

Aus dem Institut für Physiologie
der Medizinischen Fakultät Charité – Universitätsmedizin Berlin

DISSERTATION

Computergestützte Methoden in der Arzneimittel-
Repositionierung und der Naturstoffbasierten Wirkstoffforschung

Computational Methods in Drug Repurposing
and Natural Product Based Drug Discovery

zur Erlangung des akademischen Grades
Doctor rerum medicinalium (Dr. rer. medic.)

von
Renata Ewelina Abel

Datum der Promotion: 30.11.2023

LIST OF TABLES	IV
LIST OF FIGURES	V
LIST OF ABBREVIATIONS	VI
ABSTRACT	VII
ZUSAMMENFASSUNG	VIII
1. INTRODUCTION	1
1.1 AIM OF THE THESIS.....	1
1.2 THE ROLE OF NATURAL PRODUCTS IN DRUG DISCOVERY	2
1.3 THE ROLE OF COMPUTATIONAL METHODS IN DRUG DISCOVERY AND DRUG REPURPOSING	5
2. METHODS	8
2.1 LIGAND-BASED VIRTUAL SCREENING (LBVS)	9
2.2 MOLECULAR DOCKING	11
2.3 MOLECULAR DYNAMICS	12
2.4 TOXICITY AND CYTOCHROME ACTIVITY PREDICTION	13
2.5 COMPOUNDS AND DATABASES	13
2.6 IN VITRO STUDIES	14
3. RESULTS	15
3.1 PROJECT I	15
3.2 PROJECT II	22
4. DISCUSSION	25
4.1 SHORT SUMMARY OF THE RESULTS	25
4.2 CURRENT STATE OF KNOWLEDGE ON TREATMENT OF COVID-19	25
4.3 STRENGTHS AND WEAKNESSES OF THE STUDY	27
5. CONCLUSION	29
6. REFERENCES	31
STATUTORY DECLARATION	37
DECLARATION OF YOUR OWN CONTRIBUTION TO THE PUBLICATIONS	38
EXCERPT FROM JOURNAL SUMMARY	40
PRINTING COPIES OF THE PUBLICATIONS	41
CURRICULUM VITAE	79
LIST OF ALL PUBLICATIONS:	80
ACKNOWLEDGMENTS	82

List of tables

Table 1: Examples of antibacterial and anticancer natural product-based drugs (15). Abbreviations: N-natural products, ND- natural products derivatives.	4
Table 2: Top 40 compounds, potential inhibitors of SARS-CoV-2 main protease, with Tanimoto scores. The scores are calculated based on the similarity to reference molecule - N3 inhibitor. The table is modified from Abel et al., 2020 (4).	15
Table 3: Top 40 compounds, potential inhibitors of SARS-CoV-2, with Tanimoto scores. The scores are calculated based on the similarity to reference molecule - O6K inhibitor. The table is modified from Abel et al., 2020 (4).	16
Table 4: Selected candidates from SuperNatural II and SuperTCM database for inhibitors of 3CLpro, based on the molecular docking studies (4). 2D interaction diagrams were created with BIOVIA Discovery Studio (57).	17
Table 5: Selected candidates from SuperDrug 2 and WITHDRAWN database for inhibitors of 3CLpro, based on the molecular docking studies (4). 2D interaction diagrams were created with BIOVIA Discovery Studio (57).	19
Table 6: Twelve best candidates for the main protease of SARS-CoV-2 inhibitors from <i>Reynoutria</i> rhizomes. The table modified from Nawrot-Hadzik et al., 2021 (5).	22
Table 7. Selected candidates for in vitro studies - potentially good candidates for the main protease of SARS- CoV-2 inhibitors (based on interaction analyses). The table modified from Nawrot-Hadzik et al., 2021 (5).	23

List of figures

- Figure 1: Drug discovery workflow.** Schematic representation of main steps in drug discovery process from target identification to drug approval. The figure is modified from Hughes et al., 2011 (17).5
- Figure 2: Project I workflow.** Schematic representation of computational methods used in the virtual screening of compounds from 4 databases of drugs and natural compounds to select best candidates for potential Sars-CoV-2 main protease inhibitors. Own representation, based on the workflow described by Abel et al., 2020 (4).8
- Figure 3: Project II workflow.** Schematic representation of combination of computational methods and in vitro studies to select best candidates for potential Sars-CoV-2 main protease inhibitors. Own representation, based on the workflow described by Nawrot-Hadzik et al., 2021 (5).9
- Figure 4: Workflow for similarity search.** The figure is generated in KNIME software (46).10
- Figure 5: Co-crystalized ligands of SARS-CoV-2 main protease.** A - N3 ligand; B - 6OK ligand. The structures are generated in PubChem Sketcher V2.4 (49).11
- Figure 6: 3D visualization of the main SARS-CoV-2 protease in complex with ligand N3.** A - the whole enzyme; B - the binding site of the enzyme. The figure is generated in PyMOL software (58).20

List of abbreviations

COVID-19 - Coronavirus Disease 2019

Mpro/3CLpro - the main protease of Sars-Cov-2

Sars-Cov-2 - severe acute respiratory syndrome coronavirus 2

MS - mass spectroscopy

NMR - nuclear magnetic resonance

NPs- natural products

AI - artificial intelligence

ML - machine learning

DL - deep learning

CADD - Computer Aided Drug Design

LBDD - ligand-based drug design

SBDD - structure-based drug design

LBVS - ligand-based virtual screening

SBVS - Structure-based virtual screening

HTS - high-throughput screening

VS- virtual screening

IVS - inverse virtual screening

SVM - support vector machines

RF - random forest

DT - decision trees

k-NN - *k*-nearest neighbours

NB - Naive Bayes

PCA - principal component analysis

ANNs - artificial neural networks

MD - molecular dynamics

RMDS - root-mean-square deviation

WHO – World Health Organization

Abstract

For a few decades now, computation methods have been widely used in drug discovery or drug repurposing process, especially when saving time and money are important factors. Development of bioinformatics, chemoinformatics, molecular modelling techniques and machine or deep learning tools, as well as availability of various biological and chemical databases, have had a significant impact on improving the process of obtaining successful drug candidates.

This dissertation describes the role of natural products in drug discovery, as well as presents several computational methods used in drug discovery and drug repurposing. Application of these methods is presented with the example of searching for potential drug treatment options for the COVID-19 disease. The disease is caused by the novel coronavirus SARS-CoV-2, which was first discovered in December 2019 and has caused the death of more than 5.6 million people worldwide (until January 2022). Findings from two research projects, which aimed to identify potential inhibitors of main protease of SARS-CoV-2, are presented in this work. Moreover, a summary on COVID-19 treatment possibilities has been included.

In the first project, a ligand-based virtual screening of around 360,000 compounds from natural products databases, as well as approved and withdrawn drugs databases was conducted, followed by molecular docking and molecular dynamics simulations. Moreover, computational predictions of toxicity and cytochrome activity profiles for selected candidates were provided. Twelve candidates as SARS-CoV-2 main protease inhibitors were identified - among them novel drug candidates, as well as existing drugs. The second project was focused on finding potential inhibitors from plants (*Reynoutria japonica* and *Reynoutria sachalinensis*) and was based on molecular docking studies, followed by in vitro studies of the activity of selected compounds, extract, and fractions from those plants against the enzyme. Several natural compounds were identified as promising candidates for SARS-CoV-2 main protease inhibitors. Additionally, butanol fraction of *Reynoutria* rhizomes extracts also showed inhibitory activity on the enzyme.

Suggested drugs, natural compounds and plant extracts should be further investigated to confirm their potential as COVID-19 therapeutic options. Presented workflow could be used for investigation of compounds for other biological targets and different diseases in the future research projects.

Zusammenfassung

Seit einigen Jahrzehnten werden bei der Entwicklung und Repositionierung von Arzneimitteln rechenintensive computergestützte Methoden eingesetzt, insbesondere da Zeit- und Kostenersparnis wichtige Faktoren sind. Die Weiterentwicklung der Bioinformatik und Chemoinformatik und die damit einhergehende Optimierung von molekularen Modellierungstechniken und Tools für maschinelles sowie tiefes Lernen ermöglicht die Verarbeitung von großen biologischen und chemischen Datenbanken und hat einen erheblichen Einfluss auf die Verbesserung des Prozesses zur Gewinnung erfolgreicher Arzneimittelkandidaten.

In dieser Dissertation wird die Rolle von Naturstoffen bei der Entwicklung von Arzneimitteln beschrieben, und es werden verschiedene computergestützte Methoden vorgestellt, die bei der Entdeckung von Arzneimitteln und der Repositionierung von Arzneimitteln eingesetzt werden. Die Anwendung dieser Methoden wird am Beispiel der Suche nach potenziellen medikamentösen Behandlungsmöglichkeiten für die Krankheit COVID-19 vorgestellt. Die Krankheit wird durch das neuartige Coronavirus SARS-CoV-2 ausgelöst, das erst im Dezember 2019 entdeckt wurde und bisher (bis Januar 2022) weltweit mehr als 5,6 Millionen Menschen das Leben gekostet hat. In dieser Arbeit werden Ergebnisse aus zwei Forschungsprojekten vorgestellt, die darauf abzielten, potenzielle Hemmstoffe der Hauptprotease von SARS-CoV-2 zu identifizieren. Außerdem wird ein Überblick über die Behandlungsmöglichkeiten von COVID-19 gegeben.

Im ersten Projekt wurde ein ligandenbasiertes virtuelles Screening von rund 360.000 Kleinstrukturen aus Naturstoffdatenbanken sowie aus Datenbanken für zugelassene und zurückgezogene Arzneimittel durchgeführt, gefolgt von molekularem Docking und Molekulardynamiksimulationen. Darüber hinaus wurden für ausgewählte Kandidaten rechnerische Vorhersagen zur Toxizität und zu Cytochrom-P450-Aktivitätsprofilen erstellt. Es wurden zwölf Kandidaten als SARS-CoV-2-Hauptproteaseinhibitoren identifiziert - darunter sowohl neuartige als auch bereits vorhandene Arzneimittel.

Das zweite Projekt konzentrierte sich auf die Suche nach potenziellen Inhibitoren aus Pflanzen (*Reynoutria japonica* und *Reynoutria sachalinensis*) und basierte auf molekularen Docking-Studien, gefolgt von In-vitro-Studien der Aktivität ausgewählter

Verbindungen, Extrakte und Fraktionen aus diesen Pflanzen gegen das Enzym. Mehrere Naturstoffe wurden als vielversprechende Kandidaten für SARS-CoV-2-Hauptproteaseinhibitoren identifiziert. Außerdem zeigte die Butanolfraktion von *Ryenoutria* Rhizomextrakten ebenfalls eine hemmende Wirkung auf das Enzym.

Die vorgeschlagenen Arzneimittel, Naturstoffe und Pflanzenextrakte sollten weiter untersucht werden, um ihr Potenzial als COVID-19-Therapieoptionen zu bestätigen. Der vorgestellte Arbeitsablauf könnte in zukünftigen Forschungsprojekten zur Untersuchung von Verbindungen für andere biologische Ziele und verschiedene Krankheiten verwendet werden.

1. Introduction

1.1 Aim of the thesis

The aim of this dissertation is to show the application and the significance of computational approaches in developing novel drugs and also in repurposing of existing drugs, as well as to demonstrate the importance of natural products in drug discovery. The focus of the work and choice of the molecular target for drug discovery was determined by the current situation, that being the ongoing global coronavirus pandemic.

The pandemic started at the end of 2019 and was caused by the novel coronavirus (Sars-CoV-2). This virus has caused the death of more than 5.6 million people (from December 2019 until January 2022) and significantly altered everyday life around the globe. Large numbers of infected people have become an enormous challenge for healthcare providers and governments worldwide and required special measures to be introduced to help fight the pandemic. Rapid development of vaccines against COVID-19 (1) has had a huge impact on reducing deaths caused by the virus, but has not solved the problem, as a shortage of vaccines persists in many countries, and there are sizable groups of people who cannot or do not want to be vaccinated. Additional challenges have been caused by subsequent coronavirus variants, such as B.1.617.2, called Delta (2) or B.1.1.529, called Omicron (3), which are more resistant to vaccines than the original virus. For these reasons, there is still a need of searching for potential therapeutic options, as well as developing better and faster testing methods to improve the virus detection process.

This dissertation summarises two research projects focused on finding potential drug candidates against COVID-19, which can act as inhibitors of the Sars-CoV-2 main protease (called Mpro or 3CLpro). Most of the candidates that were explored were natural product-based compounds. Different computational approaches were used for screening and selecting the best candidates.

Project I ('Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2'. Published December 2020 in *Frontiers in Chemistry*) was focused on searching for potential inhibitors of the main protease of SARS-CoV-2 among compounds from four different databases: natural products, Traditional Chinese Medicine (TCM), approved drugs, and withdrawn drugs. Ligand-based virtual screening of around 360,000 compounds was followed by molecular docking, molecular dynamics studies and computational prediction of toxicity and cytochrome activity profiles (4).

Project II ('*Reynoutria* Rhizomes as a Natural Source of SARS-CoV-2 Mpro Inhibitors- Molecular Docking and In Vitro Study'. Published July 2021 in *Pharmaceuticals*) was focused on finding potential inhibitors of the SARS-CoV-2 main protease from plants (*Reynoutria japonica* and *Reynoutria sachalinensis*), based on molecular docking studies, followed by in vitro studies of the activity of selected compounds, extracts, and fractions from those plants against the enzyme. (5). The selection of these plants was influenced by the previous research projects involving *Reynoutria* species and their compounds (6).

1.2 The role of natural products in drug discovery

Natural products (NPs), because of their wide-ranging diversity, complexity of structures, unusual scaffolds and rich functionalities are very potent and important compounds for drug candidates. There are few definitions of natural products, one of which classifies NPs as secondary metabolites produced by living organisms (mostly plants, bacteria, fungi, or animals) (7). Natural products are also very challenging compounds because of their structural complexity, which also is often connected to their larger size and high flexibility. In many cases, their mode of action is not well understood as they often interact with multiple targets (8).

Nevertheless, in recent years interest in natural products as potential drugs candidates has again begun to rise, especially after 2015, when the Nobel Prize in Physiology or Medicine was awarded for the discovery of avermectins and artemisinin (8). Avermectin, which is a microbial natural product, and its derivatives caused a decrease of the incidents of River Blindness and Lymphatic Filariasis. Artemisinin, which is a plant based natural product, has significantly reduced the mortality caused by malaria parasites (8). Additionally, a few artemisinin analogues (dihydroartemisinin, artemether, and artesunate) have also been successfully used for malaria treatment (9).

Another reason for increased interest in natural products is the ongoing development of data mining and analyses of genomic, transcriptomics, proteomics and metabolomics data, as well as better understanding of the bioactivity of plant metabolites and mechanisms of action. Also, the development of synthetic biology and genetic engineering strategies gives opportunities for more efficient production of natural compounds, as well as the possibility of introducing new functionality by biosynthetic pathways modifications (10).

In recent years, more natural products databases have been developed - both commercial, such as Dictionary of Natural products (DNP) (11), or open-source, such as SuperNatural II database (12). They are often divided based on the category of geographic location: Chinese, African, or Indian traditional medicine, or other categories such as NPs found in food or marine NPs (7). Also, databases of mass spectrometry data, such as European MassBank (13), or nuclear magnetic resonance data, such as Spektraris NMR database (14), provide information about the MS and NMR spectra of natural products.

Synthetic drugs are still the majority of all approved drugs, but there is a growing percentage of natural product-based drugs. They could be classified into a few categories: unaltered natural products (N), natural products derivatives (ND), botanical drugs (NB) or biological macromolecules (B), which are usually peptides or proteins containing more than 50 residues (15). Another category represents synthetic drugs inspired by natural products (NPMs) (16). In the last nearly 40 years (from 1981 till 2019) natural products (N, NB and ND) constituted 23,5% of all approved drugs and 33,6 % of small-molecule approved drugs (15). They play an important role especially in the case of antibacterial and anticancer drugs, where respectively around 55% and 25% of all approved drugs are natural products or their derivatives (15). Some of the examples of antibacterial and anticancer drugs are shown in Table 1.

Table 1: Examples of antibacterial and anticancer natural product-based drugs (15); Abbreviations: N- natural products, ND- natural products derivatives

Generic name	Trade name	Year of introduction	Source
Antibacterial drugs			
netilmicin sulfate	Netromicine	1981	N
carumonam	Amasulin	1988	N
tazobactam sodium	Tazocillin	1992	ND
retapamulin	Altabax	2007	ND
lefamulin	Xenlita	2019	ND
Anticancer drugs			
paclitaxel	Taxol	1993	N
romidepsin	Istodax	2010	N
idarubicin HCl	Zavedos	1990	ND
gemtuzumab ozogamicin	Mylotarg	2000	ND
midostaurin	Rydapt	2017	ND

1.3 The role of computational methods in drug discovery and drug repurposing

Computational methods are widely used and are essential tools in drug discovery and drug repurposing, especially when considering factors of time and money. Continuous development of different bioinformatics and chemoinformatics methods, AI techniques, as well as development of different databases of compounds and biological targets, web servers and platforms, significantly improve the process and quality of the results.

Traditional approach to drug discovery process can take up to 15 years (17). Computational approaches, used in the first stages of drug discovery or drug repurposing, can shorten this process and lower costs of identifying targets and potential drug candidates. The standard pipeline contains such steps as: identification and validation of a target, identification of potential drug candidates, in-vitro and pre-clinical studies, and clinical studies (Figure 1) (17). Drug repurposing additionally shortens the time of introducing a new drug into the market by omitting the in-vitro and pre-clinical phase, as well as Phase I of clinical studies, if the compound has previously been approved by FDA (18).



Figure1: Drug discovery workflow. Schematic representation of main steps in drug discovery process from target identification to drug approval. The figure is modified from Hughes et al., 2011 (17).

Between 1981 and 2019 more than 70 drugs that were developed with the application of CADD (computer-aided drug design), were approved (19). Included among such drugs are tyrosine kinase inhibitors, such as: Imatinib (1990), Lorlatinib (2018) or HIV-1, protease inhibitors, such as: Saquinavir (1995), Ritonavir (1996), Lopinavir (2000), and many others (19). As for successfully repurposed drugs, two such examples can be mentioned: Raloxifene, which was first used for osteoporosis and was afterwards approved to use for breast cancer treatment; and Thalidomide, used for morning sickness, and later approved also for multiple myeloma (20). Computational approaches

are widely used, for instance, in anti-cancer drug discovery - especially for target predictions, binding site prediction, interaction network analysis etc. (21). Notably, they are also employed in antiviral drug discovery, which has been particularly visible during the COVID-19 pandemic, where researchers worldwide struggled to develop novel drugs against the coronavirus disease. There are numerous examples of publications showing the application of computational methods for screening large numbers of compounds as potential candidates for drugs against COVID-19 (22) (23) (24) (25).

Computer-aided drug design (CADD) methods can be classified into two major categories: ligand-based (LBDD) and structure-based (SBDD) methods (26). Additionally, two more categories can be differentiated: fragment-based methods, which includes fragment-based docking, fragment-based QSAR or ligand growing, and secondly, system-based methods, such as protein-ligand interactome, protein-protein interactom, and drug-target gene expression (27).

Ligand-based and structure-based methods in drug discovery play an important role in virtual screening (VS), which is widely used in early stages of drug discovery process. It is a cheaper and faster method than using experimental techniques for identifying active molecules, such as high-throughput screening (HTS), which allows the screening of large numbers of compounds for a specific target (28). LBVS is based on the assumption that molecules with a similar structure can resemble each other's biological activity (29), and it is based on the comparison of compound structures by using different descriptors and molecular fingerprints.

Structure-based virtual screening (SBVS) is a major technique used in structure-based drug discovery (30). Structure-based methods rely on a 3D structure of the target and investigation of the binding mode of targets with ligands. Molecular docking methods are used in SBVS for screening large numbers of ligands into the binding site of selected biological targets. Also, fragments of ligands can be docked and then linked together, but this approach requires using additional software to help with synthesizability of ligands produced in this manner (27). An example of such software for designing new drug-like molecules or for optimization of existing ligands could be AutoGrow4 (31). Another type of virtual screening, called IVS (inverse virtual screening), also uses molecular docking where, conversely, multiple proteins are docked against one ligand to find a potential target for this ligand (32).

Also, machine learning (ML) and deep learning (DL) methods are applied to the virtual screening process, either by a combination with other methods or as separate tools. They could enhance the ligand-based virtual screening by improving similarity searching and its data mining process, as well as structure-based virtual screening by improving scoring functions (19). The most commonly used ML techniques in drug discovery classified as supervised learning are: support vector machines (SVM), random forest (RF), decision trees (DT), *k*-nearest neighbours (*k*-NN) or Naive Bayes (NB) methods. To unsupervised learning, such methods as: *k*-means clustering, or principal component analysis (PCA) are included (33) (34). Deep learning is a subcategory of machine learning and can find application in compounds classification, targets identification or QSAR studies, for example, by using an artificial neural networks (ANNs) approach.

In structure-based methods, availability of 3D structures of a target is a limitation. When experimentally solved protein structure is not available, computational methods in structure prediction play an important role. Homology modelling is one of such methods and it is based on the alignment to the reference protein. It can be based on global or local alignment (35). Web servers, such as SWISS-MODEL can be used for this purpose (36). Structure can be also determined using ab initio modelling, which is not a simple approach as it involves building the structure from scratch, without known folding patterns and only sequence information. An example of a protein structure prediction server is Robetta (37), which predicts 3D protein structure based on the amino acid sequence and is using Rosetta modelling software (38). Robetta, aside from ab initio modelling, provides further options for structure predictions, such as deep learning and comparative modelling.

For optimization of predicted structures, especially loops and side chains - molecular dynamics (MD) methods can be used (39). MD simulations often follow virtual screening to provide additional information concerning the stability of target-ligand complexes, analyses of interaction, fluctuations and for calculation of free binding energy (40). Because of the constant increase of computational power, possibilities of conducting longer and more complex simulations has improved in recent years.

Computational methods for drug discovery can be also used in drug repurposing workflows. Particularly, methods such as ligand similarity searching or molecular docking and inverse molecular docking, used for the identification of new targets for existing

drugs. There are, however, several other approaches to drug repurposing. One of them is based on transcriptomic data (41), which provides information about both overexpressed and underexpressed genes of biological systems after treatment with different drugs. Obtained in such a way transcriptional signatures can be used to find new therapeutic connections between known pharmacologically active compounds and new diseases. One of the databases that provides the information on molecular transcriptional signatures is Connectivity Map (CMap) (42). Another approach for drug repurposing is network-based approach, where drugs, targets and pathways are represented by nodes and edges represent interactions between connected nodes. Biological network perturbations play an important role, especially in case of complex diseases (41).

2. Methods

This dissertation presents a combination of several computational methods that were used for prediction of potential Sars-CoV-2 main protease inhibitors. One of the main methods employed was molecular docking, used in both projects explored in this thesis. Additionally, in Project I molecular docking was preceded by ligand-based virtual screening and followed by molecular dynamics, as well as cytochrome activity and toxicity prediction (4). In Project II, molecular docking was followed by in vitro studies of selected compounds (5). Both workflows are presented in Figure 2 and Figure 3.

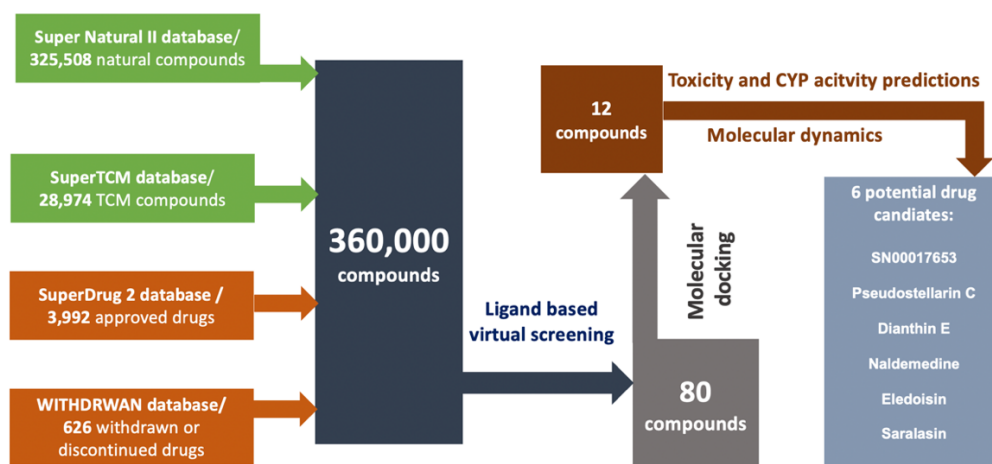


Figure 2: Project I workflow. Schematic representation of computational methods used in the virtual screening of compounds from 4 databases of drugs and natural compounds to select best candidates for potential Sars-CoV-2 main protease inhibitors. Own representation, based on the workflow described by Abel et al., 2020 (4).

