doi.org/10.26434/chemrxiv.12497693.v1

High Performance Computing Prediction of Potential Natural Product Inhibitors of SARS-CoV-2 Key Targets

Kendall Byler, Joseph Landman, Jerome Baudry

Submitted date: 17/06/2020 • Posted date: 18/06/2020

Citation information: Byler, Kendall; Landman, Joseph; Baudry, Jerome (2020): High Performance Computing Prediction of Potential Natural Product Inhibitors of SARS-CoV-2 Key Targets. ChemRxiv. Preprint. https://doi.org/10.26434/chemrxiv.12497693.v1

This work describes using a supercomputer to perform virtual screening of natural products and natural products derivatives against several conformations of 3 proteins of SARS-CoC-2 : papain-like protease, main protease and spike protein. We analyze the common chemical features of the top molecules predicted to bind and describe the pharmacophores responsible for The University of Ala predicted binding.

High Performance Computing Prediction of Potential Natural Product Inhibitors of SARS-CoV-2 Key Targets

Kendall Byler[%], Joseph Landman[#], Jerome Baudry[%*]

%: The University of Alabama in Huntsville, Department of Biological Sciences, Huntsville, AL, 35899
#: Hewlett Packard Enterprise, HPC MCS AI, Canton, MI, 48188

* Corresponding author: jerome.baudry@uah.edu

Abstract: An ensemble-docking of 50,000 natural products on a supercomputer has been performed against the papain-like protease, the main protease and the spike protein of the SARS-CoV-2 virus. The top compounds predicted to bind specifically to these protein targets are analyzed to identify common pharmacophore features. The functional groups more likely to lead to target engagement of these viral proteins are described and feature hydrophobic/resonant cores surrounded by hydrogen bonding capacities at specific locations. This work identifies natural products for immediate testing and suggests structural elements for anti COVID-19 drug development and screening.

Introduction

The fight against COVID-19 has led to an unprecedentedly large-scale effort to discover pharmaceuticals that are active against the virus or that cure or mitigate the ill health effects of the SARS-CoV-2 infection. This endeavor has been so far mostly following a strategy of repurposing existing drugs, in the hope that existing pharmaceuticals will prove efficient against the disease. This repurposing approach is powerful and can help accelerate the availability of efficient pharmaceuticals. These existing pharmaceuticals have known safety profiles and are therefore largely "de-risked". This approach not only speeds up the drug discovery process, but also limits its cost, by bypassing a large part of the development of new chemical entities (NCEs). In non-crisis, "normal" times, however, repurposing/repositioning efforts can still be lengthy processes, and typically take 1-6 years of preclinical and clinical research to get FDA approval for drugs already in use. And getting the licensing for the novel use of repurposed drugs can tack on as many as two years on the front end. It is expected and hoped that the approval of new drugs against COVID-19 will be expedited by the regulatory agencies without, naturally, sacrificing the quality and safety of the drug approval process. Repurposing is, however, not the only approach to drug discovery against COVID-19. New chemical entities are likely to be needed, even if the repurposing effort is successful. Repurposed drugs may be active against the virus itself but may not be as efficient as against the more traditional targets they were developed against. In fact, relatively few drugs have been successfully used in repurposed indications[1]. On the other hand, NCEs, even if they represent the vast majority of current drugs, are notoriously time-consuming and expensive to develop. The process also suffers from a very low success rate (contributing to both development time and cost) as most NCEs are found to fail in later-stage pre-clinical and clinical stages of drug development because of unfavorable chemical profiles and their incompatibility with the necessary biological balance between efficacy and safety of biochemicals.

This present work aims to accelerate the discovery of novel pharmaceutical approaches to the COVID-19 crisis by minimizing the issues of repurposing existing pharmaceuticals (very few efficient drugs originate from repurposing) and of developing NCEs (very long development times and very high cost). We aim to identify molecules that are targeting important and likely druggable proteins from the SARS-CoV-2 proteome. Our strategy is to work with natural product (NP) compounds, as represented by plant and animal secondary metabolites, and with chemical that are minor synthetic modifications of natural compounds. Natural products have chemical structures that have been subjected to long periods of biosynthetic natural selection in order to perform highly selective functions and have been the source of most of the known drug compounds leading into the last century. The development of synthetic drugs to treat disease, culminating in the introduction of rational drug design, has been one of the crowning achievements of modern scientific advancement. Despite this progress, compounds that result from the combined efforts of *de novo* design and synthesis and combinatorial

chemistry in modern drug discovery typically suffer problems with toxicity, off-target interactions, ADME (Absorption, Distribution, Metabolism and Excretion), and pharmacokinetic/pharmacodynamics. Natural product structures provide a substantially large and chemically diverse pool of biologically active compounds that have been fit to purpose through natural selection. As such, these compounds represent structures that frequently avoid many of the adverse biological effects observed with synthetic compounds. Furthermore, in periods of relative drought in terms of newly developed synthetic drug compounds, natural products periodically see a resurgence in popularity for drug discovery. With the number of new drug approvals (NDAs) and biologics license applications (BLAs) by the Food and Drug Administration in steady decline over the past several years (taken together, these were down 19% in 2019[2]), natural products are in an ever more significant position in the drug discovery and development process.

There are quite a number of examples of natural products being used in experimental screens of the other coronaviruses that induce deadly pneumonias in humans, such as SARS-CoV and MERS-CoV. For instance, lycorine, found in the red spider lily, *Lycoris radiata*, was reported to be a nanomolar inducer of virally-induced cytopathy of SARS-CoV in a high throughput screen of 200 plant extracts[3]. And some low-micromolar natural product inhibitors of SARS-CoV papain-like protease, such as tanshinones[4] and geranylated flavonoids[5], have also been reported. Natural products were most recently used in a small virtual screen against human Heat Shock Protein A5 substrate-binding domain β[6], which was considered at the time to be one of the sites used by the SARS-CoV-2 spike protein as a recognition site, to prevent viral attachment to the host cell.

Structure-based drug discovery and docking.
A powerful approach to develop drug candidates that are potent and efficient is to work from knowledge of the three-dimensional structure of the protein targets of importance in a particular disease and to identify small molecules that exhibit the desired chemical features needed to bind, or 'dock', to the proteins of interest. These molecular docking calculations essentially predict or estimate the binding free energy of a chemical in a protein. This involves computer programs that position a small molecule in the binding site of the target and evaluate an interaction energy between the small molecule and its protein environment. Traditionally, a structure-based drug design approach uses an X-ray crystal structure for *in silico* screening (*i.e.* docking). A recent application of this approach using chemicals from plants has identified already interesting potential modulators of SARS-CoV-2's main protease[7]. However, a growing body of evidence indicates that small molecules bind to a specific conformation dynamically sampled by the protein at ambient temperatures, a concept known as "conformational selection" (the chemical selects a conformation of the protein to bind to, among a diverse ensemble of protein substructures). The computational equivalent to conformational selection is called "ensemble docking", in which one uses several structures of a target to give small molecules the opportunity to bind to different protein target conformations. Generating multiple conformations of these macromolecules is computationally intensive and benefits from parallelized and docking software. This approach has been successfully used in our laboratory in the discovery of new molecular effectors against a variety of targets[8,9], and in particular to inhibit protein:protein complexes[10], and the superiority of ensemble docking over using a single crystal structure-based docking as has been described[11,12]. A recent use of this approach has suggested possible repurposing against the SARS-CoV-2's spike protein[13]. Simulating *in silico* such a conformational selection mechanism is much more computationally expensive than what is required for using a single structure, but contemporary massively parallel supercomputers are allowing this approach to become used in drug discovery campaigns. This work uses such a supercomputer – an HPW Cray XC50 supercomputer – to facilitate NP-based drug discovery.

Material and Methods
Database of natural products
The COlleCtion of Open NatUral producTs (COCONUT, version 4) database[14] is a curated database of 423,706 unique natural products compounds, as well as possibly synthetic variations of natural products, assembled from 117 other natural products databases cited in scientific literature since the year 2000. It contains annotations of molecular weight, atom and bond counts, source database, logP, polarizability, polar surface area and an index of natural product likeness[15,16] for each compound. The current version (4) of the COCONUT database was downloaded as a single SD file from the Zenodo website and split into 423,706 individual SD files. The 2D structure files were then converted to 3D structures using Open Babel 2.4.1[17]. All of the 3D SD files were then loaded into an MDB database using MOE 2019.01[18] and compounds with molecular weights greater than 600 a.u. filtered out, yielding a set of 340,413 structures. The resulting filtered database was clustered by molecular weight and NPL score (an

index of natural product likeness)[16], descriptors then used to select a diverse subset of 49,977 compounds from among the 200 clusters. These structures were then converted to PDBQT files with AutoDockTools (MGLTools 1.5.6)[19,20] in preparation for molecular docking.

Protein targets.
The proteases in SARS-CoV-2 are targets of choice in the search for treatments for COVID-19, as there are known HIV-1 and Hepatitis C protease inhibitors in use, indicating the druggability of these targets. The interactions between these inhibitors and the active sites of their target proteases are highly specific and the active sites, themselves, are highly conserved among viral families. The enzyme that cleaves the viral polypeptide into its component non-structural proteins is referred to as the main protease (Mpro) and is essential for viral replication. It is a homodimer with catalytic sites facing away from each other, and away from the dimer interface region. It is highly similar in sequence to the other 3-chymotrypsin-like proteases (3CLpro) found in other coronaviruses, such as SARS-CoV (96% identity)[21], MERS, and porcine epidemic diarrhea virus (PEDV). The main proteases in *Coronaviridae* all possess catalytic sites composed of a dyad of cysteine and histidine that work in concert to cleave peptide chains at specific sites, residing in a groove that selectively interacts with primarily aromatic side chains – in much the same way chymotrypsin performs the same function, although using a catalytic triad.
Along with the main protease, the papain-like protease (PLpro) is an essential enzyme in SARS-Co This work uses such a powerful supercomputer to facilitate NP-based discovery, the Cray/HPE Sentinel -2 that lies in the N-terminal region of the much larger multidomain NSP3, and is flanked by other catalytic sites that perform several functions, such as the nearby ADP-ribose-1″-phosphatase. The crystal structure used in this present work is this N-terminal region that contains the protease and two ubiquitin binding sites involved in the de-ubiquitinylation of host recognition receptors that look for pathogenic macromolecules, interfering with the innate immune response. innate immune response[22]. The catalytic site is that of a classic cysteine protease, consisting of a cysteine-histidine-aspartate triad, which resides at the junction of the "palm" and the "thumb" of the larger "hand" domain.

In addition to these two enzymes, we are also targeting in this work the so-called 'spike protein'. This protein studs the surface of the coronavirus membrane and gives the virus the appearance of a crown, or corona, under the electron microscope – hence the name *coronavirus*. It functions to anchor the virus to the angiotensin-converting enzyme 2 (ACE2) on host cell membranes to start the membrane fusion process and so invade the host cell. The spike protein is also the primary viral feature targeted by host cell antibodies. In order to evade recognition by host immune systems, this protein is partially glycosylated. Since the process of this glycosylation occurs within the host cell and some amount of serendipity is involved in the process, the pattern of glycosylation is not the same for every newly-constructed spike protein. This results in each viral coat looking slightly different to the host immune system, so that one type of antibody may not recognize the same virus twice because of variations in spike glycoslation[23,24].

Protein target conformational ensembles
Structures for the spike protein, main protease and papain-like protease were obtained from the freely available molecular dynamics simulation coordinate files from the Oak Ridge National Laboratory & collaborating laboratories SARS-CoV-2 webpage (https://coronavirus-hpc.ornl.gov/data/). These structures are representative conformations clustered from the trajectories of 100 ns restrained temperature replica exchange molecular dynamics simulations of the proteins[13] using GROMACS 4.5[25], and starting from the Protein Databank Bank entries 6Y2E (main protease), 6W41 (spike protein) and 6W9C (papain-like protease). The ten most populated conformations from each set of clustered structures were used as an ensemble for molecular docking, with each structure being prepared as individual PDBQT files using MGLTools 1.5.6.

Molecular Docking
The docking code used in this study was AutoDock Vina[26] 1.1.2, which was compiled from source for use on the Sentinel system in a massively parallel manner. Docking boxes were constructed using the AutoDock plugin for Open Source PyMOL[27] 1.8.4.0. In the case of the spike protein, a large box containing the entire ACE2 receptor binding domain (RBD), referencing PDB ID: 6VW1, was constructed. For the two proteases, smaller boxes were located around their catalytic sites, with a larger region including the S1 and S2 sites in the binding groove of the main protease. Scripts were used to submit large numbers of individual docking jobs to the Sentinel queue with a Vina exhaustiveness value of 10 and a maximum number of returned poses per docking run of 10.

Supercomputing
This work was performed on Sentinel, an HPE Cray XC50 single cabinet supercomputer located in the Microsoft Azure public cloud data center. Sentinel is a 48-node system, featuring 1920 physical Intel Skylake cores operating at 2.4GHz with 192GB ram per node.

The XC50 utilizes the Aries interconnect[28] in a Dragonfly topology. These systems have SMT/HT enabled, resulting in 3840 effective CPU cores. A shared 612 TB Cray HPE ClusterStor-based parallel file system is mounted on every node. The Altair PBS Pro scheduler is used for workload distribution.

Autodock Vina 1.1.2 was obtained from the download site (http://vina.scripps.edu/download.html). The code was updated to use the C++ Boost library (https://www.boost.org), version 1.65. Autodock Vina was statically compiled on an Ubuntu 18.04 Linux system utilizing the GCC 8.4.0-1ubuntu1 compiler. The Makefile C_OPTIONS variable, specifying the compiler options, was set to

```
-O3 -DNDEBUG -fpermissive -march=skylake
-mtune=skylake      -mavx2      -malign-
data=cacheline
```

Pharmacophore analysis

The top 500 unique natural product poses for each protein target were retained for further evaluation. In each case, the docking scores for the top 500 poses showed no correlation with molecular weight, which indicates no particular bias of the score with respect to the size of the chemicals. Compounds that appeared more than once in the top 500 unique compound lists for Mpro, PLpro and spike protein were considered promiscuous and removed from each list. In addition, compounds that exhibited more than one violation of the Lipinski "drug-likeness" criteria, as calculated with the MOE "drug-like" molecular descriptor, were not included in the final lists of results. The number of compounds that fulfilled the above criteria, *i.e.*, i) appeared in the top 500 docking scores of only one of the three proteins and ii) predicted to be "drug-like" is: 204 compounds for papain-like protease, 232 compounds for the spike protein, and 164 compounds for the main protease. Of these, the top 100 natural products were superimposed on their respective predicted protein targets binding locations in which they were predicted to bind. The compounds that were found to dock to the correct binding site (*i.e.*, in the active sites of the main protease and of the papain-like protease, and on the ACE2/spike protein interface) were kept for pharmacophoric analysis. This analysis was performed using the program MOE and identified the chemical features that at least 25% of the natural products exhibited in common in corresponding binding sites.

Results and Discussion

Performance of Vina on Sentinel

A preparation and an initial run script were written that enabled users to create and enqueue jobs. Each Vina calculation required less than 90 seconds to run on a single processor. With a test set of 12,000 calculations, running these jobs serially would have required approximately 1.08×10^6 seconds, approximately 12.5 days. Running across 10 CPU cores, we did see on average, an 8x speedup over a single CPU core. This parallelism combined with the above updated options, enabled us to complete a single calculation in approximately 6 seconds. For the 12,000 calculations, this would require approximately 72,000 seconds or 20 hours of wall clock time to complete on a single node. Over 48 nodes, we measured approximately 271 seconds for a 1% sample, or 120 runs. Typical supercomputing workloads are long running, parallel calculations, for a single large distributed task. Given that the Autodock Vina code required so few resources per calculation, yet there were many of the calculations, we adapted the supercomputer run environment to enable multiple simultaneous calculation runs per node. After this alteration, and using a modified calculation launch code, we were able to complete all 12,000 calculations in 468 seconds, or approximately 7 minutes, 48 seconds wall clock. This would work out to 25.6 calculations completed per second across the Sentinel XC50 machine. Over the course of 1 day of run time, this could provide approximately 2.22×10^6 calculations per wall clock day. With around 20 protein targets, running 20,000 ligand compounds against these targets would require about 4×10^5 calculations. This approach would require approximately 4.33 wall clock hours, or 15,600 seconds. In comparison with the original calculation mechanism provided by the UAH team, this would have required roughly 3.6×10^7 seconds, or slightly more than a single CPU-year of calculation time. Running on 10 processor cores, this would have reduced the original run mechanism to 4.5×10^6 seconds, or 1.7 months of run time. This work was faster by about 288 times. There is room for additional improvement, not simply in optimizing Vina, but environmental optimization, algorithmic shifts, and work with additional tools to reduce the search space, but these improvements were not needed to achieve the desired calculations in time and will be developed in the future.

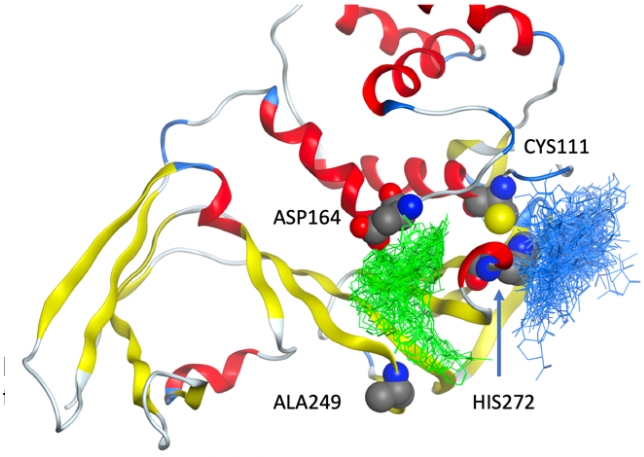
Docking results.

The docking outputs against each set of target conformations were combined and ranked by docking score. Duplicate instances of each compound were removed from the pose list, retaining only the poses with the best docking score for each compound. Computational approaches to describing physicochemical and structural properties that are

common to natural product compounds, or those that distinguish them from other classes of compounds - in a manner similar to drug-likeness - have been described in the literature over the past couple of decades. One such description of natural-product-likeness that appears in the COCONUT database derives from common chemical substructures and atom-centered fragments among the compounds in the Dictionary of Natural Products[29]. The resulting NP-likeness score[15] ranges from a value of -5 to one of +5, with +5 being the most structurally similar to compounds in the DNP. The use of such a score was suggested by the authors to be useful in the virtual screening of large chemical databases. Here, we used this score as a descriptor for clustering our diverse subset of the COCONUT database into classes of natural products (SI tables). In our diverse subset, these values range from -3.1 to 4.7, which covers the same range of values found in the complete database. The natural products predicted to be the most promising potential hits against the protein targets in the desired binding sites (see below) have NP-likeness scores that range between -0.8 to +2.9, with a distribution centered about an NP likeness score of 0.5.

Papain-like protease

Figure 1 shows the location of the top 100 compounds fulfilling the criteria above in the papain-like protease target. The compounds can be divided in two clusters, shown respectively in green (40 compounds) and blue (60 compounds) in Figure 1. These two clusters are located on two different sides of the protein, separated by a beta-sheet that is roughly between Asp164 and ALA249 on one side (green cluster), and that incorporates the binding site residues Cys111 and His272 on the other side (blue cluster). The active site of this target corresponds to the location populated by the blue cluster of docked natural products.



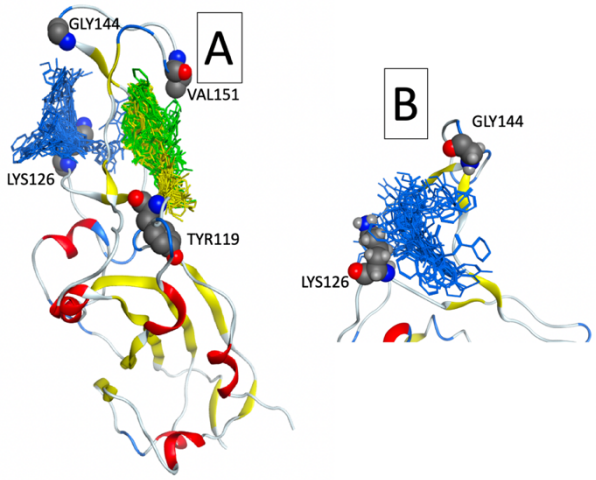


Figure 2. Top-scoring docked natural products on the spike protein.

Figure 3. Cluster of of docked molecules from Figure 2 shown on top of the spike protein (cyan) : ACE2 protein (purple) complex.

Spike protein

Figure 2 shows the location of the docked compounds in the spike protein target. As in the case of docking in the papain-like protease, two clusters of compound locations were obtained, located on two different sides of the loop that interacts with the ACE2 receptor: the blue cluster (24 compounds), located in a region roughly defined as between Glu144 and Lys216, and

the green/yellow cluster (73 compounds), located on the other side of the loop, in a region roughly defined by the residues Tyr119 and Val151. Figure 3 shows the interface between the spike protein (cyan protein backbone) and the ACE2 receptor (purple protein backbone). The blue cluster of docked compounds shown on Figure 2 is located between the two proteins of the spike:ACE2 receptor complex, and hence is likely to disrupt the protein-protein interactions between the two. On the other hand, the green/yellow cluster does not exhibit any molecule that would lead to steric clashes between the spike and the ACE2 proteins.

Main protease

Figure 4 shows the locations of the docked compounds in the main protease target. In contrast to the docked positions in the spike and papain-like protein targets, the docking locations in the main protease exhibit several different potential docking locations of the natural products (colored greed, blue, yellow, pink and orange on Figure 4). The green cluster (41 molecules) is located the closest to the catalytic site CYS145, and compounds binding in that location would potentially act as inhibitor of the enzyme by blocking or hindering substrate access.

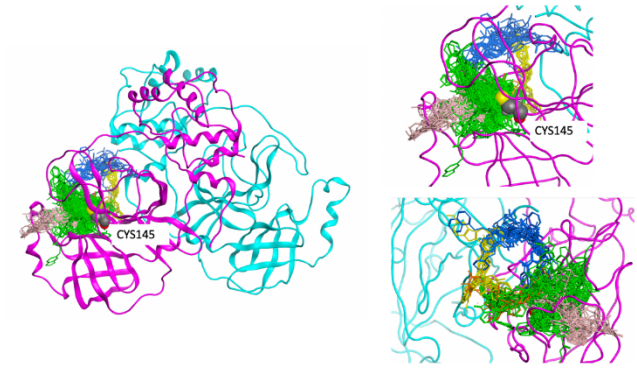


Figure 4. Top-scoring docked natural products in the main protease.

Pharmacophore analysis

The compounds binding to the desired locations in the proteins, as shown in the figures above, were analyzed to identify common pharmacophore features using the program MOE as described in Methods.

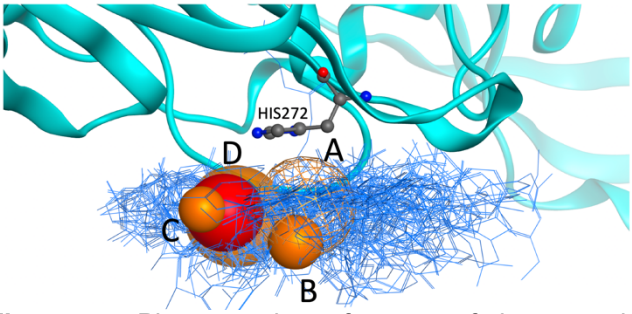


Figure 5a. Pharmacophore features of the natural compounds predicted to bind to the papain-like protease's active site and corresponding to the blue cluster of molecules on Figure 1.

