Consensus Virtual Screening of Dark Chemical Matter and Food Chemicals Uncover Potential Inhibitors of SARS-CoV-2 Main Protease

Marisa G. Santibáñez-Morán,[1] Edgar López-López,[2] Fernando D. Prieto-Martínez,[1] Norberto Sánchez-Cruz,[1] José L. Medina-Franco[1,*]

[1] DIFACQUIM research group, Department of Pharmacy, School of Chemistry, Universidad Nacional Autónoma de México, Mexico City, Mexico.

[2] Department of Pharmacology, Center of Research and Advanced Studies of the National Polytechnic Institute (CINVESTAV), Mexico City, Mexico.

[*]Contact author: medinajl@unam.mx, jose.medina.franco@gmail.com. Phone: +52 (55) 5622-3899, ext. 44458

Abstract: The COVID-19 pandemic caused by SARS-CoV-2 has claimed more than 380,000 lives Worldwide and more than 6.5 million people are infected. Unfortunately, there is no drug or vaccine for the treatment of COVID-19. The increasing information available of key molecular targets of SARS-CoV-2 and active compounds against related coronavirus facilitates computational tools to rapidly suggest drug candidates for the treatment of COVID-19. As part of a global effort to fight the COVID-19 pandemic, herein we report a consensus virtual screening of large collections of food chemicals and compounds classified as Dark Chemical Matter. The rationale is to complement global efforts and explore regions of the chemical space currently underexplored. The consensus approach included combining similarity searching with various queries and fingerprints, molecular docking with two docking programs, and ADMETox profiling. We propose three compounds commercially available that were sent to experimental testing. We disclose the full list of virtual screening hits that can be subject to additional selection for acquisition or synthesis and experimental testing. This manuscript will be updated when the experimental testing of the selected compounds becomes available.

Keywords: COVID-19; chemoinformatics; coronavirus; Dark Chemical Matter; data fusion; docking; drug discovery; food chemicals; Foodinformatics; SARS-CoV-2, similarity searching; virtual screening.

List of abbreviations: Angiotensin-converting enzyme, ACE; Angiotensin II, ANG-II; ADMETox; absorption, distribution, metabolism, excression, and toxicity; blood-brain barrier, BBB; Coronaviruses, COVs; Chemistry Development Kit, CDK; Dark Chemical Matter, DCM; Food and Drug Administration, FDA; Middle East Respiratory Syndrome, MERS; Machine learning, ML; Molecular Access System, MACCS; Molecular Operating Environment, MOE; Main protease, M^pro; Severe Acute Respiratory Syndrome, SARS; Universal Force Field, UFF; Virtual Screening, VS.

1. Introduction

Coronaviruses (COVs) *per se* have the ability to infect humans and other animal species, in fact, some of them cause a variety of previously studied diseases such as the Severe Acute Respiratory Syndrome (SARS) and Middle East Respiratory Syndrome (MERS). SARS-CoV-2 is an emergent virus that generates the COVID-19 disease [1] which is currently considered a "pandemic" according to the World Health Organization (WHO) registering more than 6.5 million people infected and causing more than 380,000 deaths Worldwide (as per June 1st, 2020) [2].

SARS-CoV-2 has a complex architecture; however, there are main proteins that are involved in the viral internalization or replication. These main targets offer a logical address for the most efficient and rational design and development of new treatments. For example, spike protein, RNA polymerase, and Chymotrypsin-like cysteine protease (3CLpro or M^{pro}) which are schematically presented in Figure 1 [3–5]. The main protease (M^{pro}) is an appropriate target for the design and proposal of new therapies due to the lack of homologous proteins in humans [6]. Several research groups are actively pursuing M^{pro} as a molecular target to identify drug candidates for the treatment of COVID-19.
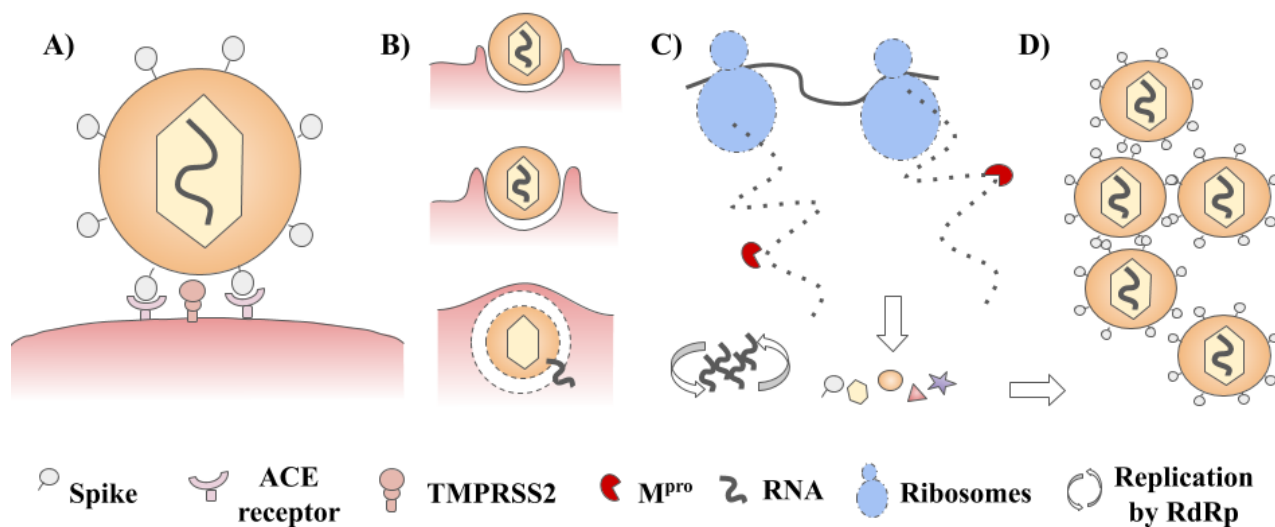

Figure 1. Schematic life cycle and main studied targets of SARS-CoV-2. A) Cellular recognition; B) Internalization and uncoating process; C) Biosynthesis of viral proteins and RNA replication; and D) Assembly of new virions.

Computational methods represent an approach with the power to efficiently filter large and diverse compound libraries to select potential candidates [7,8]. Recently published works show a tendency to explore the repurposing of approved drugs, as well as searching structurally diverse commercial libraries

(e.g., with large scaffold diversity), and natural products [9–13]. On the other hand, the search for novel compounds with the possibility of being synthesized or acquired commercially has had an important rebound (e.g. screening part or the entire ZINC database) [9,14–16]. Table 1 summarizes representative examples of virtual screening (VS) studies directed to different molecular targets, including SARS-COV-2 M[pro]. Most studies have used molecular docking and dynamics. Few others include similarity searching and quantitative structure-activity relationship (QSAR) modeling (Table 1) [17]. In this sense, there are a large number of compounds suggested by computational methods that could be easily evaluated with *in vitro* techniques, however, the use of computational consensus methodologies could improve further the performance of each method.

Table 1. Representative virtual screening studies to identify drug candidates for the treatment of COVID-19.

Target	Experimental methods	Libraries	Compounds screened / outcome	Ref.
M[pro]	Deep docking.	ZINC 15	1.3 billion / 1,000[a]	[9]
M[pro]	Pharmacophore model, molecular docking, and dynamics.	Marine natural products	14,064 / 17[a]	[10]
M[pro]	Pharmacophore screening and molecular docking.	ZINC	50,000 / 10[a]	[15]
Spike protein	Homology modeling and molecular docking.	FDA	3,300 / 12[a]	[18]
M[pro], PL[pro] and RdRp	Homology modeling, molecular docking, and dynamics.	DrugBank and Traditional Chinese Medicine	1,973 / 57[a]	[11]
ACE2	Molecular docking.	Literature compilation (natural products)	- / 5[a]	[12]
M[pro]	Molecular docking.	Literature compilation (natural products)	80 / 8[a]	[13]
M[pro]	Molecular docking.	FDA	486 / 20[a]	[19]
M[pro]	Molecular docking, and dynamics.	ZINC	606 million / 12[a]	[20]
M[pro]	Similarity search and QSAR modeling.	DrugBank (marketed, withdrawn, experimental, and investigational)	9,615 / 41[a]	[17]
M[pro]	Molecular docking and dynamics.	DrugBank (approved and drug candidates in clinical trials)	2,201 / 5[a]	[21]
M[pro] and TMPRSS2	Homology modeling and molecular docking.	ZINC	34,500 / 8[a]	[14]
M[pro]	Induced fit docking.	In-house	10,000 / 6[b]	[22]

[a] Computational hits; [b] Active hits.