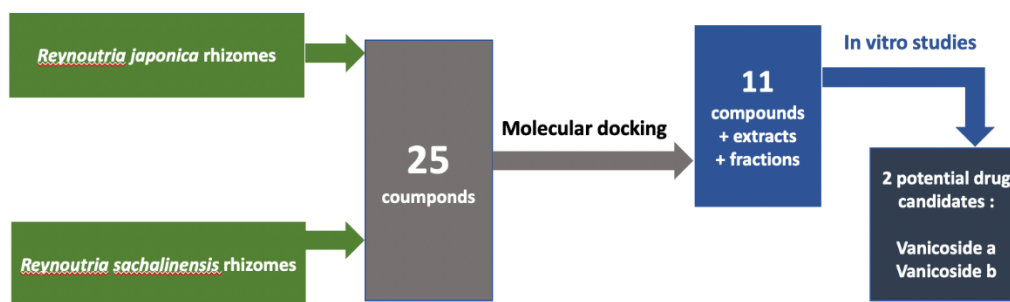


Figure 3: Project II workflow. Schematic representation of combination of computational methods and in vitro studies to select best candidates for potential Sars-CoV-2 main protease inhibitors. Own representation based on the workflow described by Nawrot-Hadzik et al., 2021 (5).

2.1 Ligand-based virtual screening (LBVS)

Ligand-based virtual screening (LBVS) and structure-based virtual screening (SPVS) are both widely used techniques in computer-aided drugs design (CADD). One of the commonly used methods in ligand-based virtual screening is searching for molecular similarity to predict properties of the compounds. This method is based on the assumption that similar molecules may have similar bioactivity (43). To analyse the similarity between molecules, firstly - the structural representation of a molecule needs to be chosen (such as physicochemical properties, pharmacophore features, molecular shapes), and secondly - the quantitative measurement of similarity between molecular structures needs to be specified (such as Tanimoto coefficient, Dice index, Tversky index) (26). There are 2D and 3D approaches in similarity search. Fingerprint similarity search is the most common 2D method, where fingerprints are encoded with the structural features of molecules. It is a fast and efficient approach but is based only on 2D structural information. The most common 2D fingerprints are: molecular access system fingerprints (MACCS), circular fingerprints (e.g., ECFP), topological (e.g., E-state), path-based fingerprints (e.g., FP2), pharmacophore fingerprints (e.g., ERG) and hybrid fingerprints (e.g., Unity 2D) (44) (45). 3D methods include 3D fingerprints, shape similarity and pharmacophore modelling (26). Molecular similarity methods are based on the global molecular view of molecules, in contrast to pharmacophores analyses or QSAR (quantitative structure–activity relationships) analyses, which are focused on local similarities during comparison of molecular structures with each other (29).

In Project I, ligand-based virtual screening was conducted using KNIME software (46). KNIME workflow for calculation of Tanimoto pairwise similarities was designed (Figure 4). RDKit nodes (47) were used for calculating different types of molecular fingerprints: MACCS, E-state and ECFP4. Based on the comparison of hits and Tanimoto scores - MACCS molecular fingerprints (48) were chosen for the following steps. Those fingerprints are bit strings with 166 structural keys, representing the presence or absence of specific fragments, functional groups, or substructures and are the most popular key-based fingerprints (44).

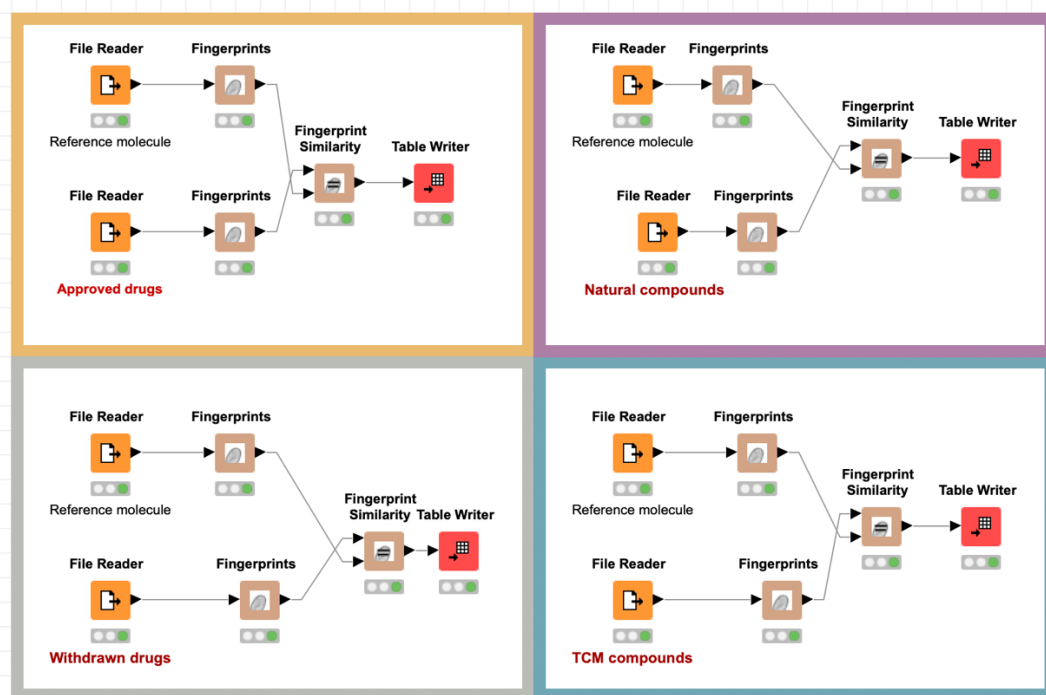
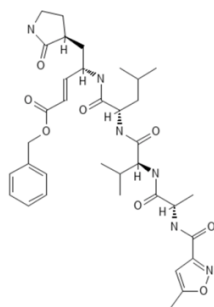


Figure 4: Workflow for similarity search. The figure is generated in KNIME software (46).

As reference compounds for virtual screening, structures of co-crystallized ligands obtained from the PDB database were used. Two PDB structures of the main Sars-Cov-2 protease (Mpro) were chosen: 6LU7 with N3 ligand as well as 6Y2F with ligand 6OK (Figure 5). Tanimoto scores were calculated for all compounds and the top 10 compounds with the highest Tanimoto score from each of four databases were selected for further steps.

A



B

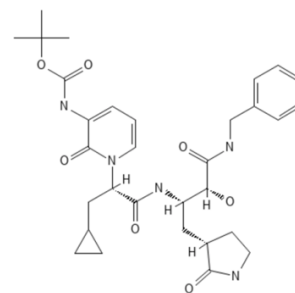


Figure 5: Co-crystallized ligands of SARS-CoV-2 main protease. A - N3 ligand; B - 6OK ligand. The structures are generated in PubChem Sketcher V2.4 (49).

2.2 Molecular docking

Molecular docking plays an important role in predicting the binding mode of the ligands in the protein binding site. Usually, docking approaches are focused on targeting orthosteric sites of the protein, but often allosteric sites need to also be considered (27). An example of such a target is cannabinoid receptor CB1, which has three allosteric sites (50). There are many types of docking software. Most known are AutoDock, AutoDock Vina, *Schrödinger* Glide or CCDC GOLD (51). Additionally, there are numerous online docking servers available, such as: Swiss Dock, for docking of small molecules to target proteins (52); blind docking servers to predict binding sites of proteins, for instance, Achilles Blind Docking Server (53) or CB-Dock server (54); and docking servers for protein-protein docking or protein–DNA/RNA docking, such as HDOCK (55).

For both projects presented in this work – GOLD docking software was used. This program uses a genetic algorithm and the GOLD scoring function. A crystal structure of the main protease of SARS-CoV-2 (PDB code: 6LU7) with its co-crystallized ligand (called N3) was downloaded from the PDB database. It was the first available crystal structure of this protease (56).

Re-docking of the co-crystallized ligand from the PDB structure was done as a first step. RMDS values were calculated to validate the docking protocol. Following this, compound structures were prepared and docked into the binding site of the protein, which was defined based on the co-crystallized ligand pose. Ten poses were generated for each molecule and each of them were then analysed in terms of interactions with crucial

residues. The crucial residues were defined based on the literature (56). Analyses of molecular interactions were conducted with the Discovery Studio (57) and PyMOL software (58). 2D diagram interactions were generated for the best poses.

Molecular docking was used with a different approach in both projects. In Project I, the application of this method was connected to virtual screening and docking of a larger number of compounds (80 compounds) from four different databases, which was preceded by ligand-based virtual screening of almost 360,000 compounds and followed by molecular dynamics to then select best candidates for inhibitors of main Sars-Cov-2 protease. Molecular docking analyses led to the choice of the best 12 candidates for MD simulations. Project II was focused on molecular docking of a smaller number of compounds (25 compounds), which are known metabolites from *Reynoutria* rhizomes in order to check their activity against Mpro. In this instance, molecular docking analysis helped to select candidates for in vitro studies as a next step. In both projects molecular interactions were analysed not only based on the docking scores, but through visual inspection of the binding modes to choose only those compounds for the following stage of the project, which created hydrogen bonds with significant residues.

2.3 Molecular dynamics

Molecular dynamics (MD) simulations were included in Project I. Conducting MD simulations is a crucial step in drug discovery workflow, as they provide a broader range of information on the stability of the protein-ligand complex than docking studies alone. Analysis of the trajectories can provide such measures as root-mean-square deviations (RMSD) or root-mean-square fluctuations (RMSF). Also, binding free energy calculation, ΔG (the difference between the free energy of an unbound ligand and a ligand bound to the target), can be calculated (59). The most commonly used software for MD simulations are AMBER, CHARMM, GROMACS, NAMD and DESMOND (59).

MD simulations were conducted with the DESMOND software, included in Schrödinger LLC Maestro suite 2019 (60). The GPU version was used with the NVIDIA QUADRO 5000 workstation. The OPLS3 force field was employed. After system minimization (2000-time steps), 100ns MD simulations were performed. Analyses of the trajectories were mainly based on RMSD calculations and analyses of protein-ligand contacts during simulation with top interacting residues. Moreover, free energy calculations were

performed using the MM-GBSA method (molecular mechanics–generalised Born surface area) (61).

2.4 Toxicity and cytochrome activity prediction

For the 40 selected compounds in Project I (10 from each database), toxicity prediction and cytochrome 450 enzymes (CYPs) inhibition profiles were additionally computed. The ProTox-II web server, which includes 40 toxicity models, was used for toxicity prediction (62). There are six toxicity classes that could be predicted with this server, where class 1 means that the compound is highly toxic ($LD_{50} \leq 5$) and class 6 that it is nontoxic ($LD_{50} > 5000$).

CYP prediction was done with SuperCYPsPred web server, which currently includes 10 models for the main CYPs isoforms, among which are: CYP1A2, CYP2C19, CYP2D6, CYP2C9 and CYP3A4 (63). These listed are five out of 57 existing isozymes in humans and are the most important cytochromes responsible for the metabolism of most drugs. CYP prediction plays an important role in predicting toxicities connected to drug-drug interactions (DDI). It is an important step in the drug design process, in addition to ADME and toxicity prediction methods (63).

Both web servers: ProTox-II and SuperCYPsPred were developed by the members of Charite Structural Bioinformatics research group. There are also many other available platforms, web servers or machine learning models for toxicity prediction (64) (65), ADME properties prediction (66) (67) or prediction of cytochrome 450 (CYP) metabolism (68), which could be used in the drug discovery process.

2.5 Compounds and databases

Application of computational methods in drug discovery and drug-repurposing would not be possible without parallel development of databases from where compound structures and protein structures could be obtained. The most commonly used compound databases used in drug discovery are: ZINC, PubChem or ChEMBL (19), but there are also many other specialised databases, such as: drugs databases (e.g. DrugBank (69)), natural product databases (e.g. COCONUT database (70)), Traditional Chinese Medicine databases (e.g. YaTCM (71)), and Indian medicinal plants databases (e.g. IMPPAT (72)),

to name a few. The extensive list of 120 natural products databases can be found in the review by Sorokina & Steinbeck (7), from which around 50 are open access. Protein structures are available in the PDB database, which currently comprises more than 180,000 structures (73). Most of the structures in this database are determined through X-Ray crystallography (80%) or nuclear magnetic resonance (NMR) spectroscopy (16%). Another technique is Cryo-Electron Microscopy, which in contrast to X-ray crystallography, enables one to observe the native environment of the structure (35).

In Project I, an approved drug database, called SuperDrug2 (74) and withdrawn and discontinued drugs database, called WITHDRAWN (75), as well as two databases of natural products, which is Super Natural II (76) and SuperTCM (77) were used for virtual screening. All four databases were developed by members of the Charite Structural Bioinformatics research group. The final dataset consisted of around 360,000 compounds.

In Project II, 25 compounds were selected, based on the knowledge of the compounds which are present in the rhizomes of *Reynoutria* Plans (78). Selected compounds belong to five phytochemical classes: anthraquinones, flavan-3-ols, phenylpropanoid disaccharide esters, procyanidins and stilbenes. The structures of these compounds were obtained from the PubChem database.

2.6 In vitro studies

In vitro studies in spectrofluorimetric assay were performed for selected compounds, extracts, and fractions from *Reynoutria* rhizomes, as a part of Project II. The procedure that was used was previously described in the other publication concerning SARS-CoV-2 inhibitors (79). Fluorescent peptide (QS1, Ac-Abu-Tle-Leu-Gln-ACC) was used as a substrate for SARS-CoV-2 Mpro (80). Measurements were done with the Molecular Devices Spectrofluorometer - SpectraMax Gemini XPS. Compounds for in vitro studies were chosen not only based on the most promising docking results, but also on the availability of the compounds from the *Reynoutria* rhizomes, as not all of the compounds of interest could be extracted or were available from vendors.

3. Results

3.1 Project I

Ligand-based virtual screening of around 360,000 compounds, based on a fingerprint's similarity with the reference molecules (N3 and 6OK), led to the selection of best 80 candidates (20 from each database), are presented in Table 2 and Table 3. The Tanimoto score for selected compounds ranged between 0.63 and 0.83 (4).

Table 2: Top 40 compounds, potential inhibitors of SARS-CoV-2 main protease, with Tanimoto scores. The scores are calculated based on the similarity to reference molecule - N3 inhibitor. The table is modified from Abel et al., 2020 (4).

SuperTCM		Super Natural II		SuperDrug2		WITHDRAWN	
Pseudostellarin C	0.77	SN00308384	0.76	Elcatonin	0.74	Saralasin	0.71
Nummularine A	0.76	SN00017653	0.75	Secretin porcine	0.74	Romidepsin	0.71
Notoamide G	0.76	SN00227324	0.74	Daptomycin	0.74	Abarelix	0.69
Segetalin D	0.76	SN00270610	0.74	Enfuvirtide	0.73	Saquinavir	0.67
Jubanine A	0.75	SN00354661	0.74	Albiglutide	0.73	Trabectedin	0.67
Notoamide R	0.75	SN00254530	0.74	Secretin human	0.73	Topotecan	0.66
Mauritine A	0.75	SN00314990	0.74	Angiotensin II	0.73	Pentagastrin	0.65
Dianthin E	0.74	SN00019468	0.74	Naldemedine	0.72	Rescinnamine	0.65
Segetalin E	0.74	SN00303378	0.73	Angiotensinamide	0.72	Dirithromycin	0.64
Mauritine B	0.74	SN00238988	0.73	Eledoisin	0.72	Aliskiren	0.63

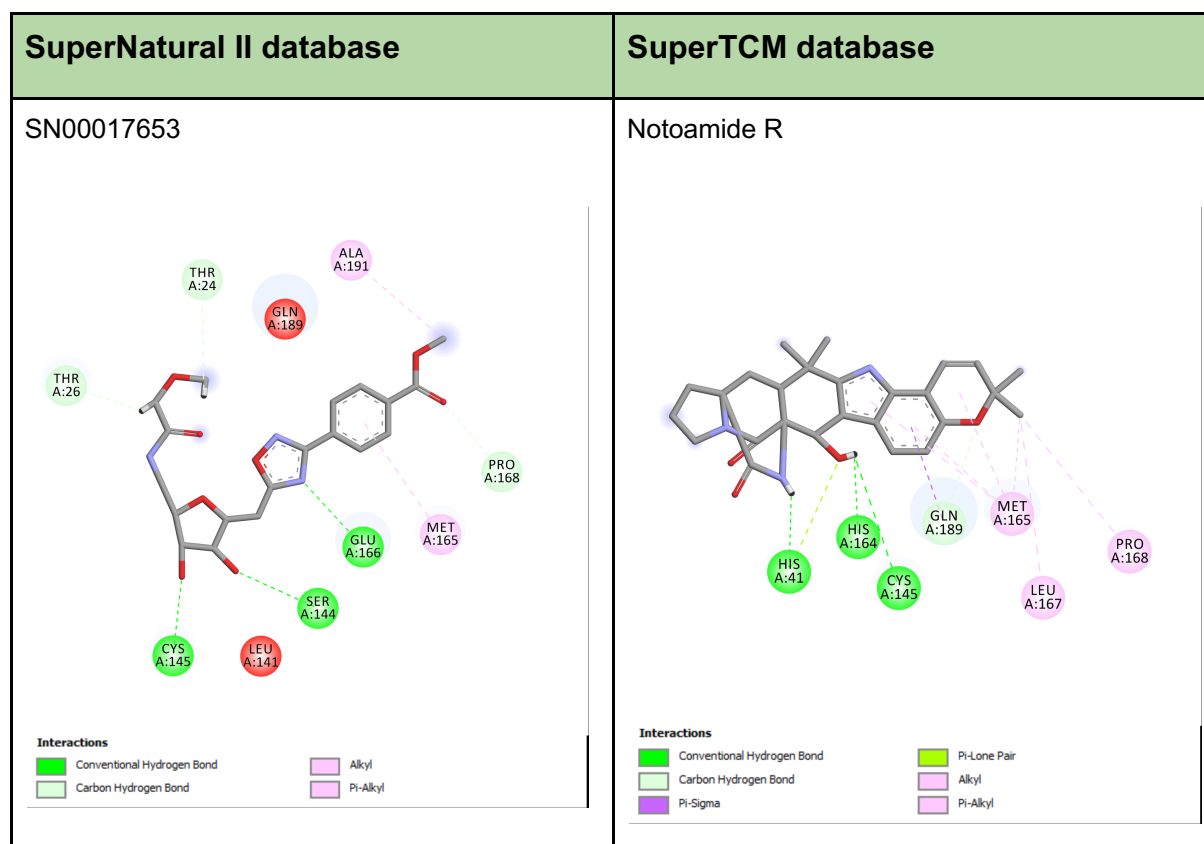
Table 3: Top 40 compounds, potential inhibitors of SARS-CoV-2, with Tanimoto scores. The scores are calculated based on the similarity to reference molecule - O6K inhibitor. The table is modified from Abel et al., 2020 (4).

SuperTCM		Super Natural II		SuperDrug2		WITHDRAWN	
Hirudin	0.82	SN00109804	0.83	Lopinavir	0.83	Saralasin	0.78
Segetalin E	0.80	SN00087725	0.82	Angiotensinamide	0.82	Saquinavir	0.77
Lyciumin C	0.80	SN00077453	0.82	Indinavir	0.82	Trabectedin	0.74
Celogenamide A	0.80	SN00077454	0.82	Dihydroergocristine	0.81	Tofacitinib Citrate	0.73
Notoamide O	0.80	SN00101312	0.82	Dihydroergocornine	0.81	Lypressin	0.73
Ergosine	0.79	SN00108593	0.82	Dihydroergocryptine	0.81	Alatrofloxacin	0.73
Ergosinine	0.79	SN00107903	0.8	Epicriptine	0.80	Nelfinavir	0.73
Ergocornine	0.79	SN00213824	0.8	Bivalirudin	0.80	Pentagastrin	0.72
Ergocorninine	0.79	SN00012917	0.79	Dihydroergotamine	0.79	Azlocillin	0.72
Ergocryptine	0.79	SN00101324	0.79	Telaprevir	0.79	Telithromycin	0.71

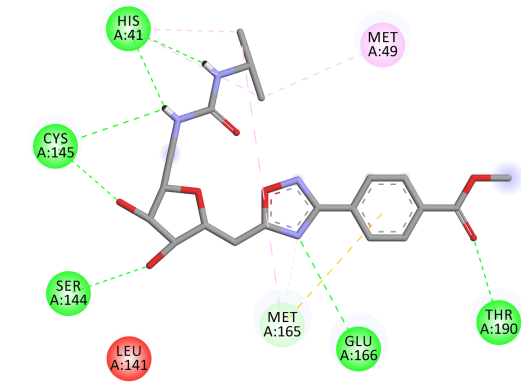
Interaction analyses of all docked 80 compounds helped to identify the 12 best candidates (3 from each database), which were later chosen for molecular dynamics simulations (Table 4 and Table 5).

Selected candidates were chosen based on the presence of interaction with crucial residues of SARS-CoV-2 main protease, which were catalytic site residues Cys145 or His41, and also with other residues from the binding site, such as Ser144, Glu166, Gln189 or Gln192. The interaction diagrams presented in Table 4 and Table 5 were created with BIOVIA Discovery Studio (57). The 3D visualization of the main protease (PDB code: 6LU7) in complex with redocked N3 ligand is presented at the Figure 6A and 6B. The 3D visualization was created with PyMOL software (58).

Table 4: Selected candidates from SuperNatural II and SuperTCM database for inhibitors of 3CLpro, based on the molecular docking studies (4). 2D interaction diagrams were created with BIOVIA Discovery Studio (57).



