Since January 2020 Elsevier has created a COVID-19 resource centre with free information in English and Mandarin on the novel coronavirus COVID-19. The COVID-19 resource centre is hosted on Elsevier Connect, the company's public news and information website.

Elsevier hereby grants permission to make all its COVID-19-related research that is available on the COVID-19 resource centre - including this research content - immediately available in PubMed Central and other publicly funded repositories, such as the WHO COVID database with rights for unrestricted research re-use and analyses in any form or by any means with acknowledgement of the original source. These permissions are granted for free by Elsevier for as long as the COVID-19 resource centre remains active.
Identification of Potential Antiviral Compounds Against SARS-CoV-2 Structural and Non Structural Protein Targets: A Pharmacoinformatics study of the CAS COVID-19 Dataset

Rolando García, Anas Hussain, Prasad Koduru, Murat Atis, Kathleen Wilson, Jason Y. Park, Inimary Toby, Kimberly Diwa, Lavang Vu, Samuel Ho, Fajar Adnan, Ashley Nguyen, Andrew Cox, Timothy Kirtek, Patricia García, Yanhui Li, Heather Jones, Guanglu Shi, Allen Green, David Rosenbaum

PII: S0010-4825(21)00158-X
DOI: https://doi.org/10.1016/j.compbiomed.2021.104364
Reference: CBM 104364

To appear in: Computers in Biology and Medicine

Received Date: 14 December 2020
Revised Date: 26 March 2021
Accepted Date: 29 March 2021

Please cite this article as: R. García, A. Hussain, P. Koduru, M. Atis, K. Wilson, J.Y. Park, I. Toby, K. Diwa, L. Vu, S. Ho, F. Adnan, A. Nguyen, A. Cox, T. Kirtek, P. García, Y. Li, H. Jones, G. Shi, A. Green, D. Rosenbaum, Identification of Potential Antiviral Compounds Against SARS-CoV-2 Structural and Non Structural Protein Targets: A Pharmacoinformatics study of the CAS COVID-19 Dataset, Computers in Biology and Medicine, https://doi.org/10.1016/j.compbiomed.2021.104364.

This is a PDF file of an article that has undergone enhancements after acceptance, such as the addition of a cover page and metadata, and formatting for readability, but it is not yet the definitive version of record. This version will undergo additional copyediting, typesetting and review before it is published in its final form, but we are providing this version to give early visibility of the article. Please note that, during the production process, errors may be discovered which could affect the content, and all legal disclaimers that apply to the journal pertain.

© 2021 The Author(s). Published by Elsevier Ltd.
Author Statement:

Author Contributions: Conceptualization, R.G; methodology, R.G, A.H, I.T, K.D; validation, R.G, A.H; formal analysis, R.G, A.H, MA, IT, KD, Y.L, T.K, L.V, SH, FA, AN, AC, HJ, PG, G.S; investigation, R.G, A.H; resources, R.G.; data curation, R.G and A.H, IT; writing—original draft preparation, R.G., A.H, A.G, D.R; writing—review and editing, R.G PK, KW, JYP; visualization, R.G.; supervision, R.G; project administration, R.G. All authors have read and agreed to the published version of the manuscript.

Funding: This research received no external funding

Conflicts of Interest: The authors declare no conflict of interest.
Identification of Potential Antiviral Compounds Against SARS-CoV-2 Structural and Non Structural Protein Targets: A Pharmacoinformatics study of the CAS COVID-19 Dataset.

Rolando García†, Anas Hussain7, Prasad Koduru1, Murat Atis2, Kathleen Wilson1, Jason Y. Park1,5,6, Inimary Toby4, Kimberly Diwa9, Lavang Vu1, Samuel Ho10, Fajar Adnan9, Ashley Nguyen4, Andrew Cox2, Timothy Kirtke1, Patricia Garcia8, Yanhui Li3, Heather Jones1, Guanglu Shi1, Allen Green1 & David Rosenbaum1

1. Department of Pathology, University of Texas Southwestern Medical Center, Dallas, TX
2. Lyda Hill Department of Bioinformatics, University of Texas Southwestern Medical Center, Dallas, TX
3. Department of Cell Biology, University of Texas Southwestern Medical Center, Dallas, TX
4. Department of Biology, University of Dallas, Dallas TX
5. Department of Pathology, Children’s Medical Center, Dallas, TX
6. Eugene McDermott Center for Human Growth and Development, University of Texas Southwestern Medical Center, Dallas, TX
7. Deccan College of Medical Sciences, Hyderabad, India
8. NCH Corporation, Irving, TX
9. Department of Computer Science, University of Dallas, Dallas TX
10. Harvard University, Department of Biology, Cambridge, MA

†Send Correspondence to:
Rolando García, PhD., FACMG.
Department of Pathology
Bio-center at UT Southwestern Medical Center
2330 Inwood Rd Suite
EB3.202-B
Dallas, TX 75235
Email: rolando.garcia@utsouthwestern.edu

Disclosures: The authors have no financial relationships relevant to this article to disclose.
Funding Sources: None

Key Words: SARS-CoV-2, COVID-19, Molecular Docking, Molecular Dynamics

Abstract: SARS-CoV-2 is a newly discovered virus which causes COVID-19 (coronavirus disease of 2019), initially documented as a human pathogen in 2019 in the city of Wuhan China, has now quickly spread across the globe with an urgency to develop effective treatments for the virus and emerging variants. Therefore, to identify potential therapeutics, an antiviral catalogue of compounds from the CAS registry, a division of the American Chemical Society was evaluated using a pharmacoinformatics approach. A total of 49,431 compounds were initially recovered. After a biological and chemical curation, only 23,575 remained. A machine learning approach was then used to identify potential compounds as inhibitors of SARS-CoV-2 based on a training dataset of molecular descriptors and fingerprints of known reported compounds to have favorable interactions with SARS-CoV-2. This approach identified 178 compounds, however, a molecular docking analysis revealed only 39 compounds with strong binding to
active sites. Downstream molecular analysis of four of these compounds revealed various non-covalent interactions along with simultaneous modulation between ligand and protein active site pockets. The pharmacological profiles of these compounds showed potential drug-likeness properties. Our work provides a list of candidate anti-viral compounds that may be used as a guide for further investigation and therapeutic development against SARS-CoV-2.

Introduction

The new SARS-CoV-2 coronavirus, responsible for causing COVID-19, was initially documented as a human pathogen in December 2019 in the city of Wuhan, Hubei province in China [1]. The virus has quickly spread across the globe, and as of December 2020, there were 119,988,220 cases reported with 2,655,612 fatalities (John Hopkins Coronavirus Resource Center 3/14/2021). Infection by the SARS-CoV-2 virus, a single-stranded RNA virus, results in a wide spectrum of illnesses from an asymptomatic carrier state to mild and severe cold-like symptoms to a fatal pneumonia. Multiple vaccines against the SARS-CoV-2 virus are available in several countries, including three in the United States [2-3]. However, concerns related to the timeline of widespread and global vaccination as well as questions about continued vaccine efficacy against newly emerging SARS-CoV-2 variants (e.g. UK and South African) continue to highlight need for development of COVID-19 treatments in parallel to vaccination efforts [4-5]. SARS-CoV-2 belongs to the beta coronavirus genus, which also includes severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East respiratory syndrome coronavirus (MERS-CoV). Rapid genomic sequencing of SARS-CoV-2 has enabled comparative analysis between the novel virus and those responsible for previous pandemics [6]. Due to significant homology between the viruses, previously curated knowledge generated through studies with SARS-CoV and MERS-CoV can be used in an attempt to find potential drug targets for SARS-CoV-2 [7]. A tremendous amount of effort has been placed in finding therapeutics for the various coronaviruses. Since the original SARS-CoV emerged in 2002, an effort has been made to target various viral structures and proteins including helicase, protease, endonuclease, exonuclease, methyltransferase, and non-structural proteins (NSPs). Researchers have continued to use traditional methods to determine antiviral activity of compounds, but these processes can be slow and cumbersome. For these reasons, many researchers have now turned to virtual screening using genomic and structural models. Past efforts have shown that using molecular docking studies as an initial step is useful for screening the most promising antiviral, antibacterial, and antiprotozoal compounds [8-9]. In April 2020, CAS, a division of the American Chemistry Society, released a database containing 49,431 chemical substances assembled from the CAS REGISTRY that have antiviral activity reported in published literature or are structurally similar to known antivirals. In an attempt to find potential anti-viral compounds as inhibitors of SARS-CoV-2, a pharmacoinformatics approach including a classifier model coupled with a multi molecular docking and dynamics analysis was performed.

Materials and Methods

To identify potential antiviral compounds as inhibitors of SARS-CoV-2, we obtained the CAS dataset of antiviral chemical compounds available at http: https://www.cas.org/covid-19-antiviral-compounds-dataset. All compounds were converted to Protein Data Bank (PDB) and AutoDock (PDBQT) format for subsequent analysis using the open source Babel package available at http://openbabel.org. The initial data-set of anti-viral compounds in SDF format was
subjected to chemical and biological curation. The Konstanz Information Miner (KNIME) workflow (https://www.knime.org/) was employed to perform these curations. We use the SDF reader node in the KNIME workflow to read chemical and biological properties of antiviral compounds. For chemical curation, modules in the KNIME workflow included the following for inorganic and organo-metallic removal: SDF reader used to read the input file, element filter (removes inorganic and organo-metallic compounds), connectivity (removes mixtures), RDKit Salt Stripper (removes salts), RDKit Optimize Geometry (geometric optimization of screened compounds), RDKit Structure Normalizer (standardizes compounds), RDKit Add Hs (adding of hydrogen), and the SDF writer (generates an output file of screened compounds in SDF format). To perform biological curation, The Duplicate Analysis Workflow using the 3D D-Similarity module was performed to identify duplicate molecules in the dataset. An activity cliff analysis using the Automated Matched Pairs module computes matched molecular pairs and understands molecular activity. A careful and manual curation of compounds with similar structure and activity values were then removed. The chemical and biological curation is well documented by by Ambure and colleagues [10]. To establish a list of standard or control compounds (i.e. reported potential compounds with favorable interactions against SARS-CoV-2), a search of the literature was performed. These latter set of compounds served to classify prospective inhibitors of SARS-CoV-2 from the CAS dataset. For developing a classification model, a set of molecular descriptors (or indices) for the CAS dataset and controls were calculated using PaDEL, a software package that calculates molecular descriptors and fingerprints available at http://padel.nus.edu.sg/software/padeldescriptor. The generated fingerprints of all compounds from each descriptor were scaled. To obtain the best fit of anti-viral compounds against SARS-CoV-2 from the optimal set of descriptors, a random forest classifier in the R environment [11] was implemented. Control and experimental biomolecules and their accompanying descriptors and fingerprint properties were obtained and divided into testing and training data sets. The training data set consisted of all 32 control molecules and an additional 32 experimental molecules, which were selected from the larger dataset by implementing the Kennard-Stone algorithm with Euclidean distance. The testing data set consisted of 23,520 experimental molecules. A receiver operating curve (ROC) and the area under the curve (AUC) was calculated to determine the reliability of the classifier. Next, four SARS-CoV-2 structural and non-structural protein targets, including NSP12 (RNA polymerase, ID: 6NUR), NSP13 (helicase, ID: 6ZSL), main protease (Mpro, ID:7BQY), spike protein (spike binding region – SBD, ID: 6LZG) from the protein databank (https://www.rcsb.org) were evaluated against potential anti-viral compounds. The angiotensin converting enzyme-2 (ACE-2) protein on the host, essential for attachment to the SBD region that leads to SARS-CoV-2 entry into the host was also evaluated (ID: 7DF4). Moreover, the emerging more transmissible South African SARS-CoV-2 variants with the following amino acid residue substitutions in the SBD region of the spike protein: LYS417ASN, GLU484LYS and ASN501TYR) were also evaluated. Target proteins were refined using AutoDock Tools by deleting water molecules, adding polar hydrogens, Kollman charges, computing the Gasteiger charges and assigning the ADT4 type atoms. The active sites of the target proteins were highlighted by using the grid box of AutoDock. Thereafter, a multi molecular docking analysis between anti-viral compounds and target proteins using AutoDock Vina [12] was subsequently performed. Only those compounds with an affinity to the target protein of less than -7.5 kcal/mol were retained for downstream analysis (binding affinities of less than -6 kcal/mol are generally considered significant values for binding). Thereafter, each of these compounds were manually curated to determine the area of the active site they covered by
using Pymol (http://pymol.org). Those ligands that covered a significant area in the active site were subsequently prepared by the LigPrep and Epik modules (Schrödinger Release 2020-4: Schrödinger, LLC, New York, NY) for docking. The AutoDock and MGL Tools were then used to dock these ligands against SARS-CoV-2 active site target structure and non-structure proteins. Subsequently, the Protein-Ligand Interaction Profiler available at https://projects.biotec.tudresden.de/plip-web/plip was used to identify the molecular noncovalent interactions of proteins and their ligands. Remdesivir obtained from the PubChem database (pubchem.ncbi.nlm.nih.gov), officially approved as an antiviral to treat SARS CoV-2 patients, was used as a control for docking and molecular interactions studies. To validate the molecular interactions, a molecular dynamic (MD) simulation was then performed. The ligand structures were parametrized, and input files were generated by using CHARM-GUI [13-15] on the docking structures for compounds 65 and remdesivir. The parametrized structures were solvated in a periodic box of TIP3 molecules. The sizes of TIP3 boxes were 90 x 90 x 90 Å³ and 85 x 85 x 85 Å³ for remdesivir-protein and compound 65-protein structures, respectively. Sodium and chloride ions were added to neutralized and obtain a salt concentration of 0.15 mol/L. CHARMM36 force field was used for protein [16]. The simulations were conducted with the NAMD 2.12 package [17]. After 50,000 minimization MD steps with Conjugate gradient algorithm at 0 K, subsequent 600,000 steps equilibration simulation with constrained protein and ligand was applied for both structure at 303.15 K. At the final production MD simulation, NPT ensemble was applied by using Nose-Hoover Langevin piston pressure control [18] at 303.15 K and 1.01325 bars for 20 ns simulation time, 2 fs integration step is applied through all simulation steps. To evaluate the interaction energies and RMSDs, VMD 1.9 software [19] was used. Thereafter, simultaneous modulation of SARS-CoV-2 target proteins was assessed using the virus-associated disease-specific chemo-genomics knowledge (Virus-CKB) available at https://www.cbligand.org/g/virus-ckb. Lastly, pharmacological characteristics, including absorption, distribution, metabolism, excretion, toxicity, drug likeness properties of potential inhibitors of SARS-CoV-2 were explored using the ADMET platform (http://admet.scbdd.com/home/index/). The general workflow illustrating steps taken in screening of the CAS dataset in search of potential inhibitor compounds of SARS-CoV-2 target proteins is shown in Figure 1.
Figure 1. Process for identifying SARS-CoV-2 inhibitors from CAS dataset.

**Results**

A total of 49,431 compounds were initially obtained from the CAS registry. After the KNIME chemical and biological curation, 23,575 compounds remained. To identify potential compounds as possible inhibitors of SARS-CoV-2 and to establish a set of standards or controls to compare against the CAS dataset, a literature search was performed to identify anti-viral compounds with potential activity against COVID-19. A list of 113 antiviral compounds was selected (see Supplemental Table 1). These compounds served as controls for subsequent classification of the CAS dataset. To construct the classification model, molecular descriptors (topo-chemical atom indices) were generated for each of the compounds (CAS dataset + controls) by using the PaDEL-molecular descriptor software. A total of 25 molecular descriptors (see Supplemental Table 2) and their molecular fingerprints for each of the compounds was calculated and their values subsequently scaled (see Figure 2).
Figure 2. Snapshot and representative illustration of descriptor calculations. First column describes all compounds (CAS antiviral compounds + controls), while remaining columns show seven of the 46 molecular descriptors with calculated fingerprints.

Thereafter, a matrix composed of 25 descriptors, 23,575 CAS compounds and 113 controls was constructed to perform a random forest tree classifier in the R environment. Figure 3 shows the random forest tree predicted plot of molecular descriptors and their importance in the classification of antiviral compounds against SARS-CoV-2, while Figure 4 shows a 0.92 reliability (area under the curve – AUC of the ROC curve) at predicting compounds as inhibitors of SARS-CoV-2.

Figure 3. Random Forest tree predicted plot of molecular descriptors. The plot shows the measure of relative unsaturation content relative to molecular size (ETA_dBetaP) as the most important classifier of potential antiviral compounds.

Figure 4. Receiver operating curve for classifier model developed to identify compounds as potential inhibitors of SARS-CoV-2 target proteins. The area under the curve (AUC) of the classifier showed a good predictive ability of 0.92 (92%).
Based on these findings, the classifier identified 178 antiviral compounds as potential inhibitors of SARS-CoV-2 (Figure 5).

![Classifier Results]

CAS dataset of antiviral compounds → KNIME chemical-biological curation → 23,575 CAS Compounds → Control compound identification → 113 Controls Identified in the literature → Descriptor Identification → 25 Descriptors Calculated for 23,575 Compounds And 113 controls

Random Forest Classifier → 178 CAS Compounds → Descriptor Results Scaled

Figure 5. Summary of results from the classifier model workflow.

To validate the compounds obtained from the classifier, a multi molecular docking analysis using AutoDock Vina was performed on 178 compounds against SARS-CoV-2 target proteins: SBD, NSP12, NSP13 (helicase) and Main protease (Mpro). Table 1 shows the protein active site residues that were targeted for multi docking.

| Target Protein | Active Site Amino Acid Residue | PMID       |
|----------------|--------------------------------|------------|
| SBD            | LEU455, ALA475, PHE486, GLN493, ARG403, GLY496, ASN501, PHE456, TYR489 | 32568012, 33190011, 33184600, 33657325, 33299995, 32623480 |
| NSP12          | ASP618, LEU758, SER759, ASP760, LEU758, SER759, ASP760, ASP761, LYS798, LIS813, SER814, LIS813, SER814 | 32283108, 32812340, 32438371, 32346490 |
| NSP13          | LYS288, SER289, ARG567, GLN404 | 32346490 |
Ligands with binding energy less than -6 kcal/mol to SARS-CoV-2 target protein active sites were then selected. A total of 39 compounds plus four duplicates (40, 137, 177, 148) were identified (Table 2).

Table 2. Scores generated by multi molecular AutoDock Vina of the top binding compounds to the active site of SARS-CoV-2 targets.

| ID | CAS registry Number | Compound Number | Binding Energy (kcal/mol) |
|----|---------------------|-----------------|--------------------------|
|    |                     |                 | Vina                     |
| 1  | 1809249-37-3        | Control         | -6.4                     |
| 2  | 1002336-72-2        | 65              | -7.1                     |
| 3  | 687967-00-6         | 137             | -12.8                    |
| 4  | 1052524-91-0        | 11              | -12                      |
| 5  | 911195-33-0         | 169             | -12                      |
| 6  | 957794-14-8         | 177             | -11.6                    |
| 7  | 774511-55-6         | 148             | -10.9                    |
| 8  | 909082-43-5         | 166             | -10.9                    |
| 9  | 134637-14-2         | 34              | -9.3                     |
| 10 | 190141-36-7         | 76              | -9.2                     |
| 11 | 918934-15-3         | 170             | -9.1                     |
| 12 | 1399805-48-1        | 40              | -8.7                     |
| 13 | 1637769-10-8        | 61              | -8.7                     |
| 14 | 687967-00-6         | 137             | -14.2                    |
From these group of compounds, a list of 4 (ID: 40, 65, 70 and 137) were selected based on visual inspection of the active site and the area ligands covered. Selected compounds underwent subsequent analysis, including docking, molecular interactions and description of pharmacological properties. Figure 6 depicts representative chemical formulas of the four potential antiviral compounds to the various active sites of the SARS-CoV-2 proteins, including the control remdesivir, whereas Table 3 describes the SMILE format and molecular weights of the four compounds. The 3D structural surface representation of SARS-CoV-2 target proteins and molecular docking of the potential candidate antiviral compound to each of the targets (including the SBD South African SARS-CoV-2 variant) is illustrated in Figure 7.
Control: Remdesivir (CAS Number 1809249-37-3)

Candidate Inhibitor of SIRD (Compound 65: CAS Number 1002336-72-2)

Candidate inhibitor of NSP12 (Compound 137: CAS Number 687967-00-6)

Candidate inhibitor of NSP13 (Compound 40: CAS Number: 1399805-48-1)

Candidate inhibitor of Mpro (Compound 70: CAS Number 1817756-68-5)
Figure 6. Chemical Structure of antiviral compounds from the CAS registry as potential inhibitors of SARS-CoV-2 target protein active sites.

Table 3. SMILES and molecular weight of candidate compounds with inhibitory potential for SARS-CoV-2.

| Ligand | SMILE Format | Molecular Weight |
|--------|--------------|------------------|
| 65     | C(OC(=O)CCC(=O)N[C@@H]1[C@@H](OCc2cccc2)[C@@H] (OCc2cccc2)[C@@H](OC1OCc1cccc1)OCc1cccc1)[C@@H] 1[C@@H]2[C@@H][(C@@H)](O1)n1nc(n1c)C(=O)N)OC(O2)OC | 907.96 |
| 137    | N(=Nc1c(C)c(=N=n2c2cc2)N=N/ccc3c4c(c(c3c(S(=O)(=O)[O-])cccc4)c(cc2N)Ncc(c(c1)c1c2c(c(c1)N=Nc1cc2(=O)(=O)[O-])cccc4)c(cc1N)nc2) | 1063.11 |
| 40     | [C@]1230[C@]1(OC1=COPO[C@@]21[C@H]3C#C)n1c2c(nc1) c(=O)[nH]c(N)n2 | 425.33 |
| 70     | C(=C1/C=C(C(=O)O)C(=O)C=C1)(c1cc(Cc2cc(C(=O)O)c (cc3)O)c3cc(C(=O)O)c3cc(Oc(=O)c(cc2O)c(=O)O)c(=O)O)c1 cc(C(=O)O)c(c1O)O | 858.71 |
Figure 7. 3-D surface representation of SARS CoV-2 target proteins and potential inhibitory compounds. Key: blue, SARS-CoV-2 target protein; yellow, target active site; magenta, antiviral compound. Panels (A): The control compound (remdesivir) and its binding relationship to the SBD region. An active site residue in the SBD region GLN493 is labeled in white for reference; (B): ligand 65 predicted binding to the SBD region; (C) ligand 65 binding to the SBD – South African variant with the N501Y (asparagine to tyrosine amino acid residue 501) substitution depicted; (D) ligand 137 bound to NSP12; (E) ligand 40 binding interaction with NSP13; (F) ligand 70 targeting Mpro and panel (G) ACE-2 predicted interaction with ligand 40. Of note, for the South African variant, the ASN417 and LYS484 substitutions are not covered by ligand 65. For the Mpro protein only the dyad active site of HIS45 and CYS145 are shown in yellow.

Covalent interactions of the four selected compounds against the active sites of SARS-CoV-2 were then determined using the Biotec molecular profiler, while representative molecular
interactions were rendered using Pymol. All molecular interactions and their distance given in angstroms is shown on Table 4, while Figure 8 illustrates representative molecular interactions between SARS-CoV-2 targets and selected compounds.

Table 4. Molecular interactions between selected compounds and SARS-CoV-2 active site target proteins.

| Target  | Compound | Interaction | Amino acid Residue | Distance (Angstroms) |
|---------|----------|-------------|-------------------|---------------------|
| SBD     | Standard | Hydrogen    | ARG403            | 3.00                |
|         |          |             | GLN493            | 2.70                |
|         |          |             | GLN493            | 2.09                |
|         |          |             | GLY496            | 3.43                |
|         |          |             | PHE497            | 3.06                |
|         |          |             | GLN498            | 3.03                |
|         |          |             | ASN501            | 2.02                |
|         |          |             | ASN501            | 3.91                |
|         |          | Hydrophobic | GLU406            | 3.92                |
|         |          |             | LYS417            | 3.30                |
|         |          |             | TYR453            | 3.45                |
|         |          |             | TYR495            | 3.41                |
| SBD     | 65       | Hydrogen    | ARG403            | 2.10                |
|         |          |             | ARG403            | 2.93                |
|         |          |             | GLN493            | 3.01                |
|         |          |             | GLN493            | 3.16                |
|         |          |             | GLY496            | 2.03                |
|         |          |             | PHE497            | 3.20                |
|         |          |             | GLN498            | 3.49                |
|         |          |             | ASN501            | 2.27                |
|         |          |             | TYR505            | 1.70                |
|         |          | Hydrophobic | TYR489            | 3.59                |
|         |          |             | PHE490            | 3.72                |
| SBD_SA  | 65       | Hydrogen    | ARG403            | 3.22                |
|         |          |             | TYR453            | 2.38                |
|         |          |             | GLY496            | 2.23                |
|         |          |             | TYR505            | 2.72                |
|         |          |             | TYR505            | 2.85                |
|         |          | Hydrophobic | ILE418            | 3.93                |
| Structure | Residue | Interaction Type | Distance (Å) |
|-----------|---------|-----------------|--------------|
| NSP12     |         | Hydrogen bond   | LYS621 3.67  |
|           |         |                 | CYS622 2.51  |
|           |         |                 | LYS798 2.97  |
|           |         | Hydrophobic     | ASP618 3.53  |
|           |         |                 | LYS798 3.59  |
|           |         |                 | TRP800 3.17  |
|           |         | Salt bridges    | ASP618 5.00  |
| NSP13     |         | Hydrogen bond   | GLY285 2.25  |
|           |         |                 | GLY287 3.17  |
|           |         |                 | LYS288 2.45  |
|           |         |                 | SER289 1.92  |
|           |         |                 | ASP374 3.46  |
|           |         |                 | GLU375 3.32  |
|           |         | Hydrophobic     | GLU375 3.95  |
|           |         | Salt bridges    | LYS288 3.97  |
|           |         |                 | ARG567 4.52  |
| Mpro      |         | Hydrogen bond   | GLU166 2.66  |
|           |         |                 | GLU166 3.11  |
|           |         | Hydrophobic     | HIS41 3.98   |
|           |         |                 | MET165 3.17  |
|           |         |                 | PRO168 3.33  |
|           |         |                 | GLU166 3.61  |
|           |         |                 | ASP187 3.70  |
|           |         | Salt bridges    | HIS41 5.43   |
|           |         |                 | HIS41 5.44   |
|           |         |                 | HIS41 5.35   |
| Interaction Type | Residue | Distance |
|-----------------|---------|----------|
| ACE-2 Hydrogen   | GLU35   | 3.60     |
| Hydrophobic     | HIS34   | 3.97     |
|                 | GLU37   | 3.70     |
| Cation interaction | GLU35 | 3.79     |
|                 | ASP38   | 4.42     |
|                 | LYS353  | 3.34     |
|                 | LYS353  | 4.60     |

Key: SBD, spike binding domain; SBD_SA, spike binding domain South African variant.
Figure 8. Molecular interactions between active site amino acid residues and anti-viral compounds. (A) SBD and control ligand remdesivir, (B) SBD with compound 65, (C) SBD South African variant interactions with ligand 65 shows hydrophobic interaction between TYR501 and ligand depicted by gray dashed lines, (D) NSP12 interaction with ligand 137, (E) NSP13 and ligand 40, (F) Mpro and ligand 70, (G) ACE-2 interactions with ligand 40. Molecular interactions represented by solid blue (hydrogen bonding) and dashed lines (gray: hydrophobic; yellow, cation interaction; orange, salt bridge).

To support these interactions, a molecular dynamic simulation was carried out between the control compound (remdesivir) and ligand 65 at the SBD interphase. Both systems were simulated for a fixed period of time (0 – 20 nanoseconds - ns). During this time interval both control and ligand 65 maintain binding affinity to active sites amino acid residues of the SBD protein (see Figure 9). The binding energies of the two ligands to the SBD region at different time intervals are described on Table 5, while contact residues at different time intervals are outlined on Table 6.
Figure 9. Molecular dynamic simulation of the control and ligand 65 against the SBD region. Panels A + B + C show the control remdesivir binding to the SBD region at multiple time interval frames (0; 506, 10.14 ns; and 986, 19.74 ns). Panels D + E + F illustrate ligand 65 binding to the SBD region at interval frames (0; 503, 10.08 ns; and 990, 19.82 ns).

Table 5. Energies of Remdesivir and ligand 65 to the SBD active site regions at different time intervals.

| Remdesivir: | Time (ns) | Elec  | VdW   | Nonbond | Total     |
|-------------|-----------|-------|-------|---------|-----------|
| Frame       |           |       |       |         |           |
| 0           | 0         | -18.804| -36.5253| -55.3293| -55.3293  |
| 506         | 10.14     | -15.3262| -29.2573| -44.5835| -44.5835  |
| 986         | 19.74     | -15.4416| -30.4299| -45.8715| -45.8715  |

=================================================================================================
### Ligand 65

| Frame | Time (ns) | Elec       | VdW        | Nonbond    | Total     |
|-------|-----------|------------|------------|------------|-----------|
| 0     | 0         | +4.4186    | -30.7731   | -26.3545   | -26.3545  |
| 503   | 10.08     | -9.6287    | -30.4741   | -40.1028   | -40.1028  |
| 990   | 19.82     | -15.0934   | -31.3019   | -46.3953   | -46.3953  |

Key: Energies are in kcal/mol; ns, nanoseconds; Elec, electrostatic; VdW, Vanderwall; Nonbond, nonbonded.

From Table 5, the total energy suggests remdesivir with an initial stronger binding energy to the SBD region, however, ligand 65 showed a stronger binding energy at the end of the 20 nanosecond simulation.

Table 6. Contact amino acid residues between ligands and the SBD region at different time (nanoseconds).

| Ligand      | Time Intervals          |
|-------------|-------------------------|
| **Remdesivir** |                     |
| Time = 0    | Time = 10.14           |
| Frame 0     | Frame 506              |
|             | Time = 19.74           |
| Frame 986   |                       |
| **ARG-403** | **ARG-403**            |
| ASP-405     | GLY-447                |
| GLU-406     | TYR-449                |
| LYS-417     | TYR-453                |
| ILE-418     | SER-494                |
| ASN-422     | TYR-495                |
| TYR-453     | **GLY-496**            |
| ARG-454     | PHE-497                |
| LEU-455     | GLN-498                |
| GLN-493     | THR-500                |
| SER-494     | **ASN-501**            |
| TYR-495     | GLY-502                |
| **GLY-496** | GLY-504                |
| PHE-497     | TYR-505                |
| GLN-498     | GLN-506                |
| **ASN-501** |                       |
| **TYR-505** |                       |
| **Ligand 65** |                     |
| Time = 0    | Time = 10.08           |
| Frame 0     | Frame 503              |
|             | Time = 19.82           |
| Frame 990   |                       |
| TYR-449     | TYR-351                |
|            | TYR-449                |
Ligand binding to active site amino acid residues in the SBD region at different time intervals are bold in black.

The selected list of ligands (40, 65, 70 and 137) were subsequently evaluated with the knowledge base Virus-CKB to determine whether these compounds showed simultaneous modulation of viral pathways in the life cycle of SARS-CoV-2. The analysis predicted all ligands with combined modulation of viral protein targets (e.g. ligand 40 to ACE-2: -9.25 kcal/mol; ligand 65 to ACE-2: -11.26 kcal/mol; ligand 70 to ACE-2: -10.15 kcal/mol, methyl transferase: -8.6 kcal/mol, Mpro: -8.77 kcal/mol, and NSP12: -8.77 kcal/mol; ligand 137 to ACE-2: -11 kcal/mol, Mpro: -9.21 kcal/mol, NSP12: -9.72 kcal/mol, methyl transferase: -10.77 kcal/mol and papain like protease: -8.68 kcal/mol). A spider plot outlining the various viral interactions pathways between selected ligands and targets across viral networks is shown in Figure 9.
Figure 9. Spider plot of selected compounds (40, 65, 70, 137) against viral targets using the knowledgebase Virus-CKB. The blue circle with yellow highlights represents the predicted targets (Spike, spike binding protein; ACE2, angiotensin converting enzyme-2; MTA SARS2: methyl transferase; PR, Mpro and RDRP, NSP12 – RNA dependent RNA polymerase; PLP, papain-like protease) of selected compounds and dashed lines represent their interactions.

The generated interactions by the spider plot were then evaluated to determine whether ligands bind to SARS CoV-2 target active site amino acid residues. This evaluation revealed only ligand 137 and 40 with a potential drug combination effect to SARS-CoV-2 active site target proteins. Table 7 shows the newly identified interactions. Ligand 137 was also evaluated for the SBD South African variant.
Table 7. Assessment of ligands with potential drug combination effect against SARS-CoV-2 active site amino acid residues.

| Target | Compound | Interaction | Amino acid Residue | Distance (Angstroms) | Binding Energy Auto-Dock Kcal/mol |
|--------|----------|-------------|--------------------|----------------------|----------------------------------|
| SBD    | 137      | Hydrogen    | ARG403             | 2.56                 | -8.6                             |
|        |          |             | TYR421             | 1.73                 |                                  |
|        |          |             | TYR453             | 2.45                 |                                  |
|        |          |             | TYR453             | 3.20                 |                                  |
|        |          |             | ALA475             | 3.39                 |                                  |
|        |          |             | TYR489             | 2.18                 |                                  |
|        |          |             | GLY496             | 2.90                 |                                  |
|        |          |             | GLY496             | 3.13                 |                                  |
|        |          |             | GLN498             | 3.05                 |                                  |
|        |          |             | ASN501             | 2.93                 |                                  |
|        |          | Hydroporphic| LEU455             | 3.27                 |                                  |
|        |          |             | PHE456             | 3.41                 |                                  |
|        |          |             | PHE456             | 3.10                 |                                  |
|        |          |             | ALA475             | 3.84                 |                                  |
|        |          |             | TYR489             | 3.74                 |                                  |
|        |          |             | GLN493             | 3.11                 |                                  |
|        |          |             | TYR495             | 3.04                 |                                  |
|        |          |             | GLN498             | 3.82                 |                                  |
|        |          |             | TYR505             | 3.49                 |                                  |
|        |          |             | TYR505             | 3.98                 |                                  |
| SBD - SA|       | Hydrogen    | ASN417             | 1.76                 | -12.41                           |
|        |          |             | LYS484             | 3.54                 |                                  |
|        |          |             | LEU492             | 2.55                 |                                  |
|        |          |             | GLN493             | 2.28                 |                                  |
|        |          |             | SER494             | 3.09                 |                                  |
|        |          | Hydroporphic| 449TYR             | 3.13                 |                                  |
|        |          |             | 455LEU             | 3.06                 |                                  |
|        |          |             | 455LEU             | 3.39                 |                                  |
|        |          |             | 456PHE             | 3.87                 |                                  |
|        |          |             | 493GLN             | 3.88                 |                                  |
|        |          |             | 493GLN             | 3.53                 |                                  |
| Property          | Compound: | 65 | 137 | 40 | 70 |
|-------------------|-----------|----|-----|----|----|
| Solubility (LogP) | 4.1       | 12.6| -0.972 | 5.43 |
| BBB               | 0.53      | 0.70| 0.819 | 0.05 |
| Absorption        | -5.24     | -6.3| -5.22 | -6.46 |

Key: SBD-SA, spike binding region South African Variant. Active site amino acid residues and South African variants are bold in black.
Based on the Lipinski’s rule of 5, ligand 65 and 40 showed the highest drug-likeness activity. Ligands 65, 137 and 70 were predicted to have poor aqueous solubility, whereas compound 40 showed poor lipid bilayer permeability. All ligands displayed blood-brain barrier permeation except for compound 70. All set of ligands showed poor absorption, however compound 40 was predicted to have greater than 30% human intestinal absorption. In terms of toxicity and elimination, all showed low toxicity and demonstrated optimal half life time.

**Discussion**

The ongoing SARS-CoV-2 (COVID-19) global epidemic outbreak, infecting people worldwide, has fast track both vaccine and drug therapeutics. In this context, computational methods may be used to decrease the time of drug discovery and development. Indeed, results from in-silico studies have advance ranking of lead compounds and reduce both time and selection of poor lead molecules in the laboratory. These studies have contributed to the development of a series of pharmaco-therapeutic drugs and has proven an effective tool for drug discovery [20-23]. In a similar manner, computational methods have the advantage of testing a wide spectrum of compounds that may be model against specific target areas. For example, a recent work by Wu and colleagues reported compounds from different sources including flavonoids, anti-bacterial, anti-HIV and anti-fungal with activity against SARS-CoV-2[24]. Likewise, Natesh et al. described culinary compounds with potential to inhibit main protease, spike protein and ACE2 receptors of SARS-CoV-2[25]. In a similar framework, Feng et al. used pattern recognition
computing to identify ten potential anti-viral compounds from the CAS COVID-19 dataset [26]. Similarly, during the first months of the COVID-19 outbreak, machine learning approaches have been applied to a wide spectrum of areas in an effort to better understand and contain the virus, including genomics, prediction of patient outcomes and infections, assay development and potential drug discovery as inhibitors of COVID-19 [27-32].

In this study, computational methods including a random forest classifier along with molecular docking, interaction profiles and molecular simulation were used to identify potential anti-viral compounds against SARS-CoV-2 structure and non-structure proteins. The random forest classifier showed 0.92 (AUC of the ROC curve) reliability at identifying potential inhibitors and selected 178 compounds with potential activity against SARS-CoV-2 from a large dataset of antiviral compounds; it also provided key chemical important features for prioritization of antiviral compounds and identified key molecular fingerprints as most important classifiers. Binding affinity of these compounds were subsequently validated and confirmed using a multi-molecular docking analysis. Of these, only 39 were validated to bind active sites of viral structural (SBD spike region) and non-structural proteins involve in RNA replication (NSP12-RNA polymerase, NSP13-helicase, and Mpro- main protease). Although the latter compounds showed coverage to SARS-CoV-2 active sites, only four (compounds 40, 65, 70 and 137) were selected for subsequent downstream analysis (molecular interactions and pharmacological properties) based on visual inspection of the active sites covered. The selected compounds showed high affinities to structure and non-structure SARS-CoV-2 targets and formed non-covalent interactions with key active site amino acid residues. All but one, to the best of our knowledge, are exclusively cited as anti-viral agents in the CAS dataset. Only compound 70 (C45 H30 O18, Benzoic acid, 5-[bis(3-carboxy-4-hydroxyphenyl)methyl]-3-[[3-carboxy-5-[(3-carboxy-4-hydroxyphenyl)(3-carboxy-4-oxo-2,5-cyclohexadien-1-ylidene)methyl]-4-hydroxyphenyl[methyl]-2-hydroxy, CAS #1817756-68-5) and its derivatives have varied applications (e.g. treating conditions including nociceptive pain, cancer, age-related macular degeneration, and hemolytic uremic syndrome) [33-36]. To support the molecular interactions identified, molecular dynamics (MD) of a control molecule (remdesivir) and ligand 65 (C48 H53 N5 O13, D-Galactopyranoside, phenylmethyl 2-[(3-carboxy-1-oxopropyl)amino]-2-deoxy-3,4,6-tris-O-(phenylmethyl)-, 5'-ester with 1-[2,3-O-(methoxymethylene)-β-D-ribofuranosyl]-1H-1,2,4-triazole-3-carboxamide, CAS #1002336-72-2) were simulated against the SBD region of the spike protein. These results show the interactions of remdesivir and ligand 65 against active site amino acid residues (ARG403, GLN 493, GLY496, TRY489, ASN 501) of the SBD spike protein region were maintained with strong binding affinity throughout the entire trajectory time interval. Only LEU455 not initially identified by docking and molecular profiles was later described by the MD simulation. We recognize that only one control and one compound were used to confirm the predicted molecular interactions by MD simulation and additional in-vitro and in-vivo experimental studies are warranted to evaluate the bioactivity effects of these compounds on target proteins. Nonetheless, our initial observations provide a foundation to pursue these candidate compounds as potential therapeutics against SARS-CoV-2.

Although the single-drug approach of selected compounds as inhibitors of SARS-CoV-2 may effectively target viral active pockets, it may not be enough to arrest the life cycle of the virus and both multi-target or combinations of drugs may be needed to treat COVID-19. This is consistent with current clinical trials [37]. In this alternative approach, we applied the four
selected compounds to the virus chemo-genomics platform described by Feng and colleagues to identify potential drug treatment combinations against SARS-CoV-2. Several combinations were identified that target structural and non-structural targets simultaneously, however, only two (ligand 40, C16 H20 N5 O7 P, Guanosine, 2'-C-1,2-propadien-1-yl-, cyclic 3',5'-(1-methylmethyl phosphate), CAS #1399805-48-1 and ligand 137, C49 H38 N14 O9 S3, 1-Naphthalenesulfonic acid, 4-[[2,4-diamino-5-][4-[[3-[[2,4-diamino-5-[(4-sulfo-1-naphthalenyl)azo]phenyl]azo]-2-methyl-5-sulfophenyl]azo]-1-naphthalenyl]azo]phenyl]azo], CAS #687967-00-6) were confirmed by docking and molecular interactions to bind to active site pockets of SARS-CoV-2. Of these, compound 40 showed binding to both NSP13 and ACE-2, while compound 137 showed affinity to NSP12 and the SBD region of the spike protein. Interestingly, compound 137 was able to bind with high affinity to the South African variant amino acid residues ASN417 and LYS484. These latter two compounds possibly generating a drug combined effect treatment strategy to both structure and replication apparatus of the virus and possibly inhibiting the newly emerged South African variant to gain access to the host may prove interesting therapeutic candidates to pursue.

In terms of the pharmacological profiles, selected compounds demonstrated acceptable drug-likeness metrics, especially compounds 40 and 65 as inhibitors of NSP13/ACE-2 and the SBD region each. Although the metric for drug likeness of compounds 70 and 137 were not optimal, it should be noted that clinical drugs may have characteristics that result them to score low on indices of drug likeness. For example, one of the compounds initially identified as a candidate inhibitor of NSP13, ligand 25 CAS #1250937-05-3, showed low drug likeness metrics, however, it has been reported as a clinical drug to inhibit viral nonstructural protein NS5A [38]. Likewise, drugs with low drug likeness scores may exhibit acceptable pharmacokinetic profiles, a prerequisite for an effective clinical drug [39]. In regards to toxicity, all selected compounds showed low level of toxicity. Although, the present work lacks in-vitro and in-vivo investigations to ascertain selected compounds toxicity and bioactivity against SARS-CoV-2, we believe our initial observations may be used as appropriate guide for the next phases of drug discovery to fast track drug development and therapeutics against SARS-CoV-2 (COVID-19).

Conclusion

With the global spread of SARS-CoV-2 and with the roll-out of the first approved vaccines, we anticipate prevention and decrease number of fatalities due to SARS-CoV-2 transmission. Nonetheless, challenges will still remain, mainly emerging variants that are more transmissible. Undoubtedly, advances in anti-viral drug development and clinical trials will lead to our arsenal of effective therapeutics against SARS-CoV-2. In this study, we have proposed use of a machine learning tool along molecular docking studies and molecular interactions to identify potential therapeutics against SARS-CoV-2. Our findings from surface 3-D structures between SARS-CoV-2 target sites and anti-viral compounds, along with molecular interactions and simulations suggests selected compounds may serve as a guide for the next phases of drug discovery to treat COVID-19. Though only a selected list of compounds was detailed for further docking and molecular interactions, the remaining compounds identified here and their affinity to the active site of viral targets suggests these compounds may also be used as a guide for further investigation. Lastly, the potential of these compounds (40, 65, 70, 137), and mainly ligand 40 and 137 to simultaneously modulate structure and non-structure protein in SARS-CoV-2 and
target variants may provide guidance to design and deliver effective therapeutics to treat COVID-19.

References

1. Benvenuto, D., et al., The 2019-new coronavirus epidemic: Evidence for virus evolution. J Med Virol, 2020. 92(4): p. 455-459.
2. FDA Moderna COVID-19 Vaccine Emergency Use Authorization (EUA) Letter of Authorization, December 18, 2020. https://www.fda.gov/media/144636/download
3. FDA Pfizer COVID-19 Vaccine Emergency Use Authorization (EUA) Letter of Authorization, reissued December 23, 2020. https://www.fda.gov/media/144412/download
4. Koyama T, Weeraratne D, Snowdon JL, Parida L. Emergence of Drift Variants That May Affect COVID-19 Vaccine Development and Antibody Treatment. Pathogens. 2020; 9(5):324. https://doi.org/10.3390/pathogens9050324
5. Wise J. Covid-19: New coronavirus variant is identified in UK. BMJ. 2020; 371. doi: https://doi.org/10.1136/bmj.m4857
6. Wu, A., et al., Genome Composition and Divergence of the Novel Coronavirus (2019-nCoV) Originating in China. Cell Host Microbe, 2020. 27(3): p. 325-328.
7. He, J., et al., Potential of coronavirus 3C-like protease inhibitors for the development of new anti-SARS-CoV-2 drugs: Insights from structures of protease and inhibitors. Int J Antimicrob Agents, 2020. 56(2): p. 106055.
8. Naik, B., et al., High throughput virtual screening reveals SARS-CoV-2 multi-target binding natural compounds to lead instant therapy for COVID-19 treatment. Int J Biol Macromol, 2020. 160: p. 1-17.
9. Lin, X., X. Li, and X. Lin, A Review on Applications of Computational Methods in Drug Screening and Design. Molecules, 2020. 25(6).
10. Ambure, P., et al., Identifying natural compounds as multi-target-directed ligands against Alzheimer's disease: an in silico approach. J Biomol Struct Dyn, 2019. 37(5): p. 1282-1306.
11. Team, R.c., R: A language and environment for statistical computing. R Foundation for Statistical Computing. Vienna, Austria. URL http://www.R-project.org/. 2013.
12. Handoko, S.D., et al., QuickVina: accelerating AutoDock Vina using gradient-based heuristics for global optimization. IEEE/ACM Trans Comput Biol Bioinform, 2012. 9(5): p. 1266-72.
13. S. Jo, T. Kim, V.G. Iyer, and W. Im (2008) CHARMM-GUI: A Web-based Graphical User Interface for CHARMM. J. Comput. Chem. 29:1859-1865.
14. B.R. Brooks, C.L. Brooks III, A.D. MacKerell, Jr., et al. (2009) CHARMM: The Biomolecular Simulation Program. J. Comput. Chem. 30:1545-1614.
15. J. Lee, X. Cheng, J.M. Swails, M.S. et al. (2016) CHARMM-GUI Input Generator for NAMD, GROMACS, AMBER, OpenMM, and CHARMM/OpenMM Simulations using the CHARMM36 Additive Force Field. J. Chem. Theory Comput. 12:405-413
16. Best, R. B.; Zhu, X.; Shim, J. et al. Optimization of the Additive CHARMM All Atom Protein Force Field Targeting Improved Sampling of the Backbone $\phi$, $\psi$ and Side-Chain $\chi(1)$ and $\chi(2)$ Dihedral Angles. J. Chem. Theory Comput. 2012, 8, 3257–3273.
17. Phillips et al., J. Comp. Chem. 26:1781-1802 (2005).
18. Martyna, G.J.; Tobias, D.J.; Klein, M.L. Constant pressure molecular dynamics algorithms. J. Chem. Phys. 1994, 101, 4177–4189.
19. Humphrey, W.; Dalke, A.; Schulten, K. VMD: Visual molecular dynamics. J. Mol. Graph. 1996, 14, 33–38.
20. Boehm, H.J., et al., Novel inhibitors of DNA gyrase: 3D structure based biased needle screening, hit validation by biophysical methods, and 3D guided optimization. A promising alternative to random screening. J Med Chem, 2000. 43(14): p. 2664-74.
21. Lu, S.J. and F.C. Chong, Combining molecular docking and molecular dynamics to predict the binding modes of flavonoid derivatives with the neuraminidase of the 2009 H1N1 influenza A virus. Int J Mol Sci, 2012. 13(4): p. 4496-507.
22. Bajorath, J., Chemoinformatics and Computational Chemical Biology. Springer, 2011.
23. Mirza, M.U., et al., Perspectives towards antiviral drug discovery against Ebola virus. J Med Virol, 2019. 91(12): p. 2029-2048.
24. Wu, C., et al., Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. Acta Pharm Sin B, 2020. 10(5): p. 766-788.
25. Natesh, J., et al., Culinary spice bioactives as potential therapeutics against SARS-CoV-2: Computational investigation. Comput Biol Med, 2020. 128: p. 104102.
26. Zhing Feng, Maozi Chen, Ying Xue et al. MCCS: a novel recognition pattern-based method for fast track discovery of anti-SARS-CoV-2 drugs. Briefings in Bioinformatics, 00(00),2020,1-17.
27. Randhawa, G.S., et al., Machine learning using intrinsic genomic signatures for rapid classification of novel pathogens: COVID-19 case study. PLoS One, 2020. 15(4): p. e0232391.
28. Alimadadi, A., et al., Artificial intelligence and machine learning to fight COVID-19. Physiol Genomics, 2020. 52(4): p. 200-202.
29. Banerjee, A., et al., Use of Machine Learning and Artificial Intelligence to predict SARS-CoV-2 infection from Full Blood Counts in a population. Int Immunopharmacol, 2020. 86: p. 106705.
30. Gussow, A.B., et al., Genomic determinants of pathogenicity in SARS-CoV-2 and other human coronaviruses. Proc Natl Acad Sci U S A, 2020. 117(26): p. 15193-15199.
31. Tarnok, A., Machine Learning, COVID-19 (2019-nCoV), and multi-OMICS. Cytometry A, 2020. 97(3): p. 215-216.
32. Mirabelli, C., et al., Morphological Cell Profiling of SARS-CoV-2 Infection Identifies Drug Repurposing Candidates for COVID-19. bioRxiv, 2020.
33. McGeer, Patrick L.; Lee, Moonhee; Kennedy, Krista; McGeer, Edith G.; Bell, Douglas N.
Methods for treating and preventing nociceptive pain using a compound selected from diacetyl salicylic acid, aurin tricarboxylic acid, aurin quadracarboxylic acid and aurin hexacarboxylic acid. From U.S. Pat. Appl. Publ. (2018), US 20180207180 A1 Jul 26, 2018.
34. McGeer, Patrick L.; Lee, Moonhee. Methods and kits for prevention and treatment of cancer comprising aurin tricarboxylic acid, aurin quadracarboxylic acid, aurin hexacarboxylic acid and aurin tricarboxylic acid complex. From U.S Pat. Appl. Publ. (2020), US 10653651 May 19, 2020.
35. McGeer, Patrick L.; Lee, Moonhee. Treatment of age-related macular degeneration. From U.S Pat. Appl. Publ. (2013), US 20130035392 Feb 7, 2013.
36. McGeer, Patrick L.; Lee, Moonhee. Treatment of atypical hemolytic uremic syndrome. From U.S Pat. Appl. Publ. (2013), US 20130035392 Feb 7, 2013.

37. ClinicalTrials.gov [Internet]. Bethesda (MD): National Library of Medicine (US). 2020 Feb. Identifier NCT04261270, A Randomized,Open,Controlled Clinical Study to Evaluate the Efficacy of ASC09F and Ritonavir for 2019-nCoV Pneumonia; 2020 Feb 7 [cited 2020 Feb 7]; Available from: www.clinicaltrials.gov/ct2/show/NCT04261270

38. Belema, M., Preparation of biphenylacetylenes, biphenylethers, terphenyls and related derivatives end-capped with amino acid or peptide derivatives as hepatitis C virus inhibitors. PCT Int. Appl. (2012), WO 2012039717 A1 20120329, 2012.

39. Ritchie, T. J. Macdonald, S. J; How drug-like are 'ugly' drugs; do drug-likeness metrics predict ADME behavior in humans? Drug Discov Today (2012), 4(19), 489-95.
Highlights

- Machine learning approaches coupled with molecular docking, interaction profiles and dynamic simulations may be used to reduce the time to identify lead therapeutic compounds to treat SARS-CoV-2 and emerging variants.

- Molecular assessment of four potential anti-viral candidate compounds from the CAS COVID-19 dataset show binding affinity to structure and non-structure SARS-CoV-2 target proteins.

- Selected compounds showed combined therapeutic effect that may simultaneously modulate both structure (SBD-ACE-2 interphase) and replication (NSP12 and NSP13) apparatus of SARS-CoV-2, while potentially inhibiting emerging variants.
Declaration of interests

☒ The authors declare that they have no known competing financial interests or personal relationships that could have appeared to influence the work reported in this paper.

☐ The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: