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.
Virtual screening and in vitro validation of natural compound inhibitors against SARS-CoV-2 spike protein

Helen Power a,b,c, Jiadai Wu a,b, Stuart Turville d, Anupriya Aggarwal d, Peter Valtchev a,b, Aaron Schindeler a,b,c, Fariba Dehghani a,b,*

a School of Chemical and Biomolecular Engineering, Faculty of Engineering, The University of Sydney, Sydney, NSW 2006, Australia
b Centre for Advanced Food Engineering, The University of Sydney, Sydney, NSW 2006, Australia
c Bioengineering and Molecular Medicine Laboratory, The Children’s Hospital at Westmead and Westmead Institute for Medical Research, Westmead, NSW 2145, Australia
d The Kirby Institute, University of NSW, Kensington, NSW 2052, Australia

A R T I C L E   I N F O

Keywords:
SARS-CoV-2
COVID-19
Molecular modeling
Virus neutralization assay
Natural compound inhibitor

A B S T R A C T

The COVID-19 pandemic caused by the SARS-CoV-2 virus has led to a major public health burden and has resulted in millions of deaths worldwide. As effective treatments are limited, there is a significant requirement for high-throughput, low resource methods for the discovery of novel antivirals. The SARS-CoV-2 spike protein plays a key role in viral entry and has been identified as a therapeutic target. Using the available spike crystal structure, we performed a virtual screen with a library of 527,209 natural compounds against the receptor binding domain of this protein. Top hits from this screen were subjected to a second, more comprehensive molecular docking experiment and filtered for favourable ADMET properties. The in vitro activity of 10 highly ranked compounds was assessed using a virus neutralisation assay designed to facilitate viral entry in a physiologically relevant manner via the plasma membrane route. Subsequently, four compounds ZINC02111387, ZINC02122196, SN00074072 and ZINC04090608 were identified to possess antiviral activity in the µM range. These findings validate the virtual screening method as a tool for identifying novel antivirals and provide a basis for future drug development against SARS-CoV-2.

1. Introduction

Since its emergence in Wuhan in December 2019, severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) has become a global health concern. As of July 2021, it has caused over 186 million infections and 4 million deaths [1]. The wild-type virus and its variants are highly transmissible with the potential to rapidly overwhelm healthcare systems [2,3]. The associated illness, known as coronavirus disease (COVID-19) has a wide range of symptoms but is typically associated with fever, cough, fatigue, loss of smell and taste, loss of appetite and muscle pain [4,5]. Serious complications have also been associated with the disease, including clotting disorders, cardiac injury, stroke, and seizures [6–8]. Moreover, long-term symptoms persisting from 3 weeks to months after disease onset occur in 10–30% of patients [9–11], including in patients no longer testing positive for the virus [12].

Since the detection of the initial Wuhan strain, several variants of concern have emerged. In particular, the variants known as alpha, beta, gamma, and delta have spread rapidly and feature mutations associated with increased transmissibility and virulence and the ability to evade the host immune response [13,14]. Currently, only one drug (remdesivir) has been approved by the FDA for non-emergency treatment of COVID-19 [15]. Remdesivir is a repurposed broad-spectrum antiviral that targets viral RNA production [16]. However, results from multiple clinical trials suggest that this drug may not improve mortality or other relevant health outcomes [17]. Monoclonal antibodies against SARS-CoV-2 have also been explored as a treatment option and are currently FDA approved for emergency use only [15]. These treatments including bamlanivimab, etesevimab, casirivimab, imdevimab and sotrovimab, are most effective as an early intervention and are not authorised for patients hospitalised for COVID-19. Moreover, supply limitations and the requirement for mAbs to be administered systemically, prevents the widespread use of these therapies [18,19]. There is also a risk of resistance to common circulating variants, as observed with bamlanivimab, etesevimab and casirivimab [20].

* Corresponding author at: Room 453, School of Chemical and Biomolecular Engineering, Faculty of Engineering, The University of Sydney, Australia.
E-mail address: fariba.dehghani@sydney.edu.au (F. Dehghani).

https://doi.org/10.1016/j.bioorg.2021.105574
Received 17 October 2021; Received in revised form 24 November 2021; Accepted 16 December 2021
Available online 21 December 2021
0045-2068/© 2021 Elsevier Inc. All rights reserved.
There is a strong interest in finding rapid and economical methods to identify new and effective anti-coronavirus treatments. Virtual screening involves computational estimation of the optimal binding of a library of potential drugs against targets of interest. This procedure can greatly reduce the time and cost of the drug development compared to in vitro screening alone and has been frequently used to identify compounds with therapeutic potential [21–25]. Natural compounds are a valuable source of molecules for drug screening. These represent a diverse range of chemical structures, many with known therapeutic properties. In fact, of the 1328 drugs approved by the US Food & Drug Administration between 1981 and 2016, 41% were natural or derived from natural products [26]. Natural compound libraries offer advantages for drug discovery due to their high structural diversity, a feature necessary for selective and specific interactions with protein targets [27].

For SARS-CoV-2, the viral spike protein represents a key therapeutic target as it mediates viral entry into host cells. The spike protein binds to the cellular receptor angiotensin-converting enzyme 2 (ACE2), leading to spike cleavage by the host enzyme transmembrane serine protease (TMPRSS2) [28]. The structure of the receptor binding domain of the Wuhan strain complexed with ACE2 has since been solved by Lan et al. and Shang et al. using X-ray crystallography [29,30]. Contacting residues have been identified, including two key regions involved in the stabilisation of the binding interaction. These two regions, known as the Lys31 and Lys353 hotspots, provide a basis for the design of targeted treatments that could prevent the early stages of infection.