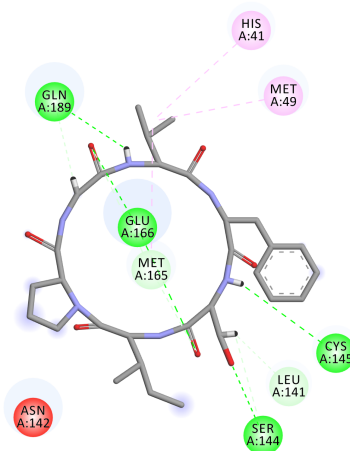
SN00019468



Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Sulfur
- Alkyl
- Pi-Alkyl

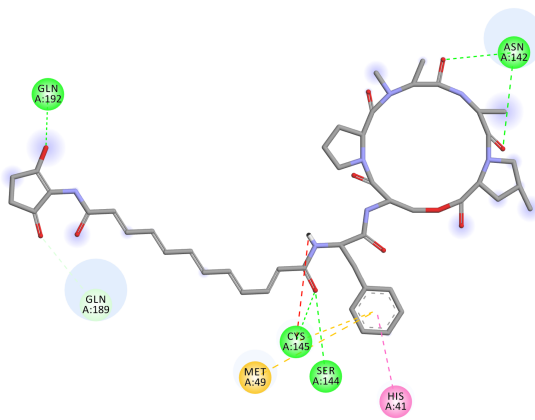
Dianthin E



Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Alkyl
- Pi-Alkyl

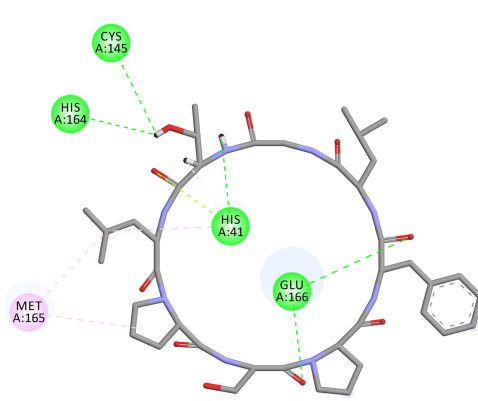
SN00303378



Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Unfavorable Donor-Donor
- Pi-Sulfur
- Pi-Pi Stacked

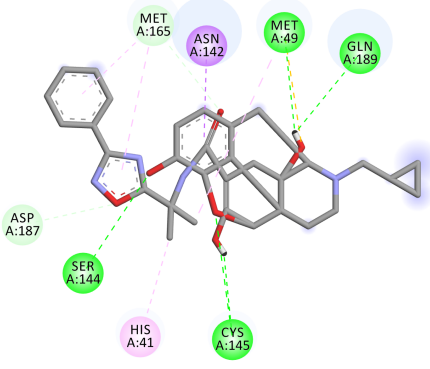
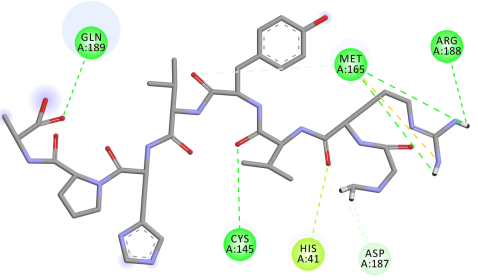
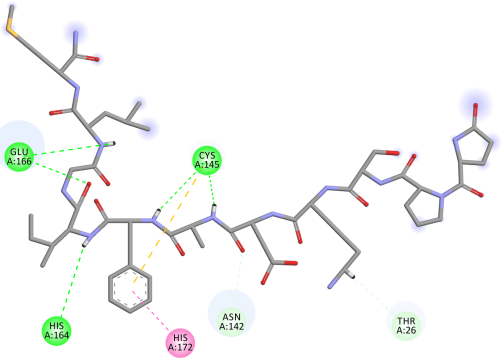
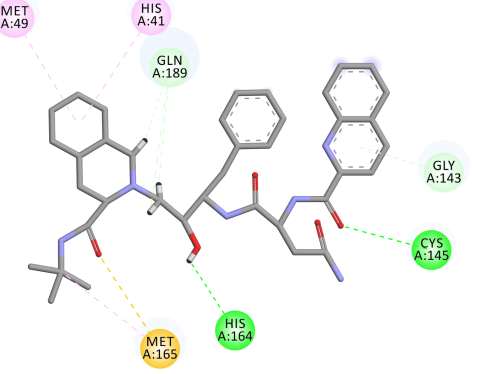
Pseudostellarin C

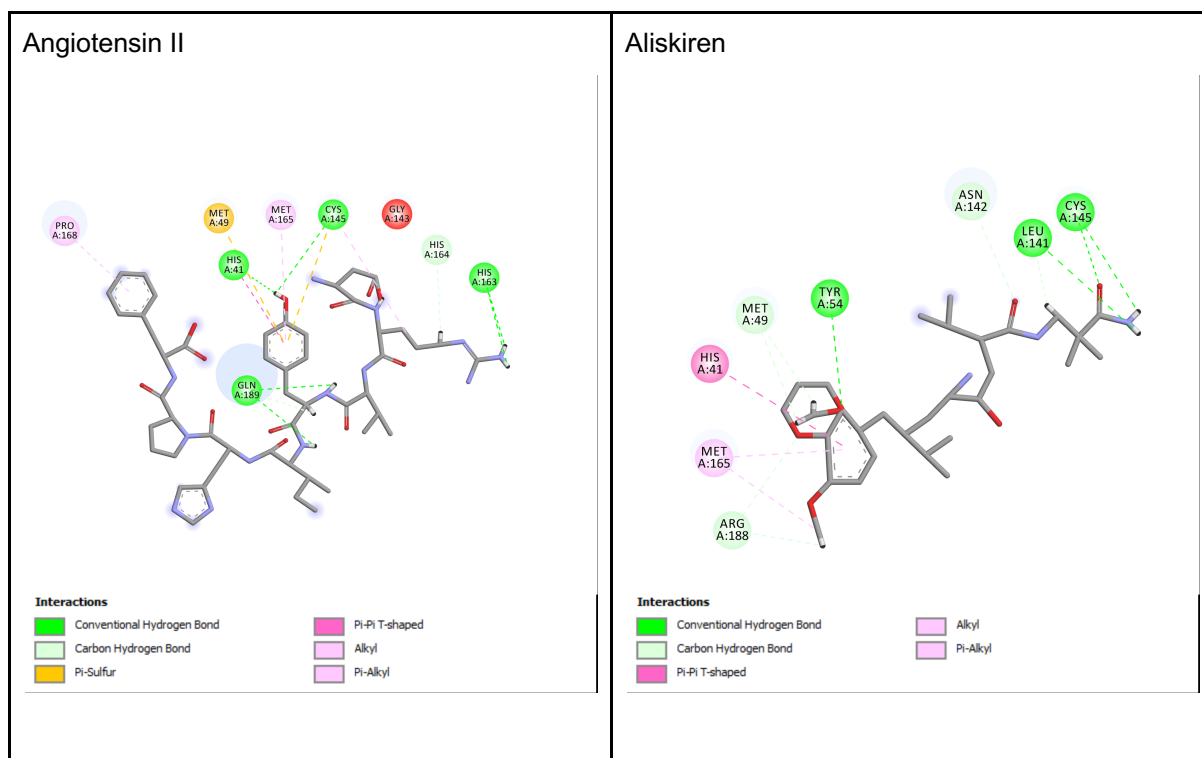


Interactions

- Conventional Hydrogen Bond
- Carbon Hydrogen Bond
- Pi-Lone Pair
- Alkyl
- Pi-Alkyl

Table 5: Selected candidates from SuperDrug 2 and WITHDRAWN database for inhibitors of 3CLpro, based on the molecular docking studies (4). 2D interaction diagrams were created with BIOVIA Discovery Studio (57).

SuperDrug2 database	WITHDRAWN Database
<p data-bbox="204 456 363 483">Naldemedine</p>  <p data-bbox="213 1057 293 1079">Interactions</p> <ul style="list-style-type: none"> ■ Conventional Hydrogen Bond ■ Carbon Hydrogen Bond ■ Sulfur-X ■ Pi-Sigma ■ Alkyl ■ Pi-Alkyl 	<p data-bbox="810 456 938 483">Saralasin</p>  <p data-bbox="820 1043 900 1066">Interactions</p> <ul style="list-style-type: none"> ■ Conventional Hydrogen Bond ■ Carbon Hydrogen Bond ■ Sulfur-X ■ Pi-Lone Pair
<p data-bbox="213 1200 325 1227">Eledoisin</p>  <p data-bbox="213 1800 293 1823">Interactions</p> <ul style="list-style-type: none"> ■ Conventional Hydrogen Bond ■ Carbon Hydrogen Bond ■ Pi-Sulfur ■ Pi-Pi T-shaped 	<p data-bbox="810 1200 948 1227">Saquinavir</p>  <p data-bbox="820 1765 900 1787">Interactions</p> <ul style="list-style-type: none"> ■ Conventional Hydrogen Bond ■ Carbon Hydrogen Bond ■ Sulfur-X ■ Pi-Donor Hydrogen Bond ■ Alkyl ■ Pi-Alkyl



A

B

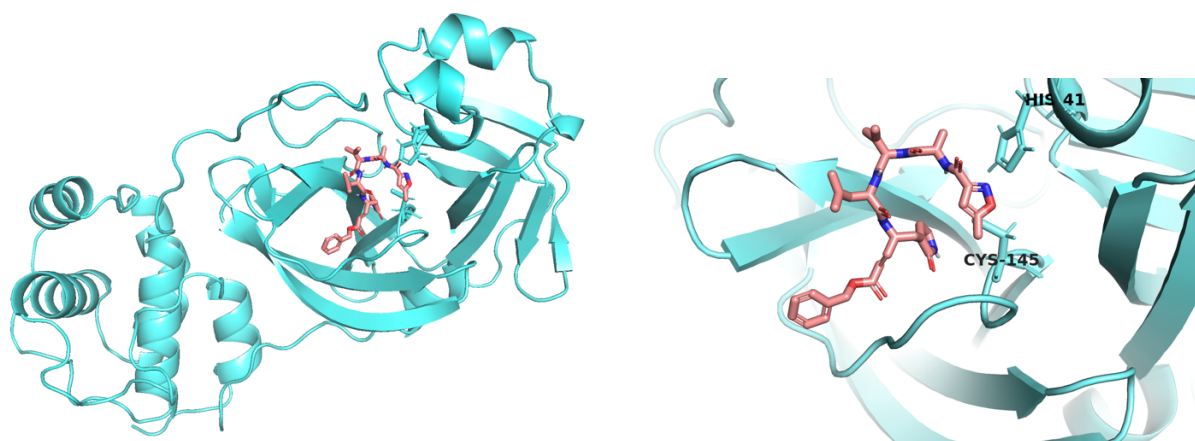


Figure 6: 3D visualization of the main SARS-CoV-2 protease in complex with ligand N3.
 A - the whole enzyme; B - the binding site of the enzyme. The figure is generated in PyMOL software (58).

Molecular dynamics simulation and analyses of MD trajectories gave additional information on the stability of the selected compounds in complex with 3CLpro. Out of 3 candidates from the natural products database, compound SN00017653 shows the

highest stability and lowest RMSD values for ligand and protein atoms. However, all three compounds have similar MM-GBSA values, around -54 kcal/mol. From the SuperTCM database, both compounds Pseudostellarin C and Dianthin E showed a high stability for the first 40ns of the simulation and after in both cases pose rearrangement could be observed. Still, there were few stable contacts observed after rearrangement, so both could be considered as good inhibitors. Notoamid R showed high instability and poor protein-ligand contacts. Additionally, this compound had poor MM-GBSA values (-21.4 kcal/mol). When considering the SuperDrug2 database, Angiotensin II shows higher RMSD values than Naldemedine and Eledoisin and might be less stable in complex with 3CLpro, even though the MM-GBSA values are low (-74.9 kcal/mol). For Naldemedine it was -64.2 kcal/mol, and a particularly high affinity value was observed in case of Eledoisin (-93.3 kcal/mol). Aliskieren, from a withdrawn database, seems not to be a good candidate when analysing the protein-ligand contacts, even though there are low MM-GBSA values (-61.2 kcal/mol). Two other compounds from this database Saquinavir and Saralasin are better candidates, as both showed good performance during MD simulation and MM-GBSA values respectively -54.3 kcal/mol and -61.8 kcal/mol.

Predictions of toxicity as well as cytochrome inhibition profiles were also conducted for selected compounds. For all three compounds from the SuperTCM database, a toxicity class 4 was predicted, as well as for most of the compounds from the SuperNatural II database (besides SN000303378, where toxicity class 3 was predicted). For approved and withdrawn drugs - toxicity class 4 or 5 was predicted. Eledoisin, Angiotensin II and Saralasin have no interaction with CYP and no toxicity endpoint predicted. Also, they all belong to toxicity class 5. Naldemedine and Saquinavir belong to toxicity class 4 and both interact with 3A4, while for saquinavir additionally with 2C8, 2C9, 2C19, 2D6 and 3A5 CYPs were predicted. CYP activity profile of this drug shows that clinically relevant drug-drug interactions with co-administered drugs may occur.

3.2 Project II

The second project concerning the search for natural product based potential inhibitors of coronavirus main protease (Mpro), combines both in silico and in vitro studies (5). Twenty-five active compounds known to be present in the *Reynoutria japonica* and *Reynoutria sachalinensis* plant's rhizomes were selected for docking into the binding site of the main protease of SARS-CoV-2. The composition of compounds of those two *Reynoutria* rhizomes was known from the previous studies (78). Molecular docking interaction analyses showed that the best 12 candidates for potential inhibitors belong to three of the phytochemical classes: procyanidins, phenylpropanoid disaccharide esters and anthranoids (Table 6).

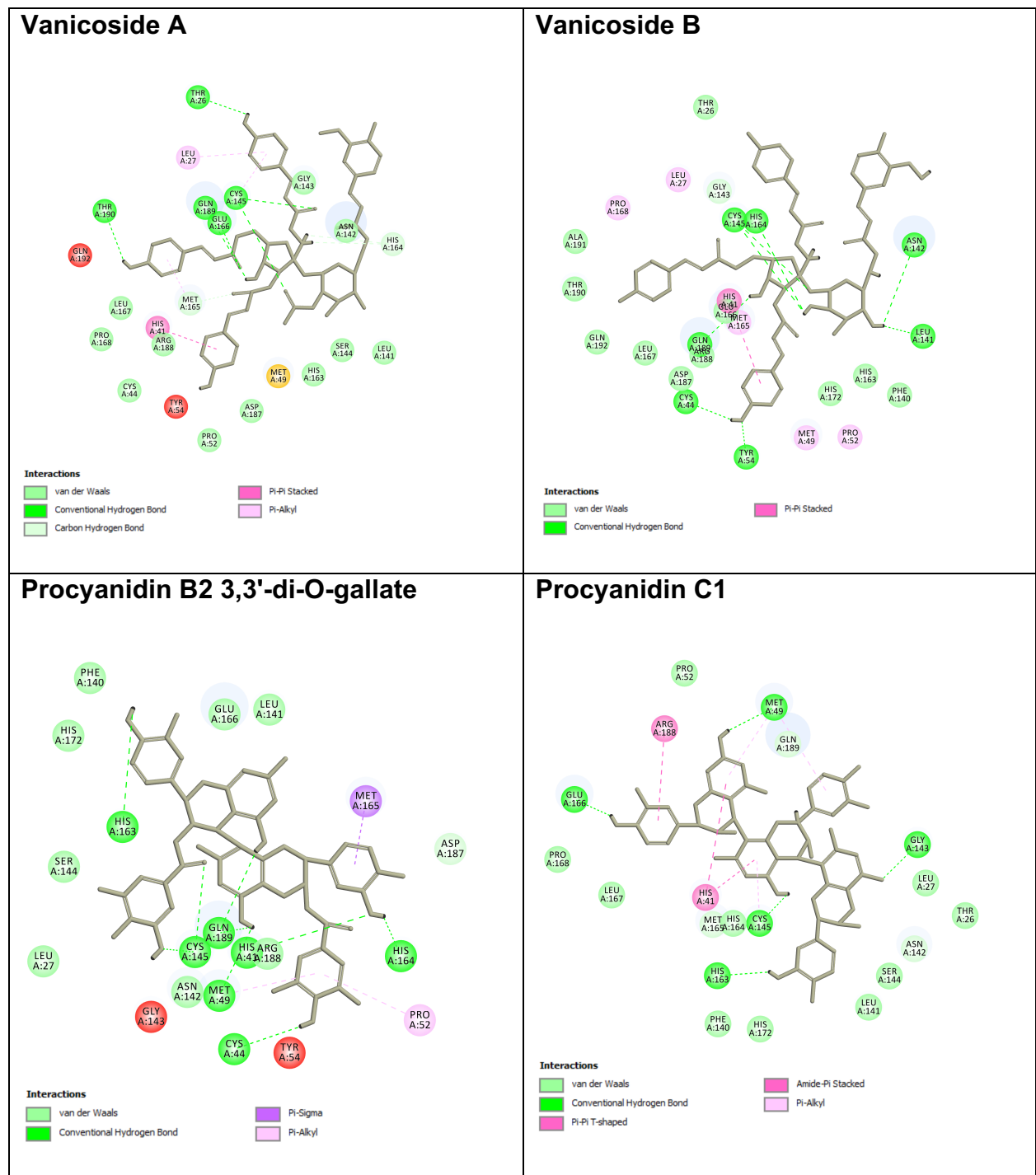
Table 6: Twelve best candidates for main protease of SARS-CoV-2 inhibitors from *Reynoutria* rhizomes. The table modified from Nawrot-Hadzik et al., 2021 (5).

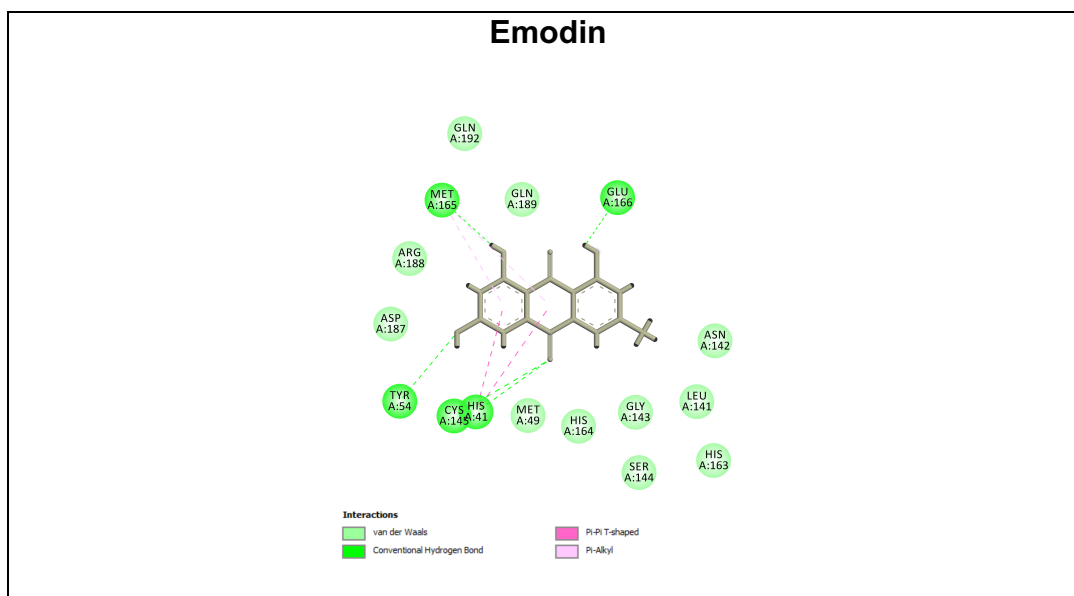
Procyanidins	Phenylpropanoid disaccharide esters	Anthranoids
Cinnamtannin A2	Vanicoside A	Emodin
Procyanidin B2 3,3'-di-O-gallate	Vanicoside B	Emodin bianthrone
Procyanidin C1	Vanicoside C	
Procyanidin C1 3',3''-di-O-gallate	Hydropiperoside	
	Lapathoside C	
	Tataroside B	

Not all of the 12 compounds were available for in vitro validation, but only five of them: vanicoside A, vanicoside B, procyanidin B2 3,3'-di-O-gallate, procyanidin C1 and emodin. Interaction analyses of those compounds, created with BIOVIA Discovery Studio, are presented in Table 7. However, an additional six compounds were selected for in vitro studies (resveratrol, piceid, (-)-epigallocatechin gallate, epicatechin, epicatechin gallate and procyanidin B2). Those compounds had less promising docking results, but as they were available for testing, they were used for validation of docking results.

In case of those compounds either interactions with catalytic residues Cys145 or His41 were not observed, or hydrogen bonds with other crucial residues such as Gly143, Ser144, Glu166, Gln189 or Thr190 were not created.

Table 7: Selected candidates for in vitro studies - potentially good candidates for main protease of SARS-CoV-2 inhibitors (based on interaction analyses). The table modified from Nawrot-Hadzik et al., 2021 (5).





Together, 11 compounds were selected for in vitro studies and nine of them (beside epicatechin and epicatechin gallate) inhibited SARS-CoV-2 Mpro enzyme at the concentration of 100 μM . The best inhibition of the enzyme - over 50%, was observed for vanicoside A, vanicoside B and emodin, and over 20% for procyanidin B2 3,3'-di-O-gallate and procyanidin C1. Weaker Inhibition - below 20%, was observed for epigallocatechin gallate, procyanidin B2, resveratrol and piceid. Those results are consistent with the docking results and selection of the best candidates by analyses of interaction analyses. Additional inhibition analyses also revealed that two best candidates for Mpro inhibitors - vanicoside A, vanicoside B - showed significant inhibition at lower concentrations (starting at 13.2 μM). The calculated IC_{50} value for vanicoside A was 23.10 μM , and for vanicoside B it was 43.59 μM .

Besides the selected 11 compounds, acetone extracts from rhizomes of *R. japonica* and *R. sachalinensis*, and five fractions (dichloromethane, diethyl ether, ethyl acetate, butanol, and water) obtained from those acetone extracts, were tested. Each of them inhibited the Mpro enzyme significantly at over 50%, with the concentration of 50 $\mu\text{g/mL}$ and the strongest inhibition was observed for the extracts and for their butanol fractions. IC_{50} for *R. sachalinensis* acetone extract was 9.42 $\mu\text{g/mL}$ and for its butanol fraction 4.031 $\mu\text{g/mL}$, whereas IC_{50} for *R. japonica* acetone extract was 16.90 $\mu\text{g/mL}$ and 7.877 $\mu\text{g/mL}$ for its butanol fraction.

4. Discussion

4.1 Short summary of the results

This work presents how computational methods were applied in the process of drug discovery and drug repurposing, and is focused on searching for potential drug candidates for COVID-19 treatment.

As a result of Project I (4), ligand virtual screening of around 360,000 compounds from four databases led to the choice of best 80 candidates for inhibitors of the main protease of SARS-CoV-2. Based on docking analyses of all those compounds, the 12 best candidates were selected. Molecular dynamics simulations, as well as prediction of toxicity and cytochrome inhibition profiles, narrowed the group of best candidates to seven compounds, which are natural products such as: compound SN0001765, Pseudostellarin C and Dianthin E, and known drugs such as: Naldemedine, Eledoisin, Saralasin and Saquinavir. Saquinavir was also identified in previous studies as a potential candidate (81), but the concern about this drug could be its cytochrome activity profile, which shows that clinically relevant drug-drug interactions may occur with co-administered drugs. The other three drugs and three natural compounds were most likely reported as potential SARS-CoV-2 main protease inhibitors for the first time.

Results of Project II (5) suggest that *R. japonica* and *R. sachalinensis* rhizomes could be the source of potential SARS-CoV-2 main protease inhibitors. Two compounds vanicoside A and vanicoside B showed best inhibition values of the enzyme among 11 selected compounds for in vitro testing. Also, interaction analyses of docking results suggested that those compounds could be a potential candidate for inhibitors. However, the highest inhibition values were obtained for a butanol fraction of the extracts from *R. japonica* and *R. sachalinensis* rhizomes, which suggests that this strong inhibition of the enzyme might be caused by compounds present in this fraction.

4.2 Current state of knowledge on treatment of COVID-19

Since the beginning of the pandemic, multiple studies that suggested candidates for therapeutic agents against COVID-19 have been published (23) (82) (83) (84) (85) (86)

(87). Each of the studies present a different approach and applications of various computational methods (such as molecular docking, molecular dynamics, homology modelling, network-based algorithms for drug repositioning etc.), uses various compound sources (such as African medicinal plants, Indian medicinal plants, ZINC database or Drug Bank), or is investigating different targets of Sars-CoV-2 (such as 3CLpro, PLpro or spike protein).

Because the process of developing new drugs is very time consuming, some existing drugs have been used for treatment of the disease. A few of them were also used in the past for treatment of SARS (severe acute respiratory syndrome) or MERS (Middle East respiratory syndrome) diseases, which also were caused by coronaviruses (88). Numerous therapeutic agents are currently in clinical trials (89). Few of the drugs already got the WHO recommendation, but also those recommendations were changing with time and not all the drugs that were as effective as previously were suggested. Current WHO recommendations (90) (published September 2020 and updated January 2022) include: corticosteroids (such as Dexamethasone, Hydrocortisone), Interleukin-6 (IL-6) receptor blockers (such as Tocilizumab, Sarilumab), Janus kinase (JAK) inhibitors (such as Baricitinib, Ruxolitinib, Tofacitinib). Conditionally, WHO recommends treatment with Monoclonal antibodies (Casirivimab and Imdevimab) and also with Sotrovimab or Remdesivir. WHO recommends against treatment with Convalescent plasma, Hydroxychloroquine, Lopinavir-Ritonavir (90).

Recently (December 2021), FDA has approved the first two antiviral agents specifically for COVID-19 treatment: PAXLOVID (91) and Molnupiravir (92). Molnupiravir targets both RNA-dependent RNA polymerase (RdRp) and PAXLOVID, which combines Ritonavir and Nirmatrelvir (also called PF-07321332), targets 3CL protease of Sars-Cov-2 (93). There is no crystal structure available of 3CL pro i complex with PF-07321332 compound, but interaction analyses of molecular dynamics simulations showed that this compound interacts with hydrogen bonds with such residues of main protease as Cys145, Glu166 and Gln189 (94). Similar results were obtained in this work and most of the selected compounds, potential 3CL pro inhibitors, were interacting with the above residues.

4.3 Strengths and weaknesses of the study

A strength of Project I, where 12 best drug candidates against SARS-CoV-2 were selected, was the implementation of four different databases in the virtual screening process. Natural products database and TCM databases offer a wide range of various plant-based compounds as potential drug candidates and existing drugs databases, provide drug candidates for COVID-19 patients much faster than in case of introduction of novel compounds into clinical application. Furthermore, using the withdrawn drug was advantageous, as not all the drugs included in the database are withdrawn because of safety reasons, but often because of marketing reasons.

Another strength of this study was to analyse docking results based on visual inspection of docked poses and selection of best candidates based on the presence of crucial interactions and not based on the docking scores only. Such an approach was helpful also in choosing the best pose for further molecular dynamics studies. MD simulations provided additional analyses of the binding mode as well as free energy calculations. Moreover, the clinical insights section was added, where safety of selected drug candidates was discussed. One of the important aspects in drug safety was the prediction of drug-drug interaction, which was based on the predicted inhibition of CYPs (isoenzymes of cytochrome 450). In cases of such diseases as COVID-19, where a combination of a few drugs needs to be administered to the patient, knowing CYP activity profile of used medications is important when selecting the best candidates for treatment. Additionally, toxicity and known side effects of those drugs were analysed in terms of potential risks for a COVID-19 patients. Such an approach can serve as a reliable source of information and help for choosing drug candidates for clinical trials.

