Research paper

In silico investigation to identify potential small molecule inhibitors of the RNA-dependent RNA polymerase (RdRp) nidovirus RdRp-associated nucleotidyltransferase domain

Eleni Pitsillou [a,b], Julia Liang [a,b], Helen Yu Meng Huang [a,c], Andrew Hung [b], Tom C. Karagiannis [a,d,*]

[a] *Epigenomic Medicine, Department of Diabetes, Central Clinical School, Monash University, Melbourne, VIC 3004, Australia*
[b] *School of Science, STEM College, RMIT University, VIC 3001, Australia*
[c] *Department of Microbiology and Immunology, The University of Melbourne, Parkville, VIC 3052, Australia*
[d] *Department of Clinical Pathology, The University of Melbourne, Parkville, VIC 3052, Australia*

ARTICLE INFO

ABSTRACT

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) RNA-dependent RNA polymerase (RdRp) is a promising target for antiviral drugs. In this study, a chemical library (n = 300) was screened against the nidovirus RdRp-associated nucleotidyltransferase (NiRAN) domain. Blind docking was performed using a selection of 30 compounds and nine ligands were chosen based on their docking scores, safety profile, and availability. Using cluster analysis on a 10 microsecond molecular dynamics simulation trajectory (from D.E. Shaw Research), the compounds were docked to the different conformations. On the basis of our modelling studies, oleuropein was identified as a potential lead compound.

1. Introduction

Severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) belongs to the betacoronavirus genus and is the infectious agent that causes coronavirus disease 2019 (COVID-19) [1]. To date, there are seven coronaviruses that can infect humans and they are divided into two groups [1]. The common human coronaviruses (hCoVs) generally cause a mild to moderate respiratory infection and include hCoV-229E, hCoV-NL63, hCoV-OC43, and hCoV-HKUI [1]. Additionally, three hCoVs have been reported to cause severe disease and they include severe acute respiratory syndrome coronavirus (SARS-CoV), Middle East respiratory syndrome coronavirus (MERS-CoV), and SARS-CoV-2 [1]. These three coronaviruses have the capacity to cause lower respiratory infections, which can result in acute lung injury (ALI), acute respiratory distress syndrome (ARDS), and multiorgan failure [2]. The long-term consequences of infection continue to be investigated. The high transmissibility of SARS-CoV-2 and its emerging variants, has resulted in strict public health measures being implemented and a significant amount of attention has been placed on developing vaccines and investigating potential antiviral drugs.

Betacoronaviruses are enveloped viruses that consist of a positive-sense, single stranded RNA genome [1]. The SARS-CoV-2 genome consists of a replicase complex that is formed by the open reading frames, ORF1a and ORF1b [3]. These two ORFs encode the polyproteins pp1a and pp1ab, which are cleaved to produce the non-structural proteins (nsp1-16) [3]. The SARS-CoV-2 genome is also comprised of structural and accessory genes. The four major structural proteins are the spike protein (S), nucleocapsid protein (N), envelope protein (E), and membrane glycoprotein (M) [3]. The receptor binding domain of the spike protein attaches to the host cell receptor angiotensin-converting enzyme 2 (ACE2) and this interaction mediates SARS-CoV-2 infection [4].

Non-structural protein 12 (nsp12), which is also known as RNA-dependent RNA polymerase (RdRp), is a crucial component of the replication-transcription complex and catalyses the synthesis of RNA from RNA templates [5]. The RdRp interacts with proteins such as nsp7, nsp8, nsp9 and nsp13 to facilitate virus replication and transcription [5]. The structure of the SARS-CoV-2 RdRp has been determined and is comprised of a right-hand RdRp domain, a nidovirus RdRp-associated nucleotidyltransferase domain (NiRAN), an interface domain and an N-terminal β-hairpin [6]. The RdRp domain consists of the fingers, palm and thumb subdomains [6].

The SARS-CoV-2 RdRp has also been identified as an ideal target for

antiviral drugs. Remdesivir and favipiravir are examples of prodrugs that are being tested for their ability to inhibit the RdRp, as they act as nucleoside analogues and are incorporated into the growing RNA chain [7]. This results in the termination of RNA synthesis. Remdesivir was the first drug to be approved by the U.S. Food and Drug Administration (FDA), despite the contradicting findings from the Solidarity Trial conducted by the World Health Organization [7]. The NiRAN domain has also been of interest and it is conserved in the *Nidovirales* [8]. This domain was first discovered in the RdRp of the equine arteritis virus (EAV) and it was hypothesised to have RNA ligase activity, nucleotidyltransferase activity, and protein priming function [8]. The potential kinase or phosphotransferase activity of the NiRAN domain is also being investigated and more recently, its interaction with nsp9 has been explored [9].

In addition to developing drugs that inhibit the catalytic activity of the RdRp through covalently binding to the RNA template, the NiRAN domain could be a potential target site for therapeutic agents [10]. Drug repurposing will continue to play an integral role in combating infectious diseases and a number of studies have utilised computational methods to identify potential lead compounds from existing drugs [11]. In this study, molecular modelling tools were used to screen a library of 300 compounds against the NiRAN domain of the SARS-CoV-2 RdRp. This consisted of pharmacological compounds and natural compounds, with antiviral, antioxidant, and anti-inflammatory properties. As aforementioned, the NiRAN domain has nucleotidylating activity and adenosine diphosphate (ADP), uridine-5′-triphosphate (UTP), and guanosine-5′-triphosphate (GTP) were used as the control ligands. Based on the results, the library was narrowed down to 30 compounds. The potential lead ligands were subsequently identified through performing blind docking on the RdRp structures and molecular docking on several conformations of the RdRp from a 10 μs trajectory [12].

2. Materials and methods

2.1. Structures of the proteins and ligands

The cryo-electron microscopy (cryo-EM) structures of the SARS-CoV-2 RdRp were obtained from the RCSB Protein Data Bank (PDB ID: 6 M71 and 6XEZ) [6,13,14]. The RdRp chain was isolated from the replication-transcription complex, the waters and ligands were removed, and the relevant ions were retained. This included the zinc (Zn^{2+}) ions in both structures. Adenosine diphosphate was the ligand present in the NiRAN domain of the 6XEZ cryo-EM structure and this was used as a control [13]. Likewise, UTP and GTP were used as control compounds. The chemical structures of ADP, UTP, GTP, and 300 ligands were obtained from the National Centre for Biotechnology Information (NCBI) PubChem Database [15]. If unavailable, the chemical structures were obtained from the ChEMBL Database [16]. The library of 300 ligands consisted of 220 phenolic compounds and 13 fatty acids from Olive-Net[TM] [17]. A number of compounds (n = 63) with antimicrobial and anti-inflammatory properties were also utilised.

2.2. Preparation of the proteins and ligands

The cryo-EM structures of the RdRp protein were imported into Maestro and were prepared using the Protein Preparation Wizard of the Schrödinger Suite (version 2020–4) [18]. Similarly, the compounds were imported into Maestro and were prepared using the LigPrep tool. The default settings were utilised and the optimised potentials for liquid simulations (OPLS3e) force field was selected [19]. A receptor grid that was 20 × 20 × 20 Å in size was generated around the conserved residues of the NiRAN domain and they were K73, E83, R116, L119, T120, T123, T206, D208, N209, Y217, D218, G220, D221, and S236 [20]. The Glide Receptor Grid Generation protocol was used for this step.

2.3. Molecular docking using the Schrödinger Suite

The 300 ligands and controls were initially screened using the Glide Ligand Docking protocol. The Glide standard precision (SP) mode was selected for this process. The SP mode allows for compounds to be docked in a timely manner and is more accurate than the high throughput screening option. The promising ligand poses were refined using the OPLS3e force field and this was followed by post-docking minimisation.

Based on the results, 30 compounds were examined further. The ligands were docked to the NiRAN domain using the quantum-mechanics-polarised ligand docking (QPLD) protocol of the Schrödinger Suite for improved docking accuracy [21]. The compounds were initially docked using Glide and energy calculations were then performed on the protein–ligand complexes generated using *ab initio* quantum mechanics (QM) methods. The ligands were re-docked using the charges that were predicted by the QSite software and the poses were ranked in the final stage. The extra precision (XP) mode was chosen for the initial docking and redocking steps, and the QM level was set to accurate for the Jaguar component [22–24]. The GlideScore (kcal/mol) was recorded and the protein–ligand interactions were visualised using the Ligand Interaction Diagram tool.