Using this information, we aimed to identify natural compounds with anti-SARS-CoV-2 activity using a combined molecular docking and in vitro approach. Most docking studies screening natural compounds against the spike protein have screened only a small number of ligands (<200) and focus on the computational aspect alone [31–40]. Even larger-scale studies do not include experimental validation [41–43]. To address these limitations, we performed a structure-based docking study targeted at the regions of the spike receptor binding domain (RBD) involved in stabilisation of the Lys31 and Lys353 hotspots. A virtual screen against this site was performed with a total of 527 209 molecules from five natural product databases: phenol explorer [44], marine natural products [45], ZINC [46], Super Natural II [47] and the Human Metabolome Database [48]. After filtering top hits for favourable drug-like characteristics, commercially available compounds were validated using a cell model designed to replicate infection in humans. This model uses human embryonic kidney (HEK) cells genetically modified to express ACE2 and TMPRSS2, thus facilitating viral entry in a physiologically relevant manner via the plasma membrane route [28]. Subsequently, we identified four compounds with anti-SARS-CoV-2 activity in the µg/ml range. We also discussed the relevance of this model in relation to the emerging variants of concern. These findings highlight the capacity of virtual screening to successfully identify inhibitory molecules and provide a basis for further studies regarding viral mechanisms and drug optimisation.

2. Methods

2.1. In silico methods

2.1.1. File preparation and virtual screening strategy

Virtual screening was conducted against the spike protein of SARS-CoV-2 using a library of 527 209 compounds from five natural compound libraries (SuperNatural II, Phenol Explorer, Human Metabolome Database, Marine Natural Products, ZINC Natural Products) available from the Miguel Hernandez University Molecular Docking site (http://docking.umh.es/). Docking was performed using the AutoDock Vina software with a rigid docking protocol [49].

To prepare the protein target for docking, the X-ray derived crystal structure of the SARS-CoV-2 spike receptor-binding domain bound with ACE2 (PDB ID: 6M0J) was downloaded in PDB format from the Research Collaboratory for Structural Bioinformatics Protein Data Bank (http://www.rcsb.org/). PyMOL was used to remove ACE2, water molecules and ions and to add polar hydrogens. The file was then converted to PDBQT format using the PyRx virtual screening tool [50]. The search space of spike was defined to include key regions of the protein involved in the stabilisation of binding to ACE2 (centre: –36.402, 24.771, 6.459; dimensions: 18.703, 29.999, 13.617).

The docking protocol was conducted in two stages. The primary round of screening was performed with all 527 209 ligands. To prepare the ligands for the initial screen, all files were downloaded in SDF format and edited to include polar hydrogens using PyMOL [51]. 3D coordinates were generated by conversion to MOL2 with Marvin Suite 6.0 from ChemAxon [52]. Ligands were then energy minimised and converted to PDBQT format using Open Babel software v 3.1.1 [53]. For the docking of each ligand, the number of docked conformations (modes) and the number of independent runs (exhaustiveness) was set to 10.

To increase the probability of identifying binding conformations with the lowest energy, a secondary round of virtual screening was conducted using the top ligands from the primary screen ranked with the leading 30 scores (n = 12,322). Ligands screened in the second round were energy minimised and converted to PDBQT format using the PyRx virtual screening tool [50]. For this stage, the number of runs was increased to 40.

The binding conformation of the ligands complexed with spike protein were visualised using PyMOL [51]. Residues involved in hydrogen-bonding and hydrophobic interactions were analysed and plotted using the LigPlot+ software [54].

2.1.2. ADMET and drug-likeness analysis

To filter compounds with favourable drug-like characteristics, the OSIRIS property explorer was used to calculate the toxicity risk, hydrophilicity (cLogP), solubility (logS), molecular weight (MW), topological surface area (TPSA), drug-likeness and overall drug-score of the top ranked ligands [55]. Compounds with no predicted toxic fragments (mutagenic, tumorigenic, irritant or reproductive effects) and favourable drug-like properties were retained.

2.2. In vitro methods

2.2.1. Chemicals

Ten commercially available compounds with docking scores ≤−7.9 and favourable drug-like properties were purchased from Vitas-M Laboratory for in vitro testing. All compounds were dissolved in DMSO at a concentration of 10 mg/ml and stored at −20 °C.

2.2.2. Cytotoxicity assay

The cytotoxicity of the purchased compounds was initially assessed in VERO cells using the water-soluble tetrazolium salt (WST-1) assay. Twenty-four hours prior to the assay, VERO cells were seeded in 96-well plates at a density of 10,000 cells per well. Cells were then treated with test compounds serially diluted 10-fold to final concentrations of 100 µg/ml, 10 µg/ml, 1 µg/ml, and 0.1 µg/ml. Vehicle controls were prepared by diluting DMSO to final concentrations corresponding to the percentage of DMSO in the prepared chemicals. Untreated cells were treated as positive controls and wells containing DMEM and DMSO only were used as blanks. Following treatment for 24 h at 37 °C, cells were washed once with DMEM and incubated with 10% WST-1 for 2 h. To determine cell viability, absorbance was measured at 450 nm and 620 nm as a reference wavelength. The relative cell viability was calculated using the following equation:

$$\text{Cell viability} = \frac{(A_{450} - A_{620}) \text{compounds} - (A_{450} - A_{620}) \text{blank}}{(A_{450} - A_{620}) \text{no treatment control} - (A_{450} - A_{620}) \text{blank}} \times 100$$

The highest concentration with a cell viability over 75% was selected...
as the safe concentration for downstream antiviral assays.

2.2.3. In vitro antiviral activity assay

The human cell line HEK clone 24 used by this study was genetically modified to express ACE2 and TMPRSS2 (PMID:34228725) [56]. Successful genetic modification was assessed based on the ability of this cell line to be infected with a strain of SARS-CoV-2 that is genetically the same as that observed in Wuhan in late 2019 (Clade A2.2).