Papain-like protease

A pharmacophore analysis of the commonalities between the compounds docked in the active site of the papain-like protease is shown in Figure 5a. There are relatively few common regions identified as common between the compounds: a large region (orange wireframe sphere labeled "A" in Figure 5a) represents a resonant and/or hydrophobic group present within that sphere in 45% of the natural products predicted to bind well in the protein's active site. A region (orange solid sphere labeled "B" on Figure 5a) indicates the perpendicular to the resonant plane found in 29% of the binding natural products. The other perpendicular to that resonant plane of region A would be the HIS 272 residue, suggesting that resonant-resonant interactions between this histidine residue and ligands are important for binding of natural products in this active site. Another hydrophobic/resonant region, indicated by the red solid sphere in Figure 5a, is found in 42% of the natural products binding in the site. The perpendiculars to the resonant plane, labelled C and D in Figure 5a, and present in 29% and 47% of the ligands, respectively, are also common features of the natural products predicted to be potential ligands and inhibitors of that protein. These commonalities of resonant / hydrophobic moieties found in ligands does not mean that the natural products predicted to bind well in that protein regions must be entirely hydrophobic, and indeed they are not.

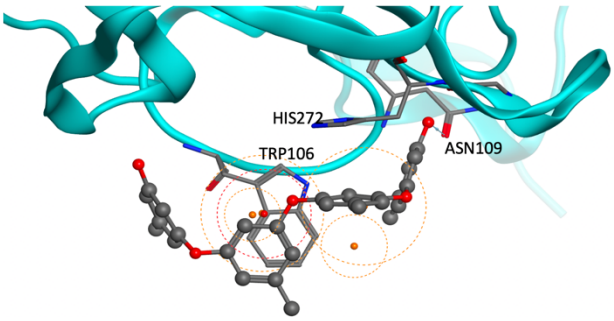


Figure 5b. Example of natural products docked to the papain-like protease's active site and fulfilling the pharmacophore binding motif of Figure 5a.

Figure 5b shows an example of a top-scoring natural product in the binding site of the papain-like protease active site. This compound, tetraorcinol A, is found in the coral-associated fungus *Aspergilus versicolor* LCJ-5-4 and has been reported to be a free radical scavenger[30]. Tetraorcinol A fits the pharmacophore description of Figure 5a, with a resonant-resonant interaction between the central benzene of the natural compound and Trp106 from the enzyme, with the location of the second benzene ring and of its perpendicular in the other resonant orange pharmacophore / normal to the resonant plane regions. In addition, this compound has two terminal hydroxyl groups, one of them making a hydrogen bond with the backbone's carbonyl of Asn109.

Spike protein

Figure 6a shows the most common pharmacophore features of the compounds predicted to bind well on the spike protein's regions corresponding to the blue

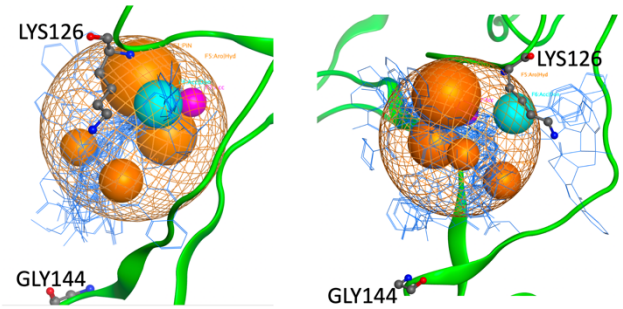


Figure 6a. pharmacophore features of the natural compounds predicted to bind to the spike protein and corresponding to the blue cluster on Figure 3. Right side is rotated 45 degrees from Left side.

cluster of Figures 2 & 3, *i.e.*, predicted to disturb the interactions between the spike protein and the ACE2 receptor. As in the case of the papain-like protease results, a large region (orange wireframe sphere in Figure 6a) represents a resonant and/or hydrophobic group present within that sphere in 96% of the compounds. Inside that large region, 4 spheres (solid orange sphere on Figure 6a) contain the perpendiculars to resonant rings of the compounds (of 83% of the compounds for the larger orange solid sphere in Figure 6a, and 79%, 46% and 29% for the three next large orange spheres, respectively.). In addition to the resonant/hydrophobic features, 39% of the compounds exhibit a hydrogen bond donor or a hydrogen bond acceptor within the location of the cyan solid sphere, and 26% of the compounds exhibit a hydrogen bond acceptor within the location of the purple sphere. Figure 6b shows the details of a top-scoring natural product that fits this pharmacophore description. This compound, prekinamycin, belonging to the kinamycin class of diazofluorene antitumor antibiotics first isolated from the bacteria *Streptomyces murayamaensis*[31–34], exhibits hydrogen bonds between its diazo moiety and the amine of Lys216, and between one of its carbonyls and the amine group of Arg122 (Figure 6b, *left*). Prekinamycin also fulfills the hydrophobic pharmacophore regions, making resonant interactions with Phe124 and Tyr141, and surrounded in a hydrophobic environment defined by the residues Leu123, Pro159 and Tyr157.

Main protease

Figure 7 shows the most common pharmacophore features of the compounds predicted to bind well in the main protein's catalytic region (close to CYS145). A large region of the binding site corresponds to a large

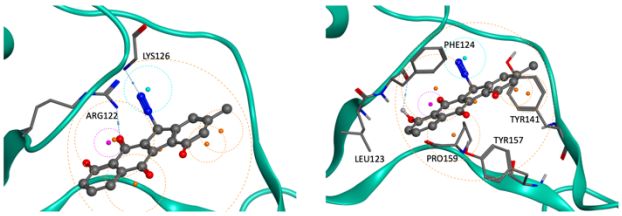


Figure 6b. Example of natural product docked in the spike protein and fulfilling the pharmacophore binding motif of Figure 6a.

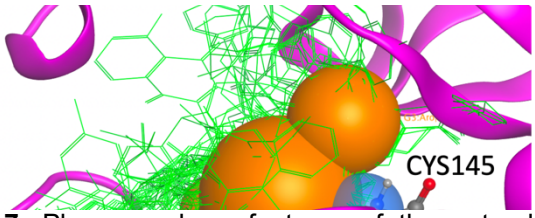


Figure 7. Pharmacophore features of the natural compounds predicted to bind to the main protease active site and corresponding to the green cluster on Figure 4.

region (largest orange solid sphere in Figure 7) representing a resonant and/or hydrophobic group in 58% of the compounds. Next to that region, another region (second largest orange solid sphere in Figure 7) contains other resonant and/or hydrophobic groups for 53% of the natural products predicted to bind in that active site. The semi-transparent blue sphere contains hydrogen bond acceptors or donors for 47% of the natural products binding in the active site, and 34% of the compounds are contained in a region perpendicular to a resonant ring in the third largest orange solid sphere. In addition to the resonant/hydrophobic features, 39% of the compounds exhibit a hydrogen bond donor or a hydrogen bond acceptor within the location of the cyan solid sphere, and 26% of the compounds exhibit a hydrogen bond acceptor within the location of the purple sphere.

Chemical space of the computational hits

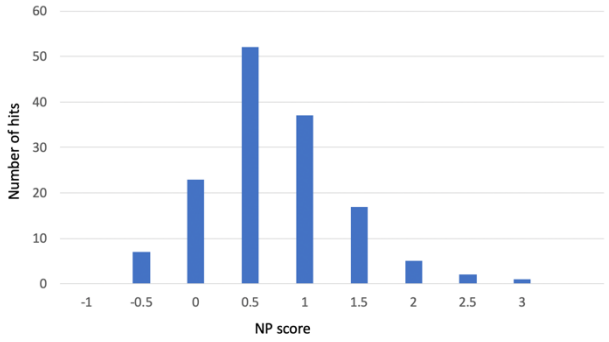


Figure 8. Distribution of NP scores in the top scoring natural products predicted to bind at the desired sites.

The distribution of NP scores in the 144 hits given in SI tables and figures is shown on Figure 8. The distribution of scores across all known natural products exhibit higher populations in the ~0.5-~1 range of scores with the most populated bin of NPL score[15] having a value of 1, with the next most populated bin at an NP value of 0.5. This finding is comparable to what is seen here, albeit in this work the most populated bin is at a NP score of 0.5 and the second most populated bin is for NP score value of 1. The NP scores for the compounds in the entire COCONUT database range from -3.08 to 4.67, (Figure SI-1).

Conclusions

Several compounds from a diverse subset of the largest natural products database currently available have been identified as potential inhibitors of the three most significant targets in the SARS-CoV-2 proteome, CLpro, PLpro and the spike protein receptor-binding domain. The compounds are within the natural product chemical space, are relatively diverse structurally, although they share common chemical features as described in the pharmacophore analysis. The common chemical features suggest that the natural product that are predicted to bind the best to the protein targets exhibit a largely hydrophobic center, with a preference for aromatic moieties, and some hydrogen bond acceptors/donors on the periphery of this resonant core. Many of the top-scoring compounds that fulfill these criteria exhibit a mix of fused rings and heterocyclic rings, *e.g.* (but not limited to) flavonoids, and a relatively high diversity of location and nature of hydrophilic attachments. We list below some of the compounds for which organismal and/or health indications could be retrieved from the COCONUT database.

Several structures found in the virtual screen of the main protease also matched at least 4 of the 6 features of the Mpro pharmacophore. One of these structures is Cassiarine E [CNP0328498], which is found in the cassia tree *Cassia siamea*, which is used in traditional medicines in Burma and Thailand. The medicinal value of the extract of this plant is attributed to a related compound, barakol, which has sedative and anxiolytic properties[35,36]. These structures belong to a class of pyranoisoquinolines known as Cassiarines (A-J), which are also found in the extracts of the *Cassia* family[37]. Yet another compound predicted to be potentially binding to the main protease is 5,6-dihydro, 5α-chloro-6β-hydroxy-Jaborosalactone 5 [CNP0252890], one of several jaborosalactones[38]; naturally-occurring Withanolide chlorohydrins found in the extracts of species in the genus *Jaborosa* that have been reported to act as insect anti-feedants[39]. These plants are members of the Solanaceae family and are widely distributed throughout South America, but primarily in the Andes region. 4^{5},3-dichloro-1(1,2),2(1,4),4,5(1,3)-tetrabenzenacycloheptaphane-2^{3},4^{4},5^{6}-triol [CNP0273258], also known as 12,10'-dichloroisoplagiochin C[40] is found in Japanese liverwort *Herbertus sakuraii*. Kadcoccilactone G [CNP0363620] is among several other Kadcoccilactones[41] found in *Kadsura coccinea*. This genus of climbing plant is found throughout southeast Asia, where *K. coccinea* itself is used in cuisine and traditional medicine to treat gastric and duodenal ulcers, gastroenteritis, rheumatism, lumbago, and dysmenorrhea[42]. Aspertryptanthrin B [CNP0110880], is found in several terrestrial and marine species in the genus Aspergillus, but it has not been shown to have biological activity, although many other indole diketopiperazines from this genus have reported cytotoxic effects[43].

Several of the natural products predicted to bind in the papain-like protease's active site possess a

tetrahydropyran ring. Structures found in the virtual screen of the papain-like protease that matched at least 3 of the 5 features of the pharmacophore were also recovered. In addition to tetraorcinol A, 6-hydroxystaurosporinone [CNP0301743] is isolated from the myxomycete *Lycogala Epidendrum*, and has been reported to be a protein tyrosine kinase inhibitor[44]. Asperlicin D [CNP0245001], isolated from the fungus *Aspergillus alliaceus,* is in the family of mycotoxins known as Asperlicins[45,46] that have been shown to be selective antagonists for the cholecystokinin receptor CCK_A[47,48]. 1-hydroxy-4-phenethyl-2-((3,4,5-trihydroxy-6-(1-hydroxycyclohexyl)tetrahydro-2H-pyran-2-yl)oxy)anthracene-9,10-dione [CNP0420259], is a 1,2-dihydroxyanthraquinone glycoside. Many naturally-occurring anthraquinone derivatives have been shown to exhibit antibacterial and antitumor activity. For instance, deoxy-2,3,3'4',6,7-hexahydro-8-(2,2-dimethyl-2H-benzopyran-6-yl)-5-hydroxy-2,2-dimethyl-2H,6H-benzo[1,2-b:5,4-b']dipyran-6-one[49] [CNP0292046] is a flavonoid found in the leaf extract of *Artocarpus fulvicortex*, a fruiting tree from Indonesia and Malaysia. Flavonoids isolated from this genus have been shown to have medicinal properties, such as antiplatelet[50,51], antimicrobial[52], anti-inflammatory[53] and cytotoxic effects[54].

Among the natural products predicted to bind to the spike protein in a way that would disrupt its interaction with ACE2, polyphenols and in particular flavonoids, are seen in several computational hits. In particular, compounds based at least in part on apigenin scaffolds are observed. Apigenin, interestingly, is hypothesized to regulate the expression of ACE2[55]. Apigenin is found in many plants, and is particularly abundant in the flowers of the chamomile plants. In addition to prekinamycin A, which has been reported in the literature to exhibit cytotoxicity to human leukemia cells[34], harunganin [CNP0334196], isolated from the dragon's blood tree, *Harungana madagascariensis*, from Madagascar and several nearby countries in Africa, where the extracts have been used in traditional medicines to treat a variety of ailments such as dysentery, diarrhea, anemia, typhoid and heart ailments[56]. Beccamarin[57,58] [CNP0132136] was also found to fit the pharmacophore. This compound is found in the bark of the ironwood tree, *Mesua beccariana*, whose extract has been used in traditional medicine to treat fever, renal diseases, poultice and dyspepsia in Malaysia. Beccamarin has recently been shown to inhibit the proliferation of the Raji (lymphoma), SK-MEL-28 (malignant melanoma cells) and HeLa (cervical cells) human cancer cell lines[59].

This list can be the basis of an experimental validation for target engagement and against SARS-CoV-2's three proteins studied here, and for phenotypic effect. The pharmacophore models derived here from the docking calculations will be used in a subsequent screening of the entire Coconut database, as well as for extremely large databases of chemicals outside of the natural products space.

Acknowledgements

We wish to thank Ryan Yates, Amar Chittiboyana and Anna Petroff for useful discussions. We thank the HPE Strategic Alliance and Communication teams for facilitating this work, in particular Joseph George, Alison Paisley, Paul Rosien and Rangan Sukumar.

References

(1) Pushpakom, S.; Iorio, F.; Eyers, P. A.; Escott, K. J.; Hopper, S.; Wells, A.; Doig, A.; Guilliams, T.; Latimer, J.; McNamee, C.; et al. Drug Repurposing: Progress, Challenges and Recommendations. *Nat. Rev. Drug Discov.* **2019**, *18* (1), 41–58. https://doi.org/10.1038/nrd.2018.168.

(2) GlobalData. FDA new drug approvals down 16% in 2019 https://www.globaldata.com/fda-new-drug-approvals-down-16-in-2019/.

(3) Li, S. Y.; Chen, C.; Zhang, H. Q.; Guo, H. Y.; Wang, H.; Wang, L.; Zhang, X.; Hua, S. N.; Yu, J.; Xiao, P. G.; et al. Identification of Natural Compounds with Antiviral Activities against SARS-Associated Coronavirus. *Antiviral Res.* **2005**, *67* (1), 18–23. https://doi.org/10.1016/j.antiviral.2005.02.007.

(4) Park, J.; Hoon, J.; Min, Y.; Jae, H.; Wook, D.; Hun, K.; Kwon, H.; Park, S.; Song, W.; Bae, Y. Tanshinones as Selective and Slow-Binding Inhibitors for SARS-CoV Cysteine Proteases. *Bioorg. Med. Chem.* **2012**, *20* (January), 5928–5935.

(5) Keun, J.; Curtis-long, M. J.; Ho, K.; Wook, D.; Won, H.; Joo, H.; Hun, K. Geranylated Flavonoids Displaying SARS-CoV Papain-like Protease Inhibition from the Fruits of Paulownia Tomentosa. *Bioorg. Med. Chem.* **2013**, *21*, 3051–3057.

(6) Elfiky, A. A. Natural Products May Interfere with SARS-CoV-2 Attachment to the Host Cell. *J. Biomol. Struct. Dyn.* **2020**, *0* (0), 1–16. https://doi.org/10.1080/07391102.2020.1761881.

(7) Tahir ul Qamar, M.; Alqahtani, S. M.; Alamri, M. A.; Chen, L.-L. Structural Basis of SARS-CoV-2 3CLpro and Anti-COVID-19 Drug Discovery from Medicinal Plants. *J. Pharm. Anal.* **2020**. https://doi.org/https://doi.org/10.1016/j.jpha.2020.03.009.

(8) Abdali, N.; Parks, J. M.; Haynes, K. M.; Chaney, J. L.; Green, A. T.; Wolloscheck, D.; Walker, J. K.; Rybenkov, V. V; Baudry, J.; Smith, J. C.; et al. Reviving Antibiotics: Efflux Pump Inhibitors That Interact with AcrA, a Membrane Fusion Protein of the AcrAB-TolC Multidrug Efflux Pump. *ACS Infect. Dis.* **2017**, *3* (1), 89–98. https://doi.org/10.1021/acsinfecdis.6b00167.

(9) Velazquez, H. A.; Riccardi, D.; Xiao, Z.; Quarles, L. D.; Yates, C. R.; Baudry, J.; Smith, J. C. Ensemble Docking to Difficult Targets in Early-Stage Drug Discovery: Methodology and Application to Fibroblast Growth Factor 23. *Chem. Biol. Drug Des.* **2018**, *91* (2), 491–504. https://doi.org/10.1111/cbdd.13110.

(10) Kapoor, K.; McGill, N.; Peterson, C. B.; Meyers, H. V; Blackburn, M. N.; Baudry, J. Discovery of Novel Nonactive Site Inhibitors of the Prothrombinase Enzyme Complex. *J. Chem. Inf. Model.* **2016**, *56* (3), 535–547. https://doi.org/10.1021/acs.jcim.5b00596.

(11) Amaro, R. E.; Baudry, J.; Chodera, J.; Demir, Ö.; McCammon, J. A.; Miao, Y.; Smith, J. C. Ensemble Docking in Drug Discovery. *Biophys. J.* **2018**, *114* (10), 2271–2278. https://doi.org/https://doi.org/10.1016/j.bpj.2018.02.038.

(12) Evangelista Falcon, W.; Ellingson, S. R.; Smith, J. C.; Baudry, J. Ensemble Docking in Drug Discovery: How Many Protein Configurations from Molecular Dynamics Simulations Are Needed To Reproduce Known Ligand Binding? *J. Phys. Chem. B* **2019**, *123* (25), 5189–5195. https://doi.org/10.1021/acs.jpcb.8b11491.

(13) Smith, M.; Smith, J. C. Repurposing Therapeutics for COVID-19: Supercomputer-Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface. **2020**. https://doi.org/10.26434/chemrxiv.11871402.v4.

(14) Sorokina, M.; Steinbeck, C. COCONUT: the COlleCtion of Open NatUral producTs http://doi.org/10.5281/zenodo.3778405 (accessed May 28, 2020).

(15) Ertl, P.; Roggo, S.; Schuffenhauer, A. Natural Product-Likeness Score and Its Application for Prioritization of Compound Libraries. *J. Chem. Inf. Model.* **2008**, *48* (1), 68–74. https://doi.org/10.1021/ci700286x.

(16) Sorokina, M.; Steinbeck, C. NaPLeS: A Natural Products Likeness Scorer—Web Application and Database. *J. Cheminform.* **2019**, *11* (1), 55. https://doi.org/10.1186/s13321-019-0378-z.

(17) O'Boyle, N. M.; Banck, M.; James, C. A.; Morley, C.; Vandermeersch, T.; Hutchison, G. R. Open Babel: An Open Chemical Toolbox. *J. Cheminform.* **2011**, *3* (1), 33.

https://doi.org/10.1186/1758-2946-3-33.

(18) Molecular Operating Environment (MOE) 2019, Chemical Computing Group. Chemical Computing Group: Montreal **2019**.

(19) Morris, G. M.; Huey, R.; Lindstrom, W.; Sanner, M. F.; Belew, R. K.; Goodsell, D. S.; Olson, A. J. AutoDock4 and AutoDockTools4: Automated Docking with Selective Receptor Flexibility. *J. Comput. Chem.* **2009**, *30* (16), 2785–2791. https://doi.org/10.1002/jcc.21256.

(20) Sanner, M. F. Python: A Programming Language for Software Integration and Development. *J. Mol. Graph. Model.* **1999**, *17* (1), 57–61.

(21) Zhang, L.; Lin, D.; Sun, X.; Curth, U.; Drosten, C.; Sauerhering, L.; Becker, S.; Rox, K.; Hilgenfeld, R. Crystal Structure of SARS-CoV-2 Main Protease Provides a Basis for Design of Improved a-Ketoamide Inhibitors. *Science (80-.).* **2020**, *368* (6489), 409–412. https://doi.org/10.1126/science.abb3405.

(22) Báez-Santos, Y. M.; St John, S. E.; Mesecar, A. D. The SARS-Coronavirus Papain-like Protease: Structure, Function and Inhibition by Designed Antiviral Compounds. *Antiviral Res.* **2015**, *115*, 21–38. https://doi.org/10.1016/j.antiviral.2014.12.015.

(23) Ou, X.; Liu, Y.; Lei, X.; Li, P.; Mi, D.; Ren, L.; Guo, L.; Guo, R.; Chen, T.; Hu, J.; et al. Characterization of Spike Glycoprotein of SARS-CoV-2 on Virus Entry and Its Immune Cross-Reactivity with SARS-CoV. *Nat. Commun.* **2020**, *11* (1), 1620. https://doi.org/10.1038/s41467-020-15562-9.

(24) Grant, O. C.; Montgomery, D.; Ito, K.; Woods, R. J. Analysis of the SARS-CoV-2 Spike Protein Glycan Shield: Implications for Immune Recognition. *bioRxiv* **2020**, 2020.04.07.030445. https://doi.org/10.1101/2020.04.07.030445.

(25) Van Der Spoel, D.; Lindahl, E.; Hess, B.; Groenhof, G.; Mark, A. E.; Berendsen, H. J. C. GROMACS: Fast, Flexible, and Free. *J. Comput. Chem.* **2005**, *26* (16), 1701–1718. https://doi.org/10.1002/jcc.20291.

(26) Trott, O.; Olson, A. J. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *J. Comput. Chem.* **2010**, *31* (2), 455–461. https://doi.org/10.1002/jcc.21334.

(27) The PyMOL Molecular Graphics System. Schrödinger, LLC.

(28) Alverson, B.; Froese, E.; Kaplan, L.; Roweth, D. *Cray® XC™ Series Network*; 2012.

(29) *CRC Dictionary of Natural Products*, 15.2.; CRC Press: Boca Raton, 2006.

(30) Zhuang, Y.; Teng, X.; Wang, Y.; Liu, P.; Wang, H.; Li, J.; Li, G.; Zhu, W. Cyclopeptides and Polyketides from Coral-Associated Fungus, Aspergillus Versicolor LCJ-5-4. *Tetrahedron* **2011**, *67* (37), 7085–7089. https://doi.org/https://doi.org/10.1016/j.tet.2011.07.003.

(31) Woo, C. M. Synthetic and Chemical Biological Studies of the Diazofluorene Antitumor Antibiotics, Yale University, **2017**.

(32) Gould, S. J.; Chen, J.; Cone, M. C.; Gore, M. P.; Melville, C. R.; Tamayo, N. Identification of Prekinamycin in Extracts of Streptomyces Murayamaensis. *J. Org. Chem.* **1996**, *61* (17), 5720–5721. https://doi.org/10.1021/jo9611024.

(33) Hata, T.; Omura, S.; Iwai, Y.; Nakagawa, A.; Otani, M.; Ito, S.; Matsuya, T. A New Antibiotic, Kinamycin : Fermentation, Isolation, Purification and Properties. *J. Antibiot. (Tokyo).* **1971**, *24* (6), 353–359.

(34) Abbott, G. L.; Wu, X.; Zhao, Z.; Guo, L.; Birman, V. B.; Hasinoff, B. B.; Dmitrienko, G. I. Prekinamycin and an Isosteric-Isoelectronic Analogue Exhibit Comparable Cytotoxicity towards K562 Human Leukemia Cells. *Medchemcomm* **2014**, *5*, 1364–1370.

(35) Thongsaard, W.; Pongsakorn, S.; Sudsuang, R.; Bennett, G. W.; Kendall, D. A.; Marsden, C. A. Barakol, a Natural Anxiolytic, Inhibits Striatal Dopamine Release but Not Uptake in Vitro. *Eur. J. Pharmacol.* **1997**, *319* (2), 157–164. https://doi.org/https://doi.org/10.1016/S0014-2999(96)00850-3.

(36) Sukma, M.; Chaichantipyuth, C.; Murakami, Y.; Tohda, M.; Matsumoto, K.; Watanabe, H. CNS Inhibitory Effects of Barakol, a Constituent of Cassia Siamia Lamk. *J. Ethnopharmacol.* **2002**, *83* (1–2), 87–94.

(37) Bondarenko, S. P.; Frasinyuk, M. S. Chromone Alkaloids: Structural Features, Distribution in Nature, and Biological Activity. *Chem. Nat. Compd.* **2019**, *55* (2), 201–234. https://doi.org/10.1007/s10600-019-02656-0.

(38) Bonetto, G. M.; Gil, R. R.; Oberti, J. C.; Veleiro, A. S.; Burton, G. Novel Withanolides from Jaborosa Sativa. *J. Nat. Prod.* **1995**, *58* (5), 705–711. https://doi.org/10.1021/np50119a008.

(39) Vaccarini, C. E.; Bonetto, G. M. Antifeedant Activity Evaluation of Withanolides from Jaborosa Integrifolia. *Molecules* **2000**, *5* (3), 422–423. https://doi.org/10.3390/50300422.

(40) Hashimoto, T.; Irita, H.; Takaoka, S.; Tanaka, M.; Asakawa, Y. New Chlorinated Cyclic Bis(Bibenzyls) from the Liverworts Herbertus Sakuraii and Mastigophora Diclados. *Tetrahedron* **2000**, *56* (20), 3153–3159. https://doi.org/https://doi.org/10.1016/S0040-4020(00)00236-2.

(41) Gao, X.-M.; Pu, J.-X.; Huang, S.-X.; Lu, Y.; Lou, L.-G.; Li, R.-T.; Xiao, W.-L.; Chang, Y.; Sun, H.-D. Kadcoccilactones A−J, Triterpenoids from Kadsura Coccinea. *J. Nat. Prod.* **2008**, *71* (7), 1182–1188. https://doi.org/10.1021/np800078x.

(42) Saunders, R. M. K. *Systematic Botany Monographs*, Monograph.; Anderson, C., Ed.; The American Society of Plant Taxonomists: Ann Arbor, 1998; Vol. 54.

(43) Ma, Y.-M.; Liang, X.-A.; Kong, Y.; Jia, B. Structural Diversity and Biological Activities of Indole Diketopiperazine Alkaloids from Fungi. *J. Agric. Food Chem.* **2016**, *64* (35), 6659–6671. https://doi.org/10.1021/acs.jafc.6b01772.

(44) Hosoya, T.; Yamamoto, Y.; Uehara, Y.; Hayashi, M.; Komiyama, K.; Ishibashi, M. New Cytotoxic Bisindole Alkaloids with Protein Tyrosine Kinase Inhibitory Activity from a Myxomycete Lycogala Epidendrum. *Bioorg. Med. Chem. Lett.* **2005**, *15* (11), 2776–2780. https://doi.org/10.1016/j.bmcl.2005.03.103.

(45) Chang, R. S.; Lotti, V. J.; Monaghan, R. L.; Birnbaum, J.; Stapley, E. O.; Goetz, M. A.; Albers-Schönberg, G.; Patchett, A. A.; Liesch, J. M.; Hensens, O. D. A Potent Nonpeptide Cholecystokinin Antagonist Selective for Peripheral Tissues Isolated from Aspergillus Alliaceus. *Science* **1985**, *230* (4722), 177–179. https://doi.org/10.1126/science.2994227.

(46) Liesch, J. M.; Hensens, O. D.; Springer, J. P.; Chang, R. S.; Lotti, V. J. Asperlicin, a Novel Non-Peptidal Cholecystokinin Antagonist from Aspergillus Alliaceus. Structure Elucidation. *J. Antibiot. (Tokyo).* **1985**, *38* (12), 1638–1641. https://doi.org/10.7164/antibiotics.38.1638.

(47) Van der Bent, A.; Ter Laak, A. M.; IJzerman, A. P.; Soudijn, W. Molecular Modelling of Asperlicin Derived Cholecystokinin A Receptor Antagonists. *Eur. J. Pharmacol.* **1992**, *226* (4), 327–334. https://doi.org/10.1016/0922-4106(92)90050-6.

(48) Lattmann, E.; Billington, D. C.; Poyner, D. R.; Howitt, S. B.; Offel, M. Synthesis and Evaluation of Asperlicin Analogues as Non-Peptidal Cholecystokinin-Antagonists. *Drug Des. Discov.* **2001**, *17* (3), 219–230.

(49) Jamil, S.; Taher, M.; Sirat, H. M.; Othman, N. A. Flavonoids and Triterpenes from the Leaves of Artocarpus Fulvicortex. *Nat. Prod. Commun.* **2012**, *7* (12), 1587–1588. https://doi.org/10.1177/1934578x1200701211.

(50) Weng, J.-R.; Chan, S.-C.; Lu, Y.-H.; Lin, H.-C.; Ko, H.-H.; Lin, C.-N. Antiplatelet Prenylflavonoids from Artocarpus Communis. *Phytochemistry* **2006**, *67* (8), 824–829. https://doi.org/10.1016/j.phytochem.2006.01.030.

(51) Jantan, I.; Mohd Yasin, Y. H.; Jamil, S.; Sirat, H.; Basar, N. Effect of Prenylated Flavonoids and Chalcones Isolated from Artocarpus Species on Platelet Aggregation in Human Whole Blood. *J. Nat. Med.* **2010**, *64* (3), 365–369. https://doi.org/10.1007/s11418-010-0410-0.

(52) Kuete, V.; Ango, P. Y.; Fotso, G. W.; Kapche, G. D. W. F.; Dzoyem, J. P.; Wouking, A. G.; Ngadjui, B. T.; Abegaz, B. M. Antimicrobial Activities of the Methanol Extract and Compounds from Artocarpus Communis (Moraceae). *BMC Complement. Altern. Med.* **2011**, *11* (1), 42. https://doi.org/10.1186/1472-6882-11-42.

(53) Wei, B.-L.; Weng, J.-R.; Chiu, P.-H.; Hung, C.-F.; Wang, J.-P.; Lin, C.-N. Antiinflammatory Flavonoids from Artocarpus Heterophyllus and Artocarpus Communis. *J. Agric. Food Chem.* **2005**, *53* (10), 3867–3871. https://doi.org/10.1021/jf047873n.

(54) Chin-Lin, H.; Shyu, M.-H.; Lin, J.-A.; Yen, G.-C.; Fang, S.-C. Cytotoxic Effects of Geranyl Flavonoid Derivatives from the Fruit of Artocarpus Communis in SK-Hep-1 Human Hepatocellular Carcinoma Cells. *Food Chem.* **2011**, *v. 127* (1), 127-134–2011 v.127 no.1. https://doi.org/10.1016/j.foodchem.2010.12.100.

(55) Salehi, B.; Venditti, A.; Sharifi-Rad, M.; Kręgiel, D.; Sharifi-Rad, J.; Durazzo, A.; Lucarini, M.; Santini, A.; Souto, E. B.; Novellino, E.; et al. The Therapeutic Potential of Apigenin. *Int. J. Mol. Sci.* **2019**, *20* (6), 1305. https://doi.org/10.3390/ijms20061305.

(56) Happi, G. M.; Tiani, G. L. M.; Gbetnkom, B. Y. M.; Hussain, H.; Green, I. R.; Ngadjui, B. T.; Kouam, S. F. Phytochemistry and Pharmacology of Harungana Madagascariensis: Mini Review. *Phytochem. Lett.* **2020**, *35*, 103–112. https://doi.org/https://doi.org/10.1016/j.phytol.2019.11.015.

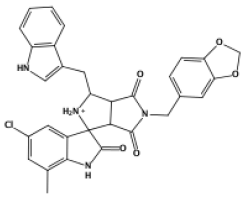
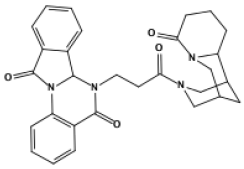
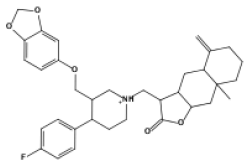
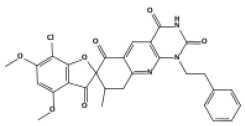
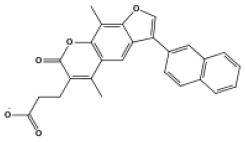
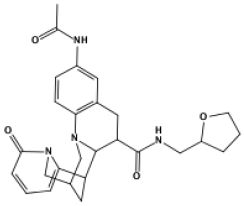
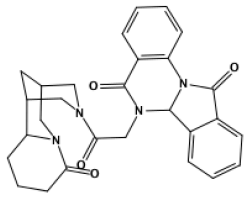
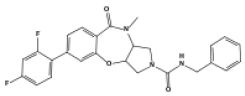
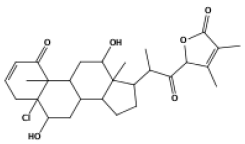
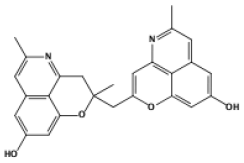
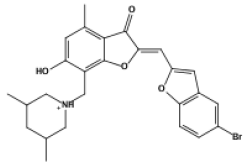
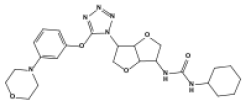
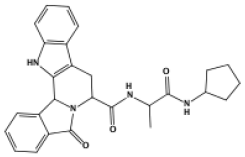
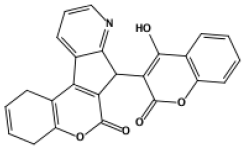
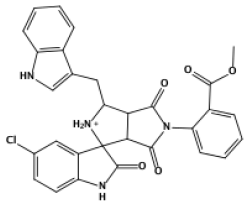
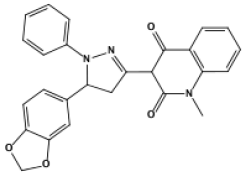
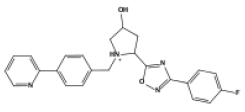
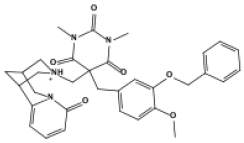
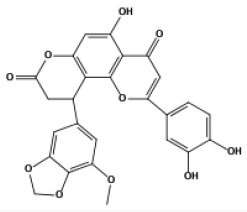
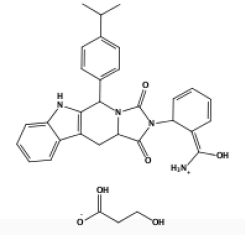
(57) Lian Ee, G. C.; Teh, S. S.; Mah, S. H.; Rahmani, M.; Taufiq-Yap, Y. H.; Awang, K. A Novel Cyclodione Coumarin from the Stem Bark of Mesua Beccariana. *Molecules* **2011**, *16* (9), 7249–7255. https://doi.org/10.3390/molecules16097249.

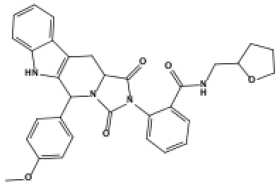
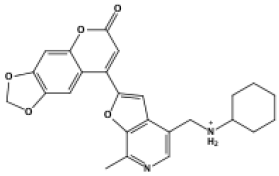
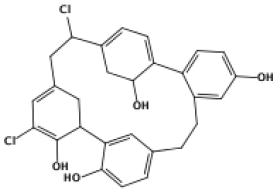
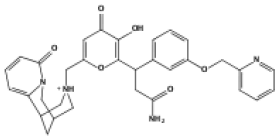
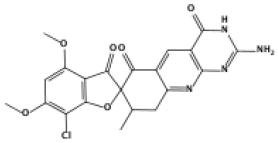
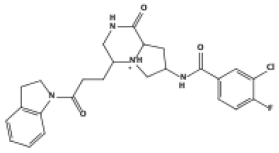
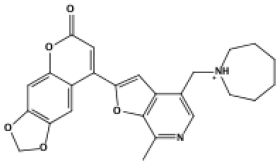
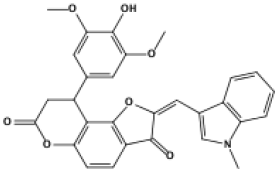
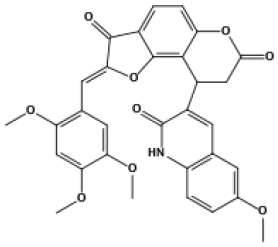
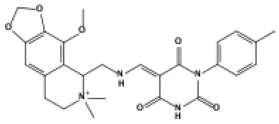
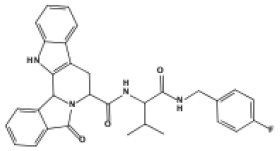
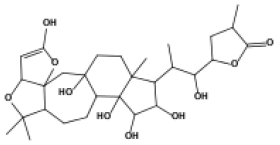
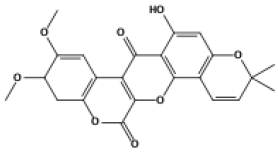
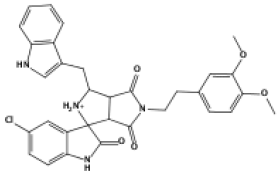
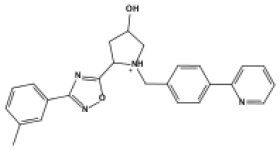
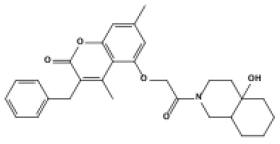
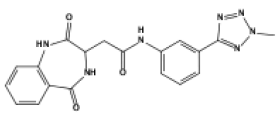
(58) Karunakaran, T.; Ee, G. C. L.; Tee, K. H.; Ismail, I. S.; Zamakshshari, N. H.; Peter, W. M. Cytotoxic Prenylated Xanthone and Coumarin Derivatives from Malaysian Mesua Beccariana. *Phytochem. Lett.* **2016**, *17*, 131–134. https://doi.org/https://doi.org/10.1016/j.phytol.2016.07.026.

(59) Teh, S. S.; Ee, G. C. L.; Mah, S. H.; Lim, Y. M.; Rahmani, M. Mesua Beccariana (Clusiaceae), a Source of Potential Anti-Cancer Lead Compounds in Drug Discovery. *Molecules* **2012**, *17* (9), 10791–10800. https://doi.org/10.3390/molecules170910791.

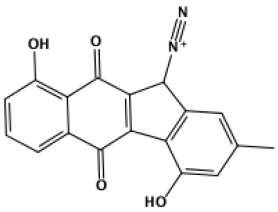
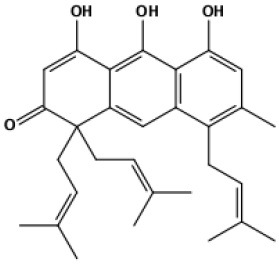
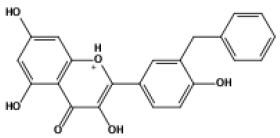
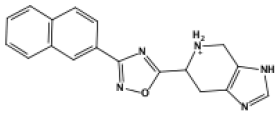
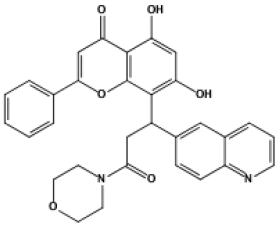
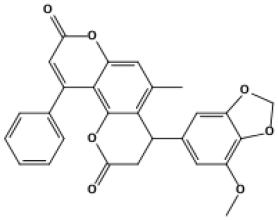
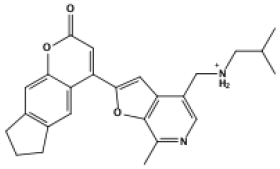
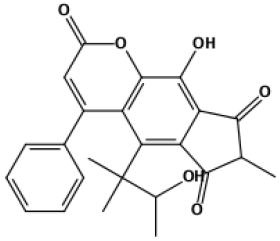
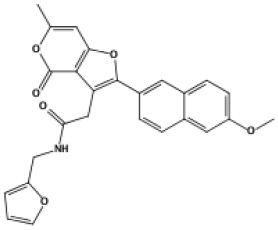
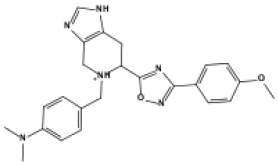
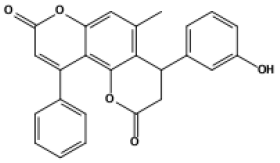
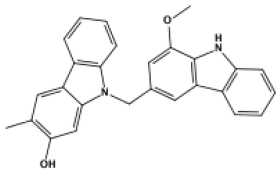
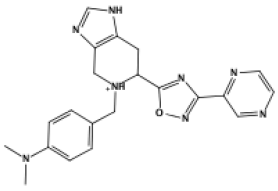
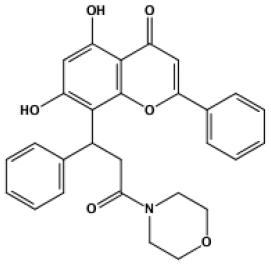
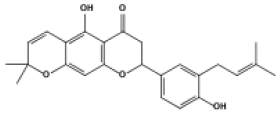
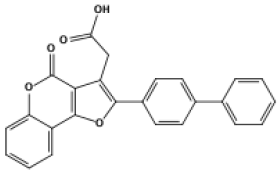
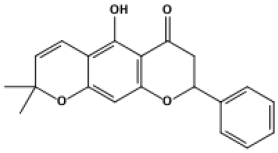
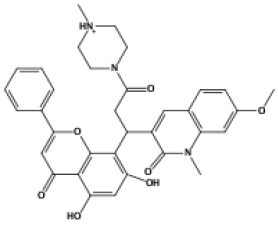
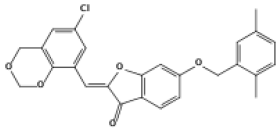
Supplementary Information

List and Coconut4 IDs of top 37 natural products predicted to be top binders in the MPRO active site

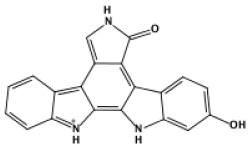
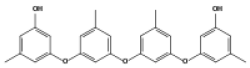
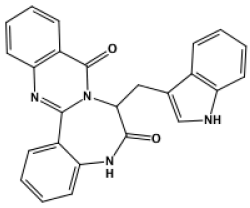
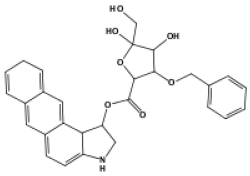
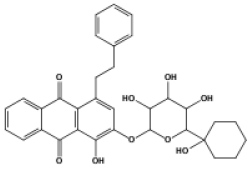
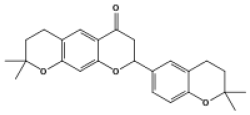
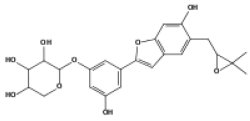
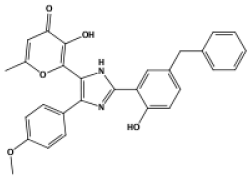
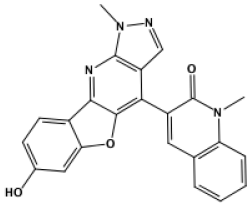
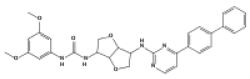
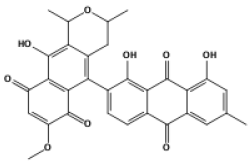
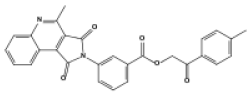
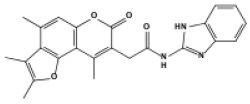
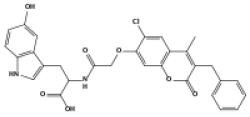
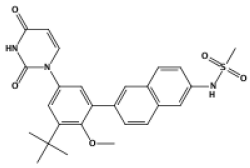
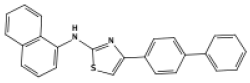
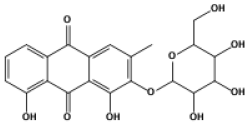
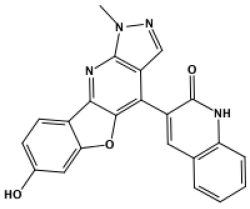
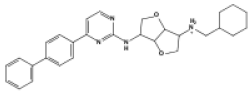
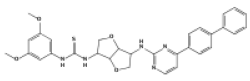
CNP0131030	CNP0228228	CNP0369441	CNP0414324
			
CNP0037401	CNP0380989	CNP0413646	CNP0120445
			
CNP0252890	CNP0328498	CNP0061637	CNP0293624
			
CNP0390225	CNP0055030	CNP0390606	CNP0201098
			
CNP0245259	CNP0362995	CNP0393487	CNP0390629
			

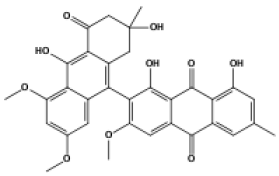
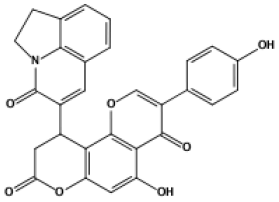
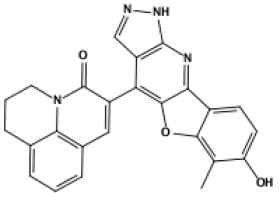
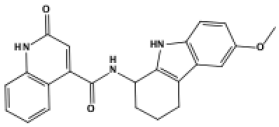
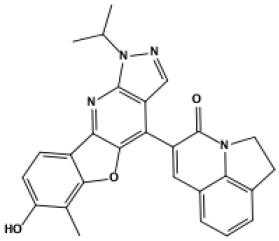
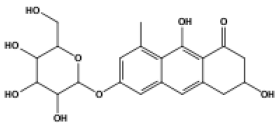
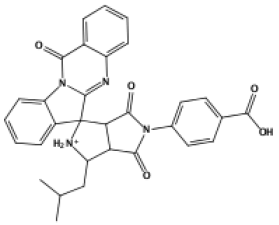
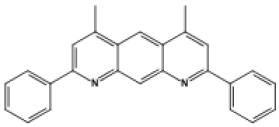
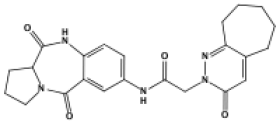
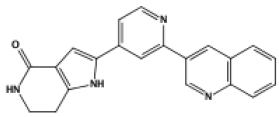
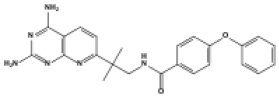
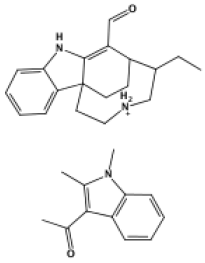
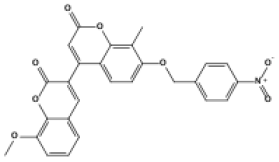
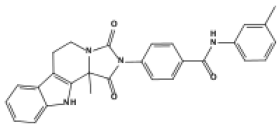
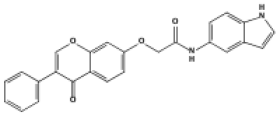
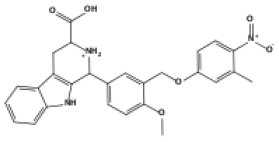
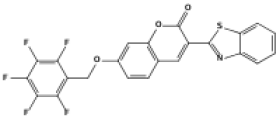
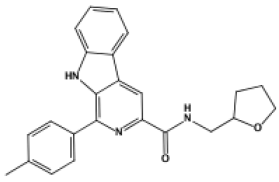
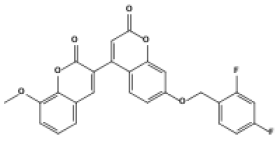
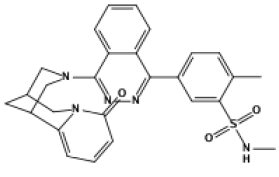
CNP0390969	CNP0401185	CNP0273258	CNP0355623
			
CNP0374052	CNP0236086	CNP0042143	CNP0403172
			
CNP0161959	CNP0370364	CNP0390684	CNP0363620
			
CNP0331417	CNP0373422	CNP0013026	CNP0381876
			
CNP0353621			
			

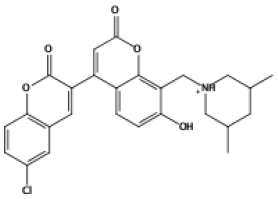
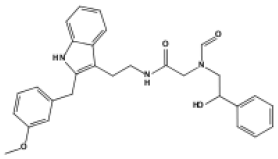
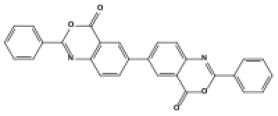
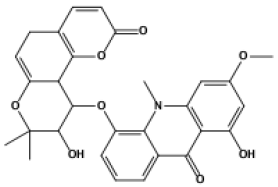
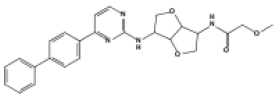
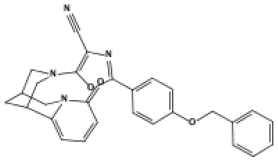
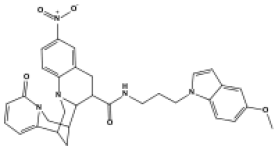
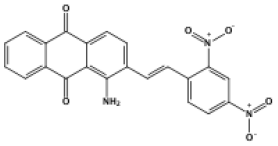
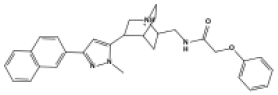
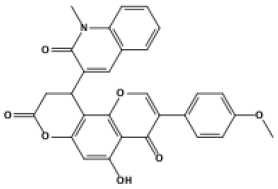
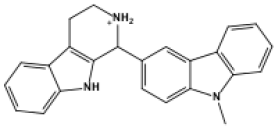
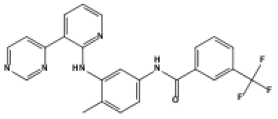
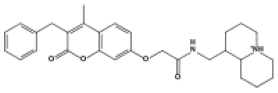
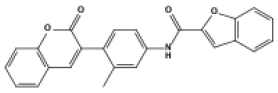
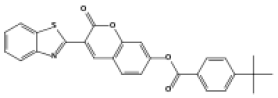
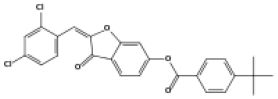
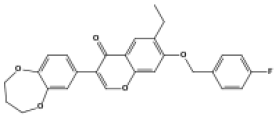
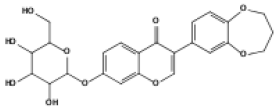
List and Coconut4 IDs of 19 top natural products predicted to be top binders in the spike protein:ACE2 interaction surface

CNP0375993 	CNP0334196 	CNP0233013 	CNP0113725 
CNP0285404 	CNP0409713 	CNP0403008 	CNP0132136 
CNP0323621 	CNP0212508 	CNP0196493 	CNP0129938 
CNP0012180 	CNP0239713 	CNP0168743 	CNP0125961 
CNP0224601 	CNP0198440 	CNP0230664 	

List and Coconut4 IDs of 58 natural products predicted to be top binders in the PLPRO binding site

CNP0301743 	CNP0265390 	CNP0245001 	CNP0093767 
CNP0420259 	CNP0292046 	CNP0332080 	CNP0403695 
CNP0408592 	CNP0110823 	CNP0228546 	CNP0368276 
CNP0393385 	CNP0386161 	CNP0376441 	CNP0077894 
CNP0322100 	CNP0382733 	CNP0183607 	CNP0287331 

CNP0273147	CNP0402381	CNP0198351	CNP0189790
			
CNP0402494	CNP0150083	CNP0257622	CNP0122189
			
CNP0256592	CNP0356119	CNP0020386	CNP0411802
			
CNP0241792	CNP0221693	CNP0409252	CNP0373611
			
CNP0069706	CNP0338566	CNP0249427	CNP0369815
			

CNP0029910 	**CNP0379817** 	**CNP0349877** 	**CNP0387628** 
CNP0309003 	**CNP0396391** 	**CNP0138847** 	**CNP0048604** 
CNP0141831 	**CNP0375273** 	**CNP0297541** 	**CNP0019832** 
CNP0041038 	**CNP0080109** 	**CNP0165409** 	**CNP0057738** 
CNP0064085 	**CNP0365565** 		

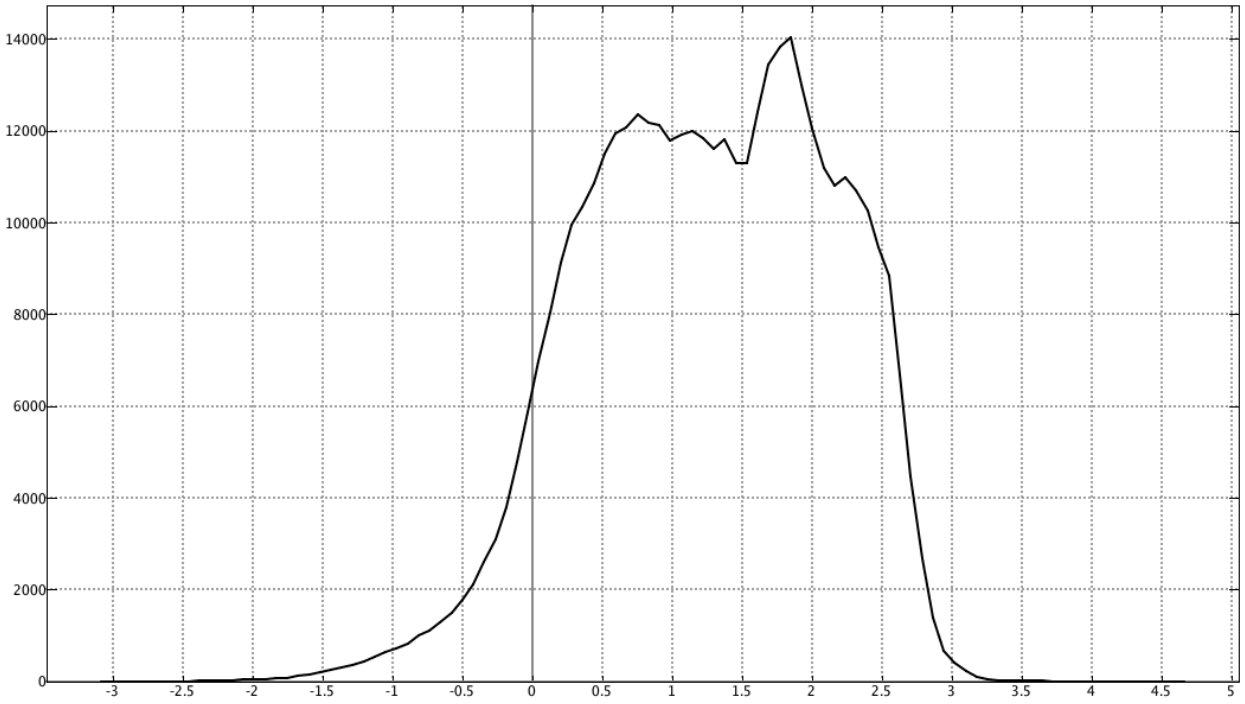


Figure SI-1: Distribution of NP scores in the top scoring natural products in the Coconut4 database

PLpro hits matching all pharmacophore features

Coconut_ID	Docking_Score	molecular_formula	textTaxa	NPL_score	found_in_databases
CNP0332080	-8.5	C24H26O9	[notax]	1.83	[supernatural2 unpd]
CNP0403695	-8.5	C29H24N2O5	[notax]	0.39	[ibs2019mar_nc]
CNP0110823	-8.5	C31H31N5O5	[notax]	0.44	[zincnp]
CNP0376441	-8.4	[C26H26N3O5S]-	[notax]	-0.21	[chembl_np]
CNP0077894	-8.4	C25H18N2S	[notax]	-0.69	[supernatural2]
CNP0382733	-8.3	C22H14N4O3	[notax]	-0.10	[ibs2019mar_nc]
CNP0183607	-8.3	C29H34N4O2	[notax]	1.09	[zincnp]
CNP0287331	-8.3	C31H31N5O4S	[notax]	0.34	[zincnp]
CNP0356119	-8.2	C21H16N4O	[notax]	-0.29	[drugbanknp]
CNP0349877	-8.1	C28H16N2O4	[notax]	-0.75	[ibs2019mar_nc supernatural2]
CNP0309003	-8.1	C25H26N4O4	[notax]	0.65	[supernatural2 zincnp]
CNP0396391	-8.1	C28H24N4O3	[notax]	-0.02	[ibs2019mar_nc supernatural2]
CNP0141831	-8.1	[C30H33N4O2]+	[notax]	0.29	[supernatural2]
CNP0297541	-8.1	C24H21N3	[notax]	0.16	[ibs2019mar_nc]

PLpro hits that matching at least 3 of 5 pharmacophore features

Coconut_ID	Docking_Score	molecular_formula	textTaxa	NPL_score	found_in_databases
CNP0332080	-8.5	C24H26O9	[notax]	1.83	[supernatural2 unpd]
CNP0403695	-8.5	C29H24N2O5	[notax]	0.39	[ibs2019mar_nc]
CNP0408592	-8.5	C23H16N4O3	[notax]	-0.03	[ibs2019mar_nc]
CNP0110823	-8.5	C31H31N5O5	[notax]	0.44	[zincnp]
CNP0228546	-8.5	C31H24O9	[notax]	1.58	[supernatural2 unpd]

CNP0368276	-8.5	C28H20N2O5	[notax]	-0.36	[ibs2019mar_nc supernatural2]
CNP0093767	-8.4	C29H25NO7	[plants]	0.06	[tcmdb_taiwan]
CNP0393385	-8.4	C24H21N3O4	[notax]	0.46	[ibs2019mar_nc]
CNP0386161	-8.4	C30H25ClN2O7	[notax]	0.57	[ibs2019mar_nc]
CNP0376441	-8.4	[C26H26N3O5S]-	[notax]	-0.21	[chembl_np]
CNP0077894	-8.4	C25H18N2S	[notax]	-0.69	[supernatural2]
CNP0420259	-8.3	C33H34O9	[plants]	0.83	[tcmdb_taiwan supernatural2 zincnp]
CNP0322100	-8.3	C21H20O10	[notax]	1.24	[supernatural2 unpd]
CNP0382733	-8.3	C22H14N4O3	[notax]	-0.10	[ibs2019mar_nc]
CNP0183607	-8.3	C29H34N4O2	[notax]	1.09	[zincnp]
CNP0287331	-8.3	C31H31N5O4S	[notax]	0.34	[zincnp]
CNP0273147	-8.3	C33H28O10	[notax]	1.35	[supernatural2 unpd]
CNP0402381	-8.3	C29H19NO7	[notax]	1.03	[ibs2019mar_nc]
CNP0198351	-8.3	C25H18N4O3	[notax]	0.28	[ibs2019mar_nc]
CNP0189790	-8.3	C23H21N3O3	[notax]	0.32	[ibs2019mar_nc]
CNP0402494	-8.3	C27H22N4O3	[notax]	0.34	[ibs2019mar_nc]
CNP0150083	-8.3	C21H24O9	[notax]	2.04	[supernatural2 unpd]
CNP0257622	-8.3	C31H26N4O5	[notax]	0.43	[ibs2019mar_nc]
CNP0122189	-8.3	C26H20N2	[notax]	-0.45	[supernatural2]
CNP0301743	-8.2	C20H13N3O2	[Fungus fungi epidendrum Lycogala]	0.20	[npatlas np_atlas_2019_12 unpd npass]
CNP0292046	-8.2	C25H28O4	[plants]	1.49	[conmednp cmaup supernatural2 tcmid unpd p-anapl]
CNP0256592	-8.2	C23H25N5O4	[notax]	0.35	[ibs2019mar_nc]
CNP0356119	-8.2	C21H16N4O	[notax]	-0.29	[drugbanknp]
CNP0020386	-8.2	C24H24N6O2	[notax]	-0.13	[drugbanknp]
CNP0411802	-8.2	[C31H32N3O2]+	[notax]	0.72	[ibs2019mar_nc]
CNP0241792	-8.2	C27H19NO8	[notax]	0.09	[ibs2019mar_nc supernatural2]
CNP0221693	-8.2	C28H24N4O3	[notax]	-0.04	[ibs2019mar_nc]
CNP0409252	-8.2	C25H18N2O4	[notax]	-0.06	[ibs2019mar_nc]
CNP0373611	-8.2	C27H25N3O6	[notax]	0.15	[ibs2019mar_nc]
CNP0069706	-8.2	C23H10F5NO3S	[notax]	-0.81	[supernatural2]
CNP0265390	-8.1	C28H26O5	[Fungus Aspergillus versicolor LCJ-5 fungi]	0.45	[npatlas np_atlas_2019_12 unpd]
CNP0245001	-8.1	C25H18N4O2	[Fungus Aspergillus alliaceus ATCC 20655]	-0.08	[np_atlas_2019_12 unpd]
CNP0338566	-8.1	C24H23N3O2	[notax]	0.30	[npedia ibs2019mar_nc]
CNP0249427	-8.1	C26H16F2O6	[notax]	0.07	[ibs2019mar_nc supernatural2]
CNP0369815	-8.1	C27H27N5O3S	[notax]	-0.30	[ibs2019mar_nc]
CNP0029910	-8.1	[C26H25ClNO5]+	[notax]	0.67	[supernatural2]
CNP0379817	-8.1	C29H27N3O4	[notax]	0.40	[ibs2019mar_nc]
CNP0349877	-8.1	C28H16N2O4	[notax]	-0.75	[ibs2019mar_nc supernatural2]
CNP0387628	-8.1	C29H25NO8	[notax]	1.12	[fooddb supernatural2 unpd]
CNP0309003	-8.1	C25H26N4O4	[notax]	0.65	[supernatural2 zincnp]

CNP0396391	-8.1	C28H24N4O3	[notax]	-0.02	[ibs2019mar_nc supernatural2]
CNP0138847	-8.1	C32H33N5O5	[notax]	0.48	[ibs2019mar_nc]
CNP0048604	-8.1	C22H13N3O6	[notax]	-0.81	[supernatural2]
CNP0141831	-8.1	[C30H33N4O2]+	[notax]	0.29	[supernatural2]
CNP0375273	-8.1	C29H21NO7	[notax]	0.79	[ibs2019mar_nc supernatural2]
CNP0297541	-8.1	C24H21N3	[notax]	0.16	[ibs2019mar_nc]
CNP0019832	-8.1	C24H18F3N5O	[notax]	-0.72	[drugbanknp]
CNP0041038	-8.1	[C29H35N2O4]+	[notax]	0.80	[supernatural2]
CNP0080109	-8.1	C25H17NO4	[notax]	-0.25	[zincnp]
CNP0165409	-8.1	C27H21NO4S	[notax]	-0.16	[supernatural2]
CNP0057738	-8.1	C26H20Cl2O4	[notax]	-0.04	[supernatural2]
CNP0064085	-8.1	C27H23FO5	[notax]	0.45	[supernatural2]
CNP0365565	-8.1	C24H24O10	[notax]	0.93	[ibs2019mar_nc supernatural2]

Spike protein hits matching at least 4 of 7 pharmacophore features (stereochemistry indicated for chiral compounds)

Coconut_ID	Docking_Score	molecular_formula	textTaxa	NPL_score	found_in_databases
CNP0113725	-9.4	C18H15N5O	[notax]	0.36	[zincnp]
CNP0285404_s	-9.3	C31H26N2O6	[notax]	0.52	[ibs2019mar_nc]
CNP0409713_s	-9.3	C27H20O7	[notax]	0.80	[ibs2019mar_nc]
CNP0403008	-9.2	C25H26N2O3	[notax]	0.95	[ibs2019mar_nc]
CNP0132136	-9.2	C24H22O6	[notax]	1.40	[unpd]
CNP0323621	-9.2	C26H21NO6	[notax]	0.30	[ibs2019mar_nc]
CNP0212508	-9.1	C24H26N6O2	[notax]	0.30	[zincnp]
CNP0196493	-9.1	C25H18O5	[notax]	0.77	[ibs2019mar_nc]
CNP0129938	-9.1	C27H22N2O2	[notax]	0.01	[npass]
CNP0012180	-9.1	C21H22N8O	[notax]	0.27	[zincnp]
CNP0239713_r	-9.0	C28H25NO6	[notax]	0.63	[ibs2019mar_nc]
CNP0168743_s	-9.0	C25H26O5	[notax]	1.56	[supernatural2 unpd]
CNP0334196	-8.9	C30H36O4	[plants]	1.56	[npact zincnp npcare]
CNP0233013	-8.9	[C22H17O6]+	[plants]	0.98	[tcmdb_taiwan]
CNP0125961	-8.9	C25H16O5	[notax]	0.37	[ibs2019mar_nc]
CNP0224601	-8.9	C20H18O4	[notax]	1.36	[unpd]
CNP0198440_s	-8.9	C34H33N3O7	[notax]	0.56	[ibs2019mar_nc]
CNP0375993	-8.8	[C18H11N2O4]+	[murayamaensis Bacterium bacteria Streptomyces]	1.05	[npatlas np_atlas_2019_12]
CNP0230664	-8.8	C26H21ClO5	[notax]	0.24	[supernatural2]

Mpro hits matching all pharmacophore features

Coconut_ID	Docking_Score	molecular_formula	textTaxa	NPL_score	found_in_databases
CNP0131030	-9.5	C31H25ClN4O5	[notax]	0.49	[ibs2019mar_nc]

CNP0414324	-9.4	C29H24ClN3O7	[notax]	0.68	[ibs2019mar_nc]
CNP0120445	-9.2	C26H23F2N3O3	[notax]	0.02	[zincnp]
CNP0061637	-9.1	C26H26BrNO4	[notax]	0.92	[supernatural2]
CNP0390225	-9.1	C27H28N4O3	[notax]	0.77	[ibs2019mar_nc]
CNP0352209	-9.0	C28H25N3O7	[Bacterium bacteria Actinomadura melliaura ATCC 39691]	-0.06	[npatlas np_atlas_2019_12 npass]
CNP0390606	-9.0	C30H23ClN4O5	[notax]	0.37	[ibs2019mar_nc]
CNP0362995	-9.0	C33H36N4O6	[notax]	0.50	[ibs2019mar_nc supernatural2]
CNP0390969	-9.0	C32H30N4O5	[notax]	0.45	[ibs2019mar_nc]
CNP0374052	-9.0	C21H17ClN4O6	[notax]	1.03	[ibs2019mar_nc]
CNP0042143	-8.9	[C25H25N2O5]+	[notax]	0.83	[supernatural2]
CNP0370364	-8.9	[C26H29N4O6]+	[notax]	0.46	[ibs2019mar_nc supernatural2]
CNP0390684	-8.9	C31H29FN4O3	[notax]	0.40	[ibs2019mar_nc]
CNP0331417	-8.9	C23H18O8	[notax]	1.29	[supernatural2 unpd]
CNP0373422	-8.9	C32H29ClN4O5	[notax]	0.55	[ibs2019mar_nc]

Mpro hits matchomg at least 3 of 4 pharmacophore features

Coconut_ID	Docking_Score	molecular_formula	textTaxa	NPL_score	found_in_databases
CNP0131030	-9.5	C31H25ClN4O5	[notax]	0.49	[ibs2019mar_nc]
CNP0228228	-9.5	C29H30N4O4	[notax]	0.78	[ibs2019mar_nc]
CNP0369441	-9.4	[C34H41FNO5]+	[notax]	1.46	[supernatural2]
CNP0414324	-9.4	C29H24ClN3O7	[notax]	0.68	[ibs2019mar_nc]
CNP0037401	-9.4	[C26H19O5]-	[notax]	0.90	[supernatural2]
CNP0380989	-9.3	C27H32N4O4	[notax]	0.90	[ibs2019mar_nc]
CNP0413646	-9.3	C28H28N4O4	[notax]	0.74	[ibs2019mar_nc]
CNP0120445	-9.2	C26H23F2N3O3	[notax]	0.03	[zincnp]
CNP0252890	-9.2	C28H33ClO6	[notax]	2.12	[supernatural2 unpd]
CNP0328498	-9.1	C26H22N2O4	[plants]	0.70	[cmaup supernatural2 unpd npass]
CNP0061637	-9.1	C26H26BrNO4	[notax]	0.92	[supernatural2]
CNP0293624	-9.1	C24H33N7O5	[notax]	0.90	[supernatural2]

CNP0390225	-9.1	C27H28N4O3	[notax]	0.77	[ibs2019mar_nc]
CNP0055030	-9.1	[C24H12NO5]-	[notax]	0.34	[supernatural2]
CNP0352209	-9.0	C28H25N3O7	[Bacterium bacteria Actinomadura melliaura ATCC 39691]	-0.06	[npatlas np_atlas_2019_12 npass]
CNP0390606	-9.0	C30H23ClN4O5	[notax]	0.37	[ibs2019mar_nc]
CNP0201098	-9.0	[C26H20N3O4]-	[notax]	-0.19	[supernatural2]
CNP0245259	-9.0	C24H21FN4O2	[notax]	0.13	[zincnp]
CNP0362995	-9.0	C33H36N4O6	[notax]	0.50	[ibs2019mar_nc supernatural2]
CNP0393487	-9.0	C26H18O10	[notax]	1.35	[ibs2019mar_nc]
CNP0390629	-9.0	C32H30N4O6	[notax]	0.56	[ibs2019mar_nc]
CNP0390969	-9.0	C32H30N4O5	[notax]	0.45	[ibs2019mar_nc]
CNP0401185	-9.0	C25H24N2O5	[notax]	0.94	[ibs2019mar_nc]
CNP0273258	-9.0	C28H20Cl2O4	[notax]	0.35	[supernatural2 unpd]
CNP0355623	-9.0	C32H32N4O6	[notax]	0.75	[ibs2019mar_nc]
CNP0374052	-9.0	C21H17ClN4O6	[notax]	1.02	[ibs2019mar_nc]
CNP0236086	-8.9	C25H26ClFN4O3	[notax]	0.41	[zincnp]
CNP0042143	-8.9	[C25H25N2O5]+	[notax]	0.83	[supernatural2]
CNP0403172	-8.9	C29H23NO7	[notax]	0.77	[ibs2019mar_nc]
CNP0161959	-8.9	C31H25NO9	[notax]	0.68	[ibs2019mar_nc]
CNP0370364	-8.9	[C26H29N4O6]+	[notax]	0.46	[ibs2019mar_nc supernatural2]
CNP0390684	-8.9	C31H29FN4O3	[notax]	0.40	[ibs2019mar_nc]
CNP0363620	-8.9	C29H44O10	[notax]	2.89	[supernatural2 unpd]
CNP0331417	-8.9	C23H18O8	[notax]	1.29	[supernatural2 unpd]
CNP0373422	-8.9	C32H29ClN4O5	[notax]	0.54	[ibs2019mar_nc]
CNP0013026	-8.9	C25H24N4O2	[notax]	0.26	[zincnp]
CNP0381876	-8.9	C29H33NO5	[notax]	1.03	[ibs2019mar_nc supernatural2]
CNP0353621	-8.8	C19H17N7O3	[notax]	-0.49	[ibs2019mar_nc]