2.4. Blind docking and binding site prediction

The P2Rank software package is a template-free tool that predicts ligand-binding sites based on machine learning and the SARS-CoV-2 RdRp cryo-EM structures were analysed using this program [25]. Blind docking was also performed with the selection of 30 compounds. The goal of blind docking was to investigate whether the ligands would preferentially bind to the NiRAN domain or any other site in the protein, which may potentially include an allosteric binding site. The structures of the proteins and compounds were imported into PyRx and they were prepared as macromolecules and ligands, respectively [26]. The protein was set as rigid, while all torsions of the ligands were activated. The receptor grid was generated around the entire protein and the exhaustiveness was increased to 2048. AutoDock Vina was used to perform the blind docking calculations and the jobs were run on Galileo, which is a cloud computing service (Hypernet Labs), and the Spartan High Performance Computing (HPC) system [27–29].

2.5. Cluster analysis of MD simulation trajectory

A 10 μs molecular dynamics (MD) simulation trajectory of the SARS-CoV-2 nsp7-nsp8-nsp12 RNA polymerase complex (PDB ID: 6 M71) was obtained from the D.E. Shaw Research group and analysed using the Gromacs 2018.2 software package with plug-ins for Visual Molecular Dynamics 1.9.3 [12,30–32]. Nsp12 was isolated from the protein complex and analysed using root mean square deviation (RMSD) and root mean square fluctuation (RMSF) analysis tools included in Gromacs 2018.2. The Gromacs clustering tool gmx cluster was utilised to calculate clusters of similar structures based on RMSD of the protein. The gromos clustering algorithm, as described by Daura et al. [33], was applied. Cluster analysis was performed on the partially disordered N-terminal region of the protein (residues 30–120) for the entire trajectory, where the time interval between frames was 1.2 ns. Using an RMSD cut-off of 0.3 nm to define two structures as neighbours, 15 clusters were obtained. Representative protein structures for the top six clusters were extracted from the trajectory based on the median frame of each group for molecular docking of compounds to the NiRAN domain of RdRp.

2.6. Protein structure alignment

The Protein Structure Alignment tool in Maestro was used to align the NiRAN domain of the conformations that were representative of each cluster, using the cluster 1 structure as the reference. The NiRAN

domain of the 6XEZ cryo-EM structure was also aligned to the conformation corresponding to cluster 1 for comparison. The RMSD values of the aligned amino acids were recorded.

3. Results and discussion

3.1. Molecular docking of 300 compounds to the NiRAN domain

In a study that was performed on the RdRp of the EAV, the nucleotidylation activity of the NiRAN domain was observed when UTP and GTP were present as substrates [8]. In a recent paper by Slanina et al. it was demonstrated that the coronavirus NiRAN domains could transfer nucleoside monophosphates to nsp9 and that there was relatively low specificity for a particular NTP substrate [9]. Residues K73, R116, T123, D126, D218 and F219 of the SARS-CoV-2 RdRp have previously been predicted to be essential for the enzymatic activity of the NiRAN domain, and multiple sequence alignment of coronavirus RdRp sequences has revealed that there are a number of conserved residues (Fig. 1) [20].

Using the Glide Ligand Docking protocol, 300 compounds were screened against the NiRAN domain of the SARS-CoV-2 RdRp (PDB ID: 6 M71) (Table S1). Lucidumoside C was the flavonoid compound that had the weakest binding affinity and the GlideScore was −0.4 kcal/mol. Delphinidin was predicted to be the strongest binding ligand and the GlideScore was found to be −7.2 kcal/mol. In addition to the library of 300 ligands, ADP, UTP and GTP were used as the control compounds. The GlideScores of these compounds were −6.3, −6.3, and −6.0 kcal/mol, respectively.

Based on this initial screen, 30 compounds with a broad range of binding affinities were selected for further analysis. This allowed for comparison of a range of ligands to ensure that there was a reasonable agreement between the ranking according to binding affinities. The commercial availability, approval by the FDA, and known side effects of these compounds were also taken into consideration. They included protease inhibitors, antibiotics, kinase inhibitors, nucleoside analogues, dietary compounds, and compounds with antioxidant and anti-inflammatory properties. For improved docking accuracy and further refinement, the 30 ligands and control compounds were subsequently docked to the NiRAN domain using the QPLD protocol. The GlideScores ranged from −3.6 to −10.8 kcal/mol and the chemical structures of these ligands are provided in Table 1. In order to evaluate whether the controls and selected compounds would preferentially bind to the active site of the NiRAN domain, blind docking was performed on the cryo-EM structure of the SARS-CoV-2 RdRp. For the 6 M71 structure, 20 ligands had poses within the NiRAN domain (Table S2).

In a study by Dwivedy et al. the NiRAN domain was found to assume a kinase-like fold and is thought that this region may have pseudokinase or phosphotransferase activity [20]. The motif search also predicted the presence of kinase-like motifs and to explore this further, they docked broad specificity kinase inhibitors to the active site of the NiRAN domain [20]. Sunitinib and sorafenib were predicted to interact with aspartate residues, while SU6656 formed a hydrogen bond with K73 [20]. Through using an ADP-Glo Kinase assay kit, Dwivedy et al. were also able to provide evidence that the SARS-CoV-2 RdRp had intrinsic kinase/phosphotransferase like activity and that the kinase inhibitors significantly reduced its kinase-like activity [20]. Sunitinib, ibrutinib, zanubrutinib, sorafenib and acalabrutinib were the kinase inhibitors examined in the current study and when examining the protein–ligand interactions, it was apparent that the ligands also formed hydrogen bonds with negatively charged aspartate residues in the NiRAN domain
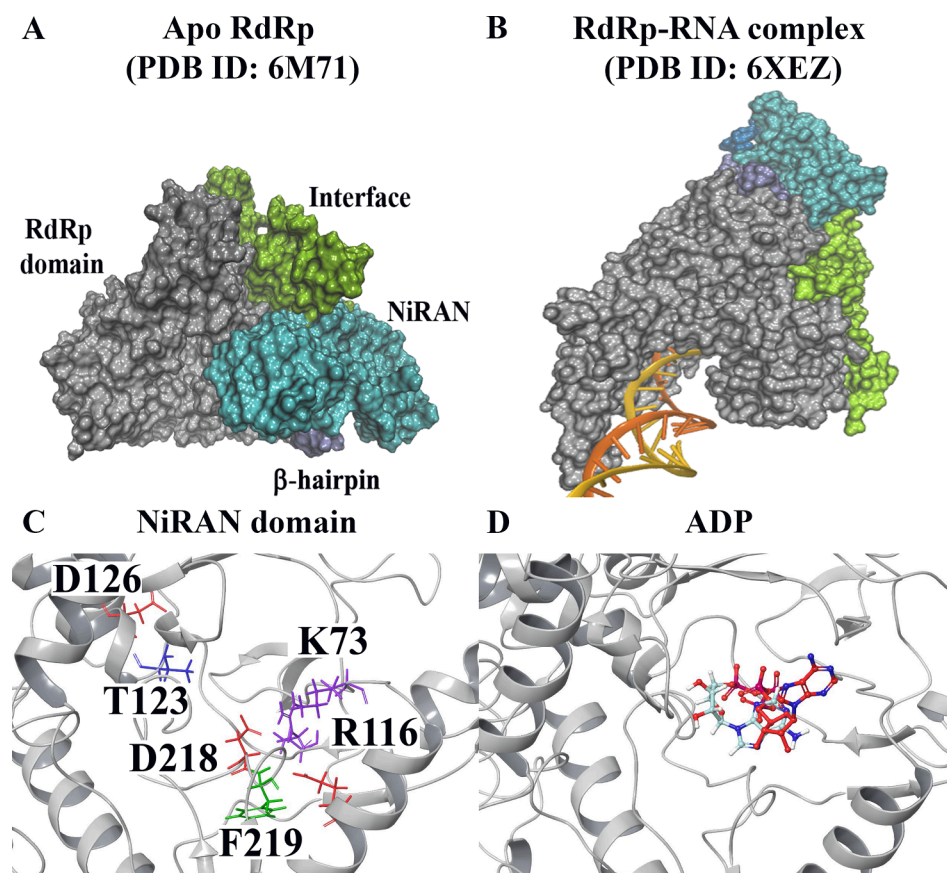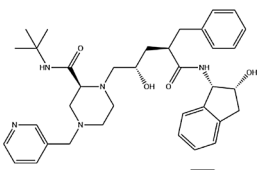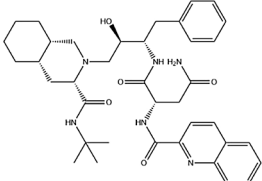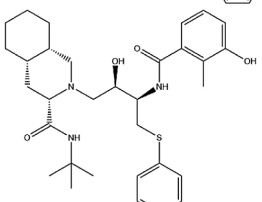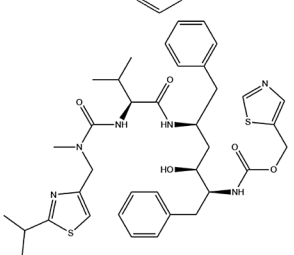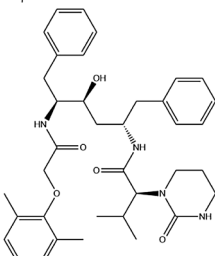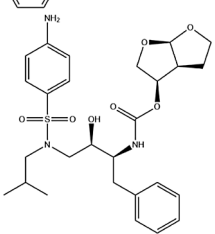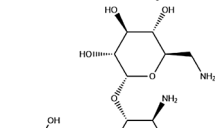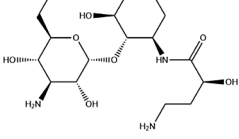