On the day of the experiment, live HEK-24 cells were stained using NucBlue dye and seeded at 15 k cells per well in a 384-well plate. Dilutions of test compounds were added to cells for 30 min at 37 °C, before an equal volume of virus solution was added. The starting concentration used for each compound was the safe amount determined by the cytotoxicity assay. Each compound was tested at eight concentrations using two-fold serial dilutions in duplicate (n = 2). No virus and virus only conditions were included as controls. Following incubation at 37 °C for 24 h, cells were imaged using the InCell 2500 high throughput microscope (Cytaba). To calculate percentage cell protection, imaged nuclei were quantified using InCarta high-content image analysis software (Cytaba). Virus neutralisation was calculated using the following formula:

\[
\text{% neutralisation} = \left[ D - \left(1 - \frac{nuclei \text{ in test well}}{nuclei \text{ in uninfected controls}}\right) \right] \times 100
\]

where D represents the percentage of cell death in the absence of test compounds and is calculated as:

\[
D = 1 - \frac{nuclei \text{ in virus only controls}}{nuclei \text{ in uninfected controls}}
\]

2.2.4. Statistical analysis

All statistical analyses were performed using the GraphPad Prism 9 software (GraphPad Software, Inc.). Non-linear regression analysis was used to generate dose–response curves and calculate EC_{50} values. Values are represented as the mean ± standard deviation (SD).

3. Results and discussion

3.1. In silico analysis

AutoDock Vina was used to perform molecular docking with a total of 527,209 ligands from five natural compound libraries (SuperNatural II, Phenol Explorer, Human Metabolome Database, Marine Natural Products, ZINC Natural Products). Vina utilises a semi-empirical scoring function based on experimental receptor-ligand conformations and affinities. The docking score generated by Vina accounts for the contribution of steric interactions, hydrophobic interactions and hydrogen bonds and is given as a measure of the Gibbs free energy of the binding interaction. Therefore, the more negative the score, the more favourable the interaction. The top 30 scores from the primary screen ranged from −4.6 to −8.1 and included 12,322 (2.3%) of ligands. A secondary screen to refine the top ligands produced scores ranging from −4.6 to −8.9, with the top 1.0% of ligands (n = 128) scoring ≤ −7.9. A flowchart of the workflow is shown in Fig. 1.

The top 128 ligands were then filtered for commercial availability and adsorption, distribution, metabolism, excretion and toxicity (ADMET) predictors using the OSIRIS property explorer [55]. This software predicts the suitability of a small molecule to function as a drug based on the properties of currently traded drugs and toxicity data from the Registry of Toxic Effects of Chemical Substances (RTECS) database. The calculated properties were toxicity risk (mutagenicity, tumorigenicity, irritating effects and reproductive effects), MW, hydrophilicity (cLogP), solubility (logS), topological polar surface area (TPSA), drug-likeness and overall drug-score cLogP (hydrophilicity), logS (solubility). Of the top 128 ligands, 22 could not be purchased from commercial vendors and a further 25 had limited availability. Of the remaining 81 compounds, 10 had pharmacokinetic profiles aligned with traded drugs (Table 1, Fig. 2). The properties for these 10 compounds ranged from

![Fig. 1. The workflow for identifying natural compound inhibitors of SARS-CoV-2.](image-url)
293.3 to 478.5 Da (MW), 1.13 – 4.33 (cLogP), (5.61) – (2.93) (LogS), 47.58 – 94.3 Å² (TPSA), (0.16) – 7.76 (druglikeness) and 0.45 – 0.79 (drug-score). All compounds had no predicted toxicity risks.

The docking scores for these top 10 compounds ranged from (8.4) – (7.9). One ligand had the top docking score of –8.4 (ZINC02726715), two ligands had a docking score of –8.3 (ZINC022122196, ZINC96115460), three ligands had a docking score of –8.1 (SN00074072, ZINC02111387, ZINC96114431) two ligands had a docking score of –8.0 (SN00092464, ZINC04090608) and two ligands had a docking score of –7.9 (ZINC06624435, ZINC85878555). The number of receptor residues involved in the binding interaction ranged from 7 to 11 and the number of hydrogen bonds ranged from 0 to 4 (Table 2). Additionally, all selected compounds contained 4 – 7 ring structures (Fig. 2). This feature is particularly notable for inhibitors of protein–protein interactions (iPPI), as iPPIs are associated with a greater number of aromatic rings than other drugs [57].

The search space used in the docking study was defined to include the spike residues involved in binding to two key regions of the ACE2 protein known as the Lys31 and Lys353 hotspots. Neutralisation of these lysine residues is required to stabilise the binding interaction between the viral RBD and its receptor. Within the spike protein, Leu455 and Gln493 are required for stabilisation of the Lys31 hotspot and Gly496