The limitation of this study is lack of experimental validation of the results. At this point no collaboration with an experimental group was established; furthermore, because of the pandemic situation, time played an important role and making the results available as quickly as possible for other scientists was crucial. The limitation in case of future testing could be also the availability of some of the compounds, especially of those selected from TCM database and from natural product databases. Compounds that are not available from vendors need to be first isolated and purified before in vitro testing, which requires more time and effort. Availability of plants can also be challenging. However, experimental

validation of all available compounds would be recommended as a next step. A limitation in case of repurposed drugs is that some of the suggested drugs may not be suitable for COVID-19 patients when their known side effects or predicted drug-drugs interactions will be taken into consideration.

Another limitation is connected to the virtual screening technique. The 2D similarity search that was used in this project is not as precise as, for instance, a combination of this method with the 3D-shape similarity methods or with the comparison of the electrostatic potential of compounds (18). Additionally, as the lower threshold for Tanimoto score was not set up, and the choice of best candidates from each database was solely based on the top 10 scores, this resulted in a selection of compounds that did not present a very high similarity score. For the best 12 compounds, the scores ranged between 0.72 till 0.77 for the compounds from SuperDrug 2, SuperTCM and SuperNatural II database and between 0.63 and 0.71 for compounds from WITHDRWAN database. Despite the low similarity scores to the reference molecule, such compounds as Aliskiren (Tanimoto score 0.63) or Saquinavir (Tanimoto score 0.67) both showed a good docking result and were selected for the next step, which was molecular dynamics studies.

The strength of the Project II is a combination of in silico methods with in vitro studies of selected compounds and also extracts and fractions. Docking analyses determined the choice of the best compound from two species of *Reynoutria* rhizomes for experimental testing and, in the early stage, also helped to decide about the selection of the *Reynoutria* species, as a plant of interest for studies of SARS-CoV-2 inhibitors. Previous research experience with these plants and knowledge of various compounds of those species was also advantageous and made the isolation and extraction process faster.

The limitation of the study was not the fully known composition of plants extracts and fractions. Butanol fraction had the best inhibition parameters and higher inhibition values than single compounds, which could indicate that most potent compounds have not yet been identified. This could also mean that a combination of compounds and their synergic effect influence the inhibition and not a single compound. Another limitation is also the availability of compounds for in vitro testing. Not all of the docked compounds that showed interactions with crucial residues were tested in vitro. Additionally, molecular dynamics simulations studies were not conducted, in order to provide more extensive information

on the stability of selected compounds in complex with the enzyme. However, as this is a time-consuming method, that step was omitted, as the experimental validation was provided.

5. Conclusion

In this dissertation, application of computational methods to identify potential candidates for inhibitors of the main protease of coronavirus Sars-CoV-2 has been presented. The coronavirus pandemic, which started at the end of 2019, impacted the lives of most of the people around the world and also changed the focus of many research groups, which started to work on development of the vaccines or drugs against this virus. One of the fastest ways of introducing a drug for the clinical trials is to find a new indication of already existing drugs. In this work two databases of existing drugs, those of approved drug database and withdrawn drugs database, were used for virtual screening. Several drug candidates (Naldemedine, Eledoisin, Saralasin) from those databases has been suggested for further investigation, as potential treatment for COVID-19 and could be considered for further clinical testing for COVID-19 patients.

Beside drug repurposing methods, the focus of this work has been focused on investigating natural products as potential drug candidates for COVID-19. Natural products database and TCM products database were used for virtual screening and three compounds were identified as potential inhibitors of Mpro, which were: compound SN0001765, Pseudostellarin C and Dianthin E. In the next step, inhibition of those compounds could be tested in vitro to confirm the results of those findings. Additionally, compounds from rhizomes of *Reynoutria* species were investigated and two natural products from those plants: vanicoside A and vanicoside B were identified as best candidates and tested in vitro. Further investigation of those compounds, as well as in vitro testing of more compounds from *Reynoutria* species, could be done in a next steps, as most likely not all of the potential inhibitors from those extracts were identified.

This dissertation demonstrates the importance of computational methods in drug design and drug repurposing, as well as the potential of natural products in drug design. Most of the selected compounds were suggested as a potential SARS-CoV-2 main protease for the first time. Recent approval by FDA of a novel SARS-CoV-2 main protease inhibitor (PF-07321332 compound) shows that inhibition of the target, which was also chosen in

this work, can be effective for COVID-19 treatment. This can increase the interest in those compounds for further investigation and importance of those findings. Workflow and computational methods presented in this work could be also useful for investigation of compounds for other biological targets and different diseases in the future.

6. References

1. Ndwandwe D, Wiysonge CS. COVID-19 vaccines. *Curr Opin Immunol*. 2021 Aug;71:111–6.
2. Pouwels KB, Pritchard E, Matthews PC, Stoesser N, Eyre DW, Vihta KD, House T, Hay J, Bell JI, Newton JN, Farrar J, Crook D, Cook D, Rourke E, Studley R, Peto TEA, Diamond I, Walker AS. Effect of Delta variant on viral burden and vaccine effectiveness against new SARS-CoV-2 infections in the UK. *Nat Med*. 2021 Dec 14;27(12):2127–35.
3. Thakur V, Ratho RK. OMICRON (B.1.1.529): A new SARS-CoV-2 variant of concern mounting worldwide fear. *J Med Virol*. 2021 Dec 30;
4. Abel R, Paredes Ramos M, Chen Q, Pérez-Sánchez H, Coluzzi F, Rocco M, Marchetti P, Mura C, Simmaco M, Bourne PE, Preissner R, Banerjee P. Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front Chem*. 2020 Dec 23;8.
5. Nawrot-Hadzik I, Zmudzinski M, Matkowski A, Preissner R, Kęsik-Brodacka M, Hadzik J, Drag M, Abel R. Reynoutria Rhizomes as a Natural Source of SARS-CoV-2 Mpro Inhibitors—Molecular Docking and In Vitro Study. *Pharmaceuticals*. 2021 Jul 29;14(8):742.
6. Nawrot-Hadzik I, Choromańska A, Abel R, Preissner R, Saczko J, Matkowski A, Hadzik J. Cytotoxic Effect of Vanicosides A and B from *Reynoutria sachalinensis* against Melanotic and Amelanotic Melanoma Cell Lines and in silico Evaluation for Inhibition of BRAFV600E and MEK1. *Int J Mol Sci*. 2020 Jun 29;21(13):4611.
7. Sorokina M, Steinbeck C. Review on natural products databases: where to find data in 2020. *J Cheminform*. 2020 Dec 3;12(1):20.
8. Shen B. A New Golden Age of Natural Products Drug Discovery. *Cell*. 2015 Dec;163(6):1297–300.
9. Tu Y. The discovery of artemisinin (qinghaosu) and gifts from Chinese medicine. *Nat Med*. 2011 Oct 11;17(10):1217–20.
10. Weber T, Kim HU. The secondary metabolite bioinformatics portal: Computational tools to facilitate synthetic biology of secondary metabolite production. *Synth Syst Biotechnol*. 2016 Jun;1(2):69–79.
11. Dictionary of Natural Products. Available from: <https://dnp.chemnetbase.com/faces/chemical/ChemicalSearch.xhtml>
12. Super Natural II. Available from: https://bioinf-applied.charite.de/supernatural_new/
13. MassBank. Available from: <https://massbank.eu/MassBank/>
14. Fishedick JT, Johnson SR, Ketchum REB, Croteau RB, Lange BM. NMR spectroscopic search module for Spektraris, an online resource for plant natural product identification – Taxane diterpenoids from *Taxus*×media cell suspension cultures as a case study. *Phytochemistry*. 2015 May;113:87–95.
15. Newman DJ, Cragg GM. Natural Products as Sources of New Drugs from 1981 to 2014. Vol. 79, *Journal of Natural Products*. American Chemical Society; 2016. p. 629–61.
16. Sparks TC, Hahn DR, Garizi N V. Natural products, their derivatives, mimics and synthetic equivalents: role in agrochemical discovery. *Pest Manag Sci*. 2017 Apr;73(4):700–15.
17. Hughes JP, Rees S, Kalindjian SB, Philpott KL. Principles of early drug discovery. *Br J Pharmacol*. 2011 Mar;162(6):1239–49.
18. Gimeno A, Ojeda-Montes M, Tomás-Hernández S, Cereto-Massagué A, Beltrán-Debón R,

- Mulero M, Pujadas G, Garcia-Vallvé S. The Light and Dark Sides of Virtual Screening: What Is There to Know? *Int J Mol Sci.* 2019 Mar 19;20(6):1375.
19. Sabe VT, Ntombela T, Jhamba LA, Maguire GEM, Govender T, Naicker T, Kruger HG. Current trends in computer aided drug design and a highlight of drugs discovered via computational techniques: A review. *Eur J Med Chem.* 2021 Nov;224:113705.
 20. Parvathaneni V, Kulkarni NS, Muth A, Gupta V. Drug repurposing: a promising tool to accelerate the drug discovery process. *Drug Discov Today.* 2019 Oct;24(10):2076–85.
 21. Cui W, Aouidate A, Wang S, Yu Q, Li Y, Yuan S. Discovering Anti-Cancer Drugs via Computational Methods. *Front Pharmacol.* 2020 May 20;11.
 22. Khan A, Ali SS, Khan MT, Saleem S, Ali A, Suleman M, Babar Z, Shafiq A, Khan M, Wei DQ. Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease (3CLpro). *J Biomol Struct Dyn.* 2021 Sep 2;39(13):4659–70.
 23. Wu C, Liu Y, Yang Y, Zhang P, Zhong W, Wang Y, Wang Q, Xu Y, Li M, Li X, Zheng M, Chen L, Li H. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm Sin B.* 2020 May;10(5):766–88.
 24. Yadav R, Imran M, Dhamija P, Chaurasia DK, Handu S. Virtual screening, ADMET prediction and dynamics simulation of potential compounds targeting the main protease of SARS-CoV-2. *J Biomol Struct Dyn.* 2021 Nov 22;39(17):6617–32.
 25. Fischer A, Sellner M, Neranjan S, Smieško M, Lill MA. Potential Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 606 Million Compounds. *Int J Mol Sci.* 2020 May 21;21(10):3626.
 26. Kumar A, Zhang KYJ. Advances in the Development of Shape Similarity Methods and Their Application in Drug Discovery. *Front Chem.* 2018 Jul 25;6.
 27. Nantasenamat C, Prachayasittikul V. Maximizing computational tools for successful drug discovery. *Expert Opin Drug Discov.* 2015 Apr 3;10(4):321–9.
 28. Lin X, Li X, Lin X. A Review on Applications of Computational Methods in Drug Screening and Design. *Molecules.* 2020 Mar 18;25(6):1375.
 29. Eckert H, Bajorath J. Molecular similarity analysis in virtual screening: foundations, limitations and novel approaches. *Drug Discov Today.* 2007 Mar;12(5–6):225–33.
 30. Batool M, Ahmad B, Choi S. A Structure-Based Drug Discovery Paradigm. *Int J Mol Sci.* 2019 Jun 6;20(11):2783.
 31. Spiegel JO, Durrant JD. AutoGrow4: an open-source genetic algorithm for de novo drug design and lead optimization. *J Cheminform.* 2020 Dec 17;12(1):25.
 32. Xu X, Huang M, Zou X. Docking-based inverse virtual screening: methods, applications, and challenges. *Biophys Reports.* 2018 Feb 1;4(1):1–16.
 33. Lavecchia A. Machine-learning approaches in drug discovery: methods and applications. *Drug Discov Today.* 2015 Mar;20(3):318–31.
 34. Lo Y-C, Rensi SE, Torng W, Altman RB. Machine learning in chemoinformatics and drug discovery. *Drug Discov Today.* 2018 Aug;23(8):1538–46.

35. Yadav LR, Thapa P, Das L, Varma AK. Structureomics in Systems-Based Drug Discovery. In: Systems Biology Application in Synthetic Biology. New Delhi: Springer India; 2016. p. 33–51.
36. Waterhouse A, Bertoni M, Bienert S, Studer G, Tauriello G, Gumienny R, Heer FT, de Beer TAP, Rempfer C, Bordoli L, Lepore R, Schwede T. SWISS-MODEL: homology modelling of protein structures and complexes. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W296–303.
37. Robetta. Available from: <https://robetta.bakerlab.org/faqs.php>
38. Rohl CA, Strauss CEM, Misura KMS, Baker D. Protein Structure Prediction Using Rosetta. In 2004. p. 66–93.
39. Muhammed MT, Aki-Yalcin E. Homology modeling in drug discovery: Overview, current applications, and future perspectives. *Chem Biol Drug Des.* 2019 Jan;93(1):12–20.
40. Hollingsworth SA, Dror RO. Molecular Dynamics Simulation for All. *Neuron.* 2018 Sep;99(6):1129–43.
41. March-Vila E, Pinzi L, Sturm N, Tinivella A, Engkvist O, Chen H, Rastelli G. On the Integration of In Silico Drug Design Methods for Drug Repurposing. *Front Pharmacol.* 2017 May 23;8.
42. Lamb J, Crawford ED, Peck D, Modell JW, Blat IC, Wrobel MJ, Lerner J, Brunet JP, Subramanian A, Ross KN, Reich M, Hieronymus H, Wei G, Armstrong SA, Haggarty SJ, Clemons PA, Wei R, Carr SA, Lander ES, Golub TR. The Connectivity Map: Using Gene-Expression Signatures to Connect Small Molecules, Genes, and Disease. *Science (80-).* 2006 Sep 29;313(5795):1929–35.
43. Maggiora G, Vogt M, Stumpfe D, Bajorath J. Molecular Similarity in Medicinal Chemistry. *J Med Chem.* 2014 Apr 24;57(8):3186–204.
44. Gao K, Nguyen DD, Sresht V, Mathiowetz AM, Tu M, Wei G-W. Are 2D fingerprints still valuable for drug discovery? *Phys Chem Chem Phys.* 2020;22(16):8373–90.
45. Muegge I, Mukherjee P. An overview of molecular fingerprint similarity search in virtual screening. *Expert Opin Drug Discov.* 2016 Feb 4;11(2):137–48.
46. Berthold MR, Cebron N, Dill F, Gabriel TR, Kötter T, Meinl T, Ohl P, Thiel K, Wiswedel B. KNIME - the Konstanz information miner. *ACM SIGKDD Explor Newsl.* 2009 Nov 16;11(1):26–31.
47. RDKit. Available from: <https://www.rdkit.org/>
48. Durant JL, Leland BA, Henry DR, Nourse JG. Reoptimization of MDL keys for use in drug discovery. *J Chem Inf Comput Sci.* 2002 Nov;42(6):1273–80.
49. PubChem Sketcher V2.4. Available from: <https://pubchem.ncbi.nlm.nih.gov/edit3/index.html>.
50. Jakowiecki J, Abel R, Orzeł U, Pasznik P, Preissner R, Filipek S. Allosteric Modulation of the CB1 Cannabinoid Receptor by Cannabidiol—A Molecular Modeling Study of the N-Terminal Domain and the Allosteric-Orthosteric Coupling. *Molecules.* 2021 Apr 23;26(9):2456.
51. Pagadala NS, Syed K, Tuszynski J. Software for molecular docking: a review. *Biophys Rev.* 2017 Apr 16;9(2):91–102.
52. Grosdidier A, Zoete V, Michielin O. SwissDock, a protein-small molecule docking web service based on EADock DSS. *Nucleic Acids Res.* 2011 Jul 1;39(suppl):W270–7.
53. Sánchez-Linares I, Pérez-Sánchez H, Cecilia JM, García JM. High-Throughput parallel blind Virtual Screening using BINDSURF. *BMC Bioinformatics.* 2012 Sep 7;13(S14):S13.
54. Liu Y, Grimm M, Dai W, Hou M, Xiao Z-X, Cao Y. CB-Dock: a web server for cavity detection-

- guided protein–ligand blind docking. *Acta Pharmacol Sin.* 2020 Jan 1;41(1):138–44.
55. Yan Y, Zhang D, Zhou P, Li B, Huang S-Y. HDock: a web server for protein–protein and protein–DNA/RNA docking based on a hybrid strategy. *Nucleic Acids Res.* 2017 Jul 3;45(W1):W365–73.
 56. 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, Bai F, Liu H, Liu X, Guddat LW, Xu W, Xiao G, Qin C, Shi Z, Jiang H, Rao Z, Yang H. Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature.* 2020 Jun 11 [cited 2020 Jul 31];582(7811):289–93.
 57. Free Download: BIOVIA Discovery Studio Visualizer - Dassault Systèmes. Available from: <https://discover.3ds.com/discovery-studio-visualizer-download>
 58. PyMOL | pymol.org. Available from: <https://pymol.org/2/>
 59. Mortier J, Rakers C, Bermudez M, Murgueitio MS, Riniker S, Wolber G. The impact of molecular dynamics on drug design: applications for the characterization of ligand–macromolecule complexes. *Drug Discov Today.* 2015 Jun;20(6):686–702.
 60. Maestro | Schrödinger. Available from: <https://www.schrodinger.com/maestro>
 61. Greenidge PA, Kramer C, Mozziconacci J-C, Wolf RM. MM/GBSA Binding Energy Prediction on the PDBbind Data Set: Successes, Failures, and Directions for Further Improvement. *J Chem Inf Model.* 2013 Jan 28;53(1):201–9.
 62. Banerjee P, Eckert AO, Schrey AK, Preissner R. ProTox-II: a webserver for the prediction of toxicity of chemicals. *Nucleic Acids Res.* 2018 Jul 2;46(W1):W257–63.
 63. Banerjee P, Dunkel M, Kemmler E, Preissner R. SuperCYPsPred—a web server for the prediction of cytochrome activity. *Nucleic Acids Res.* 2020 Jul 2;48(W1):W580–5.
 64. Ji C, Svensson F, Zoufir A, Bender A. eMolTox: prediction of molecular toxicity with confidence. Valencia A, editor. *Bioinformatics.* 2018 Jul 15;34(14):2508–9.
 65. Garcia de Lomana M, Morger A, Norinder U, Buesen R, Landsiedel R, Volkamer A, Kirchmair J, Mathea M. ChemBioSim: Enhancing Conformal Prediction of In Vivo Toxicity by Use of Predicted Bioactivities. *J Chem Inf Model.* 2021 Jul 26;61(7):3255–72.
 66. Xiong G, Wu Z, Yi J, Fu L, Yang Z, Hsieh C, Yin M, Zeng X, Wu C, Lu A, Chen X, Hou T, Cao D. ADMETlab 2.0: an integrated online platform for accurate and comprehensive predictions of ADMET properties. *Nucleic Acids Res.* 2021 Jul 2;49(W1):W5–14.
 67. Siramshetty V, Williams J, Nguyễn ĐT, Neyra J, Southall N, Mathé E, Xu X, Shah P. Validating ADME QSAR Models Using Marketed
 68. Tyzack JD, Kirchmair J. Computational methods and tools to predict cytochrome P450 metabolism for drug discovery. *Chem Biol Drug Des.* 2019 Apr 15;93(4):377–86.
 69. Wishart DS, Feunang YD, Guo AC, Lo EJ, Marcu A, Grant JR, Sajed T, Johnson D, Li C, Sayeeda Z, Assempour N, Iynkkaran I, Liu Y, Maciejewski A, Gale N, Wilson A, Chin L, Cummings R, Le D, Pon A, Knox C, Wilson M. DrugBank 5.0: a major update to the DrugBank database for 2018. *Nucleic Acids Res.* 2018 Jan 4;46(D1):D1074–82.
 70. Sorokina M, Merseburger P, Rajan K, Yirik MA, Steinbeck C. COCONUT online: Collection of Open Natural Products database. *J Cheminform.* 2021 Dec 10;13(1):2.
 71. Li B, Ma C, Zhao X, Hu Z, Du T, Xu X, Wang Z, Lin J. YaTCM: Yet another Traditional Chinese

- Medicine Database for Drug Discovery. *Comput Struct Biotechnol J*. 2018;16:600–10.
72. Mohanraj K, Karthikeyan BS, Vivek-Ananth RP, Chand RPB, Aparna SR, Mangalapandi P, Samal A. IMPPAT: A curated database of Indian Medicinal Plants, Phytochemistry And Therapeutics. *Sci Rep*. 2018 Dec 12;8(1):4329.
 73. Burley SK, Bhikadiya C, Bi C, Bittrich S, Chen L, Crichlow G V, Christie CH, Dalenberg K, Di Costanzo L, Duarte JM, Dutta S, Feng Z, Ganesan S, Goodsell DS, Ghosh S, Green RK, Guranović V, Guzenko D, Hudson BP, Lawson CL, Liang Y, Lowe R, Namkoong H, Peisach E, Persikova I, Randle C, Rose A, Rose Y, Sali A, Segura J, Sekharan M, Shao C, Tao YP, Voigt M, Westbrook John D, Young JY, Zardecki C, Zhuravleva M. RCSB Protein Data Bank: powerful new tools for exploring 3D structures of biological macromolecules for basic and applied research and education in fundamental biology, biomedicine, biotechnology, bioengineering and energy sciences. *Nucleic Acids Res*. 2021 Jan 8;49(D1):D437–51.
 74. SuperDRUG2: a one stop resource for approved/ marketed drugs - PubMed. Available from: <https://pubmed.ncbi.nlm.nih.gov/29140469/>
 75. WITHDRAWN--a resource for withdrawn and discontinued drugs - PubMed. Available from: <https://pubmed.ncbi.nlm.nih.gov/26553801/>
 76. Super Natural II--a database of natural products - PubMed. Available from: <https://pubmed.ncbi.nlm.nih.gov/25300487/>
 77. Chen Q, Springer L, Gohlke BO, Goede A, Dunkel M, Abel R, Gallo K, Preissner S, Eckert A, Seshadri L, Preissner R. SuperTCM: A biocultural database combining biological pathways and historical linguistic data of Chinese Materia Medica for drug development. *Biomed Pharmacother*. 2021 Dec;144:112315.
 78. Nawrot-Hadzik I, Ślusarczyk S, Granica S, Hadzik J, Matkowski A. Phytochemical Diversity in Rhizomes of Three Reynoutria Species and their Antioxidant Activity Correlations Elucidated by LC-ESI-MS/MS Analysis. *Molecules*. 2019 Mar 21;24(6):1136.
 79. Mikolaj Zmudzinski, Wioletta Rut, Kamila Olech, Jarosław Granda, Mirosław Giurg, Małgorzata Burda-Grabowska, Linlin Zhang, Xinyuanyuan Sun, Zongyang Lv, Digant Nayak, Malgorzata Kesik-Brodacka, Shaun K. Olsen, Rolf Hilgenfeld MD. Ebselen derivatives are very potent dual inhibitors of SARS-CoV-2 proteases - PLpro and Mpro in in vitro studies. *bioRxiv*. 2020;
 80. Rut W, Groborz K, Zhang L, Sun X, Zmudzinski M, Pawlik B, Wang X, Jochmans D, Neyts J, Młynarski W, Hilgenfeld R, Drag M. SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat Chem Biol*. 2021 Feb 22;17(2):222–8.
 81. Bello M, Martínez-Muñoz A, Balbuena-Rebolledo I. Identification of saquinavir as a potent inhibitor of dimeric SARS-CoV2 main protease through MM/GBSA. *J Mol Model*. 2020 Dec 12;26(12):340.
 82. Kwofie SK, Broni E, Asiedu SO, Kwarko GB, Dankwa B, Enniful KS, Tiburu EK, Wilson MD. Cheminformatics-Based Identification of Potential Novel Anti-SARS-CoV-2 Natural Compounds of African Origin. *Molecules*. 2021 Jan 14;26(2):406.
 83. Zhao W, Xu G, Yu Z, Li J, Liu J. Identification of nut protein-derived peptides against SARS-CoV-2 spike protein and main protease. *Comput Biol Med*. 2021 Nov;138:104937.
 84. Fiscon G, Conte F, Farina L, Paci P. SAveRUNNER: A network-based algorithm for drug

- repurposing and its application to COVID-19. Marz M, editor. *PLOS Comput Biol*. 2021 Feb 5;17(2):e1008686.
85. Borkotoky S, Banerjee M. A computational prediction of SARS-CoV-2 structural protein inhibitors from *Azadirachta indica* (Neem). *J Biomol Struct Dyn*. 2021 Jul 24;39(11):4111–21.
 86. van Breemen RB, Muchiri RN, Bates TA, Weinstein JB, Leier HC, Farley S, Tafesse FG. Cannabinoids Block Cellular Entry of SARS-CoV-2 and the Emerging Variants. *J Nat Prod*. 2022 Jan 28;85(1):176–84.
 87. Mohamed K, Yazdanpanah N, Saghadzadeh A, Rezaei N. Computational drug discovery and repurposing for the treatment of COVID-19: A systematic review. *Bioorg Chem*. 2021 Jan;106:104490.
 88. Ford N, Vitoria M, Rangaraj A, Norris SL, Calmy A, Doherty M. Systematic review of the efficacy and safety of antiretroviral drugs against SARS, MERS or COVID-19: initial assessment. *J Int AIDS Soc*. 2020 Apr 15;23(4).
 89. Drożdżał S, Rosik J, Lechowicz K, Machaj F, Szostak B, Przybyciński J, Lorzadeh S, Kotfis K, Ghavami S, Łos MJ. An update on drugs with therapeutic potential for SARS-CoV-2 (COVID-19) treatment. *Drug Resist Updat*. 2021 Dec;59:100794. Available from: <https://linkinghub.elsevier.com/retrieve/pii/S1368764621000546>.
 90. Agarwal A, Rochweg B, Lamontagne F, Siemieniuk RA, Agoritsas T, Askie L, Lytvyn L, Leo YS, Macdonald H, Zeng L, Amin W, Barragan FA, Bausch FJ, Burhan E, Calfee CS, Cecconi M, Chanda D, Dat VQ, De Sutter A, Du B, Geduld H, Gee P, Harley N, Hashmi M, Hunt B, Jehan F, Kabra S K, Kanda S, Kim Y J, Kissoon N, Krishna S, Kuppalli K, Kwizera A, Lisboa T, Mahaka I, Manai H, Mino G, Nsutebu E, Preller J, Pshenichnaya N, Qadir N, Sabzwari S, Sarin R, Shankar-Hari M, Sharland M, Shen Y, Ranganathan S S, Souza J P, Stegemann M, Swanstrom R, Ugarte S, Venkatapuram S, Vuyiseka D, Wijewickrama A, Maguire B, Zeraatkar D, Bartoszko J J, Ge L, Brignardello-Petersen R, Owen A, Guyatt G, Diaz J, Kawano-Dourado L, Jacobs M, Vandvik P O. A living WHO guideline on drugs for covid-19. *BMJ*. 2020 Sep 4;m3379.
 91. FDA. Available from: <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-first-oral-antiviral-treatment-covid-19>
 92. FDA. Available from: <https://www.fda.gov/news-events/press-announcements/coronavirus-covid-19-update-fda-authorizes-additional-oral-antiviral-treatment-covid-19-certain>
 93. Li P, Wang Y, Lavrijsen M, Lamers MM, de Vries AC, Rottier RJ, Bruno MJ, Peppelenbosch MP, Haagmans BL, Pan Q. SARS-CoV-2 Omicron variant is highly sensitive to molnupiravir, nirmatrelvir, and the combination. *Cell Res*. 2022 Jan 20.
 94. Ahmad B, Batool M, Ain Q ul, Kim MS, Choi S. Exploring the Binding Mechanism of PF-07321332 SARS-CoV-2 Protease Inhibitor through Molecular Dynamics and Binding Free Energy Simulations. *Int J Mol Sci*. 2021 Aug 24;22(17):9124.