Fig. 1. Structure of the SARS-CoV-2 RdRp. The cryo-EM structure of the apo RdRp (A) and the RdRp chain in complex with RNA (B) can be seen. The key active site residues in the NiRAN domain are labelled (C). The orientation of the ADP (red) present within the cryo-EM structure of the RdRp (PDB ID: 6XEZ) has also been compared to the docked ADP (blue) (D).
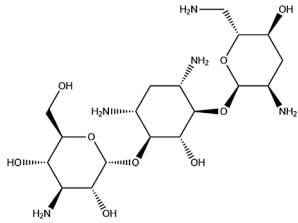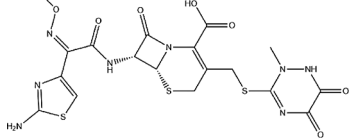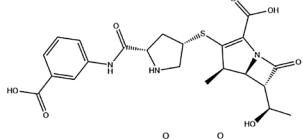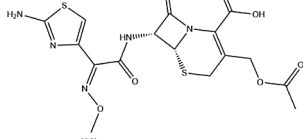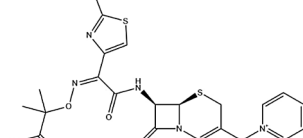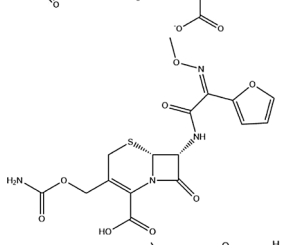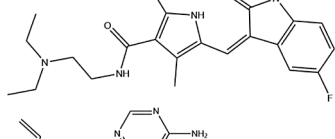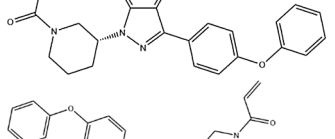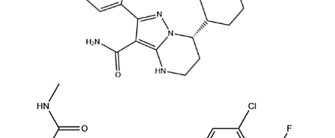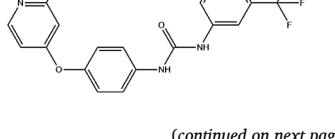
Table 1

The selection of 30 compounds that were docked to the NiRAN domain are provided, along with their GlideScores (kcal/mol) (PDB ID: 6 M71). The control compounds are also included.

Description	Ligand	GlideScore (kcal/mol)	Structure
Protease inhibitors	Indinavir: FDA approved	−7.4	
	Saquinavir: FDA approved	−6.5	
	Nelfinavir: FDA approved	−5.7	
	Ritonavir: FDA approved	−5.1	
	Lopinavir: FDA approved	−4.9	
	Darunavir: FDA approved	−4.4	
Antibiotics	Amikacin: FDA approved	−10.2	
	Tobramycin: FDA approved	−10.1	
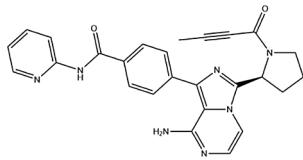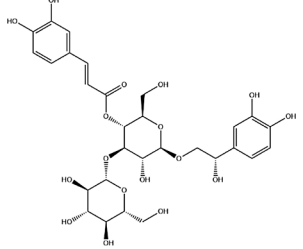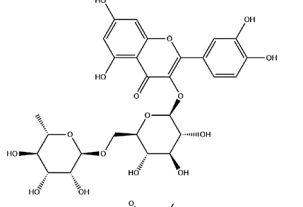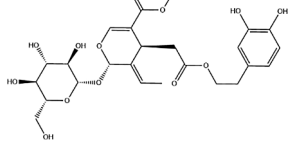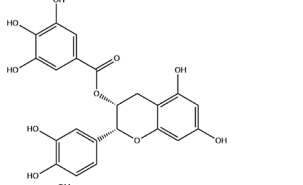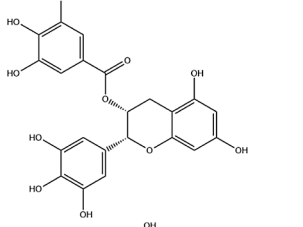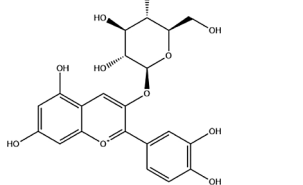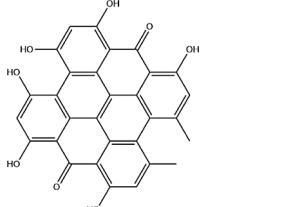
Table 1 (*continued*)

Description	Ligand	GlideScore (kcal/mol)	Structure
	Ceftriaxone: FDA approved	−6.0	
	Ertapenem: FDA approved	−5.5	
	Cefotaxime: FDA approved	−4.2	
	Ceftazidime: FDA approved	−3.8	
	Cefuroxime: FDA approved	−3.6	
Kinase inhibitors	Sunitinib: FDA approved	−5.0	
	Ibrutinib: FDA approved	−4.4	
	Zanubrutinib: FDA approved	−4.3	
	Sorafenib: FDA approved	−4.2	
	Acalabrutinib: FDA approved	−3.9	
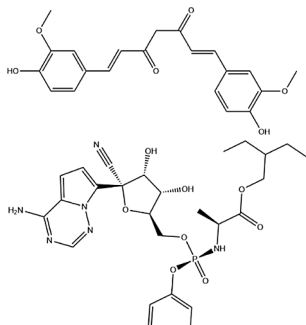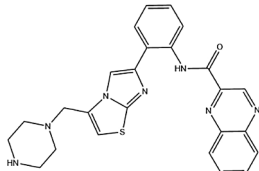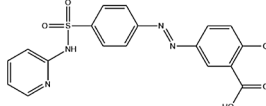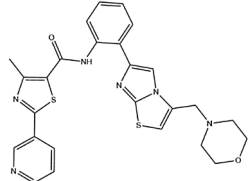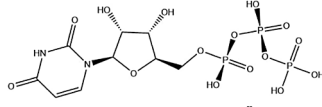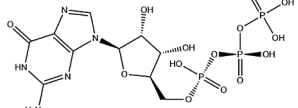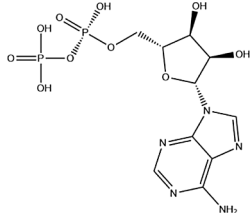
Table 1 (*continued*)

Description	Ligand	GlideScore (kcal/mol)	Structure
Dietary compounds: Antioxidant and anti-inflammatory activities	Hellicoside: Phenolic compound	−10.8	
	Rutin: Phenolic compound	−9.8	
	Oleuropein: Phenolic compound	−8.8	
	Epicatechin gallate: Polyphenol	−8.0	
	Epigallocatechin gallate: Polyphenol	−7.6	
	Cyanidin-3-O-glucoside: Phenolic compound	−7.2	
	Hypericin: Orphan drug designation for the synthetic form of hypericin (SGX301)	−5.8	
	Curcumin: Polyphenol	−5.5	

Table 1 (*continued*)

Description	Ligand	GlideScore (kcal/mol)	Structure
RdRp inhibitors	Remdesivir: FDA approved	−6.5	
Anti-inflammatory agents	SRT1720: Experimental drug – sirtuin activator	−5.2	
	Sulfasalazine: FDA approved	−4.1	
	SRT2104: Experimental drug – sirtuin activator	−4.0	
Controls	UTP	−10.0	
	GTP	−7.2	
	ADP	−7.0	

(Table S3). It is important to note that kinase inhibitors may provide clinical benefit by exerting dual antiviral and anti-inflammatory effects, and the potential side-effects should also be taken into consideration [34].

Furthermore, the antiretroviral protease inhibitors that are used to treat patients with human immunodeficiency virus/acquired immunodeficiency syndrome (HIV/AIDS) have been of interest [35]. Lopinavir, for example, is an inhibitor of the SARS-CoV main protease (M^{pro}) and *in vitro* studies have shown that this drug has inhibitory activity against SARS-CoV, SARS-CoV-2, and MERS-CoV [36]. Lopinavir is commonly used in combination with ritonavir, and these inhibitors have been tested in patients admitted to hospital with COVID-19 [37]. Broad-spectrum antibiotics have also played a role in the drug repurposing process, as they may be used for the treatment of co-infections, and their mechanisms of action require further elucidation [38].

Several dietary compounds were also part of the 30 ligands to be selected, with rutin, hellicoside, oleuropein, and cyanidin-3-O-glucoside being phenolic compounds from the OliveNet™ database [17]. Curcumin, which is the major constituent of turmeric, as well as the catechins (epicatechin gallate and epigallocatechin gallate) are also classified as polyphenols [39]. Hypericin is classified as an anthraquinone derivative and is found in St. John's Wort [40]. Over one-third of new molecular entities that are approved by the FDA are natural products and their derivatives, and numerous studies have focused on screening large libraries of phytochemicals against coronavirus proteins to identify potential lead compounds [41]. Various natural compounds have been examined for their ability to target the spike protein, M^{pro}, papain-like protease (PL^{pro}), and RdRp, and further research is required to validate their antiviral effects and pharmacokinetic properties. Curcumin, piperine, demethoxycurcumin, glycyrrhizic acid, rutin, nicotiflorin,

epigallocatechin-3-gallate, and theaflavin are natural compounds that have been identified as potential antiviral drugs against the SARS-CoV-2 RdRp based on *in silico* analysis [42–44].

3.2. Structural analysis of 10 μs MD simulation trajectory of RdRp protein complex

Due to the NiRAN domain being a flexible region of the RdRp, cluster analysis was performed on a 10 μs MD simulation trajectory of the RdRp protein complex that was made available by the D.E Shaw Research group. Ensemble docking uses MD simulations to generate conformations of the protein for docking calculations, aiming to reproduce the selection of ligands for specific protein conformations that form more thermodynamically favourable protein/ligand complexes [45]. Thus, the aim was to select a subset of conformations where a representative protein structure for each cluster could be used for further docking.

The average RMSD of nsp12 protein backbone was 0.47 nm over the duration of the trajectory. There was a slight fluctuation in backbone RMSD at approximately 3 μs before stabilising after 4.2 μs (Fig. 2A).

RMSF analysis (Fig. 2B) indicated that this may be attributed to flexibility in the partially disordered residues 30 – 120 encompassing the N-terminal region of nsp12. The most prominent peaks in this region are at residues D61 and D107 located on the outer loops of the protein with RMSF values of 0.90 nm. It is noted that due to the highly flexible nature of the N-terminal residues, the structure of this region was previously unable to be resolved in SARS-CoV nsp12 [46]. As this was the most flexible region of the protein in proximity to the proposed binding site, this region was selected for cluster analysis.

Cluster analysis was performed to identify the most prevalent conformations in the trajectory for further screening of compounds. Cut-off values were varied between 0.1 and 0.5 nm in increments of 0.1 nm, with clustering analysis performed for each of these values. Based on the distribution of structures captured by each group, a cut-off distance of 0.3 nm for the N-terminal protein was selected. 8334 frames of the trajectory were divided into 15 clusters. Representative structures from the six most prevalent structures were used as starting structures for docking with lead compounds. The majority (59.2%) of frames were assigned to cluster 1, followed by 20.7% to cluster 2. The remaining
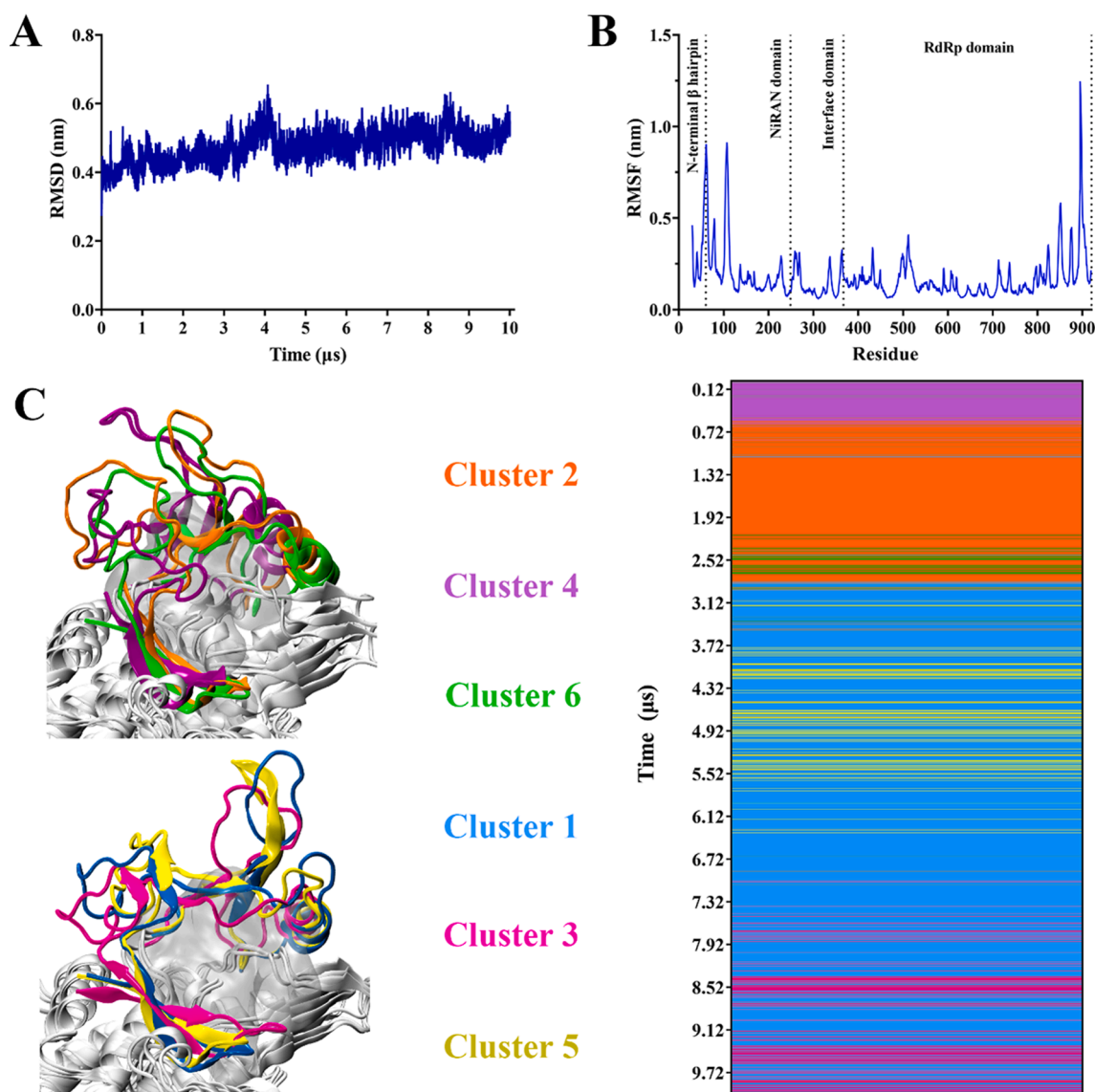
Fig. 2. Analysis of 10 μs MD simulation trajectory of the SARS-CoV-2 nsp7-nsp8-nsp12 RNA polymerase complex. A) Root mean square deviation (RMSD) of nsp12 backbone after fitting to backbone for 10 μs trajectory. B) Root mean square fluctuation (RMSF) of nsp12 backbone. C) Cluster analysis of flexible N-terminal region of NiRAN domain in SARS-CoV-2 nsp12, where the six most prevalent clusters are highlighted in representative structures. The cluster number over time throughout the trajectory is represented as a heat map.

clusters captured: 6.9% (cluster 3), 5.5% (cluster 4), 4.6% (cluster 5), 1.4% (cluster 6) of frames. Clusters 7 to 15 each captured less than 0.5% of frames, and were thus excluded from analysis.

From the 10 μs trajectory of the SARS-CoV-2 nsp12, this N-terminal region is initially partially disordered, becoming folded into a stable ordered structure resembling the N-lobe fold of protein kinases in agreement with the same complex determined in the presence of a reducing agent [6,12]. From the heatmap shown in Fig. 2C, the N-terminal residues at the beginning of the trajectory are in conformations consistent with clusters 4, 2, and 6 until approximately 3 μs. From this time point, the protein becomes stable in conformations corresponding

to cluster 1, which becomes the most common structure for the remainder of the trajectory. Conformations assigned to cluster 5 emerge at approximately 5 μs, while conformations corresponding to cluster 3 occur 8 μs into the trajectory. It is inferred from this analysis that clusters 1, 5, and 3 may represent conformations of the stable ordered N-terminal region of nsp12. However, it is acknowledged that further analysis will be required to characterise this. For the purpose of the present manuscript, representative structures for each cluster were utilised for molecular docking.
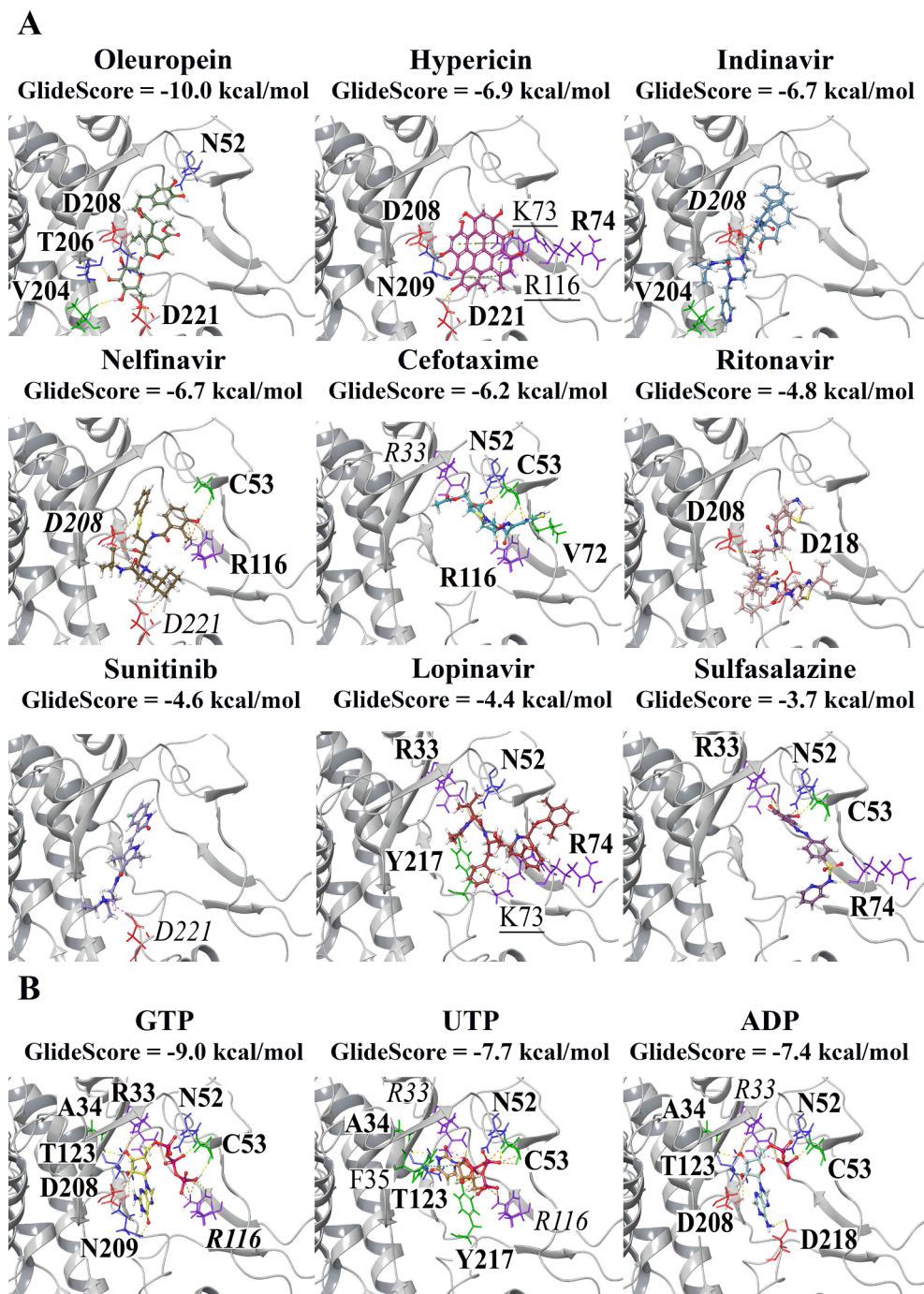
Fig. 3. Molecular docking results of nine compounds and the control ligands. A) Oleuropein, cefotaxime, hypericin, indinavir, sulfasalazine, nelfinavir, sunitinib, lopinavir and ritonavir were docked to the NiRAN domain of the conformation representative of cluster 1 from the 10 μs MD simulation trajectory. B) The docking results of ADP, GTP, and UTP can be seen for the NiRAN domain. The residues that were involved in intermolecular bonds with the ligands are labelled (hydrogen bonds: bold font, salt bridge: italics, π-π interaction: regular font, π-π cation: regular font and underline, hydrogen bonds and salt bridges: bold font and italics, salt bridge and π-π cation: regular font, underline, and italics). The GlideScores (kcal/mol) are provided. Polar residues are coloured blue, positively charged residues are coloured purple, negatively charged residues are coloured red, and hydrophobic residues are coloured green.

3.3. Molecular docking to the representative conformations of the RdRp identified from cluster analysis

The selected 30 compounds and control ligands were docked to the NiRAN domain of the representative structure for cluster 1 (Table S4). Interestingly, the phenolic compounds rutin, oleuropein, and hellicoside from the OliveNet™ database were the top three ligands with the strongest binding affinities. The GlideScores of these ligands were −10.9, −10.0, and −9.9 kcal/mol, respectively. Rutin, oleuropein, and hellicoside predominantly formed hydrogen bonds with the residues of

the NiRAN domain, and hellicoside also formed a π-π cation with the amino acid R116. The control compounds ADP, UTP, and GTP had GlideScores between −7.4 and −9.0 kcal/mol (Fig. 3).

Blind docking was performed on the representative structure for cluster 1 using the control compounds and selection of 30 ligands (Table S5). Several poses of ADP (n = 3), UTP (n = 6), and GTP (n = 7) were predicted to be positioned in the NiRAN domain (Fig. 4). Conversely, the ligands sunitinib, tobramycin, hellicoside, rutin, SRT1720, and hypericin were predicted to bind away from this region and had no poses within the NiRAN site.
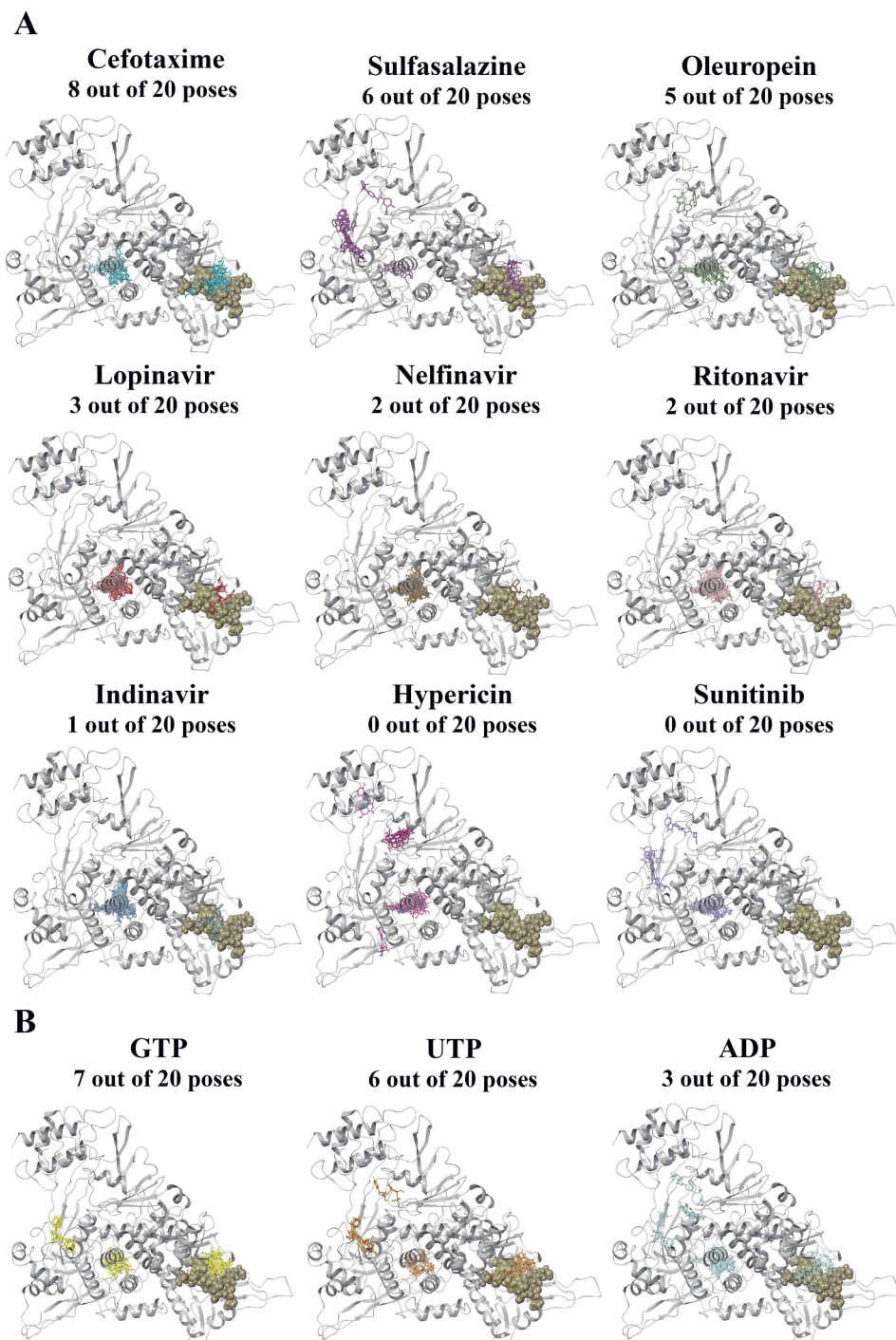
Fig. 4. Blind docking results of nine compounds and the control ligands for the conformation representative of cluster 1 from the 10 μs MD simulation trajectory. A) The blind docking results for hypericin, oleuropein, lopinavir, sunitinib, cefotaxime, ritonavir, sulfasalazine, indinavir and nelfinavir are shown. B) The blind docking results for ADP, GTP, and UTP are also depicted. The number of poses that were present in the NiRAN domain are provided for each ligand. The NiRAN domain is coloured tan and RdRp chain is coloured silver.

In addition to the control compounds, nine ligands with a range of binding affinities were selected and were docked to the conformations of the NiRAN domain that were assigned to clusters 2 to 6 for comparison (Table S6). They were indinavir, ritonavir, nelfinavir, sulfasalazine, lopinavir, hypericin, oleuropein, cefotaxime, and sunitinib. There were differences in the GlideScores for each cluster, as well as the intermolecular bonds that were formed between the ligands and the protein residues. The amino acids that participated in hydrogen bond interactions (maximum distance of 2.8 Å) with each ligand are described in Table 2.

When examining the intermolecular bonds that were present between the ligands and the protein structures for each cluster, it was evident that several residues formed part of the NiRAN domain and N-terminal β-hairpin structure. To compare this region in the representative conformations assigned to the clusters, protein structure alignment was performed using the cluster 1 structure as the reference. The RMSD value for cluster 2 was 2.315 Å, 1.970 Å for cluster 3, 2.313 Å for cluster 4, 1.903 Å for cluster 5, and 2.117 Å for cluster 6.

The RMSD values of the amino acids in the NiRAN domain and nearby β-hairpin were also evaluated. Greater RMSD values were observed for the residues in the conformations corresponding to clusters 2, 4 and 6. The larger RMSD values were mainly associated with residues

Table 2

Hydrogen bond interactions that were formed between the ligands and each conformation representative of the clusters identified from the 10 µs MD simulation trajectory.

Ligands	Cluster 1	Cluster 2	Cluster 3	Cluster 4	Cluster 5	Cluster 6
Indinavir	- V204 - D208	- G220 - N79	- D221 - N209 - R116	- G220 - D221	- Y728 - D36 - D208 - Y217	- D208
Ritonavir	- D218 - D208	- N79 - G220 - K73 - N209	- D218	No H-bonds	- K73 - R116 - N209 - D208	- R733 - D208
Nelfinavir	- D208 - R116 - C53	- R74 - D221	- R116 - N209 - D218	- T51 - C53	- N209 - D208 - D218 - E83	- R33 - Y217 - D208
Sulfasalazine	- R33 - N52 - C53 - R74	- D208 - R74	- D221 - R116	- V31 - R74	- D208 - N209 - T51 - C53	- C53 - D221 - G220
Lopinavir	- N52 - R33 - Y217 - R74	- T51 - K73 - D221	- D208 - N209 - R116 - T76	- K50	- V204 - D208 - N209	- E83 - R116 - D218
Hypericin	- D208 - N209 - D221 - R74	- N209 - D221 - E83 - K73	- N209 - T76 - E83	- N209 - D218 - K73	- D218 - L49	- N209 - D221 - D218
Oleuropein	- N209 - D208 - T206 - V204 - D221 - N52	- L49 - D208 - G220 - R74 - N79	- D218 - E83 - T120 - N52 - R33 - K73	- D36 - S236 - Y732 - C53 - D218 - N209	- D218 - V204 - D208 - N209 - R33 - N52 - K73	- D221 - N209 - D208 - Y217 - T120
Cefotaxime	- N52 - R116 - C53 - V72	- T51 - N79 - D221	- C53 - R116 - N209 - D208	- K50 - R33 - S68 - V72	- N52 - K73 - R74 - R116	- A34 - R33
Sunitinib	No H-bonds	- T51 - R74	- D208 - N209	- R74 - E58	- N209	- D208 - D218 - Y217
GTP	- C53 - N52 - R116 - R33 - A34 - T123 - D208 - N209	- D36 - K73 - R116	- H75 - R74 - K73 - C53 - N52	- E58 - C54 - T51 - D218 - N79	- R74 - R116 - D218 - C53 - N52 - R33 - N209	- D208 - D218 - D221 - K73 - V72 - R116
UTP	- Y217 - C53 - N52 - A34 - T123	No H-bonds	- H75 - R74 - K73 - C53 - R33	- V72 - K50 - N64 - R33 - V31	- Y217 - N209 - D208 - R33 - R116 - N52 - C53	- D208 - R116 - R33 - N52 - C53
ADP	- D218 - T123 - D208 - A34 - N52 - C53	No H-bonds	- D208 - R33	- R74 - R33 - V31 - L65	- D208 - N209 - Y217 - N52 - C53 - D36	- D208 - R33

K50 to Y69, and K103 to P112. The conformations corresponding to clusters 3 and 5 were found to be more similar to cluster 1. As aforementioned, cluster 1, 3 and 5 were prominent towards the end of the 10 μs MD simulation trajectory and may represent conformations of the stable ordered N-terminal region of the RdRp. Moreover, differences were observed in the RMSD values for several residues that formed intermolecular bonds with the ligands and this was more noticeable in the conformations corresponding to clusters 2, 4 and 6 (Table S7). The conformational changes that occur in this region over the course of the trajectory and the flexibility of some of the residues may consequently be contributing to the differences seen in the binding affinities of the compounds and intermolecular bonds that are formed.

Oleuropein was found to consistently bind strongly to the conformations corresponding to each cluster and was selected as a potential lead compound. The GlideScore for the cluster 1 structure was −10.0 kcal/mol and oleuropein predominantly formed hydrogen bonds with the protein residues including N209, D208, T206, V204, D221, and N52. Due to there being missing residues in the NiRAN domain of the 6 M71 structure that was originally obtained from the RCSB PDB, oleuropein and the control compounds were also docked to the RdRp chain of the cryo-EM replication-transcription complex that was determined by Chen et al (PDB ID: 6XEZ) [13]. When comparing the NiRAN domain of the 6XEZ cryo-EM structure to the conformation that was representative of cluster 1, the RMSD was found to be 1.945 Å and the RMSD values of the amino acids can be found in the Supplementary Information (Table S8). Oleuropein had a GlideScore of −8.1 kcal/mol and hydrogen bonds were present with residues R116, N52, N209, Y217, D218, and K73. Most notably, the hydroxyl groups of oleuropein were predominantly involved in hydrogen bonding. The GlideScores of GTP, ADP, and UTP were −8.1, −7.1, and −6.9 kcal/mol, respectively. The protein–ligand interactions of oleuropein and the control compounds can be seen in Fig. 5.

The molecular docking results revealed that ADP formed hydrogen bonds with N209 and K50, salt bridges with K73, K50 and R116, as well as a π-π cation with R116. The ADP that was present in the cryo-EM structure formed a π-π interaction with H75, and salt bridges with K73, R116, and K50 (Fig. 1). Blind docking revealed that oleuropein had

eight poses within the NiRAN domain, while ADP had seven poses in this region (Fig. 5, Table S9). Guanosine-5′-triphosphate and UTP had 11 poses and 10 poses positioned in the NiRAN domain, respectively. The 6 M71 and 6XEZ cryo-EM structures that were obtained from the RCSB PDB, and conformation of 6 M71 that was representative of cluster 1 from the 10 μs MD simulation trajectory were also examined using the P2Rank server. In addition to the NiRAN domain being identified as a potential ligand binding site, the results revealed that there were several other pockets that may be potential allosteric sites and this included the nsp12-nsp8 interface region (Table S10).

Oleuropein is the most prominent phenolic compound in *Olea europaea* and belongs to the secoiridoid subclass [47]. Studies have shown that oleuropein exhibits antiviral activity *in vitro* against respiratory syncytial virus and *para*-influenza type 3 virus [48]. The pharmacokinetic profile of oleuropein in humans needs to be investigated further and its use as a potential prophylactic and therapeutic agent has been discussed in the literature [47]. Oleuropein has been screened against SARS-CoV-2 protein targets using *in silico* and *in vitro* methods, namely the spike protein and cysteine proteases [49,50]. In general, polyphenols are being investigated for their antiviral activity against SARS-CoV-2 using a combination of molecular modelling and classical experimental methods. A number of studies have previously examined the role of the hydroxyl groups in the antioxidant activity of polyphenols and structure–activity relationships should also be performed to explore the function of hydroxyl groups in the antiviral activity of these compounds.

4. Conclusion

Overall, molecular docking was used to screen 300 ligands against the NiRAN domain of the SARS-CoV-2 RdRp. A selection of 30 compounds were then further investigated by high stringency blind docking before a final selection of nine potential lead compounds. These were docked to different conformations of the NiRAN domain identified through cluster analysis of a 10 μs MD simulation trajectory. By careful consideration of all analyses, oleuropein was identified as a lead compound. Given that this compound is relatively well-known and investigated, its potential antiviral effects can be relatively easily investigated

Molecular docking to the NiRAN domain and blind docking on the RdRp (PDB ID: 6XEZ)
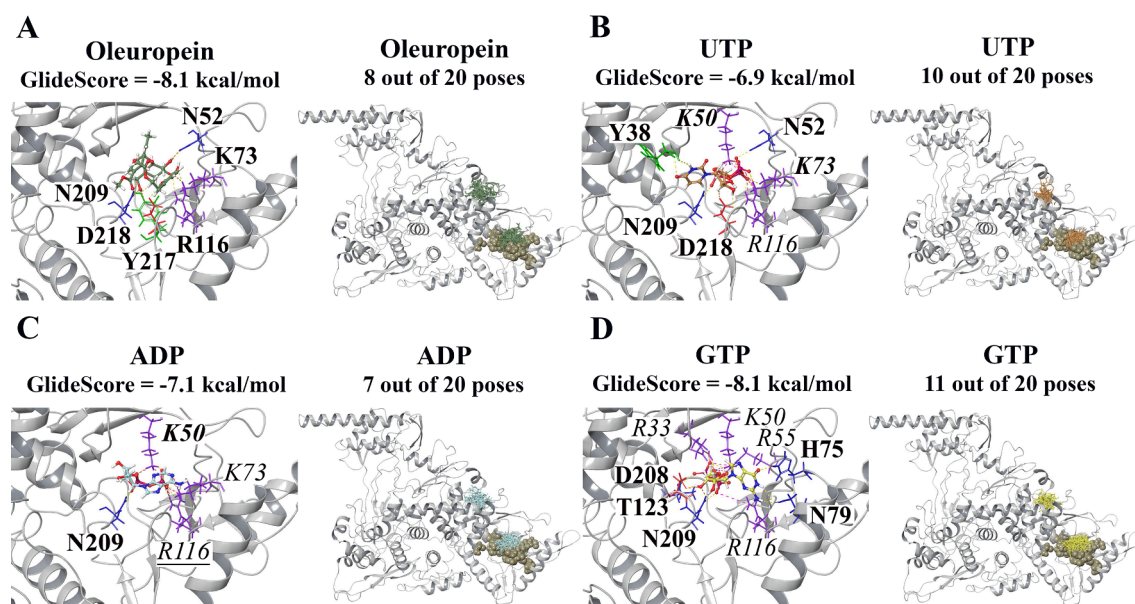

Fig. 5. Molecular docking and blind docking results for the 6XEZ cryo-EM structure. Oleuropein and the control compounds ADP, UTP, and GTP were docked to the NiRAN domain of the 6XEZ cryo-EM structure. The GlideScores (kcal/mol) are provided for each ligand. Blind docking was also performed and the number of poses that were found to be positioned in the NiRAN domain are provided (A-D).

in vitro and *in vivo.*

Author contributions statement

TCK and AH conceptualized the aims and methodology and were involved in supervision. EP performed data analysis, data curation, and was involved in production of the first draft of the manuscript. JL was involved in data analysis and curation and was involved in production of the first draft of the manuscript. HYMH performed formal data analysis and validation. All authors contributed to editing and reviewing the manuscript.

Declaration of Competing Interest

The authors declare the following financial interests/personal relationships which may be considered as potential competing interests: Epigenomic Medicine Program (TCK) is supported financially by McCord Research (Iowa, USA), which has a financial interest in dietary compounds described in this work. However, there is no conflict of interest with respect to the inhibition of the SARS-CoV-2 RNA-dependent RNA polymerase. The remaining co-authors also have no conflicts of interest.

Acknowledgements

Appendix A. Supplementary data

Supplementary data to this article can be found online at https://doi.org/10.1016/j.cplett.2021.138889.

References

[1] Z. Zhu, et al., From SARS and MERS to COVID-19: a brief summary and comparison of severe acute respiratory infections caused by three highly pathogenic human coronaviruses, Respir. Res. 21 (1) (2020) 224.

[2] K. Yuki, M. Fujiogi, S. Koutsogiannaki, *COVID-19 pathophysiology: A review*, Clin. Immunology (Orlando Fla.) 215 (2020) 108427.

[3] R.A. Khailany, M. Safdar, M. Ozaslan, Genomic characterization of a novel SARS-CoV-2, Gene Rep. 19 (2020), p. 100682-100682.

[4] J. Lan, et al., Structure of the SARS-CoV-2 spike receptor-binding domain bound to the ACE2 receptor, Nature 581 (7807) (2020) 215–220.

[5] L. Yan, et al., Cryo-EM structure of an extended SARS-CoV-2 replication and transcription complex reveals an intermediate state in cap synthesis, Cell 184 (1) (2021) 184–193.e10.

[6] Y. Gao, et al., Structure of the RNA-dependent RNA polymerase from COVID-19 virus, Science 368 (6492) (2020) 779.

[7] W. Yin, et al., Structural basis for inhibition of the RNA-dependent RNA polymerase from SARS-CoV-2 by remdesivir, Science 368 (6498) (2020) 1499.

[8] K.C. Lehmann, et al., Discovery of an essential nucleotidylating activity associated with a newly delineated conserved domain in the RNA polymerase-containing protein of all nidoviruses, Nucleic Acids Res. 43 (17) (2015) 8416–8434.

[9] H. Slanina, et al., Coronavirus replication–transcription complex: Vital and selective NMPylation of a conserved site in nsp9 by the NiRAN-RdRp subunit, Proc. Natl. Acad. Sci. 118 (6) (2021), e2022310118.

[10] G.U. Jeong, et al., Therapeutic strategies against COVID-19 and structural characterization of SARS-CoV-2: A review, Front. Microbiol. 11 (1723) (2020).

[11] S. Mahdian, A. Ebrahim-Habibi, M. Zarrabi, Drug repurposing using computational methods to identify therapeutic options for COVID-19, J. Diabetes & Metabolic Disorders 19 (2) (2020) 691–699.

[12] D.E. Shaw Research. Molecular dynamics simulations related to SARS-CoV-2. 2020; Available from: http://www.deshawresearch.com/resources_sarscov2.html..

[13] J. Chen, et al., Structural basis for helicase-polymerase coupling in the SARS-CoV-2 replication-transcription complex, Cell 182 (6) (2020) 1560–1573.e13.

[14] H.M. Berman, et al., The protein data bank, Nucleic Acids Res. 28 (1) (2000) 235–242.

[15] S. Kim, et al., PubChem 2019 update: improved access to chemical data, Nucleic Acids Res. 47 (D1) (2019) D1102–D1109.

[16] D. Mendez, et al., ChEMBL: towards direct deposition of bioassay data, Nucleic Acids Res. 47 (D1) (2019) D930–D940.

[17] N.P. Bonvino, et al., OliveNet™: a comprehensive library of compounds from Olea europaea. Database: The J. Biol. Databases Curation, 2018. 2018: p. bay016.

[18] G. Madhavi Sastry, et al., Protein and ligand preparation: parameters, protocols, and influence on virtual screening enrichments, J. Comput. Aided Mol. Des. 27 (3) (2013) 221–234.

[19] E. Harder, et al., OPLS3: A force field providing broad coverage of drug-like small molecules and proteins, J. Chem. Theory Comput. 12 (1) (2016) 281–296.

[20] A. Dwivedy, et al., Characterization of the NiRAN domain from RNA-dependent RNA polymerase provides insights into a potential therapeutic target against SARS-CoV-2. bioRxiv, 2021: p. 2021.02.03.429510.

[21] A.E. Cho, et al., Importance of accurate charges in molecular docking: quantum mechanical/molecular mechanical (QM/MM) approach, J. Comput. Chem. 26 (9) (2005) 915–931.

[22] R.A. Friesner, et al., Extra precision glide: docking and scoring incorporating a model of hydrophobic enclosure for protein−ligand complexes, J. Med. Chem. 49 (21) (2006) 6177–6196.

[23] A.D. Bochevarov, et al., Jaguar: A high-performance quantum chemistry software program with strengths in life and materials sciences, Int. J. Quantum Chem. 113 (18) (2013) 2110–2142.

[24] R.B. Murphy, D.M. Philipp, R.A. Friesner, A mixed quantum mechanics/molecular mechanics (QM/MM) method for large-scale modeling of chemistry in protein environments, J. Comput. Chem. 21 (16) (2000) 1442–1457.

[25] R. Krivák, D. Hoksza, P2Rank: machine learning based tool for rapid and accurate prediction of ligand binding sites from protein structure, J. Cheminformatics 10 (1) (2018) 39.

[26] S. Dallakyan, A. Olson, Small-molecule library screening by docking with PyRx, Methods Mol. Biol. (Clifton N.J.) 1263 (2015) 243–250.

[27] O. Trott, A.J. Olson, AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading, J. Comput. Chem. 31 (2) (2010) 455–461.

[28] Hypernet Labs. Galileo. https://galileoapp.io/. 2020; Available from: https://galileoapp.io/.

[29] L. Lafayette, et al., Spartan performance and flexibility; An hpc-cloud chimera, in OpenStack Summit 2016: Barcelona.

[30] H.J.C. Berendsen, D. van der Spoel, R. van Drunen, GROMACS: A message-passing parallel molecular dynamics implementation, Comput. Phys. Commun. 91 (1) (1995) 43–56.

[31] M.J. Abraham, et al., GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers, SoftwareX 1–2 (2015) 19–25.

[32] W. Humphrey, A. Dalke, K. Schulten, VMD: visual molecular dynamics, J Mol Graph 14 (1) (1996), pp. 33–8, 27–8.

[33] X. Daura, et al., Peptide folding: When simulation meets experiment, Angew. Chem. Int. Ed. 38 (1–2) (1999) 236–240.

[34] E. Weisberg, et al., Repurposing of kinase inhibitors for treatment of COVID-19, Pharm. Res. 37 (9) (2020) 167.

[35] G. Bolcato, et al., Targeting the coronavirus SARS-CoV-2: computational insights into the mechanism of action of the protease inhibitors lopinavir, ritonavir and nelfinavir, Sci. Rep. 10 (1) (2020) 20927.

[36] K.-T. Choy, et al., Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro, Antiviral Res. 178 (2020) 104786.

[37] P.W. Horby, et al., Lopinavir-ritonavir in patients admitted to hospital with COVID-19 (RECOVERY): a randomised, controlled, open-label, platform trial, The Lancet 396 (10259) (2020) 1345–1352.

[38] A.S. Ginsburg, K.P. Klugman, COVID-19 pneumonia and the appropriate use of antibiotics, The Lancet Global Health 8 (12) (2020) e1453–e1454.

[39] A. Khalil, D. Tazeddinova, The upshot of Polyphenolic compounds on immunity amid COVID-19 pandemic and other emerging communicable diseases: An appraisal, Natural Products and Bioprospecting 10 (6) (2020) 411–429.

[40] J.M. Jacobson, et al., Pharmacokinetics, safety, and antiviral effects of hypericin, a derivative of St. John's wort plant, in patients with chronic hepatitis C virus infection, Antimicrob. Agents Chemother. 45 (2) (2001) 517–524.

[41] E. Patridge, et al., An analysis of FDA-approved drugs: natural products and their derivatives, Drug Discovery Today 21 (2) (2016) 204–207.

[42] O.M. Ogunyemi, et al., Alkaloids and flavonoids from African phytochemicals as potential inhibitors of SARS-Cov-2 RNA-dependent RNA polymerase: an in silico perspective, Antiviral Chem. Chemother. 28 (2020), 2040206620984076-2040206620984076.

[43] S. Vardhan, S.K. Sahoo, In silico ADMET and molecular docking study on searching potential inhibitors from limonoids and triterpenoids for COVID-19, Comput. Biol. Med. 124 (2020) 103936.

[44] S. Singh, et al., Plant-derived natural polyphenols as potential antiviral drugs against SARS-CoV-2 via RNA-dependent RNA polymerase (RdRp) inhibition: an in-silico analysis, J. Biomol. Structure Dynamics (2020) 1–16.

[45] W. Evangelista Falcon, et al., Ensemble docking in drug discovery: how many protein configurations from molecular dynamics simulations are needed to reproduce known ligand binding? J. Phys. Chem. B 123 (25) (2019) 5189–5195.

[46] R.N. Kirchdoerfer, A.B. Ward, Structure of the SARS-CoV nsp12 polymerase bound to nsp7 and nsp8 co-factors, Nat. Commun. 10 (1) (2019) 2342.

[47] C. Nediani, et al., Oleuropein, a bioactive compound from Olea europaea L., as a potential preventive and therapeutic agent in non-communicable diseases. Antioxidants (Basel, Switzerland), 2019. 8(12): p. 578.

[48] S.-C. Ma, et al., In vitro evaluation of secoiridoid glucosides from the fruits of Ligustrum lucidum as antiviral agents, Chem. Pharm. Bull. 49 (11) (2001) 1471–1473.

[49] A. Romeo, F. Iacovelli, M. Falconi, Targeting the SARS-CoV-2 spike glycoprotein prefusion conformation: virtual screening and molecular dynamics simulations applied to the identification of potential fusion inhibitors, Virus Res. 286 (2020) 198068.

[50] E. Pitsillou, et al., Identification of small molecule inhibitors of the deubiquitinating activity of the SARS-CoV-2 papain-like protease: in silico molecular docking studies and in vitro enzymatic activity assay 2020. 8(1171).