---

**Table 1**

| Database ID   | Chemical Name                                                                 | MW (Da) | cLogP  | LogS  | TPSA (Å²) | Druglikeness | Drug-score |
|---------------|-------------------------------------------------------------------------------|---------|--------|-------|-----------|--------------|------------|
| ZINC02726715  | 9,10-dimethoxy-2-(naphthalen-2-yloxy)-6,7-dihydro-4H-pyrimido[6,1-a]isoquinolin-4-one | 400.4   | 4.33   | –5.49 | 60.4      | 3.62         | 0.50       |
| ZINC022122196 | (12aS)-2-benzyl-6-(2-methylphenyl)-2,3,6,7,12,12a-hexahydropyrazino[1',2':1,6]pyrido[3,4-b]indole-1,4-dione | 435.5   | 1.28   | –4.57 | 58.4      | 6.64         | 0.58       |
| ZINC96115460  | 2-hydroxy-3-[2-oxo-2-(1,3,4,5-tetrahydro-2H-pyrido[3,4-b]indol-2-yl)ethyl]-3,4-dihydro-5H-1,4-benzodiazepin-5-one | 388.4   | 1.28   | –3.78 | 94.3      | 7.76         | 0.78       |
| SN00074072    | (Z)-1-(1-benzofuran-2-yl)-2-(7,8-dihydro[1,3]dioxolo[4,5-g]isoquinolin-5(6H)-ylidene)ethanone | 333.1   | 3.78   | –5.61 | 60.7      | 3.85         | 0.58       |
| ZINC02111387  | (12aS)-2-cycloheptyl-6-(3-methoxyphenyl)-2,3,6,7,12,12a-hexahydropyrazino[1',2':1,6]pyrido[3,4-b]indole-1,4-dione | 457.6   | 4.11   | –4.78 | 65.6      | 0.70         | 0.45       |
| ZINC96114431  | [1-9H]purin-6-yl]piperidin-4-yl][1,3,4,9-tetrahydro-2H-beta-carbolin-2-yl]methanone | 401.5   | 1.98   | –4.67 | 93.8      | 5.63         | 0.68       |
| SN00092464    | (8S,9S,11S)-10-(1,3-benzodioxol-5-ylmethyl)-8-(phenylcarbonyl)-11a,11b-dihydro-8H-pyrrrolo[3',4';3,4]pyrrolo[2,1-a]isoquinoline-9,11(l8aH,10H)-dione | 478.5   | 3.07   | –5.50 | 76.2      | 4.68         | 0.50       |
| ZINC04090608  | (Z)-2-(7,8-dihydro[1,3]dioxolo[4,5-g]isoquinolin-5(6H)-ylidene)-1-phenylethanol | 293.3   | 3.27   | –4.43 | 47.6      | 2.49         | 0.70       |
| ZINC06624435  | (1S,5R)-3-[(2-oxo-2H-chromen-7-yl]oxy]acetyl]-1,2,3,4,5,6-hexahydro-8H-1,5-methanopyrido[1,2-a][1,5]diazocin-8-one | 392.4   | 1.55   | –2.93 | 76.2      | –0.16        | 0.61       |
| ZINC85878555  | 3-[(1H-indazol-3-yl)-5-oxopyrrolidin-3-yl]carbonyl]-1,2,3,4,5,6-hexahydro-8H-1,5-methanopyrido[1,2-a][1,5]diazocin-8-one | 417.5   | 1.13   | –3.32 | 89.6      | 5.61         | 0.79       |

Fig. 2. Structural formulas of the top 10 lead compounds from the virtual screen after filtering for commercial availability and suitable pharmacokinetic profiles.
and Asn501 are required for stabilisation of the Lys353 hotspot [30].
Lys417 forms two salt bridges with ACE2 and is also important for re-
ceptor binding [29]. Therefore, the interaction of small molecules with
these residues may interfere with stable receptor binding and reduce
viral entry. Notably, the docking conformation of all 10 ligands included
interactions with both Gly496 and Asn501. Three ligands interacted
with all five key residues (Fig. 3, Table 2).

Several approaches may be considered to expand the number of can-
didate compounds identifiable using computational methods. Firstly,
screening larger compound libraries has been shown to increase the
number of true-positive hits within the top ranked ligands [59]. How-
ever, this requires access to large amounts of computational resources
and relies on the availability of extensive compound libraries. When
focusing on natural products, even the largest collections contain only
hundreds of thousands of molecules [60], compared to other synthetic
libraries with over one billion compounds [61]. Another approach is to
reduce stringency when filtering for drug-like properties or to eliminate
this step entirely. This approach, however, may increase the presence
of undesirable properties in lead compounds including poor bioavailability
and high cytotoxicity and hence may require further research into
appropriate delivery methods or modification of functional groups using
techniques such as bioisosterism.

### 3.2. In vitro analysis

The top 10 ranked ligands from the virtual screen that were both
commercially available and had favourable ADMET profiles were pur-
chased for in vitro assessment of their anti-SARS-CoV-2 activity. The
concentration of these compounds employed in antiviral assays was
guided by the outcomes of a cytotoxicity assay; this was defined as 100
µg/ml or the highest concentration in which cell viability was over 75%.
The highest tested concentration of 100 µg/ml showed >75% viability
for six compounds, 10 µg/ml showed >75% viability for three com-
pounds, and 0.1 µg/ml showed >75% viability for one compound
(Table 3).

A SARS-CoV-2 infection assay in HEK-24 cells genetically modified to
express ACE2 and TMPRSS2 was used to assess the candidate natural
compounds. Due to the expression of the TMPRSS2 protease in the
human respiratory tract, viral particles largely utilise the plasma mem-
brane route for entry into these tissues. Some alternative cell lines such
as VeroE6, do not express the required proteases and entry is predomi-
nantly endosomal [28,62]. Therefore, based on physiological relevance,
the genetically modified HEK-24 cell line was selected for experimental
validation. Using this in vitro model, anti-SARS-CoV-2 activity was
determined at 100 µg/ml and exhibiting an EC50 values in the range of
1-1.2 µg/ml. The other active compounds ZINC02122196, SN00074072
(maximal dose 100 µg/ml and exhibiting an EC50 value of 1.12 µg/ml.
The other active compounds ZINC02122196, SN00074072 and ZINC04090608, demonstrated 29.8%, 24.9% and
18.3% virus inhibition at their highest tested concentrations respec-
tively. EC50 values could not be calculated for these compounds due to
the absence of a sigmoidal dose response curve. A dose response may be
observable at higher test concentrations, although these were not