Statutory Declaration

"I, Renata Abel, by personally signing this document in lieu of an oath, hereby affirm that I prepared the submitted dissertation on the topic 'Computational Methods in Drug Repurposing and Natural Product Based Drug Discovery' / 'Computergestützte Methoden in der Arzneimittel-Repositionierung und der Naturstoffbasierten Wirkstoffforschung', independently and without the support of third parties, and that I used no other sources and aids than those stated.

All parts which are based on the publications or presentations of other authors, either in letter or in spirit, are specified as such in accordance with the citing guidelines. The sections on methodology (in particular regarding practical work, laboratory regulations, statistical processing) and results (in particular regarding figures, charts and tables) are exclusively my responsibility.

Furthermore, I declare that I have correctly marked all of the data, the analyses, and the conclusions generated from data obtained in collaboration with other persons, and that I have correctly marked my own contribution and the contributions of other persons (cf. declaration of contribution). I have correctly marked all texts or parts of texts that were generated in collaboration with other persons.

My contributions to any publications to this dissertation correspond to those stated in the below joint declaration made together with the supervisor. All publications created within the scope of the dissertation comply with the guidelines of the ICMJE (International Committee of Medical Journal Editors; www.icmje.org) on authorship. In addition, I declare that I shall comply with the regulations of Charité – Universitätsmedizin Berlin on ensuring good scientific practice.

I declare that I have not yet submitted this dissertation in identical or similar form to another Faculty.

The significance of this statutory declaration and the consequences of a false statutory declaration under criminal law (Sections 156, 161 of the German Criminal Code) are known to me."

Date

Signature

Declaration of your own contribution to the publications

I, Renata Abel, contributed the following to the below listed publications:

Publication 1:

Renata Abel, Maria Paredes Ramos, Qiaofeng Chen, Horacio Perez-Sanchez, Flaminia Coluzzi, Monica Rocco, Paolo Marchetti, Cameron Mura, Maurizio Simmaco, Philip E. Bourne, Robert Preissner, Priyanka Banerjee;

Computational prediction of potential inhibitors of the main protease of SARS-CoV-2;
Frontiers in Chemistry; 2020

Contribution:

I designed the project workflow and implemented the project. I have chosen databases for ligand based virtual screening, compared different types of molecular fingerprints, chosen most suitable fingerprints for the virtual screening, designed the KNIME workflow for similarity search, and performed similarity search. I have chosen the target for molecular docking studies, established molecular docking protocol with GOLD software, conducted molecular docking of chosen 80 compounds, and analysed the docked compounds through visual inspection of interaction with BIOVIA Discovery Studio and PyMOL software. I selected 12 compounds for the molecular dynamics simulations studies, contributed to the analyzing of obtained molecular dynamics results, and selected the best candidates for inhibitors of SARS-CoV-2 main protease. I generated 2D and 3D visualizations of compounds interactions with the enzyme, which were used to create Figures (Figure 1, 2, 3, 4, 5), provided list of interacting residues and structures of the compounds for Tables 2 and 3 in the manuscript and created Tables in Supplementary Materials (Table 1, 2, 3, 4, 5, 7, 8, 9 and 10). I searched for literature and participated in manuscript writing, manuscript editing and in responding to reviewers. I coordinated the research with collaborators from Spain and Italy.

Publication 2:

Izabela Nawrot-Hadzik, Mikołaj Zmudziński, Adam Matkowski, Robert Preissner, Małgorzata Kęsik-Brodacka, Jakub Hadzik, Marcin Drąg, **Renata Abel**;

Reynoutria rhizomes as a natural source of SARS-CoV-2 Mpro inhibitors - molecular docking and in vitro study;

Pharmaceuticals; 2021

Contribution:

I was responsible for conceptualization of the project, project management and supervision. I developed the methodology, chose natural compounds for molecular docking studies, conducted molecular docking with GOLD software, analysed the docking results with BIOVIA Discovery Studio and PyMOL software, and selected the best candidates for in-vitro studies. I generated 2D interaction diagrams and 3D visualizations of the best compounds in complex with enzyme, and prepared Figures 1, 2, 3 and Table 1, 2 in the manuscript, as well as Figure S1 and Tables S1 and S2 in Supplementary Materials. I searched for literature and participated in manuscript writing and manuscript editing.

Signature, date and stamp of first supervising university professor / lecturer

Signature of doctoral candidate

Excerpt from Journal Summary

Category: "PHARMACOLOGY and PHARMACY"

Journal: **Pharmaceuticals**, IF: 5.863, Rank: 38

JCR Year: 2020

Journal: **Frontiers in Chemistry**, IF: 3.782

JCR Year: 2019

Journal Frontiers in Chemistry was not included in Charité list of Journals Impact Factors 2019

IF of this journal for 2019 was taken from the report from the publisher website:

https://reports.frontiersin.org/progress/2019/?utm_source=fweb&utm_medium=frep&utm_campaign=ba-pr19



Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2

Renata Abel¹, María Paredes Ramos², Qiaofeng Chen¹, Horacio Pérez-Sánchez³, Flaminia Coluzzi^{4,5}, Monica Rocco^{5,6}, Paolo Marchetti⁶, Cameron Mura⁷, Maurizio Simmaco^{8,9}, Philip E. Bourne⁷, Robert Preissner¹⁰ and Priyanka Banerjee^{1*}

¹ Institute of Physiology, Charité–University Medicine Berlin, Berlin, Germany, ² METMED Research Group, Physical Chemistry Department, Universidade da Coruña (UDC), A Coruña, Spain, ³ Structural Bioinformatics and High-Performance Computing (BIO-HPC) Research Group, Universidad Católica de Murcia (UCAM), Murcia, Spain, ⁴ Department of Medical and Surgical Sciences and Biotechnologies, Sapienza University of Rome, Latina, Italy, ⁵ Unit of Anesthesia and Intensive Care Medicine, Sant' Andrea University Hospital, Rome, Italy, ⁶ Department of Clinical and Surgical Translational Medicine, Sapienza University of Rome, Rome, Italy, ⁷ Department of Biomedical Engineering and School of Data Science, University of Virginia, Charlottesville, VA, United States, ⁸ Department of Neurosciences, Mental Health and Sensory Organs, Sapienza University of Rome, Rome, Italy, ⁹ Advanced Molecular Diagnostic Unit, Sant' Andrea University Hospital, Rome, Italy, ¹⁰ Institute of Physiology and Science-IT, Charité–Universitätsmedizin Berlin, Corporate Member of Freie Universität Berlin, Humboldt-Universität zu Berlin, and Berlin Institute of Health, Berlin, Germany

OPEN ACCESS

Edited by:

Chandrabose Selvaraj,
Alagappa University, India

Reviewed by:

Kshiti Verma,
Genentech, Inc., United States
Mahmoud A. A. Ibrahim,
Minia University, Egypt

*Correspondence:

Priyanka Banerjee
priyanka.banerjee@charite.de

Specialty section:

This article was submitted to
Medicinal and Pharmaceutical
Chemistry,
a section of the journal
Frontiers in Chemistry

Received: 03 August 2020

Accepted: 13 November 2020

Published: 23 December 2020

Citation:

Abel R, Paredes Ramos M, Chen Q, Pérez-Sánchez H, Coluzzi F, Rocco M, Marchetti P, Mura C, Simmaco M, Bourne PE, Preissner R and Banerjee P (2020) Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front. Chem.* 8:590263. doi: 10.3389/fchem.2020.590263

The rapidly developing pandemic, known as coronavirus disease 2019 (COVID-19) and caused by the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), has recently spread across 213 countries and territories. This pandemic is a dire public health threat—particularly for those suffering from hypertension, cardiovascular diseases, pulmonary diseases, or diabetes; without approved treatments, it is likely to persist or recur. To facilitate the rapid discovery of inhibitors with clinical potential, we have applied ligand- and structure-based computational approaches to develop a virtual screening methodology that allows us to predict potential inhibitors. In this work, virtual screening was performed against two natural products databases, Super Natural II and Traditional Chinese Medicine. Additionally, we have used an integrated drug repurposing approach to computationally identify potential inhibitors of the main protease of SARS-CoV-2 in databases of drugs (both approved and withdrawn). Roughly 360,000 compounds were screened using various molecular fingerprints and molecular docking methods; of these, 80 docked compounds were evaluated in detail, and the 12 best hits from four datasets were further inspected *via* molecular dynamics simulations. Finally, toxicity and cytochrome inhibition profiles were computationally analyzed for the selected candidate compounds.

Keywords: virtual screening (VS), drug repurposing and molecular docking, SARS-CoV-2, COVID-19, computational drug discovery, molecular dynamics

INTRODUCTION

A novel coronavirus (CoV), known as the severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), began spreading among humans in December 2019 in the city of Wuhan, China, causing a major outbreak of often-fatal pneumonia (Wu et al., 2020). The rapid expansion of SARS-CoV-2 has been labeled a pandemic by the World Health Organization (WHO), and the global crisis

has continued to devastate both human health and national economies (WHO¹). The symptoms associated with most instances of this infection include fever, dry cough, fatigue, shortness of breath and respiratory issues (Wu et al., 2020), and deterioration of some sensory modalities (e.g., taste, smell); a smaller fraction of cases also present with other symptoms, e.g., conjunctivitis (presumably another mode of transmission, too) (Scasso et al., 2018). With SARS-CoV-2, aggressive human-human transmission has occurred, yielding exponential growth in the number of detected cases. The disease has now been termed as “coronavirus disease 2019” (COVID-19) (Zhang L. et al., 2020). At present, the number of confirmed cases reported internationally has reached 15,581,009, with 635,173 deaths reported². As of yet, no potent drug or vaccine has been reported (or approved) to treat individuals infected with SARS-CoV-2; only symptomatic treatment has been given to the most critically ill patients. A surge in activity among the scientific community has advanced research efforts toward the development of therapeutic intervention and finding viral drug targets; currently, 36 repurposed drugs are already used in experimental (unapproved) treatments for COVID-19, and 432 drugs are being tested in ongoing clinical trials³. In addition, there are 23 candidate vaccines in clinical evaluation and 140 candidate vaccines in preclinical evaluation⁴. Initial results from a phase 1 clinical trial are already available for a vaccine known as mRNA-1273 (Jackson et al., 2020). Recent reports suggest that some U.S. Food & Drug Administration (FDA)-approved drugs, specifically remdesivir (which inhibits viral RNA polymerase) (Al-Tawfiq et al., 2020) and lopinavir and ritonavir (HIV protease inhibitors) (Cao et al., 2020), may be effective against SARS-CoV-2. Remdesivir exhibits an antiviral activity with an EC₅₀ of 0.77 μM against SARS-CoV-2, and shorter recovery times (vs. a placebo group) were found for adults hospitalized with COVID-19 and treated with remdesivir; additionally, those patients showed fewer infections of the respiratory tract (Beigel et al., 2020). In March 2020, the WHO launched a “solidarity clinical trials” of repurposed drugs and experimental candidates, wherein testing of the three aforementioned drugs was supplemented with testing of the antimalarial compounds chloroquine and

hydroxychloroquine⁵. In July 2020, WHO decided to discontinue the hydroxychloroquine and lopinavir/ritonavir trials, as these compounds yielded little to no reduction in the mortality of hospitalized COVID-19 patients when compared to standard of care⁵. Currently, dexamethasone—an anti-inflammatory drug approved to treat COVID-19 patients in the UK and Japan—also has been reported, in an unpublished study, to reduce mortality among COVID-19 patients hospitalized with severe infection (Horby et al., 2020).

The earliest-discovered CoVs do not correspond to those strains that are the causative infectious agents in recent outbreaks, including COVID-19 (Khedkar and Patzak, 2020). The first “coronavirus” (to be termed as such) was isolated from chicken in 1937; human CoVs were identified years later, in the mid-1960s⁶. These viruses belong to the taxonomic family *Coronaviridae*, which are single-stranded, positive-sense RNA viruses of ~29.9 Kb genomic length (Khedkar and Patzak, 2020). The CoVs encode more than a dozen proteins, some of which have been identified as critical for viral entry and replication (Muramatsu et al., 2016). Among the structural proteins encoded by the CoV genome, four proteins are of special interest from the perspective of therapeutics and drug design—namely, the *spike* (S), *envelope* (E), *membrane* (M), and *nucleocapsid* (N) proteins. The S, E, and M proteins are housed in the membranes of these enveloped virions. The M and E proteins are actively involved in viral coat assembly, while the N protein is involved in compacting the RNA genome. The most-studied proteins thus far have been a papain-like protease (PLpro), a 3C-like protease (3CL^{Pro}), an endosomal protease, and the spike protein (Yang and Wang, 2020).

At the molecular level, CoVs are known to gain cellular entry *via* the S protein (Anand et al., 2003). Viral entry depends on the binding of the surface unit S1 of the S protein to a surface-exposed cellular receptor in the host, thereby supporting the process of viral attachment to target cell surfaces (Muramatsu et al., 2016). The 3CL protease (3CL^{Pro}), also known as M^{Pro}, is the main protease produced by the CoV; it plays a key role in viral replication (Wu et al., 2020). Most of the functional proteins of CoVs are encoded by specific genes, which are first translated into polyproteins that are then cleaved by the viral 3CL^{Pro} or by PLpro. This stage of the viral replication cycle yields the RNA-dependent RNA polymerase (RdRp), along with multiple other proteins that play roles in virus replication, transcription, and translation. Inhibiting the activity of the main CoV protease would presumably block viral replication (Yang and Wang, 2020). Thus, 3CL^{Pro} is considered a potential drug target for COVID-19. In addition, targeting the main protease for inhibition is an appealing strategy because it may well be thought that this would inactivate the virus in different cell types and

Abbreviations: SARS-CoV-2, severe acute respiratory syndrome coronavirus 2; WHO, World Health Organization; COVID-19, coronavirus disease 2019; FDA, Food & Drug Administration; EMA, European Medicines Agency; TCM, Traditional Chinese Medicine; MD, molecular dynamics; VS, virtual screening; 2D, two-dimensional; 3D, three-dimensional; LBVS, ligand-based virtual screening; SBVS, structure-based virtual screening; ADMET, absorption, distribution, metabolism, elimination, and toxicity; LD, lethal dose; RMSD, root-mean-square deviation; CYP, cytochrome P450; ACE, angiotensin-converting-enzyme.

¹https://www.who.int/health-topics/coronavirus#tab=tab_1 (accessed July 28, 2020).

²WHO Coronavirus Disease (COVID-19) Dashboard | WHO Coronavirus Disease (COVID-19) Dashboard. Available online at: https://covid19.who.int/?gclid=CjwKCAjwsO_4BRBBEiwAyagRTezBE06lkOleffHglbg0emog5Zo38YH6hOcVz3YyPc15LkKi0DCIxwhoCkGgQAvD_BwE (accessed July 28, 2020).

³DrugBank. Available online at: <https://www.drugbank.ca/covid-19> (accessed July 28, 2020).

⁴Draft landscape of COVID-19 candidate vaccines. Available online at: <https://www.who.int/publications/m/item/draft-landscape-of-covid-19-candidate-vaccines> (accessed July 28, 2020).

⁵Solidarity clinical trial for COVID-19 treatments. Available online at: <https://www.who.int/emergencies/diseases/novel-coronavirus-2019/global-research-on-novel-coronavirus-2019-ncov/solidarity-clinical-trial-for-covid-19-treatments> (accessed July 28, 2020).

⁶Coronavirus Disease 2019 (COVID-19) | CDC. Available online at: <https://www.cdc.gov/coronavirus/2019-ncov/index.html> (accessed July 30, 2020).

in different organs—regardless of the various matches between receptors/host proteases (on the cell membrane) that underlie viral entry in a cell- or tissue-specific manner (Zhang L. et al., 2020).

Because of its mechanistic significance, 3CL^{Pro} is now a central target for the development of effective inhibitors (antiviral drugs) against both SARS-CoV-2 and other known CoVs (Anand et al., 2003). The X-ray crystal structure of 3CL^{Pro}(M^{Pro}) from SARS-CoV-2 (PDB code: 6LU7) reveals a protein comprised of three primary domains (Jin et al., 2020). The first domain (Domain I) consists of amino acid residues 8–101; the second domain (Domain II) maps to residues 102–184; and the third domain (Domain III) mainly consists of residues 201–306, largely as a cluster of α -helical conformations (Jin et al., 2020). The substrate-binding region of 3CL^{Pro}, located between Domains I and II, includes residues His41 and Cys145. Visual inspection of the binding site structure confirms that peptide-type inhibitors attach to the active site cysteine: a previously identified peptidomimetic inhibitor, “N3,” interacts irreversibly with this site and engages in supporting interactions with subsites (S1, S2, and S4) (Jin et al., 2020). The S1 subsite contains residues His163, Glu166, Cys145, Gly143, His172, and Phe140, while the S2 subsite consists mainly of Cys145, His41, and Thr25; these amino acid types are compatible with favorable non-bonded contacts such as electrostatic interactions and van der Waals (apolar/dispersive) forces. There are two additional subsites (S3–S5), consisting of Thr190, Gln192, Glu166, Met49, Leu167, Gln189, and Met165 (Anand et al., 2003). A ligand interaction diagram drawn from the 3D crystal structure (Figure 1) illustrates that this particular N3 inhibitor engages in multidentate hydrogen bonds with Glu166 (as both donor and acceptor). In addition, there are close—and presumably energetically favorable—contacts between moieties of N3 and Gly143 and the catalytic Cys145 (both of the S1 subsite).

The traditional drug discovery and development pipeline is generally a quite time-consuming endeavor, taking upward of \approx 10–15 years (Turanli et al., 2019; Kupferschmidt and Cohen, 2020). Computational drug “repurposing” is an effective approach to accelerate this timescale by finding new uses for existing (and already approved) drugs (McNamee et al., 2017). Computational approaches to drug discovery, particularly as part of a repurposing strategy/framework, can fasten the drug development process and alleviate the burdens of traditional approaches—features that are especially critical in the context of a pandemic. Such computational approaches have been used to identify candidate drugs for several infectious diseases, including Ebola, Zika, dengue, and influenza (Cha et al., 2018). Many computational methods are available to examine the key interrelationships between chemical structure, biological/physiological systems, interactions between chemicals and (bio) molecular targets, and, finally, the ultimate therapeutic endpoints and diseases (Metushi et al., 2015).

For COVID-19, several recent studies have reported the computational screening of inhibitors for specific single targets, such as either the main protease (the aforementioned 3CL^{Pro}/M^{Pro}) or the spike protein (Shiryaev et al., 2017; Botta et al., 2018; Pizzorno et al., 2019; Ton et al., 2020; Wang, 2020;

Zhang D.-h. et al., 2020). To the best of our knowledge, the work reported here is the first study to report screening results including databases like Super Natural II⁷, as well as the TCM (Chen, 2011)⁸, and repurposing-focused databases, such as SuperDRUG2⁹ and WITHDRAWN database¹⁰. Additionally, our final selection of lead compounds is based on visual inspection and analysis of ligand interactions with crucial residues of the target (3CL^{Pro}), thereby implicitly incorporating human expertise and clinical insights into our workflow (once the number of candidates becomes manageable for manual analysis). The final 12 best candidates were further evaluated using molecular dynamics (MD) simulation studies, and clinical feasibility for repurposed drugs was further investigated.

The general approach used in this study is based on an integrated pipeline: we include a virtual screening of cheminformatics-driven databases, and we employ molecular similarity and molecular docking to identify promising drug target pairs. Additionally, MD simulation studies of selected compounds were performed to select the best candidates for main protease inhibitors and evaluate their stability and strength of interactions. Our screening is aimed at dual objectives: (i) first, to find potential new candidates in the Super Natural II⁷ and TCM (Chen, 2011)⁸, and (ii) second, to identify promising repurposing candidates in the SuperDRUG2⁸ and WITHDRAWN databases¹⁰.

MATERIALS AND METHODS

To facilitate the rapid discovery of target inhibitors with real clinical potential, we have employed a prediction strategy that is based on the fundamental principle of “neighborhood behavior” (Patterson et al., 1996) implemented as a computational pipeline that utilizes both *ligand-based* [two-dimensional (2D) chemical space of small molecules] and *structure-based* [protein three-dimensional (3D) spaces and features]. Our computational predictions use the first resolved crystal structure of SARS-CoV-2 main protease (at a resolution 2.16 Å). Currently, several crystal structures of the SARS-CoV-2 main protease have been experimentally determined (approximately 175 structures)¹¹. The methodological details of our approach and computational methods are described below.

Databases

In order to broadly screen, we utilized four different database resources: (i) SuperNatural II, a database of natural products⁷; (ii) The Traditional Chinese Medicine (SuperTCM) database, an in-house (unpublished) database that is manually created

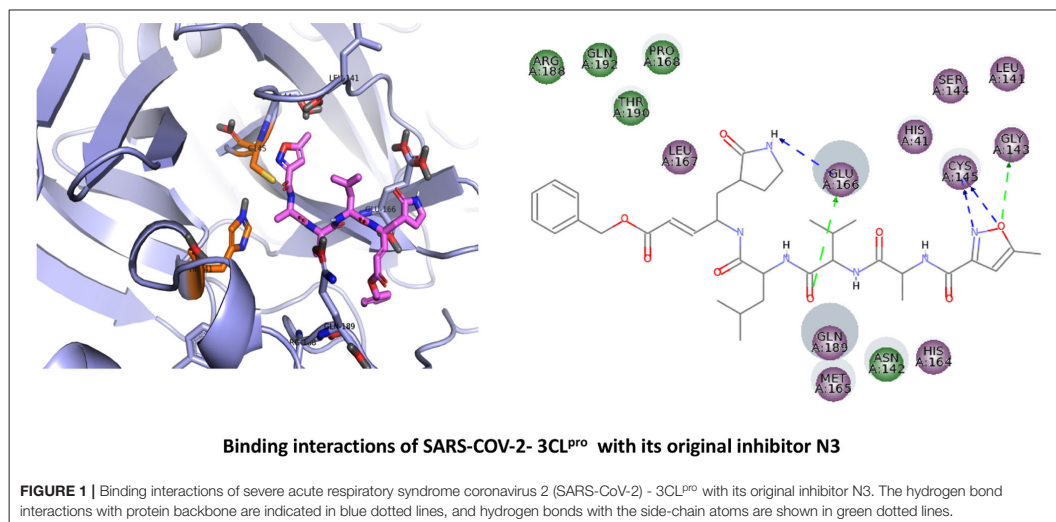
⁷Super Natural II—a database of natural products - PubMed. Available at: <https://pubmed.ncbi.nlm.nih.gov/25300487/> (accessed July 28, 2020).

⁸ETCM: an encyclopaedia of traditional Chinese medicine. Available online at: <https://pubmed.ncbi.nlm.nih.gov/30365030/> (accessed July 28, 2020).

⁹SuperDRUG2: a one stop resource for approved/marketed drugs. Available online at: <https://pubmed.ncbi.nlm.nih.gov/29140469/> (accessed July 28, 2020).

¹⁰WITHDRAWN—a resource for withdrawn and discontinued drugs - PubMed. Available online at: <https://pubmed.ncbi.nlm.nih.gov/26553801/> (accessed July 28, 2020).

¹¹RCSB PDB: Homepage. Available online at: <https://www.rcsb.org/> (accessed July 31, 2020).



from other databases and Chinese literature (Chen, 2011)⁸; this comprehensive database built at the Charité-University of Medicine^{12,13} covers all aspects of traditional Chinese medicine mainly derived from medicinal plants, and it encompasses pharmaceutical recipes up to molecular ingredients. The database was manually curated by domain experts, and Chinese plant-based drugs were mapped to their plant of origin, common non-scientific names and scientific names, targets, Kyoto Encyclopedia of Genes and Genomes (KEGG) pathways¹⁴, as well as the traditional Chinese recipes. (iii) We also used SuperDRUG2, a one-stop resource for approved/ marketed drugs⁸; (iv) And, finally, we also used WITHDRAWN, a resource for withdrawn and discontinued drugs¹⁰. Overall, we utilized more than 360,000 compounds for virtual screening and initial filtering purposes (Table 1).

The potential inhibitors are screened mainly from phytochemical databases because the literature suggests that seven out of 10 synthetic agents approved by the U.S. FDA are modeled on a natural product parent (Newman and Cragg, 2016). There is an urgent need to identify novel active chemotypes as leads for effective antiviral therapy for COVID-19 infections. Similarly, several thousands of plant extracts have been shown to possess *in vitro* antiviral activity with little overlap in species between studies (Chen, 2011). Promising docking outcomes have been executed in this study, which evidenced the

TABLE 1 | Databases and total number of compounds used in this study.

Database	Compounds	Total number of compounds used in this study
SuperDRUG2	Approved and marketed drugs	3,992
WITHDRAWN	Withdrawn or discontinued drugs	626
TCM	In-house database of compounds related to Traditional Chinese Medicine	28,974
Super Natural II	Natural compounds	325,508

worth of these selected chemical compounds from Super Natural II and TCM databases for future drug development to combat CoV diseases.

Additionally, drug databases from both approved and withdrawn chemical spaces were screened to support reuse of already available drugs for new indications such as COVID-19 therapy when they have been originally developed for specific diseases.

Ligand-Based Screening

We screened the molecular libraries from these databases based on ligand similarity, which rests upon the assumption that “structurally similar compounds might have similar biological properties” (Stumpfe and Bajorath, 2011). As many chemical-based fingerprint methods are in widespread use, based on the performances of various structural similarity measures, a 2D similarity screening protocol was designed, and three different types of molecular fingerprints were initially chosen for comparison and performance evaluation: MACCS (Durant et al.,

¹²Structural Bioinformatics Group. Available online at: <http://bioinf-apache.charite.de/main/index.php> (accessed July 31, 2020).

¹³Traditional Chinese Medicine: Institut für Theorie, Geschichte, Ethik Chinesischer Lebenswissenschaften - Charité - Universitätsmedizin Berlin Available online at: https://icl.charite.de/forschung/buch_details_2011_2020/traditional_chinese_medicine (accessed July 31, 2020).

¹⁴KEGG PATHWAY Database. Available online at: <https://www.genome.jp/kegg/pathway.html> (accessed July 31, 2020).

2002), Extended Connectivity Fingerprints (ECFP-4) (Rogers and Hahn, 2010), and E-state (Hall and Kier, 1995). We found that the performance of the MACCS fingerprint surpasses that of ECFP-4 and E-state in an approved drug dataset, with a higher Tanimoto score threshold (from 0.73 to 0.83) than ECFP-4 (Supplementary Table 1). MACCS ranked the similarity scores higher for co-crystallized ligands and drugs like lopinavir, angiotensin II, and indinavir (compounds that were previously reported as potential inhibitors of the main protease 3CL^{Pro}) (Contini, 2020; Nutho et al., 2020). Using this small set of active compounds, we tried to find the optimal similarity cutoff as well as optimal chemical fingerprint to yield the best balance of precision vs. recall.

Chemical-based fingerprints were calculated using RDKit¹⁵ nodes in KNIME (Berthold et al., 2008), and pairwise similarities were calculated for all the datasets.

The lead compounds used in the similarity search for targeting the 3CL^{Pro} of SARS-CoV-2 were obtained from the Protein Data Bank (PDB)⁵. The original ligands from PDB structures 6LU7 and 6Y2F were considered as lead compounds. The first ligand (“N3”) is a peptidomimetic irreversible inhibitor and was found covalently bonded to Cys145. This interaction is reported to be essential for preserving the protease’s S1 pocket in the right shape and also for the active conformation of the enzyme (Zhang L. et al., 2020). The N3 ligand interacts with the catalytic center of the target proteases through two hydrogen bond interactions. It was observed that the pyridine ring might have some steric clash with the side chain of Gln 189 (Zhang L. et al., 2020). The reported α -ketoamide ligand-bound X-ray crystal structure of SARS-CoV-2 Mpro (PDB ID: 6Y2F) forms hydrogen bonds with the Ser of chain B and Glu166 of chain A (Zhang L. et al., 2020). This interaction is reported to be essential for keeping the S1 pocket in the right shape for ligand–receptor interactions and also for the active conformation of the enzyme (Zhang L. et al., 2020). It interacts with the catalytic center of the target proteases through two hydrogen bond interactions. It was observed that the pyridine ring might have some steric clash with the side chain of Gln189 (Zhang L. et al., 2020). The compounds that showed the highest similarity considering structural properties were finally chosen for molecular docking studies. For each of the dataset and each of the lead compounds, the 10 best hits were chosen (Supplementary Tables 7, 8). The Tanimoto score values for selected compounds range between 0.63 and 0.83. It is also possible that a “false similar” or “false active” pair of molecules could occur, featuring structural similarity but dissimilarity in terms of their biological activities; to assess this possibility regarding activity profiles, further molecular docking and MD simulation studies were conducted.

Structure-Based Screening

To further refine the list of candidate compounds and select the top hits, molecular docking calculations were carried out using the GOLD software (version 5.7.2) (Jones et al., 1997). This code uses a genetic algorithm to sample the ligand’s conformational space, making it particularly suitable for docking flexible ligands

with numerous rotational degrees of freedom (Jones et al., 1997). In addition, the GOLD scoring function was used to rank the compounds, with the number of docked poses to be generated set to 10. In the present study, the co-crystal structures of 3CL^{Pro} of SARS-CoV-2 (PDB 6LU7) was selected as the starting structure for docking calculations. Residues that are proximal (within a 10 Å radius) to the original ligand co-crystallized in 6LU7, along with binding site residues (as defined in the literature) were taken to be the active site for docking calculations. Thus, the final docking protocol incorporates information from the successful re-docking of the original ligand to the target.

A total of 80 compounds (top 10 screened compounds for each lead compound, selected using ligand-based screening from four different databases) were docked into the main protease protein 3CL^{Pro}. Based on interactions with key residues (His41 or Cys145), present in the binding cavity of the 6LU7 crystal structure (and based on visual inspection of the ligand–receptor binding interactions), the top 3 best candidates were selected per database. Therefore, our final list includes the 12 best candidates based on this computational screening protocol. Two-dimensional ligand–receptor binding interaction maps were computed using Accelrys Discovery Studio (version 4.5)¹⁶. The 3D interactions and structural illustrations were created using PyMOL (version 2.3.5)¹⁷. After the molecular docking analyses, MD simulations were performed for the best 12 screened candidates, enabling us to evaluate the dynamical stability of the bound/docked complexes, at least on the timescale of the MD trajectories.

Molecular Dynamics Simulation

Analyzing the dynamic evolution of the ligand-bound system helps us predict the stability of those interactions that we first detected via ligand-based virtual screening (LBVS) and structure-based virtual screening (SBVS); ideally, the MD faithfully recapitulates the real (physiological) environment of such interactions. Ligand–protein interactions at the binding site can be monitored for a period of time, so that ligands with more temporally stable poses can be detected and proposed as better candidates for 3CL^{Pro} inhibition. Thus, after our database screening stages, MD simulations of the different complexes were computed using the GPU version of Desmond included with Maestro suite 2019.4 (Schrödinger LLC)¹⁸ on a workstation with a NVIDIA QUADRO 5000. The various drug/receptor complexes were solvated in an aqueous environment in a cubic box with a minimal distance of 10 Å between the biomolecule and the box boundary (for periodic boundary conditions). Next, systems were neutralized and maintained in 0.15 M NaCl. The OPLS3 force field and the TIP3P-TIP4P water model were employed (Mark and Nilsson, 2001). Initially, the systems were simply

¹⁵RDKit. Available online at: <https://www.rdkit.org/> (accessed July 31, 2020).

¹⁶Free Download: BIOVIA Discovery Studio Visualizer - Dassault Systèmes. Available online at: <https://discover.3ds.com/discovery-studio-visualizer-download> (accessed July 31, 2020).

¹⁷PyMOL | pymol.org. Available online at: <https://pymol.org/2/> (accessed July 31, 2020).

¹⁸Maestro | Schrödinger. Available online at: <https://www.schrodinger.com/maestro> (accessed July 31, 2020).

energy-minimized for 2,000 time steps. Next, systems were allowed to execute free dynamics in the NPT ensemble; pressure was controlled using the Martyna–Tobias–Klein methodology, and the Nose–Hoover thermostat was employed to maintain the system near 310 K. Production-grade MD trajectories were extended to a total duration of 100 ns per system. MD trajectories were characterized in terms of the root-mean-square deviation (RMSD) of fluctuations of ligand and enzyme, particularly in terms of the main interactions with the top interacting residues. The trajectories were also used to assess the stabilities of the protein secondary structures (in complex with potential inhibitor) by plotting RMSDs. Additionally, to estimate the relative binding free energies of the 12 final compounds and also N3 ligand to the macromolecule, molecular mechanics–generalized Born surface area (MM–GBSA) method was applied. The MM–GBSA method is based on the difference between the free energies of the protein, ligand, and the complex in solution. The free energy for each species involved in the reaction (ligand, protein, and ligand–protein complex) is described as a sum of a gas-phase energy, polar and non-polar solvation terms, and an entropy term. In our computational protocol, the MM–GBSA method is used to calculate the free energy (ΔG) related to all poses obtained in the MD simulation by using the OPLS3 as implemented in the Small-Drug Design Suite of Schrodinger (Kollman et al., 2000; Greenidge et al., 2013).

Absorption, Distribution, Metabolism, Elimination, and Toxicity Properties

The docked compounds were further filtered using the standard ADMET (Absorption, Distribution, Metabolism, Elimination, and Toxicity) pharmacokinetic properties. Computational toxicity analysis was performed using the ProTox-II methods¹⁹. The ProTox-II web server currently holds 40 different predictive models, incorporating chemical similarity, fragment propensities, most frequent features, pharmacophores, and machine learning for toxicity prediction. The acute toxicity value of the ProTox-II method is divided into six classes based on a globally harmonized system of classification of labeling of chemicals (GHS). The classes are described as: Class I: fatal if swallowed ($LD_{50} \leq 5$); Class 2: fatal if swallowed ($5 < LD_{50} \leq 50$); Class 3: toxic if swallowed ($50 < LD_{50} \leq 300$); Class 4: harmful if swallowed ($300 < LD_{50} \leq 2,000$); Class 5: may be harmful if swallowed ($2,000 < LD_{50} \leq 5,000$); Class 6: non-toxic ($LD_{50} > 5,000$)³⁸. Additionally, cytochrome (CYP) inhibition profiles of each compound were computed using the SuperCYPsPred web server²⁰. Currently, the SuperCYPsPred web server includes 10 models for five major CYPs isoforms (including 3A4, 2C9, 2C19, 2D6, and 1A2). These cytochrome predictive models are based on machine learning methods (see **Tables 2–4** in the Results section).

¹⁹ProTox-II: a webserver for the prediction of toxicity of chemicals. Available online at: <https://pubmed.ncbi.nlm.nih.gov/29718510/> (accessed July 28, 2020).

²⁰SuperCYPsPred—a web server for the prediction of cytochrome activity. Available online at: <https://academic.oup.com/nar/article/48/W1/W580/5809167> (accessed July 28, 2020).

RESULTS

Overview

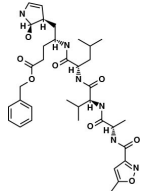
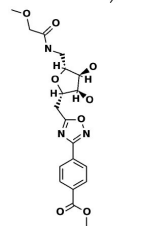
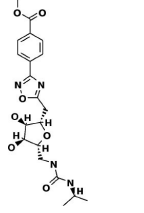
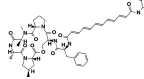
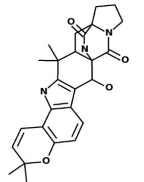
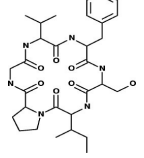
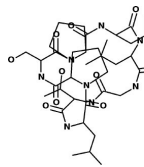
A similarity-based approach, using structural chemical fingerprints, was used for initial screening of compounds from different databases. The crystal structure of the main protease was used as the molecular target for computational docking and protein–ligand interaction analyses; in total, visual inspections were performed for 80 compounds. The top 40 compounds for the N3 inhibitor-based docking are shown in **Supplementary Table 7**, and the top 40 compounds for the 6OK inhibitor-based efforts are given in **Supplementary Table 8**. Using our integrated approach—i.e., chemical similarity with molecular docking—a list of 12 top lead compounds were identified from four different databases. Visual inspection and selection of best candidates were based on analyses of the interaction with at least one of the two catalytic residues Cys154 or His41. The results for these compounds are shown in **Tables 2, 3**. Besides the selected compounds, the tables also include information on therapeutic endpoints, toxicity endpoints, and cytochrome activity profiles of the compounds; interactive residues in the protease receptor are also given. For compounds from the Traditional Chinese Medicine database, information on their respective plants (as curated from our database) is also provided, with additional information on their indications and effects (**Supplementary Table 6**). Furthermore, to elucidate potential molecular mechanisms in terms of ligand interactions, the top 3 compounds from each database were selected and examined via all-atom MD simulations. The 12 best candidates are described below, including the RMSD plots obtained from the MD trajectories. For the validation of the MD protocol, simulation of N3 ligand was also performed. Results are presented in **Supplementary Figures 5, 6**. Additional information on binding interactions from molecular docking studies (for a selected subset of 56 compounds) is provided in **Supplementary Table 9**, and docking scores for the N3 and O6K ligands are provided in **Supplementary Table 10**.

Super Natural II Database

The top 3 compounds we identified in the Super Natural II database are SN00017653, SN00019468, and SN00303378. The compound SN00017653 interacts *via* hydrogen bonds with the side-chain atoms of Glu166, Ser144, Leu141, Gly143, and Cys145 (**Figure 2A**). The second compound, SN00019468, interacts with Cys145, Gly143, and Ser144 and hydrogen-bonds with the backbone of His41 (**Figure 2B**). The third compound, SN00303378, interacts with His41, Asn142, Thr190, Gln192, and Ser144 (**Figure 2C**). In addition, the top 10 compounds from the Super Natural database are reported in **Supplementary Table 2**.

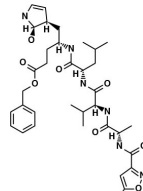
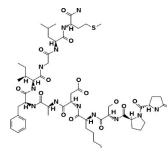
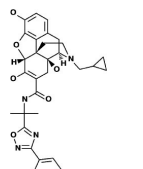
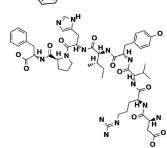
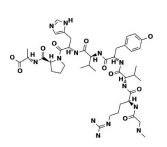
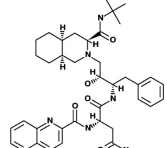
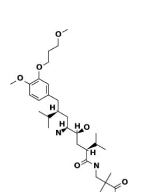
MD simulations reveal that SN00017653 exhibits a stable pose when interacting with 3CL^{Pro}, at least in terms of RMSD values of ≈ 2.0 and 2.8 Å for ligand and protein atoms, respectively. This compound exhibits quite stable contacts with the S2 subsite, where it interacts with Thr26 and His41. The time evolution of protein–ligand contacts shows also a weak interaction with Cys145, but this contact is highly transient. Furthermore, SN00017653 interacts with the S3 and S5 sub-pockets by binding

TABLE 2 | Potential inhibitors for the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from Super Natural II and SuperTCM databases.

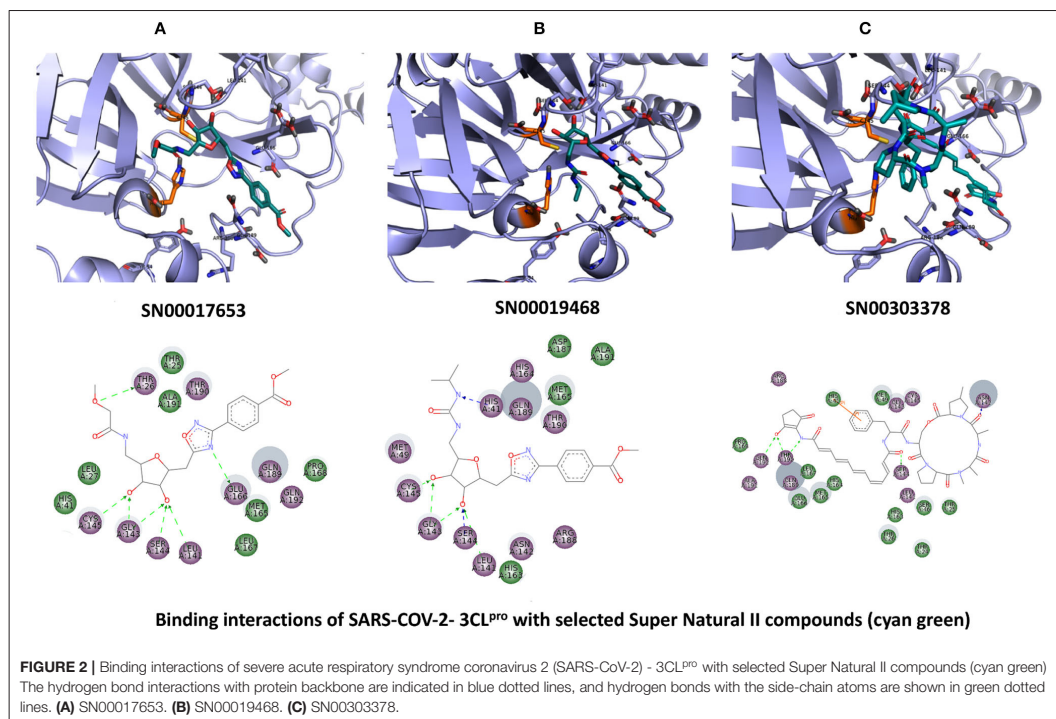
Compound	Interacting residues (3CL)	Acute toxicity	Toxicity endpoints	CYP activity	Tanimoto score	Structure*
N3-inhibitor	Glu166, Cys145, Gly143	Class 5	NA	CYP3A4	1	
Super Natural II database						
SN00017653	Cys145, Gly143, Glu166, Ser144, Leu141, Thr26	Class 4	NA	NA	0.75	
SN00019468	His41, Cys145, Gly143, Ser144, Leu141	Class 4	NA	NA	0.74	
SN00303378	His41, Thr190, Gln192, Ser144	Class 3	Immunotoxic	CYP3A4	0.73	
TCM database						
Notoamide R	Cys145, His41, His164, Gln189	Class 4	Immunotoxic	3A4	0.75	
Dianthin E	Cys145, Gln189, Glu166, Ser144, Gly143	Class 4	None	None	0.74	
Pseudostellarin C	Cys145, His164, His41, Glu166	Class 4	Immunotoxic	None	0.77	

*Structures of compounds generated with PubChem Sketcher V2.4 (Hillenfeldt et al., 2009).

TABLE 3 | Potential inhibitors for the main protease of severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) from SuperDRUG2 and WITHDRAWN drug databases.

Compound	Interacting residues (3CL)	Therapeutic endpoints	Acute toxicity	Toxicity endpoints	CYP activity profile	Tanimoto score	Structure*
N3-inhibitor	Glu166, Cys145, Gly143		Class 5	NA	CYP3A4	1	
Approved drugs database							
Eledoisin	Ser144, Arg188, Asn142, Cys145, His41	Vasodilator	Class 5	None	None	0.72	
Naldemedine	His41, Cys145, Gln192, His164	Alimentary tract and metabolism	Class 4	Immunotoxic	3A4	0.72	
Angiotensin II	Cys145, Gln189, Asn142	Cardiac therapy	Class 5	None	None	0.73	
Withdrawn drugs database							
Saralasin	Cys145, Met165, Gln189, Arg188	Cardiac therapy	Class 5	None	None	0.71	
Saquinavir	Cys145, Gly143, His41, Glu166	Antiviral	Class 4	None	2C8, 2C9, 2C19, 2D6, 3A4, 3A5	0.67	
Aliskiren	Tyr54, Cys145, Ser144	Cardiac therapy	Class 5	Immunotoxic	3A4	0.63	

*Structures of compounds generated with PubChem Sketcher V2.4 (Ihlenfeldt et al., 2009).



Glu166, Gln189, and Thr190. Interactions with Glu166 persist after 100 ns, and it shows a high interaction score in comparison with the other key residue, so it is considered the most important interaction (Figure 6, Supplementary Figure 1).

Compound SN00019468 shows high ligand RMSD fluctuations, especially during 50–100-ns trajectory. Examination of protein-ligand interaction contacts reveals that SN00019468 samples two distinct positions. During the first 50 ns, this ligand interacts with Thr26, His164, and Gln189, being only contact important for 3CL^{pro} inhibition, so it is part of the S3–S5 subsite. During the last 50 ns, these contacts evolve to an interaction with the S2 subsite by means of His41, but this interaction is not persistent, maintained during 50–90 ns, but lost after the 100-ns simulation. The poor stability of SN00019468 in the binding site suggests it is less likely to be an effective 3CL^{pro} inhibitor (Figure 6).

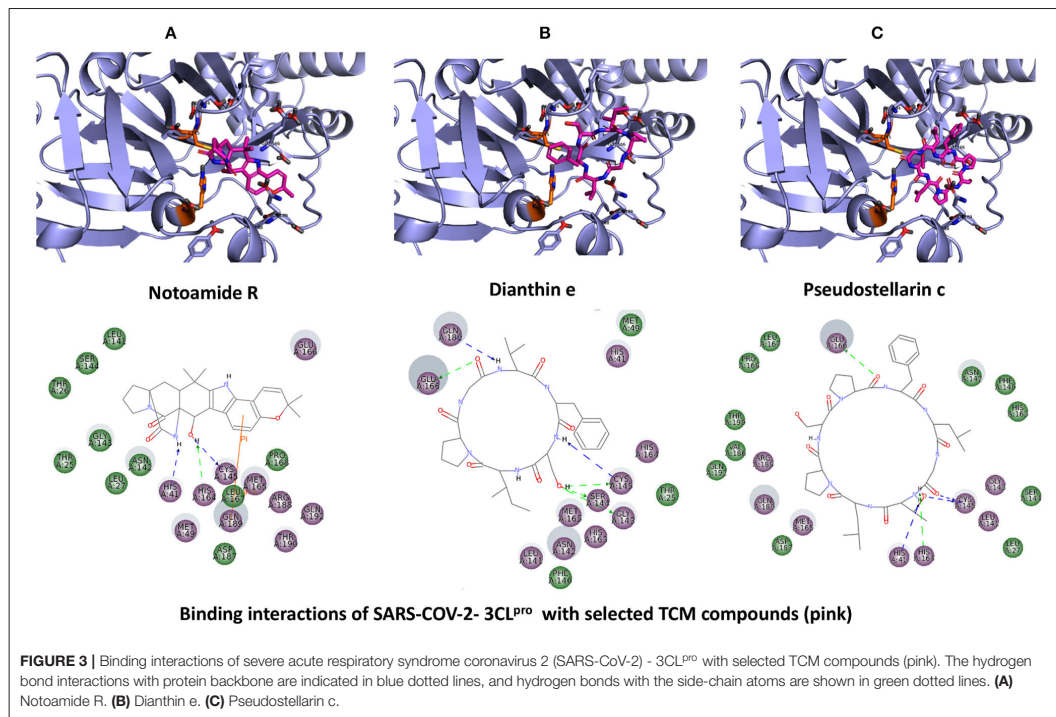
The ligand and the protein RMSDs are stable for compound SN00303378, indicating a stable position of this compound when bound to the protein. The pattern of protein contacts (Figure 6) shows an interaction with His41, but it fluctuates along the 100-ns simulation (being more important in the \approx 50–70-ns range). Also, the RMSD is stabilized during this period, with closely matched values for protein and ligand atoms. Glu166 engages in the clearest interactions, persisting along the full trajectory. Nevertheless, SN00303378 is not considered a strong candidate

to inhibit 3CL^{pro} because of its paucity of contacts with other key residues of the main protease (Figure 6).

Traditional Chinese Medicine Database (SuperTCM)

The top 3 compounds from the SuperTCM database are notoamide R, dianthin E, and pseudostellarin C. Notoamide R interacts with Cys145, His41, His164, and Gln189 (Figure 3A). Dianthin E interacts with the backbone atoms of Cys145 and Gln189, and the side-chain atoms of Glu166, Ser144, and Gly143 chiefly *via* hydrogen bonds (Figure 3B). Pseudostellarin C interacts with Cys145, His164, His41, and Glu166 (Figure 3C). Further information on the top 10 compounds from the SuperTCM database is reported in Supplementary Table 3.

MD simulations do not suggest that Notoamide R engages in notable interactions with the binding site and, in fact, reveals an unstable pattern of protein–ligand contacts. Regarding the time evolution of the RMSDs and protein–ligand contacts, this ligand appears to adopt rather different conformations along the trajectory. During the first 30 ns, it occupies a stable conformation and nominally interacts with His41 and, to a lesser extent, with Cys145. A conformational change then occurs, and the positions adopted by the ligand from 30 to 100 ns preclude contacts with binding-site residues. Thus, this ligand



is not considered as a suitable candidate for 3CL^{Pro} inhibition (Figure 7).

The RMSD graphic of Dianthin E shows a highly stable pose along the first ≈ 40 ns. After that, a peak in the plot indicates a shift in the pose; nevertheless, the simulation concludes with stable RMSD values for protein and ligand (granted, these values are higher than for the other ligands simulated here). The patterns of protein–ligand contacts are in concordance with the RMSD fluctuations. The first ≈ 40 ns saw some interaction with Cys145 and, in a clearer way, with Glu166 and Gln189. The ligand reoriented in the next ≈ 60 ns, corresponding to a weak bond with His41 and to a stabilization of the Gln189 bond (Figure 7).

The RMSD trace for Pseudostellarin C shows a high stable pose from the first 40 ns. After that, there is a peak that indicates a pose change, but the simulation ends with stable RMSD values for protein and ligand. The protein–ligand contacts are aligned with the RMSD fluctuation. During the first 40 ns, there was some interaction with Cys145 and with higher interaction score values with Glu166 and Gln189 (Supplementary Figure 2). The ligand reorientation after the following 60 ns was translated to an intermittent bond with His41 and to a persistent bond with Gln189 (Figure 7).

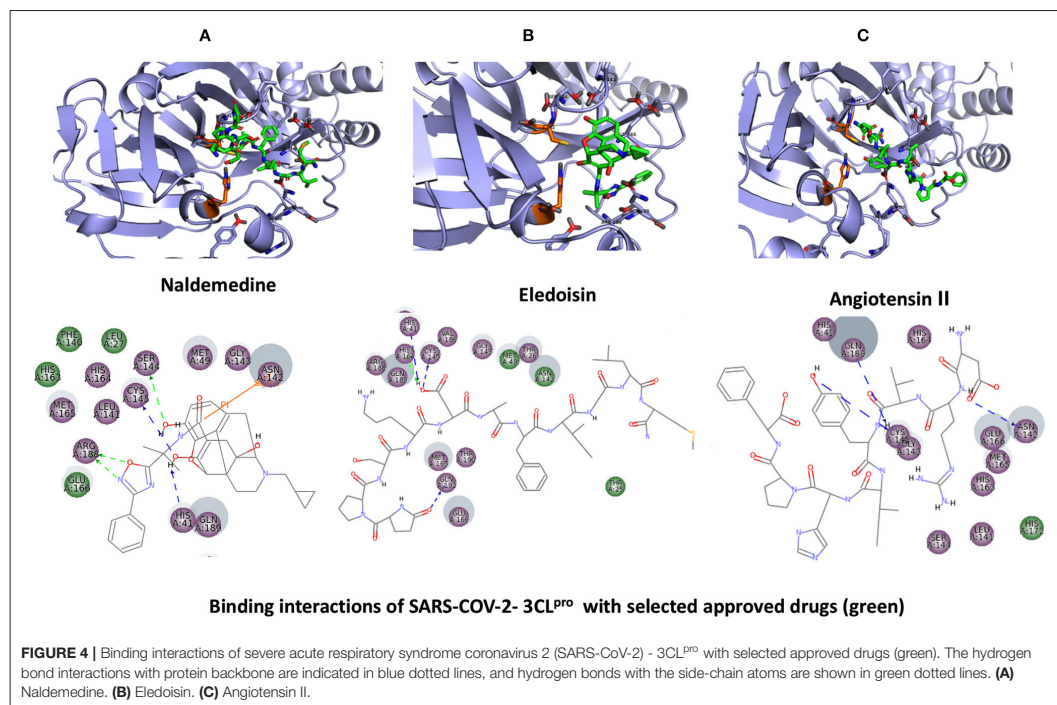
SuperDrug2 Drug Database

The top three selected compounds from the approved drug database include naldemedine, eledoisin, and angiotensin II.

Naldemedine interacts with Ser144, Arg188, and Asn142 (Figure 5A). It also interacts with the catalytic residues Cys145 and His41. Eledoisin also interacts with the catalytic residues His41 and Cys145 and with two other residues, Gln192 and His164 (Figure 5B). Angiotensin II interacts with Cys145, Gln189, and Asn142 (Figure 5C). More information on the 10 best drug candidates from the SuperDRUG2 database is reported in Supplementary Table 4.

Naldemedine shows close values for protein and ligand RMSDs, and the ligand pose is considered stable. Regarding protein–ligand contacts, the residues Glu166 and Gln189 were stable in contact throughout the simulation. Initially, also Cys145 showed a continuous interaction, so the S2 sub-pocket would be the main location responsible for 3CL^{Pro} inhibition. After the first 60 ns, this interaction becomes less stable and the contacts with Glu166 and Gln189 become more prominent, being the S3–S5 subunits responsible for the inhibition. Hence, an inhibition mode that is initially dominated by interactions at the S2 sub-pocket (and a relatively minor S3–S5 presence) evolves over the course of the trajectory to feature dominant S3–S5 inhibition (Figure 8).

Eledoisin shows persistent interaction with Thr26. Also, it interacts with His41 and, to a lesser extent, with Cys145. This means that this ligand would efficiently bind the S2 subsite of 3CL^{Pro}. In addition, it interacts with Glu166 and Gln189. However, the glutamine contacts, despite being a



strong (hydrogen bond) interaction, are relatively transient in the simulation; therefore, interactions at the S3–S5 site are considered less important than at the S2 site. Although the RMSD values are slightly higher than with saquinavir, we suspect that eledoisin could be a viable candidate to inhibit 3CL^{pro} (Figure 8).

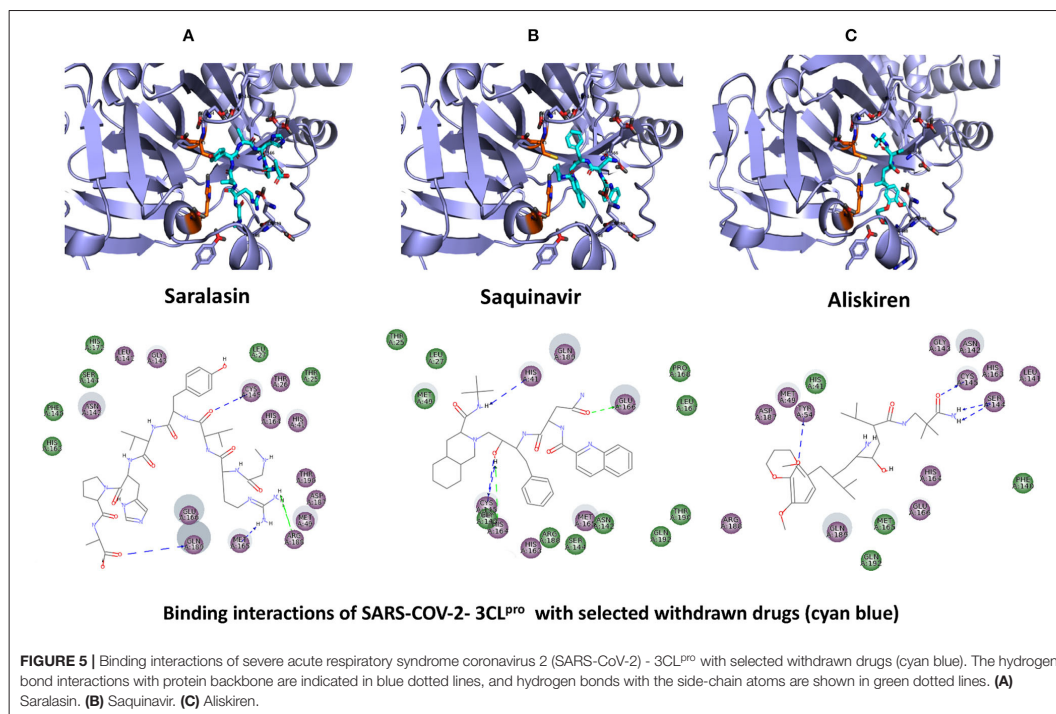
The protein–ligand complex with angiotensin II exhibits somewhat elevated RMSD values. Angiotensin II reliably contacts the S1 subunit, being Gly143 and Glu166 responsible for the stronger and most stable interactions during the simulation. The S2 site was also the location of interactions with Thr26, but neither His41 nor Cys145 contacts were relevant. Furthermore, it interacts with the S3 and S5 sub-pockets by binding Glu166 and Gln189. The continuity and interaction scores of these contacts were remarkable during the whole simulation, so they are considered the most important key residues (Figure 8, Supplementary Figure 3).

WITHDRAWN Drug Database

The top 3 compounds from the WITHDRAWN database are saralasin, saquinavir, and aliskiren. Saralasin interacts with Cys145, Met165, Gln189, and Arg188 (Figure 6). Saquinavir interacts with Cys145 (backbone), Gly143, His41 (backbone), and Glu166 (similar to original N3 ligand) (Figure 6). Aliskiren interacts with Tyr54, Cys145, and Ser144 (Figure 6). More

information on the 10 potential drug candidates from the WITHDRAWN database is reported in Supplementary Table 5.

The protein–ligand complex with saralasin exhibits overall structural stability, although the RMSD values are slightly higher than for the other systems. Saralasin has no relevant interactions within the S1 sub-pocket, formed by Thr25, His41, and Cys145 (it interacts with Cys145 during the first 10 ns, but this key contact was lost thereafter). Saralasin does show clear, persistent interactions with the S3–S5 subunit, mediated by interactions with Glu166 and Gln189. The relatively low protein and ligand RMSD values for the saquinavir trajectory reflect the structural rigidity of this system, which maintains a stable conformation during the full 100-ns simulation. The ligand engages in energetically favorable contacts with His41 and Glu166 during the whole simulation. The role of these two key residues as primary positions of interaction/attachment could strongly anchor saquinavir to the protein, so this compound is considered a potential candidate to inhibit 3CL^{pro}. The protein–ligand complex with aliskiren has low RMSD values. Nevertheless, despite maintaining the same conformation during the simulation, a sparse and transient pattern of interactions with protein residues would likely correspond to an unstable complex with aliskiren; this compound is not expected to be a good candidate for 3CL^{pro} inhibition (Figure 9, Supplementary Figure 4).



Calculation of Relative Protein–Ligand Binding Free Energies Using the Molecular Mechanics–Generalized Born Surface Area Method

Notoamide R, which showed an inefficient performance to inhibit 3CL^{Pro} during the MD analysis due to poor contacts with binding site residues and high instability, also showed poor MM-GBSA values, with -21.4 kcal/mol.

SN00017653, which is considered a strong inhibitor considering the RMSD analysis from MD results, shows a strong affinity, with -54.5 kcal/mol, and low deviation values. This MM-GBSA value is comparable to SN000303378, but this ligand, despite having a strong affinity value, does not interact with key binding site residues. SN00019468 shows also a strong affinity value, but this ligand does not interact significantly with the binding site residues and also has a high instability, which is translated to an elevated standard deviation value (16.5 kcal/mol), which is comparable to notoamide R; these two ligands are less suitable for inhibition of 3CL^{Pro}.

Both dianthin E and pseudostellarin C suffer a pose rearrangement during the 100-ns simulation, but the interaction within the binding site persists during the whole simulation despite of interacting with different residues. Thus, they were considered possible 3CL^{Pro} inhibitors, and the MM-GBSA values indicate their strong affinity.

Aliskiren showed a stable conformation during the 100-ns simulation, but with intermittent contacts with binding site residues. MM-GBSA values show a strong interaction, but this is not relevant due to the discontinuous contacts.

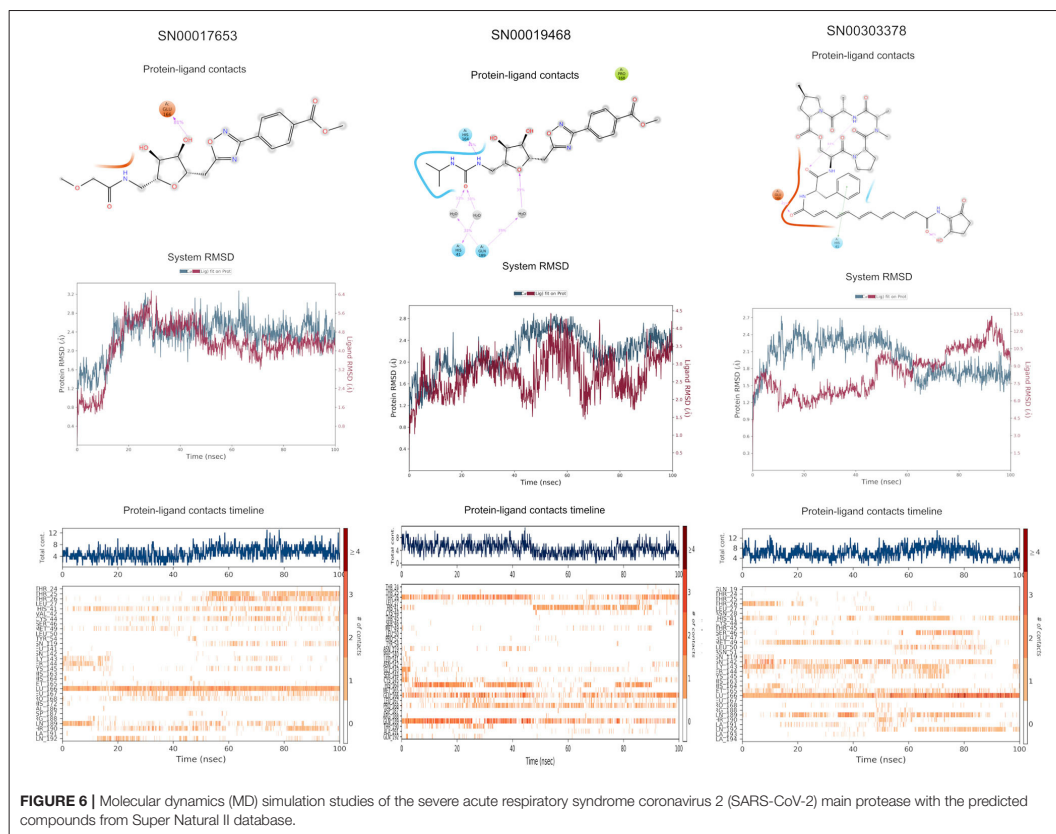
Both saquinavir and saralasin showed a good performance during the MD analysis, and the MM-GBSA values also indicate a strong affinity of these two ligands within the binding site.

Eledoisin showed an important inhibitory potential during the MD analysis, and this is correlated to the high MM-GBSA affinity value, which shows the best score for all studied ligands (-93.3 kcal/mol).

Angiotensin II also showed persistent contacts with the main key residues of the binding site during the 100-ns simulation, and this is also extrapolated to the MM-GBSA values, which show a high-affinity energy for this ligand (-74.9 kcal/mol).

Naldemedine, which interacts within the key residues of the binding site, but has less persistent contacts than eledoisin and angiotensin II, also shows a lower MM-GBSA affinity (-64.2 kcal/mol) but comparable to the N3 ligand (-64.3 kcal/mol).

Excluding notoamide R, all these ligands showed good MM-GBSA affinity values in comparison to the previously identified N3 inhibitor. Nevertheless, the MD analysis showed poor interaction of SN00019468, SN000303378, and aliskiren within the binding site, so their corresponding MM-GBSA affinities are not considered relevant. Dianthin E, pseudostellarin C, saquinavir, saralasin, eledoisin, angiotensin II, and naldemedine



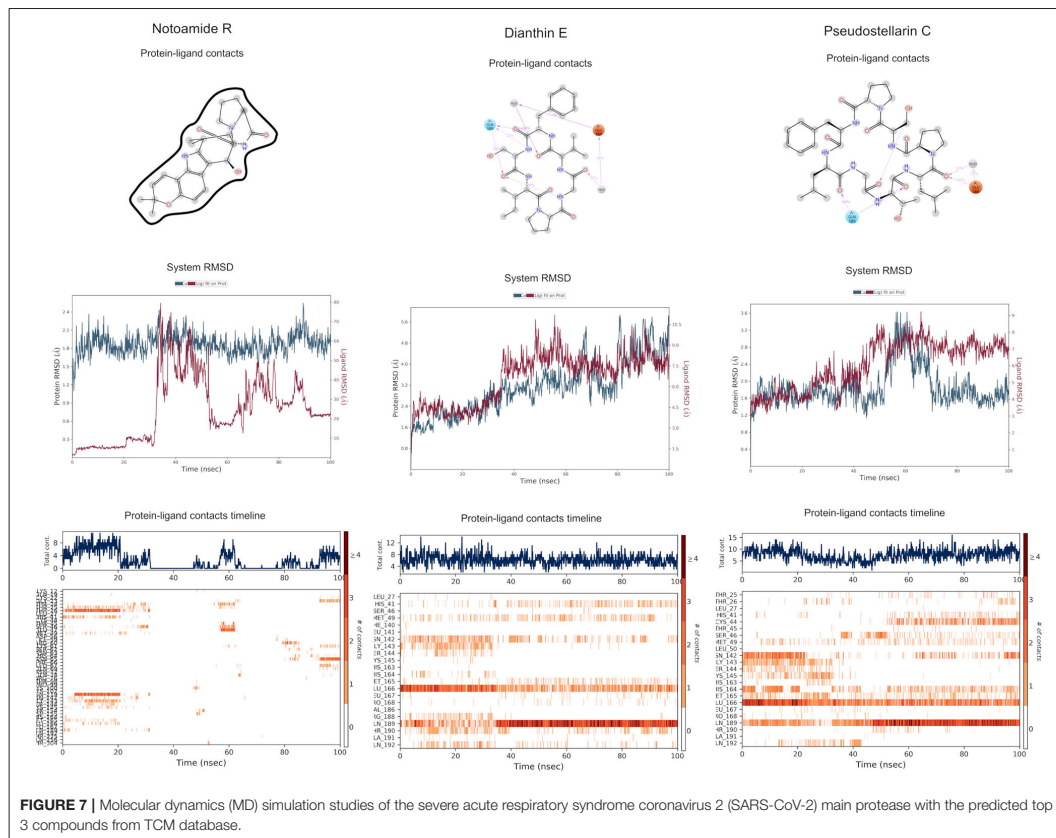
are considered as potential 3CL^{Pro} inhibitors, though eldoisoin and angiotensin II are the most promising ones due to their lower MM-GBSA values (Table 4). The obtained results (Table 4) indicate the mean of the energy calculated for each pose of the simulation and its corresponding standard deviation.

CLINICAL INSIGHTS

Our ultimate goal is to identify putative drug compounds that can be used safely and efficaciously while mitigating risks—deleterious side effects likely could not be reliably or robustly tolerated in severe COVID-19 cases. The general strategy of “drug repurposing” involves identifying existing compounds (both approved and withdrawn drugs) via their biological plausibility/rationale (e.g., mechanism-based inhibitors), via *in vitro*, *in vivo*, and *in silico* studies, or via serendipitous clinical observations. Much clinical pharmacological data, and clinical trial knowledge, are required in order to really elucidate (and extend) the use of a given chemical for a new indication; such efforts can stem from clinical expertise or

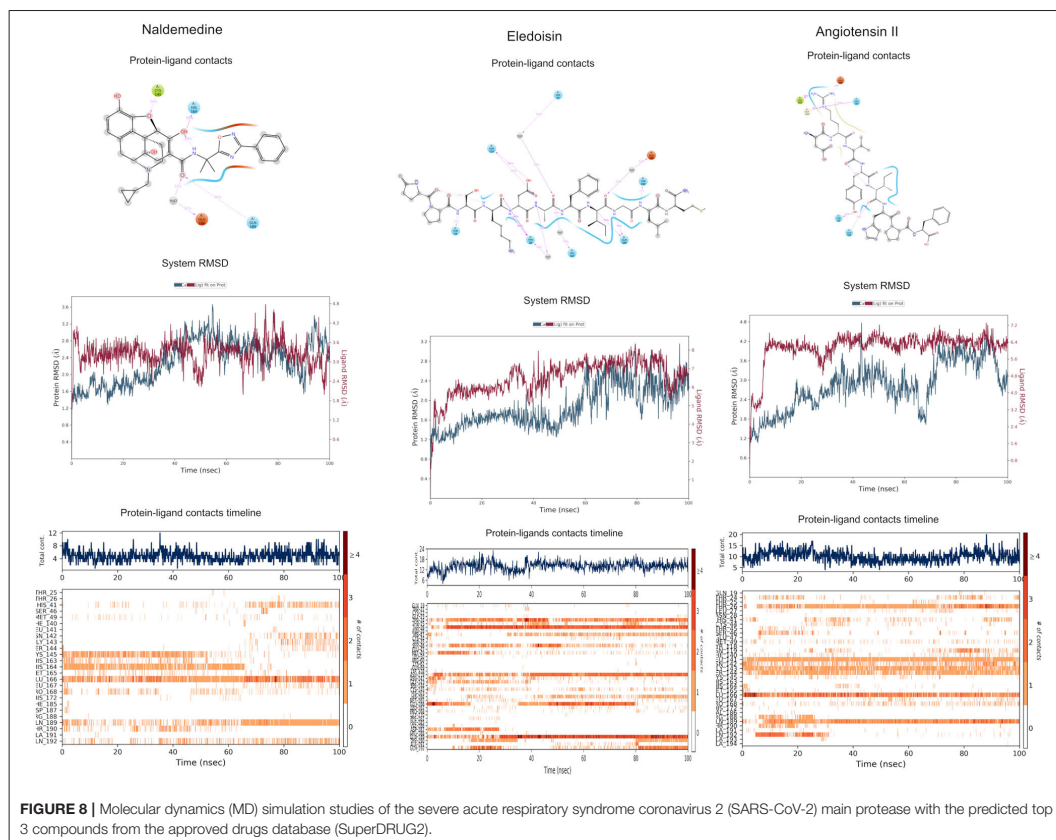
smaller-scale studies (before a fully systematic, population-wide study). A consideration of the possible strengths and weaknesses of the drug candidates predicted herein will require expertise in translational drug development, including clinical pharmacologists and infectious disease specialists conducting clinical trials related to COVID-19. As an alternative to traditional drug development strategies, which are often slow, financially costly, and failure-prone, drug repositioning approaches, though computationally intricate, can be especially useful in emergency situations such as the COVID-19 pandemic. Identifying and selecting molecular candidates for drug repositioning entails numerous factors—e.g., pharmacokinetics, clinical indications, drug-related adverse events, drug–drug interactions, toxicity profiles, and available formulations.

With regard to pharmacokinetics, metabolism plays a key role in this selection. The Phase I metabolism, through the CYP450 family, significantly increases the risk of drug–drug interactions when coadministered with CYP inhibitors or inducers. Moreover, idiosyncratic genetic variability in the CYP gene family may affect an individual’s response to the



administered treatment, both in terms of effectiveness and tolerability. By these considerations, we suggest that eldoisin, daptomycin, and angiotensin II, which have no interaction with the CYP system (**Supplementary Table 4**), could be potential inhibitors of the SARS-CoV-2 main protease. However, the vasodilatory activity of eldoisin may render it unsuitable in critically ill patients because severe cases (e.g., septic shock) often require vasopressor support; in contrast, angiotensinamide, which is an oligopeptide used to increase blood pressure by vasoconstriction, could be an interesting option in septic patients. Drugs not specifically indicated for cardiovascular disease may also affect hemodynamics. For instance, dihydroergocornine, a dopamine agonist used as an anti-Parkinson agent, presents large hypotensive effects; therefore, it should not be considered as a potential medication for COVID-19-infected patients. A consensus seems to be emerging that cardiovascular drugs (or other medications with significant hemodynamic effects) should not be considered promising candidates for targeting SARS-CoV-2, as their overall efficacy (and potential side effects) is too coupled to other clinical aspects of a patient's condition.

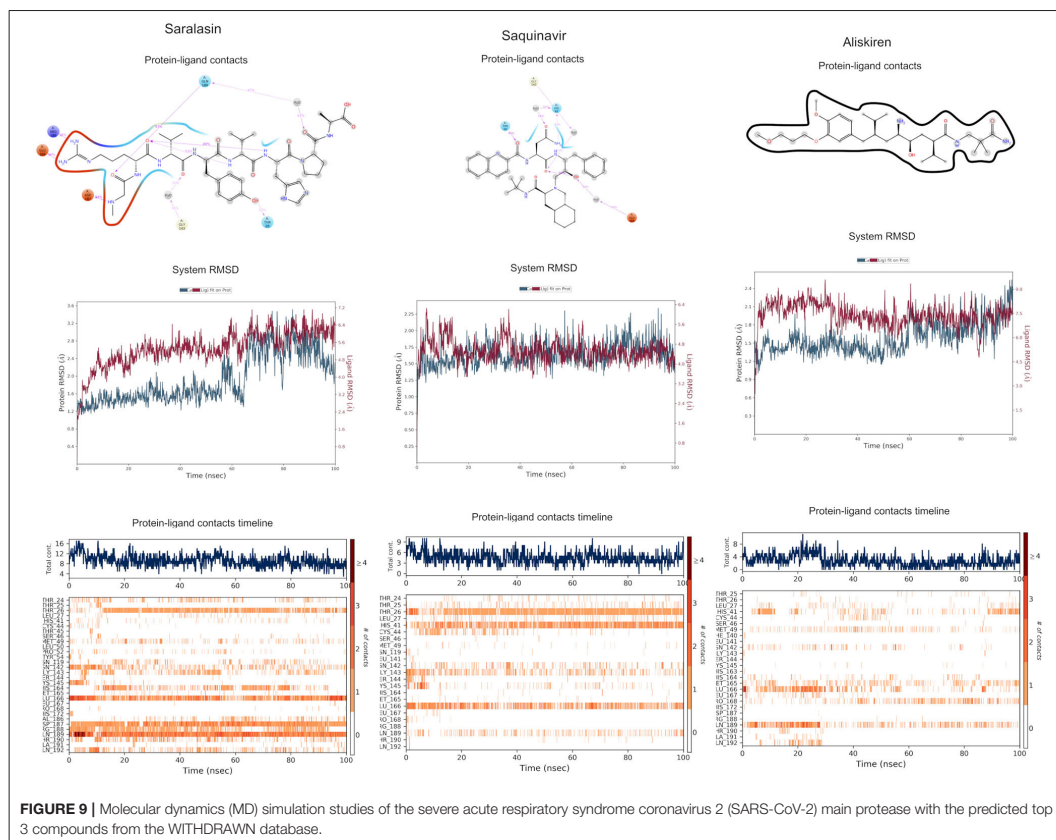
There are also additional considerations in the drug selection process, such as the environments of use and how a patient will interact with the drug, and these factors may influence the chosen administration route. Most infected COVID-19 patients have been managed in non-critical areas; therefore, in this context, the oral route of administration can be considered feasible, alongside with intravenous injections. Naldemedine, which is an oral peripherally acting μ -opioid receptor antagonist (PAMORA), indicated for opioid-induced constipation (Coluzzi et al., 2020), may represent a practicable alternative in less severe COVID-19 patients. Its use, in healthy subjects, was associated with a slight increase in the incidence of diarrhea (Fukumura et al., 2018). Conversely, in critically ill patients, even when the enteral nutrition is guaranteed, oral formulations could be unsuitable if they cannot be crushed or dissolved and administered through the enteral feeding tube, as is the case with [CM1] naldemedine. From a clinical point of view, the approved drug indication could be a relevant criterion for selection, analogous to what holds true for antimicrobials and antiviral drugs. Indeed, the COVID-19 pandemic has seen



remdesivir, an established drug with broad-spectrum antiviral activity, receive emergency use authorization from the FDA and the European Medicines Agency (EMA) (Grein et al., 2020). Similarly, the antiviral medication telaprevir, a hepatitis C virus protease inhibitor, could represent an alternative. Among antibacterial agents, daptomycin, which is a lipopeptide antibiotic with *in vitro* bactericidal activity against Gram-positive bacteria, could be interesting potential inhibitors for SARS-CoV-2 targets. Daptomycin remains one of the main treatment options for methicillin-resistant *Staphylococcus aureus* (MRSA) infections; however, sporadic cases of resistance have been noted (Barros et al., 2019). In drug repositioning approaches, even withdrawn drugs can become reborn—particularly if the reason for market withdrawal was commercial (i.e., not safety issues). The top three withdrawn compounds that we have identified here include two cardiovascular drugs and an antiviral. Saralasin, an old partial agonist of angiotensin II receptors, has been withdrawn from sale for commercial reasons. Similarly, aliskiren, a direct renin inhibitor—which failed to show

benefit over angiotensin-converting-enzyme (ACE) inhibitors in heart failure²¹—was withdrawn from the European market, without intention to market it in the future. Saquinavir, indicated for treatment of HIV-1-infected adult patients, also could be an appealing potential drug for COVID-19 patients. However, apart from its withdrawal from the market, the CYP activity profile causes significant potential clinically relevant drug–drug interactions with a number of coadministered drugs. When considering the top 10 withdrawn compounds (**Supplementary Table 5**)—excluding cardiovascular drugs (for concerns expressed above) and antineoplastic agents (for reasons of toxicity)—we suggest that antibacterial drugs could be considered. Azlocillin, a wide-spectrum acylated form of ampicillin with antibacterial activity, has been recently proposed as a potential drug candidate for Lyme disease (targeting drug-tolerant *Borrelia burgdorferi*). Besides its efficacy, the safety of

²¹End of the road for Aliskiren in heart failure. Available online at: <https://academic.oup.com/eurheartj/article/38/5/312/2990021> (accessed July 28, 2020).



azlocillin was one of the main criteria for selecting this drug (Pothineni et al., 2020). In terms of drug repositioning, note that azlocillin has also been investigated as a potential new therapeutic agent for prostate cancer (Turanli et al., 2019). Among antibacterial agents, azlocillin does not present toxicity issues and has no interactions with the CYP family; therefore, it could be considered a first-choice molecule in this pharmacologic class of agents.

CONCLUSIONS

This study reports potential inhibitors for the SARS-CoV-2 main protease, 3CL^{Pro}, via an integrated computational approach to drug repositioning. After our docking trials, a divergent pose of ligands was generated, and the pose with the optimal docking score and binding interactions was considered as the best pose for further processing and manual analysis. The docking of compounds to the 3CL^{Pro} protease was visualized in terms of interactions in the substrate recognition pockets

of the protein, and the dynamical stability of drug–protein contacts was evaluated *via* MD simulations of each putative drug–3CL^{Pro} pair. We identified compounds from four different sources—namely, the Super Natural II, TCM, approved drugs, and WITHDRAWN drugs databases. Most of the compounds identified in our present work exhibit favorable interactions with the main protease residues (Cys145, Ser144, Glu166, His41, Gln189, and Gln192), suggesting that enthalpically optimal interactions at least *can* occur (see Results section). Our proposed compounds, as bound in the active site (pocket) of the 3CL^{Pro} protein, are shown in Figures 2–5. The steric accommodation of selected compounds in 3CL^{Pro} hinges upon particular amino acid residues that engage in interactions, as shown in Figures 6–9. Our analyses elucidate, at least *in silico*, these potential drug–3CL^{Pro} interactions. Some drugs, like naldemedine and candidates from Super Natural II (SN00017653) and TCM (pseudostellarin C), have not been identified in previous studies as potential inhibitors of SARS-CoV-2 main protease and are suggested here for the first time. Interestingly, pseudostellarin C, which is a compound found in the roots of a traditional Chinese

TABLE 4 | Molecular mechanics–generalized Born surface area (MM-GBSA) calculation.

Drug	dG (kcal/mol)	SD (kcal/mol)
SN00019468	−54.0	13.5
SN00017653	−54.5	5.8
SN00303378	−54.3	6.4
Pseudostellarin C	−53.6	12.1
Notoamide R	−21.4	17.5
Dianthin E	−61.70	10.6
Angiotensin II	−74.9	6.8
Eledoisin	−93.3	8.9
Naldemedine	−64.2	6.7
Saquinavir	−54.3	6.4
Aliskiren	−61.2	9.9
Saralasin	−61.8	10.6
N3 inhibitor	−64.3	12.3

dG, free energy; SD, standard deviation. The obtained results indicate the mean of the energy calculated for each pose of the simulation and its corresponding standard deviation.

plant (*Pseudostellaria heterophylla*), is known medicinally for its application in dry cough arising from “lung dryness” (Hu et al., 2019).

Naturally occurring compounds are a rich resource for drug innovation and development. We suggest that the COVID-19-related leads reported here can support the discovery and development of high-potency inhibitors *in vitro* and *in vivo*. New leads from the Super Natural II databases (Figure 2, Table 2) are novel promising candidates, as they have similar binding interaction profiles [including overall good structural similarity (0.73 and above)] as compared to the main N3 inhibitors. Besides that, the compounds from the TCM database have also shown good interactions with the main protease (Figures 3, 7, Table 2). Specially, compound pseudostellarin C is a promising candidate from the TCM chemical space, sharing similar binding interactions as the N3 ligand and showing a high stable pose from the first 40 ns (Figure 7) on MD simulation studies.

Additionally, the repurposed drugs (as shown in Table 3, Figures 4, 5) have shown good interactions; in particular, naldemedine from the approved drug set, which is a PAMORA recently approved for the treatment of opioid-induced constipation in adult patients (Hu and Bridgeman, 2018). This drug is also supported as a clinically valid alternative due to its safe profile. Furthermore, based on the MD simulation studies, naldemedine shows close values for protein and ligand RMSD, so the ligand pose is considered stable. Another interesting candidate is saquinavir from the withdrawn dataset, which interacts with the main protease in a similar manner as compared to its original ligand (Figure 6). It has a structural similarity of 0.67 with the original N3 ligand. Saquinavir is an antiretroviral protease (peptidomimetic) inhibitor that is used in the therapy and prevention of human immunodeficiency virus (HIV) infection and the acquired immunodeficiency

syndrome (AIDS). This drug is discontinued in Europe¹⁰, and due to its CYP activity profile, which causes significant potential clinically relevant drug–drug interactions with a number of coadministered drugs, this drug becomes clinically less preferable as a COVID-19 potential drug candidate. Saquinavir is also reported as a potential repurposed drug for COVID-19 disease by other studies (Montenegro et al., 2020). On the other hand, drugs such as angiotensin II, aliskiren, and phytochemicals like notoamide R, SN00019468, dianthin E, and SN00303378 cannot be considered as optimal candidates based on this study, as some of these compounds were stable but showed poor contacts with crucial residues of the main protease, as well as some showed higher RMSD values on MD simulation studies, and sometimes both poor contacts and high ligand RMSD fluctuations were observed (see Results section). Additionally, in this study, we have also addressed the toxicity and cytochrome activity of the reported compounds. Most of the resulting compounds predicted to be immunotoxic (that is, cytotoxicity of the B and T cells). Thus, we believe gaining insight into the molecular mechanism responsible for protein–ligand recognition through this study will facilitate the development of drugs for the treatment of COVID-19 disease.

The work reported here addresses an important concern and urgent need for drugs for the treatment of SARS-CoV-2 infection. As demonstrated via this integrated approach, computational prediction of approved drugs, withdrawn drugs, and phytochemicals for inhibition of SARS-CoV-2 main protease has resulted in some promising leads for further experimental validation. We hope that the *in silico* results and predictions obtained in this study, including the potential clinical insights, could facilitate the discovery of highly potent inhibitors of the SARS-CoV-2 main protease. Overall, our computational drug repositioning strategy predicts some promising drug candidates that, if borne out *via* experimental and clinical approaches, could contribute toward resolving the global crisis of the COVID-19 pandemic.

DATA AVAILABILITY STATEMENT

The raw data supporting the conclusions of this article will be made available by the authors, without undue reservation.

AUTHOR CONTRIBUTIONS

RP, MS, and PB conceived the project. RA, MP, HP-S, QC, FC, MR, RP, and PB designed the project workflow. RA, QC, MP, and PB prepared the databases. QC performed SuperTCM data curations. RA, MP, HP-S, and PB implemented the project. RA performed molecular docking. RA and PB analyzed and selected the compounds. MP and HP-S performed the molecular dynamics. FC, MP, PC, and MS contributed to the clinical insights and clinical trial study. RA, MP, and PB provided the figures. RA and QC provided the **Supplementary Material**. PB, RA, MP,

FC, CM, and PB contributed to manuscript writing. All authors proofread the final manuscript.

FUNDING

This work has been partially funded by grants from the Spanish Ministry of Economy and Competitiveness (CTQ2017-87974-R) and by the Fundación Séneca del Centro de Coordinación de la Investigación de la Región de Murcia under Project 20988/PI/18. Supercomputing resources in this work have been supported by the infrastructures of Poznan Supercomputing Center, the e-infrastructure program of the Research Council of Norway, and the supercomputer center of UiT-the Arctic University of Norway, by the Plataforma Andaluza de Bioinformática of the University of Málaga, by the supercomputing infrastructure of the NLHPC (ECM-02, Powered@NLHPC), and by the Extremadura Research Centre for Advanced Technologies (CETA-CIEMAT), funded by the European Regional Development Fund (ERDF). CETA-CIEMAT belongs to CIEMAT and the Government of Spain. Portions of this work were also supported by the University of Virginia School of Data Science and by NSF Career award MCB-1350957. We acknowledge support from the German Research Foundation

(DFG) (KFO 339 and TRR 295) and the Open Access Publication Fund of Charité-Universitätsmedizin Berlin.

ACKNOWLEDGMENTS

Currently, clinical study is aimed and approved at evaluating the activity of drug Naldemedine in asymptomatic subjects positive for SARS-COV-2 in terms of reducing the number of subjects presenting symptoms or reducing the number of positive days of the test and a second study in symptomatic patients to assess the reduction in the number and extent of symptoms and the time to recovery. Both studies will be phase II studies, with a control group consisting of patients treated according to the clinical practice of the various participating centers in the project. The clinical study is approved by the Shionogi pharmaceutical study, and will be conducted at Sant'Andrea University Hospital, Rome, Italy.

SUPPLEMENTARY MATERIAL

The Supplementary Material for this article can be found online at: <https://www.frontiersin.org/articles/10.3389/fchem.2020.590263/full#supplementary-material>

REFERENCES

- Al-Tawfiq, J. A., Al-Homoud, A. H., and Memish, Z. A. (2020). Remdesivir as a possible therapeutic option for the COVID-19. *Travel Med. Infect. Dis.* 34, 101615. doi: 10.1016/j.tmaid.2020.101615
- Anand, K., Ziebuhr, J., Wadhvani, P., Mesters, J. R., and Hilgenfeld, R. (2003). Coronavirus main proteinase (3CLpro) structure: basis for design of anti-SARS drugs. *Science* 300, 1763–1767. doi: 10.1126/science.1085658
- Barros, E. M., Martin, M. J., Selleck, E. M., Lebreton, F., Sampaio, J. L. M., and Gilmore, M. S. (2019). Daptomycin resistance and tolerance due to loss of function in *Staphylococcus aureus* dsp1 and asp23. *Antimicrob. Agents Chemother.* 63:e01542-18. doi: 10.1128/AAC.01542-18
- Beigel, J. H., Tomashek, K. M., Dodd, L. E., Mehta, A. K., Zingman, B. S., Kalil, A. C., et al. (2020). Remdesivir for the treatment of Covid-19 — preliminary report. *N. Engl. J. Med.* 383, 1813–1826. doi: 10.1056/nejmoa2007764
- Berthold, M. R., Cebron, N., Dill, F., Gabriel, T. R., Kötter, T., Meinl, T., et al. (2008). *KNIME: The Konstanz Information Miner*. Berlin: Springer. doi: 10.1145/1656274.1656280
- Botta, L., Rivara, M., Zuliani, V., and Radi, M. (2018). Drug repurposing approaches to fight Dengue virus infection and related diseases. *Front. Biosci.* 23, 997–1019. doi: 10.2741/4630
- Cao, B., Wang, Y., Wen, D., Liu, W., Wang, J., Fan, G., et al. (2020). A trial of lopinavir-ritonavir in adults hospitalized with severe covid-19. *N. Engl. J. Med.* 382, 1787–1799. doi: 10.1056/NEJMoa2001282
- Cha, Y., Erez, T., Reynolds, I. J., Kumar, D., Ross, J., Koytiger, G., et al. (2018). Drug repurposing from the perspective of pharmaceutical companies. *Brit. J. Pharmacol.* 175, 168–180. doi: 10.1111/bph.13798
- Chen, C. Y. C. (2011). TCM Database@Taiwan: The world's largest traditional Chinese medicine database for drug screening *in silico*. *PLoS ONE* 6:e15939. doi: 10.1371/journal.pone.0015939
- Coluzzi, F., Scerpa, M. S., and Pergolizzi, J. (2020). Naldemedine: a new option for OIBD. *J. Pain Res.* 13, 1209–1222. doi: 10.2147/JPR.S243435
- Contini, A. (2020). Virtual screening of an FDA approved drugs database on two COVID-19 coronavirus proteins. *ChemRxiv*. doi: 10.26434/CHEMRXIV.11847381.V1
- Durant, J. L., Leland, B. A., Henry, D. R., and Nourse, J. G. (2002). Reoptimization of MDL keys for use in drug discovery. *J. Chem. Inform. Comput. Sci.* 42, 1273–1280. doi: 10.1021/ci010132r
- Fukumura, K., Yokota, T., Baba, Y., and Arjona Ferreira, J. C. (2018). Phase I, randomized, double-blind, placebo-controlled studies on the safety, tolerability, and pharmacokinetics of naldemedine in healthy volunteers. *Clin. Pharmacol. Drug Dev.* 7, 474–483. doi: 10.1002/cpdd.387
- Greenidge, P. A., Kramer, C., Mozziconacci, J.-C., and Wolf, R. M. (2013). MM/GBSA binding energy prediction on the pdbbind data set: successes, failures, and directions for further improvement. *J. Chem. Inform. Model.* 53, 201–209. doi: 10.1021/ci300425v
- Grein, J., Ohmagari, N., Shin, D., Diaz, G., Asperges, E., Castagna, A., et al. (2020). Compassionate use of remdesivir for patients with severe covid-19. *N. Engl. J. Med.* 382, 2327–2336. doi: 10.1056/NEJMoa2007016
- Hall, L. H., and Kier, L. B. (1995). Electrotopological state indices for atom types: a novel combination of electronic, topological, and valence state information. *J. Chem. Inform. Comput. Sci.* 35, 1039–1045. doi: 10.1021/ci00028a014
- Horby, P., Lim, W. S., Emberson, J., Mafham, M., Bell, J., Linsell, L., et al. (2020). Dexamethasone for COVID-19—preliminary report effect of dexamethasone in hospitalized patients with COVID-19 - preliminary report. *medRxiv*. doi: 10.1101/2020.06.22.20137273
- Hu, D. J., Shakerian, F., Zhao, J., and Li, S. P. (2019). Chemistry, pharmacology and analysis of *Pseudostellaria heterophylla*: a mini-review. *Chin. Med.* 14, 1–8. doi: 10.1186/s13020-019-0243-z
- Hu, K., and Bridgeman, M. B. (2018). Naldemedine (Symproic) for the treatment of opioid-induced constipation. *P T* 43, 601–605.
- Ihlenfeldt, W. D., Bolton, E. E., and Bryant, S. H. (2009). The PubChem chemical structure sketcher. *J. Cheminform.* 1:20. doi: 10.1186/1758-2946-1-20
- Jackson, L. A., Anderson, E. J., Roupael, N. G., Roberts, P. C., Makhene, M., Coler, R. N., et al. (2020). An mRNA vaccine against SARS-CoV-2 — preliminary report. *N. Engl. J. Med.* 383, 1920–1931. doi: 10.1056/NEJMoa2022483
- Jin, Z., Du, X., Xu, Y., Deng, Y., Liu, M., Zhao, Y., et al. (2020). Structure of Mpro from SARS-CoV-2 and discovery of its inhibitors. *Nature* 582, 289–293. doi: 10.1038/s41586-020-2223-y

- Jones, G., Willett, P., Glen, R. C., Leach, A. R., and Taylor, R. (1997). Development and validation of a genetic algorithm for flexible docking. *J. Mol. Biol.* 267, 727–748. doi: 10.1006/jmbi.1996.0897
- Khedkar, P. H., and Patzak, A. (2020). SARS-CoV-2: what do we know so far? *Acta Physiol.* 229:e13470. doi: 10.1111/apha.13470
- Kollman, P. A., Massova, I., Reyes, C., Kuhn, B., Huo, S., Chong, L., et al. (2000). Calculating structures and free energies of complex molecules: combining molecular mechanics and continuum models. *Acc. Chem. Res.* 33, 889–897. doi: 10.1021/ar000033j
- Kupferschmidt, K., and Cohen, J. (2020). Race to find COVID-19 treatments accelerates. *Science* 367, 1412–1413. doi: 10.1126/science.367.6485.1412
- Mark, P., and Nilsson, L. (2001). Structure and dynamics of the TIP3P, SPC, and SPC/E water models at 298 K. *J. Phys. Chem. A* 105, 9954–9960. doi: 10.1021/jp003020w
- McNamee, L. M., Walsh, M. J., and Ledley, F. D. (2017). Timelines of translational science: from technology initiation to FDA approval. *PLoS ONE* 12. doi: 10.1371/journal.pone.0177371
- Metushi, I. G., Wriston, A., Banerjee, P., Gohlke, B. O., English, A. M., Lucas, A., et al. (2015). Acyclovir has low but detectable influence on HLA-B*57:01 specificity without inducing hypersensitivity. *PLoS ONE* 10:e0124878. doi: 10.1371/journal.pone.0124878
- Montenegro, M. F., Al-Abed, Y., He, M., Tracey, K. J., and Billiar, T. R. (2020). HIV protease inhibitors saquinavir and nelfinavir are potent inhibitors of cathepsin L activity: a potential treatment for COVID-19 patients. doi: 10.21203/rs.3.rs-37258/v1
- Muramatsu, T., Takemoto, C., Kim, Y. T., Wang, H., Nishii, W., Terada, T., et al. (2016). SARS-CoV 3CL protease cleaves its C-terminal autoproteolytic site by novel subsite cooperativity. *Proc. Natl. Acad. Sci. U.S.A.* 113, 12997–13002. doi: 10.1073/pnas.1601327113
- Newman, D. J., and Cragg, G. M. (2016). Natural products as sources of new drugs from 1981 to 2014. *J. Nat. Prod.* 79, 629–661. doi: 10.1021/acs.jnatprod.5b01055
- Nutho, B., Mahalapbutr, P., Hengphasatporn, K., Pattarangoon, N. C., Simanon, N., Shigeta, Y., et al. (2020). Why are lopinavir and ritonavir effective against the newly emerged coronavirus 2019? Atomistic insights into the inhibitory mechanisms. *Biochemistry* 59, 1769–1779. doi: 10.1021/acs.biochem.0c00160
- Patterson, D. E., Cramer, R. D., Ferguson, A. M., Clark, R. D., and Weinberger, L. E. (1996). Neighborhood behavior: a useful concept for validation of “molecular diversity” descriptors. *J. Med. Chem.* 39, 3049–3059. doi: 10.1021/jm960290n
- Pizzorno, A., Padey, B., Terrier, O., and Rosa-Calatrava, M. (2019). Drug repurposing approaches for the treatment of influenza viral infection: reviving old drugs to fight against a long-lived enemy. *Front. Immunol.* 10:531. doi: 10.3389/fimmu.2019.00531
- Pothineni, V. R., Potula, H. H. S. K., Ambati, A., Mallajosyula, V. V. A., Sridharan, B., Inayathullah, M., et al. (2020). Azlocillin can be the potential drug candidate against drug-tolerant *Borrelia burgdorferi* sensu stricto JLB31. *Sci. Rep.* 10, 1–15. doi: 10.1038/s41598-020-59600-4
- Rogers, D., and Hahn, M. (2010). Extended-connectivity fingerprints. *J. Chem. