methodologies used have been used before, this study's findings highlight the importance of leveraging multiple bioinformatics tools to identify host proteins, interacting partners, and potential drug candidates. Nonetheless, further experimental validation is necessary to confirm this study's findings and facilitate the translation of these findings into clinically applicable interventions. This study has some strengths and some limitations. One of the major strengths of this study is the comprehensive and systematic approach toward identifying the host proteins that interact with SARS-CoV-2 proteins and their potential roles in COVID-19 pathogenesis. This study used multiple databases and tools to identify the interactors and DEGs in COVID-19 patients. This approach provides a more comprehensive understanding of the host-virus interactions and their potential roles in the disease pathogenesis. This study also identified



**FIGURE 4** (A) Three-dimensional EM structure of the human 40S ribosomal subunit (PDB ID: 6LZW), bounded to the ribosomal subcomplex and SARS-CoV-2 NSP1 protein (in red). (B) Magnified figure in cartoon representation of the ribosomal subcomplex comprised RPS3 (in blue), RPS2 (in green), and the phosphate backbone of rRNA helix 18 (h18) (in yellow) of 40S ribosomal protein. (C) Schematic representation of the two possible ligand-binding site pockets in the ribosomal subcomplex (I) in the center of RPS3 and (II) the head of rRNA h18 identified through blind docking.

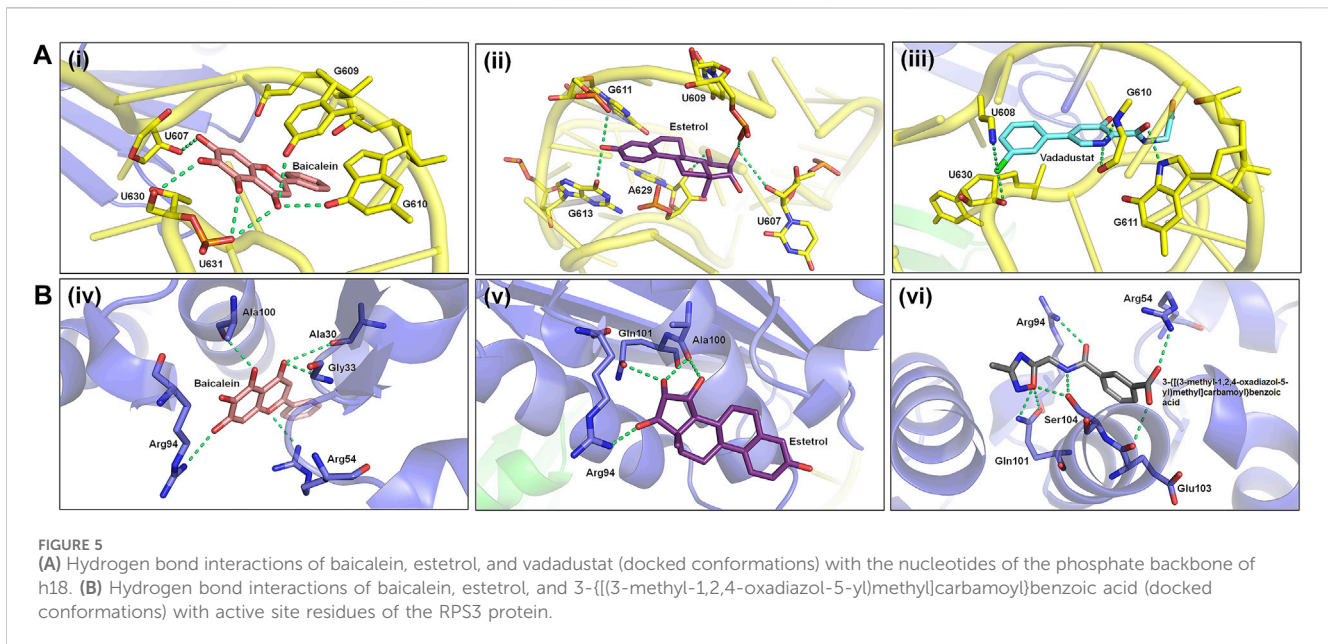
**TABLE 4** Top three docking pose prediction results for each ligand-binding sites: (I) the head of rRNA helix 18 and (II) the center of RPS3 protein.

| Ligand Binding site    | S. No. | Compound  | PubChem ID | Binding Affinity (kcal/mol) | H-bond interactions ( $\leq 3.50 \text{ \AA}$ )   |
|------------------------|--------|---|------------|-----------------------------|---|
| I. rRNA helix 18 (h18) | 1      | Baicalein   | 5281605    | -8.24                       | U607 <sup>O3'</sup> (3.12), U607 <sup>O2'</sup> (3.09), U630 <sup>O3'</sup> (3.25), U630 <sup>O2'</sup> (3.27), U630 <sup>O3'</sup> (3.07), U631 <sup>OP1</sup> (2.85), U609 <sup>O4</sup> (3.33), U631 <sup>OP1</sup> (3.30), G610 <sup>OP1</sup> (2.90) |
|                        | 2      | Vadadustat  | 23634441   | -8.01                       | G611 <sup>N7</sup> (3.00), G610 <sup>N7</sup> (3.07), C608 <sup>N4</sup> (3.57), U630 <sup>O3'</sup> (2.93)   |
|                        | 3      | Estetrol  | 27125      | -7.98                       | G613 <sup>O6</sup> (3.14), G611 <sup>OP2</sup> (3.19), A629 <sup>O2'</sup> (3.33), U609 <sup>OP2</sup> (3.10), U607 <sup>O2'</sup> (2.92)   |
| II. RPS3 protein       | 1      | Baicalein   | 5281605    | -7.54                       | Arg54 <sup>NH2</sup> (3.29), Arg94 <sup>NH1</sup> (3.22), Ala100 <sup>OB</sup> (2.95), Gly33 <sup>NP</sup> (3.60), Ala30 <sup>OB</sup> (3.55)   |
|                        | 2      | Estetrol  | 27125      | -7.33                       | Arg94 <sup>NH1</sup> (2.68), Gln101 <sup>OE1</sup> (3.04), Ala100 <sup>OB</sup> (2.70), Ala100 <sup>OB</sup> (3.31)   |
|                        | 3      | 3-[(3-methyl-1,2,4-oxadiazol-5-yl)methyl]carbamoyl]benzoic acid | 61865966   | -7.24                       | Ser104 <sup>OG</sup> (2.88), Gln101 <sup>OE1</sup> (3.59), Gln101 <sup>NE2</sup> (2.87), Arg54 <sup>NH2</sup> (3.04), Glu103 <sup>OB</sup> (3.05), Ser104 <sup>OG</sup> (3.27), Arg94 <sup>NH1</sup> (3.00)   |

significant pathways and GO terms that could provide potential targets for drug development. However, this study also has some limitations that need to be addressed. This study relied solely on *in silico* analysis and did not validate the identified DEGs or hub genes

through experimental studies. Additionally, this study used only one mRNA expression profile dataset, which may not be sufficient to capture the diversity of gene expression patterns in COVID-19 patients. This study also did not investigate the functional roles





of the identified hub gene containing RPS3, which limits the understanding of its potential importance in COVID-19 pathogenesis.

## 5 Conclusion

In conclusion, the current study provides a comprehensive and systematic approach to identifying potential therapeutic targets for COVID-19 by exploring the structural similarity of human host proteins to SARS-CoV-2 proteins and investigating their interactions and pathways. This study identified 19,051 unique interactors for SARS-CoV-2 proteins, 1,120 differentially expressed genes, and a significant hub gene containing RPS3. The screening of the PubChem library identified three compounds, baicalein, vadadustat, and estetrol, with the best binding energies and residual interaction with the target molecule. Overall, current findings suggest that exploring the human host proteins' interactions and pathways can provide valuable insights for developing effective treatments for COVID-19.

## Data availability statement

The dataset used in our study is publicly available and can be downloaded from the National Center for Biotechnology Information Gene Expression Omnibus under accession number GSE164805 (<https://www.ncbi.nlm.nih.gov/geo/query/acc.cgi?acc=GSE164805>).

## Author contributions

AT: conceptualization, data curation, formal analysis, methodology, software, validation, visualization, writing—original draft, and writing—review and editing. AS: writing—original draft. PS: software, validation, and writing—review and editing. HB: validation and writing—review and editing. HA: writing—review and editing. AYA: writing—review and editing. AB: writing—review and editing. MH: writing—review and editing. BA: writing—review and editing. AAE: writing—review and editing. ASA: writing—review and editing. MH: writing—review and editing. KD: writing—review and editing. RD: conceptualization, project administration, supervision, validation, writing—original draft, and writing—review and editing.

## Funding

The authors declare that no financial support was received for the research, authorship, and/or publication of this article.

## Acknowledgments

Authors would like to acknowledge the support of the Deputy for Research and Innovation Ministry of Education, Kingdom of Saudi Arabia for this research through a grant (NU/IFC/02//MRC/-004) under the Institutional Funding Committee at Najran University, Kingdom of Saudi Arabia.



## Conflict of interest

The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

## Publisher's note

All claims expressed in this article are solely those of the authors and do not necessarily represent those of their affiliated

organizations, or those of the publisher, the editors, and the reviewers. Any product that may be evaluated in this article, or claim that may be made by its manufacturer, is not guaranteed or endorsed by the publisher.

## Supplementary material

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fgene.2024.1292280/full#supplementary-material>

## References

- Alanis-Lobato, G., Andrade-Navarro, M. A., and Schaefer, M. H. (2017). HIPPIE v2.0: enhancing meaningfulness and reliability of protein-protein interaction networks. *Nucleic Acids Res.* 45 (D1), D408–D414. doi:10.1093/nar/gkw985
- Ali, M. J., Hanif, M., Haider, M. A., Ahmed, M. U., Sundas, F., Hirani, A., et al. (2020). Treatment options for COVID-19: a review. *Front. Med.* 7, 480. doi:10.3389/fmed.2020.00480
- Baolin, L., and Bo, H. (2007). "HPRD: a high performance rdf database," in *Network and parallel computing*. Editors K. Li, C. Jesshope, H. Jin, and J. L. Gaudiot (Berlin, Heidelberg: Springer Berlin Heidelberg), 4672, 364–374. Lecture Notes in Computer Science. doi:10.1007/978-3-540-74784-0\_37
- Blanco-Melo, D., Nilsson-Payant, B. E., Liu, W. C., Uhl, S., Hoagland, D., Möller, R., et al. (2020). Imbalanced host response to SARS-CoV-2 drives development of COVID-19. *Cell* 181 (5), 1036–1045. doi:10.1016/j.cell.2020.04.026
- Burki, T. K. (2022). The role of antiviral treatment in the COVID-19 pandemic. *Lancet Respir. Med.* 10 (2), e18. doi:10.1016/S2213-2600(22)00011-X
- Colovos, C., and Yeates, T. O. (1993). Verification of protein structures: patterns of nonbonded atomic interactions. *Protein Sci.* 2 (9), 1511–1519. doi:10.1002/pro.5560020916
- Dai, W., Zhang, B., Jiang, X. M., Su, H., Li, J., Zhao, Y., et al. (2020). Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main protease. *Science* 368 (6497), 1331–1335. doi:10.1126/science.abb4489
- Daina, A., Michielin, O., and Zoete, V. (2017). SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* 7 (1), 42717. doi:10.1038/srep42717
- Davis, F. P., Barkan, D. T., Eswar, N., McKerrow, J. H., and Sali, A. (2007). Host-pathogen protein interactions predicted by comparative modeling. *Protein Sci.* 16 (12), 2585–2596. doi:10.1110/pro.073228407
- Dinda, B., Dinda, M., Dinda, S., and De, U. C. (2023). An overview of anti-SARS-CoV-2 and anti-inflammatory potential of baicalin and its metabolite baicalin: insights into molecular mechanisms. *Eur. J. Med. Chem.* 258, 115629. doi:10.1016/j.ejmech.2023.115629
- Dolgin, E. (2021). The race for antiviral drugs to beat COVID — and the next pandemic. *Nature* 592 (7854), 340–343. doi:10.1038/d41586-021-00958-4
- Dong, H. J., Zhang, R., Kuang, Y., and Wang, X. J. (2021). Selective regulation in ribosome biogenesis and protein production for efficient viral translation. *Arch. Microbiol.* 203 (3), 1021–1032. doi:10.1007/s00203-020-02094-5
- Doolittle, J. M., and Gomez, S. M. (2011). Mapping protein interactions between Dengue virus and its human and insect hosts. *PLoS Negl. Trop. Dis.* 5 (2), e954. doi:10.1371/journal.pntd.0000954
- Fan, Y., Li, X., Zhang, L., Wan, S., Zhang, L., and Zhou, F. (2022). SARS-CoV-2 Omicron variant: recent progress and future perspectives. *Signal Transduct. Target Ther.* 7 (1), 141. doi:10.1038/s41392-022-00997-x
- Franzosa, E. A., and Xia, Y. (2011). Structural principles within the human-virus protein-protein interaction network. *Proc. Natl. Acad. Sci.* 108 (26), 10538–10543. doi:10.1073/pnas.1101440108
- Gordon, D. E., Jang, G. M., Bouhaddou, M., Xu, J., Obernier, K., White, K. M., et al. (2020). A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* 583 (7816), 459–468. doi:10.1038/s41586-020-2286-9
- Guex, N., and Peitsch, M. C. (1997). SWISS-MODEL and the Swiss-Pdb Viewer: an environment for comparative protein modeling. *Electrophoresis* 18 (15), 2714–2723. doi:10.1002/elps.1150181505
- Halgren, T. A. (1996). Merck molecular force field. I. Basis, form, scope, parameterization, and performance of MMFF94. *J. Comput. Chem.* 17 (5–6), 490–519. doi:10.1002/(sici)1096-987x(199604)17:5/6<490::aid-jcc1>3.0.co;2-p
- Holm, L. (2022). Dali server: structural unification of protein families. *Nucleic Acids Res.* 50 (W1), W210–W215. doi:10.1093/nar/gkac387
- Hoof, R. W. W., Sander, C., and Vriend, G. (1997). Objectively judging the quality of a protein structure from a Ramachandran plot. *Bioinformatics* 13 (4), 425–430. doi:10.1093/bioinformatics/13.4.425
- Huang, S., Liu, Y., Zhang, Y., Zhang, R., Zhu, C., Fan, L., et al. (2020). Baicalin inhibits SARS-CoV-2/VSV replication with interfering mitochondrial oxidative phosphorylation in a mPTP dependent manner. *Signal Transduct. Target Ther.* 5 (1), 266. doi:10.1038/s41392-020-00353-x
- Jha, P., Singh, P., Arora, S., Sultan, A., Nayek, A., Ponnusamy, K., et al. (2022). Integrative multiomics and *in silico* analysis revealed the role of ARHGEF1 and its screened antagonist in mild and severe COVID-19 patients. *J. Cell Biochem.* 123 (3), 673–690. doi:10.1002/jcb.30213
- Keskin, O., Tuncbag, N., and Gursoy, A. (2016). Predicting protein-protein interactions from the molecular to the proteome level. *Chem. Rev.* 116 (8), 4884–4909. doi:10.1021/acs.chemrev.5b00683
- Kim, S., Chen, J., Cheng, T., Gindulyte, A., He, J., He, S., et al. (2019). PubChem 2019 update: improved access to chemical data. *Nucleic Acids Res.* 47 (D1), D1102–D1109. doi:10.1093/nar/gky1033
- Kuleshov, M. V., Jones, M. R., Rouillard, A. D., Fernandez, N. F., Duan, Q., Wang, Z., et al. (2016). Enrichr: a comprehensive gene set enrichment analysis web server 2016 update. *Nucleic Acids Res.* 44 (W1), W90–W97. doi:10.1093/nar/gkw377
- Laskowski, R. A., MacArthur, M. W., Moss, D. S., and Thornton, J. M. (1993). PROCHECK: a program to check the stereochemical quality of protein structures. *J. Appl. Crystallogr.* 26 (2), 283–291. doi:10.1107/s0021889892009944
- Lee, S. A., hsiung, C. C., Tsai, C. H., Lai, J. M., Wang, F. S., Kao, C. Y., et al. (2008). Ortholog-based protein-protein interaction prediction and its application to interspecies interactions. *BMC Bioinforma.* 9 (S12), S11. doi:10.1186/1471-2105-9-S12-S11
- Licata, L., Briganti, L., Peluso, D., Peretto, L., Iannuccelli, M., Galeota, E., et al. (2012). MINT, the molecular interaction database: 2012 update. *Nucleic Acids Res.* 40 (D1), D857–D861. doi:10.1093/nar/gkr930
- Liu, H., Ye, F., Sun, Q., Liang, H., Li, C., Li, S., et al. (2021b). *Scutellaria baicalensis* extract and baicalin inhibit replication of SARS-CoV-2 and its 3C-like protease *in vitro*. *J. Enzyme Inhib. Med. Chem.* 36 (1), 497–503. doi:10.1080/14756366.2021.1873977
- Liu, X., Huuskonen, S., Laitinen, T., Redchuk, T., Bogacheva, M., Salokas, K., et al. (2021a). SARS-CoV-2–host proteome interactions for antiviral drug discovery. *Mol. Syst. Biol.* 17 (11), e10396. doi:10.15252/msb.202110396
- Mariano, R., and Wuchty, S. (2017). Structure-based prediction of host–pathogen protein interactions. *Curr. Opin. Struct. Biol.* 44, 119–124. doi:10.1016/j.sbi.2017.02.007
- Martín-Sánchez, F. J., Martínez-Sellés, M., Molero García, J. M., Moreno Guillén, S., Rodríguez-Artalejo, F., Ruiz-Galiana, J., et al. (2023). Insights for COVID-19 in 2023. *Rev. Esp. Quimioter.* 36 (2), 114–124. doi:10.37201/req/122.2022
- Morris, G. M., Huey, R., Lindstrom, W., Sanner, M. F., Belew, R. K., Goodsell, D. S., et al. (2009). AutoDock4 and AutoDockTools4: automated docking with selective receptor flexibility. *J. Comput. Chem.* 30 (16), 2785–2791. doi:10.1002/jcc.21256
- Morse, J. S., Lalonde, T., Xu, S., and Liu, W. R. (2020). Learning from the past: possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *ChemBioChem* 21 (5), 730–738. doi:10.1002/cbic.202000047
- O Boyle, N., Banck, M., James, C., Morley, C., Vandermeersch, T., and Hutchison, G. (2011). Open Babel: an open chemical toolbox. *J. Cheminformatics* 3, 33. doi:10.1186/1758-2946-3-33
- Oughtred, R., Stark, C., Breitkreutz, B. J., Rust, J., Boucher, L., Chang, C., et al. (2019). The BioGRID interaction database: 2019 update. *Nucleic Acids Res.* 47 (D1), D529–D541. doi:10.1093/nar/gky1079
- Prasad, K., AlOmar, S. Y., Alqahtani, S. A. M., Malik, M. Z., and Kumar, V. (2021). Brain disease network analysis to elucidate the neurological manifestations of COVID-19. *Mol. Neurobiol.* 58 (5), 1875–1893. doi:10.1007/s12035-020-02266-w

- Pundir, S., Martin, M. J., O Donovan, C., and UniProt Consortium (2016). UniProt tools. *Curr. Protoc. Bioinforma.* 53 (1), 1.29.1–1.29.15. doi:10.1002/0471250953.bi0129s53
- Rajasekharan, S., and Gupta, S. (2016). “Bioinformatics based approaches to study virus–host interactions during chikungunya virus infection,” in *Chikungunya virus*. Editors J. J. H. Chu and S. K. Ang (New York, NY: Springer New York), 1426, 195–200. *Methods in Molecular Biology*. doi:10.1007/978-1-4939-3618-2\_17
- Reyaz, S., Tasneem, A., Rai, G. P., and Bairagya, H. R. (2021). Investigation of structural analogs of hydroxychloroquine for SARS-CoV-2 main protease (Mpro): a computational drug discovery study. *J. Mol. Graph Model* 109, 108021. doi:10.1016/j.jmgm.2021.108021
- Ritchie, M. E., Phipson, B., Wu, D., Hu, Y., Law, C. W., Shi, W., et al. (2015). Limma powers differential expression analyses for RNA-sequencing and microarray studies. *Nucleic Acids Res.* 43 (7), e47. doi:10.1093/nar/gkv007
- Rose, P. W., Plić, A., Altunkaya, A., Bi, C., Bradley, A. R., Christie, C. H., et al. (2017). The RCSB protein data bank: integrative view of protein, gene and 3D structural information. *Nucleic Acids Res.* 45 (D1), D271–D281. doi:10.1093/nar/gkw1000
- Sagar, S. K., Kumar, M., Singh, P., Sankhwar, S., and Dohare, R. (2021). Prediction of putative protein interactions between zika virus and its hosts using computational techniques. *J. Commun. Dis.* 53 (04), 84–96. doi:10.24321/0019.5138.202178
- Sander, T. (2001). *OSIRIS property explorer*. Org Chem Portal. Available at: <https://www.organic-chemistry.org/prog/peo/>.
- Shannon, P., Markiel, A., Ozier, O., Baliga, N. S., Wang, J. T., Ramage, D., et al. (2003). Cytoscape: a software environment for integrated models of biomolecular interaction networks. *Genome Res.* 13 (11), 2498–2504. doi:10.1101/gr.1239303
- Singh, U. C., and Kollman, P. A. (1984). An approach to computing electrostatic charges for molecules. *J. Comput. Chem.* 5 (2), 129–145. doi:10.1002/jcc.540050204
- Song, J., Zhang, L., Xu, Y., Yang, D., Zhang, L., Yang, S., et al. (2021). The comprehensive study on the therapeutic effects of baicalin for the treatment of COVID-19 *in vivo* and *in vitro*. *Biochem. Pharmacol.* 183, 114302. doi:10.1016/j.bcp.2020.114302
- Su, H., Yao, S., Zhao, W., Li, M., Liu, J., Shang, W. J., et al. (2020). Anti-SARS-CoV-2 activities *in vitro* of Shuanghuanglian preparations and bioactive ingredients. *Acta Pharmacol. Sin.* 41 (9), 1167–1177. doi:10.1038/s41401-020-0483-6
- Sultan, A., Ali, R., Ishrat, R., and Ali, S. (2022). Anti-HIV and anti-HCV small molecule protease inhibitors *in-silico* repurposing against SARS-CoV-2 M<sup>pro</sup> for the treatment of COVID-19. *J. Biomol. Struct. Dyn.* 40 (23), 12848–12862. doi:10.1080/07391102.2021.1979097
- Sultan, A., Ali, R., Sultan, T., Ali, S., Khan, N. J., and Parganiha, A. (2021). Circadian clock modulating small molecules repurposing as inhibitors of SARS-CoV-2 M<sup>pro</sup> for pharmacological interventions in COVID-19 pandemic. *Chronobiol Int.* 38 (7), 971–985. doi:10.1080/07420528.2021.1903027
- Szklarczyk, D., Gable, A. L., Nastou, K. C., Lyon, D., Kirsch, R., Pyysalo, S., et al. (2021). The STRING database in 2021: customizable protein–protein networks, and functional characterization of user-uploaded gene/measurement sets. *Nucleic Acids Res.* 49 (D1), D605–D612. doi:10.1093/nar/gkaa1074
- Tripathi, D., Sodani, M., Gupta, P. K., and Kulkarni, S. (2021). Host directed therapies: COVID-19 and beyond. *Curr. Res. Pharmacol. Drug Discov.* 2, 100058. doi:10.1016/j.crphar.2021.100058
- Trott, O., and Olson, A. J. (2009). AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* 31, 455–461. doi:10.1002/jcc.21334
- Wiederstein, M., and Sippl, M. J. (2007). ProSA-web: interactive web service for the recognition of errors in three-dimensional structures of proteins. *Nucleic Acids Res.* 35, 407–410. doi:10.1093/nar/gkm290
- Zheng, W., Zhang, C., Li, Y., Pearce, R., Bell, E. W., and Zhang, Y. (2021). Folding non-homologous proteins by coupling deep-learning contact maps with I-TASSER assembly simulations. *Cell Rep. Methods* 1 (3), 100014. doi:10.1016/j.crmeth.2021.100014