### Table 2

| Database ID | Docking Score | No. of Interacting Residues | No. of H-bonds | Interacting Residues |
|-------------|---------------|-----------------------------|---------------|---------------------|
| ZINC02726715 | -8.4          | 11                          | 3             | Arg403, Tyr449,     |
|              |               |                             |               | Tyr453, Gln493,    |
|              |               |                             |               | Ser494, Tyr485,    |
|              |               |                             |               | Gly496, Gln498,    |
|              |               |                             |               | Asn501, Gly502,    |
|              |               |                             |               | Tyr505             |
| ZINC02122196 | -8.3          | 10                          | 1             | Arg403, Tyr449,     |
|              |               |                             |               | Tyr453, Ser494,    |
|              |               |                             |               | Tyr495, Gly496,    |
|              |               |                             |               | Gln498, Asn501,    |
|              |               |                             |               | Gly502, Tyr505     |
| ZINC96115460 | -8.3          | 11                          | 2             | Arg403, Tyr453,     |
|              |               |                             |               | Leu455, Phe456,    |
|              |               |                             |               | Gln484, Tyr489,    |
|              |               |                             |               | Gln493, Tyr495,    |
|              |               |                             |               | Gly496, Asn501,    |
|              |               |                             |               | Tyr505             |
| SN00074072  | -8.1          | 7                           | 1             | Tyr453, Ser494,     |
|              |               |                             |               | Tyr495, Gly496,    |
|              |               |                             |               | Asn501, Gly502,    |
|              |               |                             |               | Tyr505             |
| ZINC02111387 | -8.1          | 11                          | 1             | Arg403, Tyr417,     |
|              |               |                             |               | Tyr449, Tyr453,    |
|              |               |                             |               | Leu455, Gln493,    |
|              |               |                             |               | Ser494, Tyr495,    |
|              |               |                             |               | Gly496, Asn501,    |
|              |               |                             |               | Tyr505             |
| ZINC96114431| -8.1          | 9                           | 0             | Arg403, Gln406,     |
|              |               |                             |               | Tyr453, Ser494,    |
|              |               |                             |               | Tyr495, Gly496,    |
|              |               |                             |               | Asn501, Gly502,    |
|              |               |                             |               | Tyr505             |
| SN000924644 | -8.0          | 10                          | 2             | Arg403, Gln406,     |
|              |               |                             |               | Tyr453, Leu455,    |
|              |               |                             |               | Gln493, tyr495,    |
|              |               |                             |               | Gly496, Phe497,    |
|              |               |                             |               | Asn501, Tyr505,    |
|              |               |                             |               | Tyr505             |
| ZINC04090608 | -8.0          | 7                           | 2             | Arg403, Gln406,     |
|              |               |                             |               | Tyr453, Ser494,    |
|              |               |                             |               | Tyr495, Gly496,    |
|              |               |                             |               | Asn501, Gly502,    |
|              |               |                             |               | Tyr505             |
| ZINC06624435 | -7.9          | 8                           | 1             | Arg403, Tyr453,     |
|              |               |                             |               | Ser494, Tyr495,    |
|              |               |                             |               | Gly496, Asn501,    |
|              |               |                             |               | Gly502, Tyr505     |
| ZINC85878553 | -7.9          | 8                           | 4             | Arg403, Gln493,     |
|              |               |                             |               | Ser494, Tyr495,    |
|              |               |                             |               | Gly496, Gln498,    |
|              |               |                             |               | Asn501, Tyr505     |

Bolded residues are involved in hydrogen bonding.

An advantage of molecular docking is the ability of this method to
predict the effect of viral mutations on the binding conformation of lead
compounds. Changes to residues in the RBMD may decrease or increase
binding affinity and may also change the structural conformation of the
entire domain.

The four main variants of concern: alpha, beta, gamma, and delta,
contain mutations in five amino acid residues found in the spike RBD
(Fig. 4)[14]. The N501Y mutation has evolved independently in the
alpha, beta and gamma variants and has been shown to affect the
structural conformation of the RBD and increase binding affinity to
ACE2. Notably, Asn501 has been identified in stabilisation of the Lys353
hotspot and as mentioned above, it is predicted to interact with all 10
lead compounds. All variants of concern carry a mutation in Gln484,
which is only predicted to interact with ZINC96115460 and two vari-
ants, alpha and beta contain a mutation in Lys417, which is only pre-
dicted to interact with ZINC02111387. The two mutations unique to the
delta variant, L452R and T478K, are not present in any of the inter-
actions. To provide a better prediction of how these mutations may
affect binding to specific compounds, further modelling is required,
although this depends on the availability of high-resolution protein
structures. Currently, molecular structures of the spike protein have been
solved by cryo-electron microscopy for the alpha, beta and gamma
variants [58]. However, the delta variant spike structure has not yet
been solved and high resolution (<2.5 Å) X-ray crystallography struc-
tures are currently not available for any variants.

In summary, the molecular docking scores and interacting residues of
the top 10 commercially available lead compounds with suitable ADMET
profiles. The four main variants of concern: alpha, beta, gamma, and delta,
and Asn501 are required for stabilisation of the Lys353 hotspot [30].
Lys417 forms two salt bridges with ACE2 and is also important for re-
ceptor binding [29]. Therefore, the interaction of small molecules with
these residues may interfere with stable receptor binding and reduce
viral entry. Notably, the docking conformation of all 10 ligands included
interactions with both Gly496 and Asn501. Three ligands interacted
with the remaining 26 forming hydrophobic contacts. One compound, ZINC02111387 interacted with all five key residues (Fig. 3, Table 2).

An advantage of molecular docking is the ability of this method to
predict the effect of viral mutations on the binding conformation of lead
compounds. Changes to residues in the RBMD may decrease or increase
binding affinity and may also change the structural conformation of the
entire domain.

The four main variants of concern: alpha, beta, gamma, and delta,
investigated due to predicted cytotoxicity. The inhibition seen with SN00074072 was not dose responsive; virus neutralisation ranged from 11.3 to 24.8% for all tested concentrations (0.78–100 µg/ml), suggesting that maximum activity has been reached in the low µg/ml range.

For the four compounds that were inhibitory for SARS-CoV-2, docking scores ranged from (−8.3)–(−8.0), the number of predicted interacting residues in the spike protein ranged from 7 to 11, and the number of hydrogen bonds formed ranged from 1 to 2. Notably, the most potent compound ZINC02111387, was predicted to interact with all five key spike residues identified as important for binding to the native ACE2 receptor (Lys417, Leu455, Gln493, Gly496 and Asn501). The remaining three compounds were predicted to interact with two key residues, Gly496 and Asn501. Although there was no calculable association between docking score and virus neutralisation, the two lowest scoring compounds ZINC06624435 and ZINC85878555 did not exhibit any antiviral activity.