The goal of this work was to propose active compounds against M^{pro} of SARS-CoV-2 and related coronavirus. One of the novelties of the present study relies on the region of the chemical space explored: food chemicals and molecules in the Dark Chemical Matter (DCM) which to the best of our knowledge, have been explored for SARS-CoV-2 on a limited basis. Thus, the rationale was to expand the search of chemical space and suggest molecules for experimental screening. Active compounds could be later optimized to increase activity. As a screening strategy we implemented consensus approaches combining similarity searching using different fingerprints and select hits using data fusion strategies. Selected compounds from similarity searching were further screened *in silico* using molecular docking with two docking programs. The final selection of computational hits was based on consensus scoring, information of protein-ligand contacts, and absorption, distribution, metabolism, excretion, and toxicity (ADMETox) profile. Additional criteria used to guide the selection of hit candidates for testing included predictions by machine learning (ML) models for SARS-CoV-2 activity developed by Collaborations Pharmaceuticals, Inc and freely available [23].

2. Materials and Methods

Herein we combined ligand- and structure-based methods to virtually screen compounds from two main molecular databases and select hit candidates for testing. Ligand-based methods were based on similarity searching using the principles of data fusion [24,25]. Structure-based approaches were based on molecular docking and consensus scoring [26]. The selection of hit compounds was also made considering the predicted ADMETox profile as well as prediction by ML models made freely available by Collaborations Pharmaceuticals. Figure 2 outlines the main VS strategy and hit selection. The workflow is described in the next subsections.
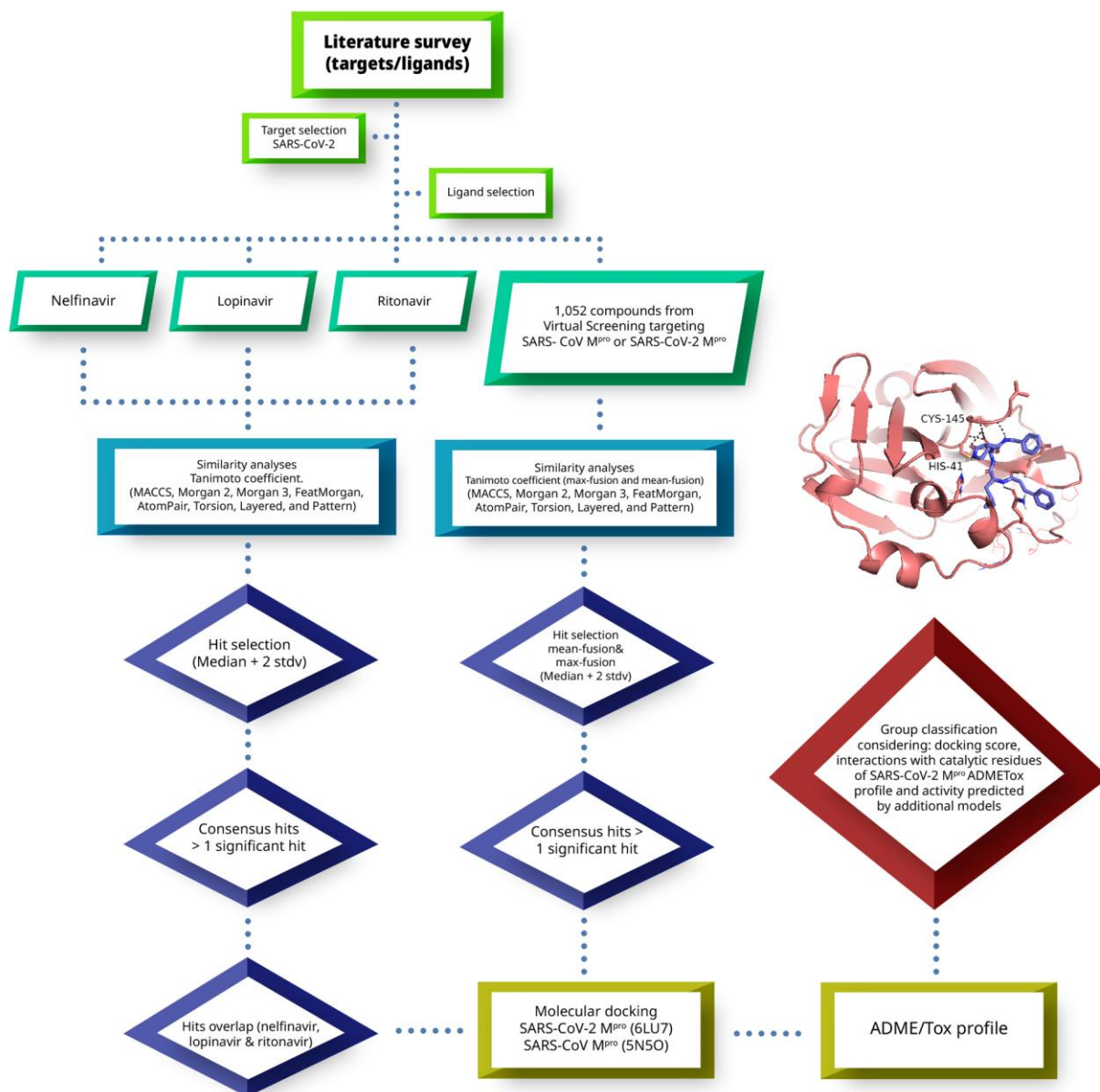
Figure 2. General workflow of the virtual screening approach used in this work.

2.1 Screening and Reference Databases

Table 2 summarizes the four major types of data sets considered in this study.

Table 2. Main screening data sets and reference compounds considered in this work.

Dataset	Content overview and size[a]	Rationale	Ref
Actives	N3, alpha-ketoamides **11a**, **11r**, and **11s**, carmofur, cinaserin, disulfiram, ebselen, PX12, shikonin, and tideglusib.	Reference compounds used in docking to compare docking scores and predicted binding modes.	[22,27]
FooDB	22,880 compounds	Large library of food chemicals. Smaller food chemical data sets have been screened.	[28]
Dark Chemical Matter (DCM)	139,329 compounds	Large screening library underexplored. Likelihood to shade light into the darkness of the COVID-19 pandemia.	[29]
ZINC (top ranked hits)	10 Top ranked virtual screening hits of ZINC using deep docking/Glide and SARS-CoV M^{pro} (PDB ID: 6LU7).	Further consensus of published computational hits with other docking programs (Vina and MOE).	[9]

[a] After data curation.

One of the screening databases was the public food chemical database, FooDB, with 23,883 compounds [28]. The chemical diversity and coverage of chemical space of FooDB has been reported revealing that food chemicals in this database are structurally diverse and have in general large molecular complexity [30]. The other main screening database was DCM that is a collection with 139,352 compounds that had shown no activity when tested in at least 100 screening assays [29]. Despite the fact that DCM has a low activity profile against common targets and biological endpoints, the rationale of screening DCM in this work was to explore regions in chemical space currently overlooked and complement global efforts. Moreover, DCM has yielded active molecules in other assays [31,32] probing the value to screen this region of the chemical space. The structures of FooDB and DCM were curated and standardized employing RDKit, CDK (Chemistry Development Kit), and ChemAxon tools. The largest component of molecules with more than one fragment was retained, compounds containing an atom type other than H, C, O, N, S, P, F, Cl, Br, I, B, Si, and Se were removed, and the tautomer with the lowest energy was generated.

Reference data sets were active compounds from the study of Jin et al. [22] including the peptide-like inhibitor N3, carmofur, cinaserin, disulfiram, ebselen, PX12, shikonin, tideglusib, and alpha-ketoamides (**11a**, **11r**, **11s**) [27]. Other reference compounds were lopinavir, nelfinavir, and ritonavir.

In order to identify additional potential hit compounds, we included the top 10 ranked virtual screening hits from the study of Ton et al. [9]. Authors of that work screened the ZINC database with the SARS-CoV-2 M^{pro} using the docking program Glide and the structure PDB ID: 6LU7. The rationale of using this set was to further explore the predicted profile of top-ranked compounds using different docking programs (i.e., Vina and MOE, *vide infra*).

2.2 Similarity Searching

Eight two-dimensional molecular fingerprints (Molecular ACCess System -MACCS- keys (166-bits), Morgan 2 [ECFP4-like], Morgan 3 [ECFP6-like], FeatMorgan, AtomPair, Torsion, Layered, and Pattern) were generated for all the queries, the 22,880 compounds in FooDB and the 139,329 molecules in DCM.

In the first virtual screening approach (Figure 2), nelfinavir, lopinavir, and ritonavir were used as independent queries. The molecular similarity between each of the queries and each of the molecules in FooDB and DCM was estimated with the Tanimoto coefficient [33]. The compounds with a Tanimoto coefficient higher than the median plus two standard deviations were considered as a hit. Of these, the molecules labeled as hits according to more than one molecular fingerprint (consensus hits), were selected. The consensus hits for the three queries were further analyzed by molecular docking.

In the second approach (Figure 2), 1,052 compounds with potential affinity for SARS-CoV-2 M^{pro} or SARS-CoV M^{pro} were selected from published molecular docking studies [9,19,34,35] and used as queries. Mean-fusion similarity scores and max-fusion similarity scores were determined using the eight molecular fingerprints and the Tanimoto coefficient [36]. Compounds with max-fusion similarity scores and mean-fusion similarity scores higher than the median plus two standard deviations for more than one fingerprint were selected as consensus hits and evaluated by molecular docking.

The molecular similarity analyses were generated in KNIME employing the RDKit node for molecular fingerprints generation and the CDK node for the similarity calculation [37,38].

2.3 Molecular Docking

In order to enhance the likelihood of finding active compounds, two docking programs with different docking and scoring algorithms were used, namely; Autodock Vina, version 1.1.2 [39] and Molecular Operating Environment (MOE) v.2019 [40].

Docking with Autodock Vina was conducted with two crystallographic structures obtained from the Protein Data Bank (PDB), [41] namely; SARS-CoV-2 M[pro] PDB ID: 6LU7 [22] and the structurally related SARS-CoV M[pro] PDB ID: 5N5O [42]. Both structures are co-crystallized with a peptide-like (**N3**) and an alpha-ketoamide (**11s**) inhibitor, respectively. The crystal structures were prepared in Autodock Tools. The grid-box was constructed based on the binding site of the alpha-ketoamide inhibitors **11a** and **11s**. The ligands were normalized, their clean 3D form was generated, hydrogens were added, and molecules were optimized using the Universal Force Field (UFF) in KNIME. The results were visualized in PyMol (version 2.3).

Induced fit docking protocol for the M[pro] (PDB ID: 6LU7) of SARS-CoV-2 was carried out with MOE software v.2019. The protein was prepared with the "Quick prepare" tool using the parameters established by default and the force field AMBER10:EHT. The peptide-like inhibitor **N3** was removed and their binding site was used to direct the docking. Triangle Matcher method was refined with the induced fit protocol, the other parameters were established by default. This protocol was validated using experimental información recently published by Jin et. al. [22]. The binding poses were successfully reproduced, and the binding score showed a correlation of 0.703 with the *in vitro* inhibition values of the data set.

2.4 ADME/Tox Profiling

Early consideration of ADMET/Tox properties is fundamental in current drug discovery efforts. Taking the advantage of the availability or several free chemoinformatic resources, [43] herein we employed the free server SwissADME[44] to calculate more than 40-related properties including, but not limited to, descriptors associated with drug-likeness, solubility, blood-brain barrier (BBB) permeability, Pgp substrate, Inhibition of CYPs, Bioavailability Score, PAINS alerts, and the number of violations to empirical rules (Lipinski, Veber, Egan, Brenk), among others. The full list of ADME/Tox related properties calculated with SwissADME is in the Supplementary Materials. We have used SwissADME to profile other compound databases of pharmaceutical relevance [45].

3. Results and Discussion

First, we describe the results of similarity searching, molecular docking, and ADMETox followed by the combined analysis to select hit compounds for experimental testing.

3.1 Similarity Searching

As discussed in the Introduction, M^{pro} is a promising drug target considering its importance in COVs life cycle (Figure 1, *vide supra*). The recent publication of the SARS-CoV-2 M^{pro} crystal structure showed a 96 % similarity with the SARS-CoV M^{pro} and the conservation of the active binding site. To search for SARS-Cov-2 M^{pro} inhibitors in underexplored regions of the chemical space, we assessed the molecular similarity of FooDB and DCM databases with compounds that potentially inhibit SARS-CoV M^{pro} or SARS-CoV-2 M^{pro}. As a first approach, three HIV-1 protease inhibitors approved for clinical use, namely; lopinavir, ritonavir, and nelfinavir were used as queries or reference compounds. Lopinavir and ritonavir have shown activity against SARS-CoV [46,47] and are currently under clinical trials for the treatment of COVID-19. Moreover, molecular dynamics predicted binding affinity of both molecules for the active site of SARS-CoV M^{pro} [48] and there is recent evidence of lopinavir *in vitro* activity against SARS-CoV-2 [49]. Another protease inhibitor with *in vitro* activity against SARS-CoV [50], nelfinavir, has been predicted to have a higher binding affinity to the SARS-CoV-2 M^{pro} by molecular dynamics [51,52]. Thus, nelfinavir was also included as a reference for the similarity search. Despite those observations, there is still no conclusive evidence of the effectiveness of these drugs in the treatment of COVID-19 [53–57] which encourage the identification of other existing molecules that target SARS-CoV-2.

After ligand preparation (described in the Methods section 2.3), 143 consensus hits in FooDB were found with Tanimoto coefficient values above the median plus two standard deviations for nelfinavir, lopinavir and ritonavir. From the 143 consensus hits, 40 compounds with drug-like properties were selected for further analyses. 500 compounds were selected from the DCM database with significantly high Tanimoto coefficient values for nelfinavir, lopinavir, and ritonavir. DCM compounds are constantly tested in HTS assays and therefore, they were considered to have suitable physicochemical properties for drug development. In this sense, it is not surprising that a larger number of consensus hits for the three drugs were found in DCM considering that molecular and physicochemical properties of DCM do not greatly differ from approved drugs, whereas FooDB was not assembled to be "drug-like".

To broaden the search of potential M^{pro} inhibitors, a small dataset of 1,052 compounds with predicted affinity to SARS-CoV-2 M^{pro} was assembled. Despite the fact these alternative reference compounds are potentially (but not confirmed) active, it has been suggested that they can increase the likelihood to identify active molecules. Such approach is reminiscent of what has been described as "turbo-similarity searching" [58]. As more data becomes available, a more chemically diverse and larger set could be integrated. Meanwhile, the top hits reported in six peer-reviewed molecular docking studies were included [9,19,34,35,59]. After ligand preparation, 178 and 174 consensus hits from FooDB and DCM were recovered, respectively. Significant hits were found for five of the eight molecular fingerprints, highlighting the advantages of using multiple molecular fingerprints [60].

Four compounds were overlapping consensus hits from both similarity searching methods. **DBB13044** and **DBB18117** from FooDB, and **DCM33835** and **DCM97265** from DCM database.

The total number of consensus hits further analyzed by molecular docking and ADMETox *in silico* profiling was 888 compounds (including stereoisomers).

3.2 Molecular Docking

Molecular docking of SARS-CoV M^{pro} was done with Autodock Vina (PDB ID: 5N5O). The docking scores for the reference compounds ranged from -8.5 to -4.1 kcal/mol, with a mean value of -6.8 kcal/mol. 393 compounds, from the hits selected by molecular similarity, fell above (less favorable) the mean docking score. However, reference compounds with docking scores above the mean value, such as ebselen (-6.2 kcal/mol) bound to the active site of SARS-CoV M^{pro} by four hydrogen bonds with residues Lys141, Gly143, Ser144, and Cys145. Hence, a hard cut-off value purely based on docking scores was not established. The docking scores for the reference compounds docked to SARS-CoV-2 M^{pro} carried out in MOE ranged from -9.4 to -5.16 kcal/mol.

Figure 3 shows the predicted binding mode with M^{pro} of representative hits compounds. As discussed hereunder in the section 3.4 Hit Selection (*vide infra*), the selected hit compounds shown in Figure 3 had favorable docking scores with Vina and MOE, and had at least one interaction with the catalytic residues His41, Cys145 and/or Glu 166 (key interactions reported) [22]. According to the docking models, other important key interactions were observed. **DBB2790** makes Pi-H interactions with sidechain of His 41, H-bond interaction with the sidechain of Cys 145 and H-bonds interactions with the sidechain and

backbone of Glu 166; **DCM78683** makes H-bond interactions with the sidechain of Asn 142 and Cys 145; and **DCM111769** makes Pi-H interactions with Glu 166. Our proposed compounds preferentially bind to the P1, P2 and P3 regions.

Of note, a literature survey revealed that the VS hit from our work **DBB2790** has a high structural similarity to compound **GC373** that has a nanomolar activity against M^{pro} from SARS-CoV-2 [61]. (the scaffold similarity of **GC373** and **DBB2790** is 0.96 based on the fingerprint MACCS keys and the Tanimoto coefficient). **GC373** was initially identified in 2013 as an inhibitor of M^{pro} of feline coronavirus [62]. Moreover, the protein-ligand interactions of both compounds and M^{pro} are similar. These observations support further the potential antiviral activity of **DBB2790**.
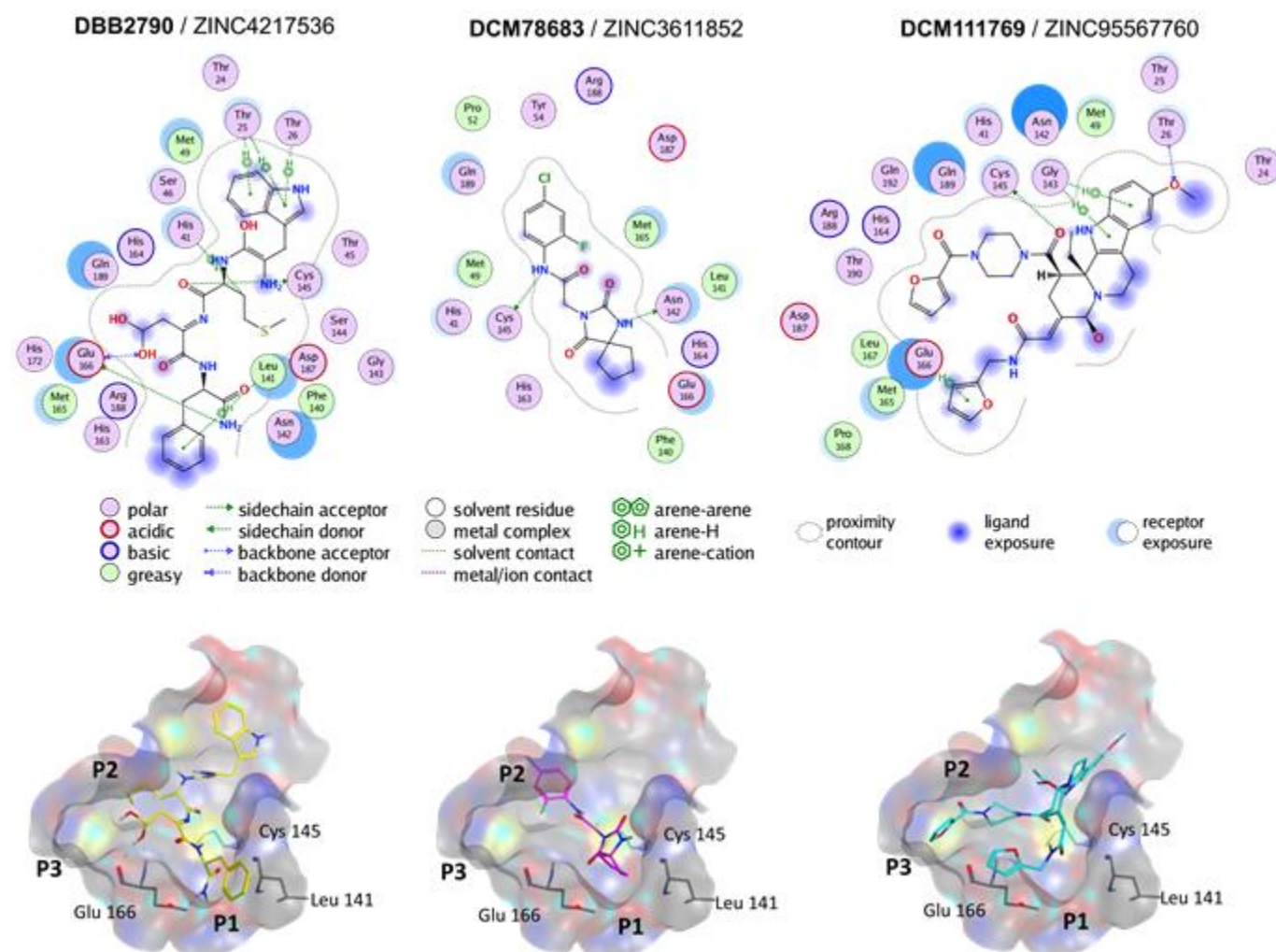

Figure 3. Binding modes predicted with Molecular Operating Environment of three selected hits with SARS-CoV-2 M^{pro} PDB ID: 6LU7.

3.3 ADMETox

For 888 selected hits the ADMETox-related descriptors were computed with SwissADME. As described in the next section, some of these descriptors were used as a guide for the classification of hit compounds in different priority groups. The main type of ADMETox descriptors that were considered were those associated with drug-likeness, solubility, and inhibition of cytochromes.

3.4 Hit Selection

Instead of establishing stringent (and arguably heuristic and hard) cut-off values, the compounds selected by molecular similarity were classified into four major groups considering their interactions with the catalytic residues of the SARS-CoV-2 M^{pro} (H41 and C145), their commercial availability, ADMETox characteristics, and their predicted activity by ML. Thereby, more compounds with suitable profiles were classified into one of the groups. The group number is associated with the priority to acquire and test the compounds so that compounds in group 1 would have the top priority, and so on. Table 3 summarizes the group classification strategy and the number of compounds that were classified into each group. A further description of each group is presented below

Table 3. Summary of the classification criteria to prioritize the compounds in four groups for testing. The number of compounds in each group is indicated.

Group	Number of compounds	Commercial availability[a]	*In silico* safety criteria[b]	Hydrogen bonds with H41 or C145	Active according to Machine Learning
1	41	Available	Safe	Present	Active/inactive
2	10	Available	Safe	Not present	Active
		Available	Not safe	Present	Active
3	34	Not available	Safe	Present	Active/inactive
4	20	Not available	Safe	Not present	Active
		Not available	Not safe	Present	Active

[a] Compounds reported as "in-stock" in the ZINC database were considered commercially available.
[b] Compounds that do not have PAINS alerts, do not pass through the BBB, and are predicted to not inhibit CYP1A2, CYP2C19, CYP2C9, CYP2D6 or CYP3A4.

Group 1 includes commercially available compounds that meet our safety criteria (based on the predictions of SwissADME), i.e. they do not have PAINS alerts, do not pass through the BBB, and do not inhibit CYP1A2, CYP2C19, CYP2C9, CYP2D6 or CYP3A4. The molecules in this group are also predicted to form hydrogen bonds with at least one of the catalytic residues of PDB ID: 6LU7. Table 4 summarizes the 41 molecules that fell into this top priority group .

Group 2 comprises 10 commercially available compounds that are predicted to be active by ML but violate one of the other two criteria. They can meet our safety criteria and do not form hydrogen bonds with the catalytic residues or they can form hydrogen bonds with the catalytic residues and do not meet our safety criteria.

Group 3 consists of 34 molecules that are not commercially available but meet the safety criteria and form hydrogen bonds with at least one of the catalytic residues. These compounds would be suited for synthesis and testing.

Group 4 contains 20 molecules that are not commercially available and are predicted to be active by ML but do not meet the safety criteria or do not form hydrogen bonds with the catalytic residues. According to our classification, compounds in this group would have the lowest priority for acquiring (synthesizing since they are not commercially available) and testing.

Table 4 summarizes the *in silico* profile of representative hit compounds selected for experimental validation.

Table 4. Virtual screening hits selected for experimental validation. The complete hit list (including those not currently on-stock in vendor catalogs) is available in the Supplementary Materials.

Set	ID	ZINC ID	Score 5N5O kcal/mol	Score 6LU7 kcal/mol	GI[a] absorption	Pgp[b] substrate	Ali[c] log_S	Ali Class	Lipinski violations	Brenk violations	Bioavailability[d]
foodb_mfsm	DBB9450	169676920	-6.6	-10.9	Low	Yes	-8.39	Poorly soluble	3	4	0.17
foodb_mfsm	DBB5554	85545908	-7.9	-10.9	Low	Yes	-6.76	Poorly soluble	3	2	0.17
foodb_mfsm	DBB2790	4217536	-7.8	-9.3	Low	Yes	-6.4	Poorly soluble	3	3	0.17
dcm_ch	DCM110214	34805301	-7.4	-9.2	Low	Yes	-2.7	Soluble	1	1	0.55

dcm_ch	DCM122034	15990331	-7	-8.9	High	Yes	-3.55	Soluble	0	1	0.55
dcm_ch	DCM73598	8918473	-7.2	-8.7	Low	Yes	-4.01	Moderately soluble	1	1	0.55
foodb_mfsm	DBB24555	53057130	-7.6	-8.6	Low	Yes	-4.76	Moderately soluble	1	1	0.55
dcm_ch	DCM22279	38144961	-6.8	-8.5	Low	Yes	-3.84	Soluble	1	1	0.55
dcm_ch	DCM82216	4270581	-7.1	-8.3	High	Yes	-2.36	Soluble	0	1	0.55
dcm_ch	DCM55533	8917865	-6.4	-8.3	High	Yes	-1.82	Very soluble	0	1	0.55
dcm_ch	DCM119353	9409555	-7.8	-8.2	Low	Yes	-4.18	Moderately soluble	0	0	0.56
dcm_ch	DCM65267	10077195	-6.2	-8.2	Low	Yes	-1.75	Very soluble	0	0	0.55
foodb_mfsm	DBB13825	4228235	-7.4	-8.1	Low	No	0.85	Highly soluble	2	4	0.17
dcm_ch	DCM131779	9159501	-6.4	-8	Low	Yes	-3.37	Soluble	1	1	0.55
dcm_ch	DCM65270	10077859	-7.2	-7.9	High	Yes	-0.63	Very soluble	0	0	0.55
dcm_ch	DCM82831	9109751	-7.8	-7.8	Low	No	-2.37	Soluble	0	1	0.55
foodb_mfsm	DBB13483	5283951	-6.3	-7.8	High	No	-3.7	Soluble	0	2	0.55
foodb_mfsm	DBB13002	2005305	-7.3	-7.8	Low	No	-3.27	Soluble	2	4	0.11
foodb_mfsm	DBB14163	8577218	-7.4	-7.7	Low	No	-2.11	Soluble	2	2	0.11
dcm_ch	DCM131783	15954557	-6.9	-7.7	High	Yes	-2.24	Soluble	0	1	0.55
dcm_ch	DCM93255[e]	32980237	-7.2	-7.7	High	No	-3.22	Soluble	0	1	0.55
foodb_mfsm	DBB13917	2036915	-7.5	-7.7	Low	No	-2.74	Soluble	2	2	0.11
dcm_mfsm	DCM116923[e]	2970717	-6.5	-7.7	High	Yes	-2.54	Soluble	0	1	0.55

dcm_ch	DCM10478	4083870	-6.6	-7.6	Low	No	0.02	Highly soluble	1	3	0.55
dcm_ch	DCM28770	100778693	-6.7	-7.6	High	Yes	-2.52	Soluble	0	0	0.55
dcm_ch	DCM33486[e]	1181094	-6.6	-7.5	High	Yes	-3.62	Soluble	0	1	0.55
dcm_ch	DCM30682[e]	1577795	-6.4	-7.5	High	No	-3.76	Soluble	0	2	0.55
dcm_ch	DCM110206[e]	12652624	-7.2	-7.5	High	Yes	-4.15	Moderately soluble	0	2	0.55
dcm_mfsm	DCM91011[e]	6754750	-7.7	-7.4	High	No	-1.49	Very soluble	0	1	0.55
foodb_mfsm	DBB13919	4228265	-7.7	-7.4	Low	No	-1.97	Very soluble	2	2	0.17
foodb_mfsm	DBB17132[e]	20431033	-6.2	-7.1	High	No	-1.7	Very soluble	0	2	0.55
dcm_ch	DCM131782[e]	2126038	-7.1	-7.1	High	No	-0.08	Very soluble	0	1	0.55
dcm_mfsm	DCM71724	18056800	-6.2	-7.1	Low	Yes	-4.37	Moderately soluble	0	2	0.55
dcm_mfsm	DCM94188[e]	18143600	-7.1	-6.9	High	No	-2.63	Soluble	0	0	0.55
foodb_mfsm	DBB20185	2242693	-6.1	-6.6	Low	No	0.98	Highly soluble	0	2	0.55
foodb_mfsm	DBB17114	4090721	-7	-6.5	High	No	-1.38	Very soluble	0	2	0.55
foodb_mfsm	DBB18961	4321512	-6.8	-6.5	Low	No	-0.96	Very soluble	0	0	0.55
foodb_mfsm	DBB18947	1303441	-6.1	-6.1	High	No	-0.42	Very soluble	0	0	0.55
foodb_mfsm	DBB19736	2040854	-5.4	-6.1	High	No	2.05	Highly soluble	0	0	0.55
foodb_mfsm	DBB19719	1532770	-5.6	-5.9	High	No	1.67	Highly soluble	0	0	0.55
foodb_mfsm	DBB21857[ef]	895813	-5.8	-5.6	High	No	-1.75	Very soluble	0	0	0.56

[a] GI Gastrointestinal
[b] Pgp P-glycoprotein
[c] Ali Topological method implemented from Ali J. et. al. 2012. [63]
[d] Probability that the compound will have F > 10%.
[e] Compounds that do not violate any of the following rules: Lipinski, Ghose, Veber, Egan, and Muegge.
[f] Compounds predicted to be active by the ML model.

Table 5 summarizes the information of 18 compounds listed in group 1 from FooDB with their corresponding IDs and annotated sources. Interestingly, some of the selected hits that were structurally similar to potential M^{pro} inhibitors were from endogenous sources. For instance, angiotensin II (**DBB9450**) and one of its degradation products, angiotensin IV (**DBB5554**) were predicted to bind to the active site of SARS-CoV-2 M^{pro} through an hydrogen-bond with His41, Ser46, Cys145, and Gln189 (**DBB9450**) and Thr26, Met49, Cys145, and Glu166 (**DBB5554**). Angiotensin II (ANG-II) is an octapeptide hormone product of the cleavage of angiotensin I by the angiotensin-converting enzyme (ACE). ANG-II binds to AT_1 and AT_2 receptors; the activation of AT_1 receptors by ANG-II induces vasoconstriction, vasopressin and aldosterone release, thirst, renal sodium reabsorption, angiogenesis, vascular aging, and inflammation. ANG-II can be converted to angiotensin 1-7 by the angiotensin-converting enzyme II (ACE2). The consecutive action of aminopeptidase A and aminopeptidase N produces angiotensin III and angiotensin IV, respectively.

Angiotensin 1-7 has opposite actions to ANG-II. Because ACE2 mediates the entry of SARS-CoV-2 to the host cells and ACE2 activity may be downregulated after virus infection, the accumulation of ANG-II could be linked to the development of severe symptoms of COVID-19 disease. If M^{pro} inhibitors are structurally similar to ANG-II, their potential binding affinity for the active site of ACE2 should be evaluated. Some studies have assessed the ability of ACE2 inhibitors to prevent SARS-CoV from entering into the cells [64]. However, the inhibition of the ACE2 function could cause overaccumulation of ANG-II and promote its undesired effects.

Nonetheless, probably, DCM compounds may not elicit a dual inhibition of SARS-CoV-2 M^{pro} and ACE2, considering that these molecules had shown no activity against common targets evaluated in HTS assays.

Food folates like 5-MTHF, folic acid, dihydrofolic acid, and tetrahydrofolate (Table 5) were also among the compounds in the top priority group with observed hydrogen bonds with the catalytic residues of the SARS-CoV-2 M^{pro} and favorable docking scores below -7.4 kcal/mol. Folates are cofactors in many one-carbon transfer reactions including nucleotide synthesis for DNA and RNA synthesis, interconversion of serine and glycine, methionine generation and methylation of histones, DNA, proteins, phospholipids, and neurotransmitters. Folate deficiency has been linked to neural tube defects, brain

dysfunction, coronary heart disease, and increased risk of colorectal and breast cancer [65]. Since mammalian cells cannot synthesize *de novo* folate, naturally occuring food folates and synthetic folic acid are used in dietary supplements and fortified food. Nevertheless, recent studies have shown that a higher intake of folic acid may be associated with a higher risk of developing leukemia and other cancers, arthritis, lower natural cell killer activity, higher insulin resistance, and masking vitamin B12 deficiency [66]. Accordingly, the implications of low and high plasma levels of folates in COVID-19 patients must be evaluated. Our results suggest that folates could inhibit SARS-CoV-2 M^{pro}, but their effect in *in vitro* and *in vivo* assays remains to be evaluated. To the extend our knowledge of the effects of a healthy diet, and the specific mechanisms through which food chemicals participate in the progression of COVID-19 disease could be a simple but ready-to-use approach for the prevention and combat of the current pandemic.

Intriguingly, aminopterin (**DBB13917**), a folic acid analog that inhibits the dihydrofolate reductase enzyme was also a potential M^{pro} inhibitor. Aminopterin is one of the so-called antifolates that interfere with folate metabolism, and in turn nucleotide synthesis. Currently, an aminopterin analog with less toxic effects, methotrexate, is under clinical trials for the treatment of COVID-19 disease (NCT04352465). Methotrexate is an immunosuppressant used in the treatment of cancer and inflammatory conditions that is often concurrently administered with folic acid.

Table 5. Representative food chemicals as hits in the virtual screening.

IDs	FooDB annotation
DBB9450 / FDB022383	Angiotensin II, endogenous
DBB5554 / FDB022385	Angiotensin IV
DBB2790 / FDB023765	Tetragastrin, endogenous
DBB2455 / FDB023767	Morphiceptin, endogenous
DBB13825 / FDB031192	Tetrahydrofolate
DBB13483 / FDB013079	Neotame, artificial sweetener
DBB13002 / FDB022600	5 methyltetrahydrofolic acid (5-MTHF)
DBB14163 / FDB014504	Folic acid
DBB13917 / FDB022702	Aminopterin
DBB13919 / FDB022395	Dihydrofolic acid

DBB17132 / FDB028374	Phenylbutyrylglutamine, metabolite of phenylbutyrate
DBB20185 / FDB003618	gamma-L-Glutamyl-L-phenylalanine, Soft-necked garlic
DBB17114 / FDB029352	Indole acetyl glutamine, endogenous
DBB18961 / FDB023789	N4-Acetylcytidine, endogenous
DBB18947 / FDB022917	5-Methyldeoxycytidine (5-mdc)
DBB19736 / FDB012937	Carnosine 44A
DBB19719 / FDB022217	Homocarnosine, metabolite
DBB21857 / FDB022212	Hydroxyphenylacetylglycine, endogenous human metabolite

3.5 Top ranked hits from deep docking of ZINC

The 10 top-ranked compounds from the analysis conducted by Ton et. al. were included in this study (*vide supra*).

Even though the ML model did not predict activity against the main protease for these molecules, they represent interesting hits selected from billions of compounds in the ZINC database. They had good docking scores in our analyses and three of them **ZINC1218583693**, **ZINC1186058814**, and **ZINC1655436520** met our safety criteria and had interactions with the catalytic residues of SARS-CoV-2 M^{pro}. Furthermore, **ZINC1655436520** also formed hydrogen bonds with residues Phe140, Leu141, Gly143, Ser144, Cys145, and Glu166 of SARS-CoV M^{pro}, it is predicted to have good water solubility and high GI absorption, and it does not violate Lipinski's, Ghose, Veber, Egan or Muegge rules.

4. Conclusions and Perspectives

Herein we report a consensus structure- and ligand-based virtual screening of two large chemical databases, namely; food chemicals (22,880 compounds) and Dark Chemical Matter (139,329 compounds) to identify potential drug candidates for the treatment of COVID-19 targeting the SARS-CoV-2 M^{pro}. This work is part of our continued effort to identify systematically bioactive food chemicals [67]. We also screened top-ranked hits identified in a separate VS of 1.6 billion molecules from ZINC using Glide [9]. The similarity searching was done following two major approaches. The first method yielded 40 molecules from drug-like food chemicals and 500 molecules from the DCM collection that were significantly similar to nelfinavir, lopinavir, and ritonavir. While the data fusion approach returned

178 food chemicals and 174 DCM compounds. In total, 888 hit compounds were subject to molecular docking with Vina and MOE. The hit compounds were analyzed and selected considering several criteria including docking score, predicted interactions with major residues including the catalytic His41, Cys145, and Glu166, and ADMETox profiling. An additional criterion that was used as a guide was a prediction by ML models developed by collaborators in North Carolina, USA [68]. After the selection criteria, 105 hits in total were identified, of which, three are commercially available (and with reasonable prices) and were sent for experimental testing. The full list of hits compounds annotated with the *in silico* profile is available in the Supplementary Materials. This manuscript will be updated when the experimental information becomes available.

Supplementary Materials: Excel file with ten worksheets that report all similarity values, docking scores, and ADMETox profile of the hit compounds outlined in Figure 2.

Conflicts of Interest: The authors declare no conflict of interest.

References

1. de Wit, E.; van Doremalen, N.; Falzarano, D.; Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **2016**, *14*, 523–534, doi:10.1038/nrmicro.2016.81.

2. Worldometer Coronavirus update Available online: https://www.worldometers.info/coronavirus/ (accessed on May 8, 2020).

3. Neuman, B. W.; Buchmeier, M. J. Supramolecular architecture of the coronavirus particle. *Adv. Virus Res.* **2016**, *96*, 1–27, doi:10.1016/bs.aivir.2016.08.005.

4. Morse, J. S.; Lalonde, T.; Xu, S.; Liu, W. R. Learning from the past: Possible urgent prevention and treatment options for severe acute respiratory infections caused by 2019-nCoV. *Chembiochem* **2020**, *21*, 730–738, doi:10.1002/cbic.202000047.

5. Lee, H.; Mittal, A.; Patel, K.; Gatuz, J. L.; Truong, L.; Torres, J.; Mulhearn, D. C.; Johnson, M. E. Identification of novel drug scaffolds for inhibition of SARS-CoV 3-Chymotrypsin-like protease using virtual and high-throughput screenings. *Bioorg. Med. Chem.* **2014**, *22*, 167–177, doi:10.1016/j.bmc.2013.11.041.

6. Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S.-H. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL rotease inhibitors: peptidomimetics and small molecule chemotherapy. *J. Med. Chem.* **2016**, *59*, 6595–6628, doi:10.1021/acs.jmedchem.5b01461.

7. Prieto-Martínez, F. D.; López-López, E.; Eurídice Juárez-Mercado, K.; Medina-Franco, J. L. Computational drug design methods—current and future perspectives. In *In silico drug design*; Elsevier, 2019; pp. 19–44 ISBN 9780128161258.

8. López-López, E.; Barrientos-Salcedo, C.; Prieto-Martínez, F. D.; Medina-Franco, J. L. In silico tools to study molecular targets of neglected diseases: inhibition of TcSir2rp3, an epigenetic enzyme of Trypanosoma cruzi. In *Advances from Computational and Experimental Studies of Chromatin Structure and Epigenetics Mechanisms*; 2020 ISBN 1876-1623.

9. Ton, A.-T.; Gentile, F.; Hsing, M.; Ban, F.; Cherkasov, A. Rapid Identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol. Inform.* **2020**, doi:10.1002/minf.202000028.

10. Gentile, D.; Patamia, V.; Scala, A.; Sciortino, M. T.; Piperno, A.; Rescifina, A. Putative inhibitors of SARS-CoV-2 Main protease from a library of marine natural products: A virtual screening and molecular modeling study. *Mar. Drugs* **2020**, *18*, doi:10.3390/md18040225.

11. Virtual Screening and molecular dynamics on blockage of key drug targets as treatment for COVID-19 caused by SARS-CoV-2 [v1] Preprints Available online: https://www.preprints.org/manuscript/202003.0239/v1 (accessed on Jun 1, 2020).

12. Chen, H.; Du, Q. Potential natural compounds for preventing SARS-CoV-2 (2019-nCoV) infection. **2020**, doi:10.20944/preprints202001.0358.v3.

13. Adem, S.; Eyupoglu, V.; Sarfraz, I.; Rasul, A.; Ali, M. Identification of potent COVID-19 main protease (mpro) inhibitors from natural polyphenols: An in silico strategy unveils a hope against

CORONA. **2020**, doi:10.20944/preprints202003.0333.v1.

14. Elmezayen, A. D.; Al-Obaidi, A.; Şahin, A. T.; Yelekçi, K. Drug repurposing for coronavirus (COVID-19): in silico screening of known drugs against coronavirus 3CL hydrolase and protease enzymes. *J. Biomol. Struct. Dyn.* **2020**, 1–12, doi:10.1080/07391102.2020.1758791.

15. Andrade, B.; Ghosh, P.; Barth, D.; Tiwari, S.; José Santana Silva, R.; Rodrigues de Assis Soares, W.; Silva Melo, T.; dos Santos Freitas, A.; González-Grande, P.; Sousa Palmeira, L.; Carlos Junior Alcantara, L.; Giovanetti, M.; Góes-Neto, A.; Ariston de Carvalho Azevedo, V. Computational screening for potential drug candidates against sars-cov-2 main protease. **2020**, doi:10.20944/preprints202004.0003.v1.

16. Fischer, A.; Sellner, M.; Neranjan, S.; Lill, M. A.; Smieško, M. Inhibitors for novel coronavirus protease identified by virtual screening of 687 million compounds. **2020**, doi:10.26434/chemrxiv.11923239.v1.

17. Bobrowski, T.; Alves, V.; Melo-Filho, C. C.; Korn, D.; Auerbach, S. S.; Schmitt, C.; Muratov, E.; Tropsha, A. Computational models identify several FDA approved or experimental drugs as putative agents against SARS-CoV-2. **2020**, doi:10.26434/chemrxiv.12153594.v1.

18. Senathilake, K.; Samarakoon, S.; Tennekoon, K. Virtual screening of inhibitors against spike glycoprotein of SARS-CoV-2: A drug repurposing approach. **2020**, doi:10.20944/preprints202003.0042.v2.

19. Kandeel, M.; Al-Nazawi, M. Virtual screening and repurposing of FDA approved drugs against COVID-19 main protease. *Life Sci.* **2020**, *251*, 117627, doi:10.1016/j.lfs.2020.117627.

20. Fischer, A.; Sellner, M.; Neranjan, S.; Smieško, M.; Lill, M. A. Potential inhibitors for novel coronavirus protease identified by virtual screening of 606 million compounds. *ijms* **2020**, *21*, 3626, doi:10.3390/ijms21103626.

21. Wang, J. Fast Identification of possible drug treatment of coronavirus disease-19 (COVID-19) through computational drug repurposing study. *J. Chem. Inf. Model.* **2020**, doi:10.1021/acs.jcim.0c00179.

22. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; Duan, Y.; Yu, J.; Wang, L.; Yang, K.; Liu, F.; Jiang, R.; Yang, X.; You, T.; Liu, X.; Yang, X.; Yang, H. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, doi:10.1038/s41586-020-2223-y.

23. Molecular Materials Informatics, Inc. Assay Central Predictions Available online: https://assaycentral.github.io/ (accessed on May 19, 2020).

24. Willett, P. Combination of similarity rankings using data fusion. *J. Chem. Inf. Model.* **2013**, *53*, 1–10, doi:10.1021/ci300547g.

25. Chen, B.; Mueller, C.; Willett, P. Combination rules for group fusion in similarity-based virtual screening. *Mol. Inform.* **2010**, *29*, 533–541, doi:10.1002/minf.201000050.

26. Feher, M. Consensus scoring for protein-ligand interactions. *Drug Discov. Today* **2006**, *11*, 421–428, doi:10.1016/j.drudis.2006.03.009.

27. Zhang, L.; Lin, D.; Kusov, Y.; Nian, Y.; Ma, Q.; Wang, J.; von Brunn, A.; Leyssen, P.; Lanko, K.; Neyts, J.; de Wilde, A.; Snijder, E. J.; Liu, H.; Hilgenfeld, R. α-Ketoamides as broad-spectrum inhibitors of coronavirus and enterovirus replication: structure-based design, synthesis, and activity assessment. *J. Med. Chem.* **2020**, *63*, 4562–4578, doi:10.1021/acs.jmedchem.9b01828.

28. Computer software, Canada: The Metabolomics Innovation Centre. The Metabolomics Innovation Centre: FooDB (Version 1). Available online: https://foodb.ca/ (accessed on May 19, 2020).

29. Wassermann, A. M.; Lounkine, E.; Hoepfner, D.; Le Goff, G.; King, F. J.; Studer, C.; Peltier, J. M.; Grippo, M. L.; Prindle, V.; Tao, J.; Schuffenhauer, A.; Wallace, I. M.; Chen, S.; Krastel, P.; Cobos-Correa, A.; Parker, C. N.; Davies, J. W.; Glick, M. Dark chemical matter as a promising starting point for drug lead discovery. *Nat. Chem. Biol.* **2015**, *11*, 958–966, doi:10.1038/nchembio.1936.

30. Naveja, J. J.; Rico-Hidalgo, M. P.; Medina-Franco, J. L. Analysis of a large food chemical database: chemical space, diversity, and complexity. [version 2; peer review: 3 approved]. *F1000Res.* **2018**, *7*, doi:10.12688/f1000research.15440.2.

31. Ballante, F.; Rudling, A.; Zeifman, A.; Luttens, A.; Vo, D. D.; Irwin, J. J.; Kihlberg, J.; Brea, J.; Loza, M. I.; Carlsson, J. Docking finds GPCR ligands in dark chemical matter. *J. Med. Chem.* **2020**, *63*, 613–620, doi:10.1021/acs.jmedchem.9b01560.

32. Wassermann, A. M.; Tudor, M.; Glick, M. Deorphanization strategies for dark chemical matter. *Drug Discov. Today Technol.* **2017**, *23*, 69–74, doi:10.1016/j.ddtec.2016.11.004.

33. Bajusz, D.; Rácz, A.; Héberger, K. Why is Tanimoto index an appropriate choice for fingerprint-based similarity calculations? *J. Cheminform.* **2015**, *7*, 20, doi:10.1186/s13321-015-0069-3.

34. Chen, Y. W.; Yiu, C.-P. B.; Wong, K.-Y. Prediction of the SARS-CoV-2 (2019-nCoV) 3C-like protease (3CL pro) structure: virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates. [version 2; peer review: 3 approved]. *F1000Res.* **2020**, *9*, 129, doi:10.12688/f1000research.22457.2.

35. Liu, S.; Zheng, Q.; Wang, Z. Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus. *Bioinformatics* **2020**, doi:10.1093/bioinformatics/btaa224.

36. Medina-Franco, J. L.; Maggiora, G. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. A similarity-based data-fusion approach to the visual characterization and comparison of compound databases. *Chem. Biol. Drug Des.* **2007**, *70*, 393–412, doi:10.1111/j.1747-0285.2007.00579.x.

37. Beisken, S.; Meinl, T.; Wiswedel, B.; de Figueiredo, L. F.; Berthold, M.; Steinbeck, C. KNIME-CDK: Workflow-driven cheminformatics. *BMC Bioinformatics* **2013**, *14*, 257, doi:10.1186/1471-2105-14-257.

38. Berthold, M. R.; Cebron, N.; Dill, F.; Gabriel, T. R.; Kötter, T.; Meinl, T.; Ohl, P.; Sieb, C.; Thiel, K.; Wiswedel, B. KNIME: The Konstanz Information Miner. In *Data Analysis, Machine Learning and Applications*; Preisach, C., Burkhardt, H., Schmidt-Thieme, L., Decker, R., Eds.; Studies in Classification, Data Analysis, and Knowledge Organization; Springer Berlin Heidelberg: Berlin, Heidelberg, 2008; pp. 319–326 ISBN 978-3-540-78239-1.

39. 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*, 455–461, doi:10.1002/jcc.21334.

40. Prieto-Martínez, F. D.; Arciniega, M.; Medina-Franco, J. L. Acoplamiento molecular: avances recientes y retos. *TIP RECQB* **2018**, *21*, doi:10.22201/fesz.23958723e.2018.0.143.

41. Berman, H. M.; Westbrook, J.; Feng, Z.; Gilliland, G.; Bhat, T. N.; Weissig, H.; Shindyalov, I. N.; Bourne, P. E. The Protein Data Bank. *Nucleic Acids Res.* **2000**, *28*, 235–242, doi:10.1093/nar/28.1.235.

42. Zhang, L.,; Hilgenfeld, R. Structure of SARS coronavirus main protease in complex with the alpha-ketoamide (S)-N-benzyl-3-((S)-2-cinnamamido-3-cyclopropylpropanamido)-2-oxo-4-((S)-2-oxopyrrolidin-3-yl)butanamide (Cinnamoyl-cyclopropylalanine-GlnLactam-CO-CO-NH-benzyl) Available online: http://doi.org/10.2210/pdb5N5O/pdb (accessed on May 18, 2020).

43. González-Medina, M.; Naveja, J. J.; Sánchez-Cruz, N.; Medina-Franco, J. L. Open chemoinformatic resources to explore the structure, properties and chemical space of molecules. *RSC Adv.* **2017**, *7*, 54153–54163, doi:10.1039/C7RA11831G.

44. Daina, A.; Michielin, O.; Zoete, V. SwissADME: a free web tool to evaluate pharmacokinetics, drug-likeness and medicinal chemistry friendliness of small molecules. *Sci. Rep.* **2017**, *7*, 42717, doi:10.1038/srep42717.

45. Durán-Iturbide, N. A.; Díaz-Eufracio, B. I.; Medina-Franco, J. L. In silico ADME/Tox profiling of natural products: A focus on BIOFACQUIM. *ACS Omega* **2020**.

46. Treatment of severe acute respiratory syndrome with lopinavir/ritonavir: a multicentre retrospective matched cohort study | HKMJ Available online: https://www.hkmj.org/abstracts/v9n6/399.htm (accessed on Jun 1, 2020).

47. Chu, C. M.; Cheng, V. C. C.; Hung, I. F. N.; Wong, M. M. L.; Chan, K. H.; Chan, K. S.; Kao, R. Y. T.; Poon, L. L. M.; Wong, C. L. P.; Guan, Y.; Peiris, J. S. M.; Yuen, K. Y.; HKU/UCH SARS Study Group role of lopinavir/ritonavir in the treatment of SARS: initial virological and clinical findings. *Thorax* **2004**, *59*, 252–256, doi:10.1136/thorax.2003.012658.

48. Nukoolkarn, V.; Lee, V. S.; Malaisree, M.; Aruksakulwong, O.; Hannongbua, S. Molecular dynamic simulations analysis of ritonavir and lopinavir as SARS-CoV 3CL(pro) inhibitors. *J. Theor. Biol.* **2008**, *254*, 861–867, doi:10.1016/j.jtbi.2008.07.030.

49. Choy, K.-T.; Wong, A. Y.-L.; Kaewpreedee, P.; Sia, S. F.; Chen, D.; Hui, K. P. Y.; Chu, D. K. W.; Chan, M. C. W.; Cheung, P. P.-H.; Huang, X.; Peiris, M.; Yen, H.-L. Remdesivir, lopinavir, emetine, and homoharringtonine inhibit SARS-CoV-2 replication in vitro. *Antiviral Res.* **2020**, *178*, 104786, doi:10.1016/j.antiviral.2020.104786.

50. Yamamoto, N.; Yang, R.; Yoshinaka, Y.; Amari, S.; Nakano, T.; Cinatl, J.; Rabenau, H.; Doerr, H. W.; Hunsmann, G.; Otaka, A.; Tamamura, H.; Fujii, N.; Yamamoto, N. HIV protease inhibitor nelfinavir inhibits replication of SARS-associated coronavirus. *Biochem. Biophys. Res. Commun.* **2004**, *318*, 719–725, doi:10.1016/j.bbrc.2004.04.083.

51. Khaerunnisa, S.; Kurniawan, H.; Awaluddin, R.; Suhartati, S.; Soetjipto, S. Potential inhibitor of

COVID-19 main protease (M^pro) from several medicinal plant compounds by molecular docking study. **2020**, doi:10.20944/preprints202003.0226.v1.

52. Xu, Z.; Peng, C.; Shi, Y.; Zhu, Z.; Mu, K.; Wang, X.; Zhu, W. Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. *BioRxiv* **2020**, doi:10.1101/2020.01.27.921627.

53. Cai, Q.; Huang, D.; Ou, P.; Yu, H.; Zhu, Z.; Xia, Z.; Su, Y.; Ma, Z.; Zhang, Y.; Li, Z.; He, Q.; Fu, Y.; Liu, L.; Chen, J. COVID-19 in a Designated Infectious Diseases HospitalOutside Hubei Province,China. *medRxiv* **2020**, doi:10.1101/2020.02.17.20024018.

54. Cao, B.; Wang, Y.; Wen, D.; Liu, W.; Wang, J.; Fan, G.; Ruan, L.; Song, B.; Cai, Y.; Wei, M.; Li, X.; Xia, J.; Chen, N.; Xiang, J.; Yu, T.; Bai, T.; Xie, X.; Zhang, L.; Li, C.; Yuan, Y.; Wang, C. A Trial of lopinavir-ritonavir in adults hospitalized with severe COVID-19. *N. Engl. J. Med.* **2020**, *382*, 1787–1799, doi:10.1056/NEJMoa2001282.

55. Hu, L.; Chen, S.; Fu, Y.; Gao, Z.; Long, H.; Ren, H.; Zuo, Y.; Li, H.; Wang, J.; Xv, Q.; Yu, W.; Liu, J.; Shao, C.; Hao, J.; Wang, C.; Ma, Y.; Wang, Z.; Yanagihara, R.; Wang, J.; Deng, Y. Risk factors associated with clinical outcomes in 323 COVID-19 patients in Wuhan, China. *medRxiv* **2020**, doi:10.1101/2020.03.25.20037721.

56. Li, Y.; Xie, Z.; Lin, W.; Cai, W.; Wen, C.; Guan, Y.; Mo, X.; Wang, J.; Wang, Y.; Peng, P.; Chen, X.; Hong, W.; Xiao, G.; Liu, J.; Zhang, L.; Hu, F.; Li, F.; Li, F.; Zhang, F.; Deng, X.; Li, L. An exploratory randomized, controlled study on the efficacy and safety of lopinavir/ritonavir or arbidol treating adult patients hospitalized with mild/moderate COVID-19 (ELACOI). *medRxiv* **2020**, doi:10.1101/2020.03.19.20038984.

57. Yan, D.; Liu, X.; Zhu, Y.; Huang, L.; Dan, B.; Zhang, G.; Gao, Y. Factors associated with prolonged viral shedding and impact of Lopinavir/Ritonavir treatment in patients with SARS-CoV-2 infection. *medRxiv* **2020**, doi:10.1101/2020.03.22.20040832.

58. Gardiner, E. J.; Gillet, V. J.; Haranczyk, M.; Hert, J.; Holliday, J. D.; Malim, N.; Patel, Y.; Willett, P. Turbo similarity searching: Effect of fingerprint and dataset on virtual-screening performance. *Stat. Anal. Data Min.* **2009**, *2*, 103–114, doi:10.1002/sam.10037.

59. Liu, X.; Wang, X.-J. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *J. Genet. Genomics* **2020**, *47*, 119–121, doi:10.1016/j.jgg.2020.02.001.

60. Medina-Franco, J. L.; Martínez-Mayorga, K.; Bender, A.; Marín, R. M.; Giulianotti, M. A.; Pinilla, C.; Houghten, R. A. Characterization of activity landscapes using 2D and 3D similarity methods: consensus activity cliffs. *J. Chem. Inf. Model.* **2009**, *49*, 477–491, doi:10.1021/ci800379q.

61. Vuong, W.; Khan, M. B.; Fischer, C.; Arutyunova, E.; Lamer, T.; Shields, J.; Saffran, H. A.; McKay, R. T.; van Belkum, M. J.; Joyce, M.; Young, H. S.; Tyrrell, D. L.; Vederas, J. C.; Lemieux, M. J. Feline coronavirus drug inhibits the main protease of SARS-CoV-2 and blocks virus replication. *BioRxiv* **2020**, doi:10.1101/2020.05.03.073080.

62. Kim, Y.; Mandadapu, S. R.; Groutas, W. C.; Chang, K.-O. Potent inhibition of feline coronaviruses

with peptidyl compounds targeting coronavirus 3C-like protease. *Antiviral Res.* **2013**, *97*, 161–168, doi:10.1016/j.antiviral.2012.11.005.

63. Ali, J.; Camilleri, P.; Brown, M. B.; Hutt, A. J.; Kirton, S. B. Revisiting the general solubility equation: in silico prediction of aqueous solubility incorporating the effect of topographical polar surface area. *J. Chem. Inf. Model.* **2012**, *52*, 420–428, doi:10.1021/ci200387c.

64. Huentelman, M. J.; Zubcevic, J.; Hernández Prada, J. A.; Xiao, X.; Dimitrov, D. S.; Raizada, M. K.; Ostrov, D. A. Structure-based discovery of a novel angiotensin-converting enzyme 2 inhibitor. *Hypertension* **2004**, *44*, 903–906, doi:10.1161/01.HYP.0000146120.29648.36.

65. Ames, B. N. Micronutrient deficiencies. A major cause of DNA damage. *Ann. N. Y. Acad. Sci.* **1999**, *889*, 87–106, doi:10.1111/j.1749-6632.1999.tb08727.x.

66. Selhub, J.; Rosenberg, I. H. Excessive folic acid intake and relation to adverse health outcome. *Biochimie* **2016**, *126*, 71–78, doi:10.1016/j.biochi.2016.04.010.

67. Peña-Castillo, A.; Méndez-Lucio, O.; Owen, J. R.; Martínez-Mayorga, K.; Medina-Franco, J. L. Chemoinformatics in food science. In *Applied Chemoinformatics: Achievements and Future Opportunities*; Engel, T., Gasteiger, J., Eds.; Wiley-VCH Verlag GmbH & Co. KGaA: Weinheim, Germany, 2018; pp. 501–525 ISBN 9783527342013.

68. Ekins, S.; Mottin, M.; Ramos, P. R. P. S.; Sousa, B. K. P.; Neves, B. J.; Foil, D. H.; Zorn, K. M.; Braga, R. C.; Coffee, M.; Southan, C.; Puhl, A. C.; Andrade, C. H. Déjà vu: Stimulating open drug discovery for SARS-CoV-2. *Drug Discov. Today* **2020**, doi:10.1016/j.drudis.2020.03.019.