Fig. 3. Interaction diagrams of the top 10 lead compounds from the virtual screen filtered for commercial availability and suitable ADMET profiles. Diagrams were generated using LigPlot+ software. Ligand bonds are drawn in purple, receptor bonds are drawn in brown, red lines represent hydrophobic contacts and broken green lines represent hydrogen bonds with distances in Angstroms. (For interpretation of the references to colour in this figure legend, the reader is referred to the web version of this article.)
The modest number of compounds tested in vitro reflects the low off-the-shelf availability of natural compounds with drug-like ADMET profiles. Natural products are advantageous for drug discovery due to their known ability to carry out diverse functions in biological systems and their exceptional structural diversity which is often not observed in large chemical databases [63]. However, many natural compounds are not readily available and if not manufactured synthetically, may require complex extraction and purification procedures. Curation of large-scale purchasable natural product libraries would allow for a greater number of top ranked ligands to be tested in vitro and hence increase the probability of identifying antiviral compounds. The structural diversity of natural compounds also results in these molecules exhibiting unique chemical properties which can lead to their removal by traditional filtering methods that focus on ADMET profiles of current drugs, such as Lipinski’s rule of five.

Whilst only 10 ligands were tested in vitro, four active compounds against SARS-CoV-2 were identified. These experimental results support the suitability of the docking approach used. Previous large-scale docking studies of 69,000–150,000,000 compounds have reported hit rates of 11–41% [64]. Many factors can influence the success of such studies including availability of high-resolution crystal structures for the target protein, presence of co-factors that mediate ligand binding, presence of protein post-translational modifications and the flexibility, size and composition of the targeted binding site. For example, docking against enclosed binding pockets is usually more successful compared to docking against larger, flat and solvent-exposed regions [64].

### Table 3

| Database ID | Docking Score | Highest tested concentration (µM) | Inhibition (%) at highest concentration |
|-------------|---------------|-----------------------------------|---------------------------------------|
| ZINC02122196 | -8.3          | 229.6                             | 29.8                                  |
| ZINC00074072 | -8.1          | 300.2                             | 24.9                                  |
| ZINC02111337 | -8.3          | 218.5                             | 40.3                                  |
| ZINC00092646 | -8.0          | 209.0                             | no antiviral activity                 |
| ZINC06624435 | -7.9          | 254.8                             | no antiviral activity                 |
| ZINC85878555 | -7.9          | 239.5                             | no antiviral activity                 |

**Tested at 100 µg/ml**

| Database ID | Docking Score | Highest tested concentration (µM) | Inhibition (%) at highest concentration |
|-------------|---------------|-----------------------------------|---------------------------------------|
| ZINC02122196 | -8.3          | 229.6                             | 29.8                                  |
| ZINC00074072 | -8.1          | 300.2                             | 24.9                                  |
| ZINC02111337 | -8.3          | 218.5                             | 40.3                                  |
| ZINC00092646 | -8.0          | 209.0                             | no antiviral activity                 |
| ZINC06624435 | -7.9          | 254.8                             | no antiviral activity                 |
| ZINC85878555 | -7.9          | 239.5                             | no antiviral activity                 |

**Tested at 10 µg/ml**

| Database ID | Docking Score | Highest tested concentration (µM) | Inhibition (%) at highest concentration |
|-------------|---------------|-----------------------------------|---------------------------------------|
| ZINC02122196 | -8.3          | 229.6                             | 29.8                                  |
| ZINC00074072 | -8.1          | 300.2                             | 24.9                                  |
| ZINC02111337 | -8.3          | 218.5                             | 40.3                                  |
| ZINC00092646 | -8.0          | 209.0                             | no antiviral activity                 |
| ZINC06624435 | -7.9          | 254.8                             | no antiviral activity                 |
| ZINC85878555 | -7.9          | 239.5                             | no antiviral activity                 |

**Tested at 1 µg/ml**

| Database ID | Docking Score | Highest tested concentration (µM) | Inhibition (%) at highest concentration |
|-------------|---------------|-----------------------------------|---------------------------------------|
| ZINC02122196 | -8.3          | 229.6                             | 29.8                                  |
| ZINC00074072 | -8.1          | 300.2                             | 24.9                                  |
| ZINC02111337 | -8.3          | 218.5                             | 40.3                                  |
| ZINC00092646 | -8.0          | 209.0                             | no antiviral activity                 |
| ZINC06624435 | -7.9          | 254.8                             | no antiviral activity                 |
| ZINC85878555 | -7.9          | 239.5                             | no antiviral activity                 |

The modest number of compounds tested in vitro reflects the low off-the-shelf availability of natural compounds with drug-like ADMET profiles. Natural products are advantageous for drug discovery due to their known ability to carry out diverse functions in biological systems and their exceptional structural diversity which is often not observed in large chemical databases [63]. However, many natural compounds are not readily available and if not manufactured synthetically, may require complex extraction and purification procedures. Curation of large-scale purchasable natural product libraries would allow for a greater number of top ranked ligands to be tested in vitro and hence increase the probability of identifying antiviral compounds. The structural diversity of natural compounds also results in these molecules exhibiting unique chemical properties which can lead to their removal by traditional filtering methods that focus on ADMET profiles of current drugs, such as Lipinski’s rule of five.

Whilst only 10 ligands were tested in vitro, four active compounds against SARS-CoV-2 were identified. These experimental results support the suitability of the docking approach used. Previous large-scale docking studies of 69,000–150,000,000 compounds have reported hit rates of 11–41% [64]. Many factors can influence the success of such studies including availability of high-resolution crystal structures for the target protein, presence of co-factors that mediate ligand binding, presence of protein post-translational modifications and the flexibility, size and composition of the targeted binding site. For example, docking against enclosed binding pockets is usually more successful compared to docking against larger, flat and solvent-exposed regions [64].
Though the antiviral activity of four compounds was confirmed experimentally, validation of the binding mechanism cannot be determined from the virus neutralisation assay. Whilst molecular docking provides a prediction of the binding conformation, viral infection may potentially occur through allosteric effects or interactions with biomolecules other than spike. Additional techniques are required to validate the docked conformation, such as SPR for investigating binding kinetics and cryo-electron microscopy or X-ray co-crystallography for analysis of the interaction at an atomic level. From a treatment perspective, however, viral neutralisation assays provide a better indication of drug efficacy in a biological system. Although beyond the scope of this study, further research into the binding mechanism of the anti-SARS-CoV-2 compounds may help to identify key spike residues for drug targeting and to serve as the basis for designing chemical modifications to improve the activity of these compounds.

4. Conclusions

This study demonstrates the ability of a targeted structure-based molecular docking approach combined with in vitro validation for the identification of novel antiviral drugs. Despite testing a limited number of compounds in vitro (n = 10), we were able to identify four compounds with previously unknown activity against SARS-CoV-2. These active compounds showed 18–40% viral inhibition using an established live virus infection assay. While their activity is low when compared to neutralising antibodies (which have been shown to inhibit 100% of viral particles in the nM range [65]), it nonetheless supports our modelling approach and its capacity to identify compounds with binding affinity to the target site. It may be possible to further augment the binding affinity and antiviral activity of these candidates by targeted chemical modification (bioisosterism), although this is beyond the scope of this modelling project.

Declaration of Competing Interest

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.

Acknowledgement

The authors acknowledge the technical assistance of Nathaniel Butternorth of the Sydney Informatics Hub, a Core Research Facility of the University of Sydney and the use of University of Sydney’s high performance computing cluster, Artemis. This research/project was also undertaken with the assistance of the high-performance computing system Gadi from the National Computational Infrastructure (NCI), which is supported by the Australian Government.

Investigators J Wu and H Power received salary support from JAT Energy (Toorak, VIC, Australia).

References

[1] Weekly epidemiological update on COVID-19 – 13 July 2021, in: Emergency Situational Updates, World Health Organisation, 2021.
[2] T.L. Dao, V.T. Hoang, P. Colson, J.C. Lagier, M. Million, D. Raoult, A. Levasseur, F. Geller, S.R. Ostrowski, K. Groothuis, C.H. Nielsen, S.D. Nielsen, U. Feldt-Rasmussen, Acute and persistent symptoms in non-hospitalized PCR-confirmed COVID-19 patients, Sci. Rep. 11 (1) (2021). https://doi.org/10.1038/s41598-021-92045-x.
[3] J.K. Logue, N.M. Frenko, D.J. McCulloch, D. McDonald, A. Magedson, C.R. Wolf, H. V. Chu, Sequence in adults at 6 months after COVID-19 infection, JAMA Netw. Open 4 (2) (2021) e210830. https://doi.org/10.1001/jamanetworkopen.2021.0830.
[4] A. Carfi, R. Bernardi, F. Landi, Persistent symptoms in patients after acute COVID-19, JAMA 324 (6) (2020) 602. https://doi.org/10.1001/jama.2020.12665.
[5] J. Khateeb, Y. Li, H. Zhang, Emerging SARS-CoV-2 variants of concern and potential intervention approaches, Crit. Care 25 (1) (2021) 244. https://doi.org/10.1186/s13054-021-04135-7.
[6] M.Z. Saleh, J.P. Derrick, Z.Z. Deris, Structural evaluation of the spike glycoprotein variants on SARS-CoV-2 transmission and immune evasion, Int. J. Mol. Sci. 22 (14) (2021) 7425. https://doi.org/10.3390/ijms22147425.
[7] FDA, Coronavirus (COVID-19) | Drugs. Emergency Preparedness | Drugs, 2021. (cited 2021 19th July). Available from: https://www.fda.gov/drugs/emergency-preparedness.
[8] A.J. Brown, J. Won, R.L. Graham, K.H. Dinnon, A. Sims, J.Y. Feng, T. Chihlar, M. R. Denison, R.S. Baric, T.P. Sheahan, Broad spectrum antiviral remdesivir inhibits human endemic and zoonotic deltacoronaviruses with a highly divergent RNA dependent RNA polymerase, Antiviral Res. 169 (2020) 104541. https://doi.org/10.1016/j.antiviral.2019.104541.
[9] WHO, Therapeutics and COVID-19: living guideline, 17 December 2020, 2020 (cited 2019 19th July). Available from: https://apps.who.int/iris/handle/10665/337876.
[10] P.C. Taylor, A.C. Adams, M.M. Huford, I. de la Torre, K. Winthrop, R.L. Gottlieb, Neutralizing monoclonal antibodies for treatment of COVID-19, Nat. Rev. Immunol. 21 (6) (2021) 382–393, https://doi.org/10.1038/s41577-021-00542-x.
[11] A. Casadevall, E. Dashchova, L.-A. Pirofski, Passive antibody therapy for infectious diseases, Nat. Rev. Microbiol. 2 (9) (2004) 695–703. https://doi.org/10.1038/nrmicro9774.
[12] P. Wang, M.S. Nair, L. Liu, S. Iabetani, Y. Luo, Y. Guo, M. Wang, J. Yu, B. Zhang, D. P. Kwong, B.S. Graham, J.R. Mascola, J.Y. Chang, M.T. Yin, M. Sobieszczak, C.A. Kyratsous, L. Shapiro, Z. Sheng, Y. Huang, D.D. Ho, Antibody resistance of SARS-CoV-2 variants B.1.351 and B.1.1.7, Nature 593 (7877) (2021) 130–135, https://doi.org/10.1038/s41586-021-03398-2.
[13] S. Schmidli, O. Ursu, C.V. Iancu, M. Oreš, T.I. Oprea, J.Y. Choe, Identification of new GLUT2-selective inhibitors through in silico ligand screening and validation in eukaryotic expression systems, Sci. Rep. 11 (1) (2021). https://doi.org/10.1038/s41598-021-93063-x.
[14] Z.P. Lin, N.N. Al Zoubabi, M.L. Xu, N.E. Bowen, T.I. Wu, E.S. Lavi, P.H. Huang, Y.-L. Zhu, B. Kim, E.S. Ratner, In silico screening identifies a novel small molecule inhibitor that counteracts PARP inhibitor resistance in ovarian cancer, Sci. Rep. 11 (1) (2021). https://doi.org/10.1038/s41598-021-87285-5.
[15] M. Awasthi, S. Singh, V.P. Pandey, U.N. Dwivedi, Alzheimer’s disease: An overview of amyloid related pathological gene expressions and its therapeutic implications along with in silico approaches emphasizing the role of natural products, J. Neurol. Sci. 361 (2021) 256–271, https://doi.org/10.1016/j.jns.2021.01.008.
[16] S. Ferla, N.E. Netzler, S. Ferla, S. Veronese, D.E. Tuipulotu, S. Guccione, A. Brancalle, P.A. White, M. Basetto, In silico screening for human norovirus antivirals reveals a non-nucleoside inhibitor of the viral polymerase, Sci. Rep. 8 (1) (2018). https://doi.org/10.1038/s41598-018-22303-y.
[17] V. Katrich, V.-P. Jaakola, J.R. Lane, J. Lin, A.P. Uzman, M. Yeager, I. Kufareva, R.C. Stevens, R. Abagyan, Structure-based discovery of novel chemotypes for adenosine A2A receptors, J. Med. Chem. 53 (4) (2010) 1799–1809. https://doi.org/10.1021/jm901647p.
[18] D.J. Newman, G.M. Cragg, Natural products as sources of new drugs from 1981 to 2014, J. Nat. Prod. 79 (3) (2016) 629–661. https://doi.org/10.1021/acs.jnatprod.5b00310.
A phytochemical-based medication search for the SARS-CoV-2 infection by molecular docking models towards spike glycoproteins and main proteases, RSC Adv. 11 (2021) 12003-12014, https://doi.org/10.1039/D0RA10458B.

S. Zhang, L. Wang, J. Wang, Structural basis of receptor recognition by SARS-CoV-2, Nature 581 (2020) 221-224, https://doi.org/10.1038/s41586-020-2179-y.

T.F. Zhao, L. Fu, B. Liu, L. Wang, The potential of natural products as SARS-CoV-2 main protease (Mpro) and Spike (S) glycoprotein inhibitors: a molecular docking study, Sciencinga (2020) 2020: 6307457, https://doi.org/10.1055/a-1342-0649.

A.C. Pushkaran, P. Nath EN, A.R. Melge, R. Patilivedath, C.G. Mohan, A phytochemical-based medicinal search for the SAR-CoV-2 infection by molecular docking models towards spike glycoproteins and main proteases, RSC Adv. 11 (2021) 12003-12014, https://doi.org/10.1039/D0RA10458B.

N. Thangavel, M.A.I. Bratty, H.A.I. Hamzi, A. Najmi, R.O.A. Alqu, Molecular docking and molecular dynamics aided virtual search of olivetree for secoiridoids to combat SARS-CoV-2 infection and associated hyperinflammatory responses, Front. Mol. Biol. 7 (2020) 627776, https://doi.org/10.3389/fmolb.2020.627776.

S. Maiti, A. Banerjee, Epigallocatechin gallate and theallavin gallate interaction in SARS-CoV-2 spike-protein central channel with reference to the hydroxylchloroquine interaction: bioinformatics and molecular docking study, Drug Dev. Res. 82 (1) (2021) 86-96, https://doi.org/10.1002/ddr.21730.

A. Basu, A. Sarkar, U. Maulik, Molecular docking study of potential phytochemicals and their effects on the complex of SARS-CoV2 spike protein and human ACE2, Sci. Rep. 10 (1) (2020) 17699, https://doi.org/10.1038/s41598-020-74715-4.

A. Ubani, F. Agwom, O.R. Morenikeji, S. Nathan, P. Luka, U.U. Marin, N.E.S. OmaleNadinji, J.C. Aguiyi, Molecular docking analysis of some phytochemicals on two SARS-CoV-2 spike targets: potential lead compounds against two target sites of SARS-CoV-2 obtained from plants, bioRxiv (2020), https://doi.org/10.1101/2020.03.31.017657.

C.N. Patel, D. Goswami, D.G. Jaishwal, R.M. Parmar, H.A. Solanki, H.A. Pandya, Pinpointing the potential hits for hindering interaction of SARS-CoV-2 spike protein with ACE2 from the pool of antiviral phytochemicals utilizing molecular docking and molecular dynamics (MD) simulations, J. Mol. Graph. Model. 105 (2021) 107874, https://doi.org/10.1016/j.jmgm.2021.107874.

G.Y. Chen, T.Y. Yao, A. Ahmed, Y.C. Pan, J.G. Yang, Y.C. Wu, The discovery of potential natural products for targeting SARS-CoV-2 spike protein by virtual screening, bioRxiv (2020), https://doi.org/10.1101/2020.06.25.170639.

A.G. Al-Sehemi, F.A. Alattar, S. Dev, M. Pinnapuri, M.E. Soliman, S. Carradori, B. Mathew, Natural products database screening for the discovery of naturally occurring SARS-CoV-2 spike glycoprotein blockers, ChemistrySelect 5 (2020) 13309-13131, https://doi.org/10.1002/slct.202003349.

K. Lapin, E.M. Jokinen, D.T. Kurkinen, Screening of natural products targeting SARS-CoV-2-ACE2 receptor complex – a VAD-based HTVS pipeline, Front. Chem. (2020) 8, https://doi.org/10.3389/fchem.2020.589769.

J.A. Rothwell et al., Phenol-Explorer 3.0: a major update of the Phenol-Explorer database to incorporate data on the effects of food processing on phenolic compounds, Database (2021) e20200163, https://doi.org/10.1093/database/btab024.

V. Ruiz-Torres, J. Encinar, M. Herranz-López, A. Pérez-Sánchez, V. Galindo, E. Barragán-Catalán, V. Micó, An updated review on marine anticancer compounds: the use of virtual screening for the discovery of small-molecule cancer drugs, Molecules 22 (7) (2017) 1037, https://doi.org/10.3390/molecules22071037.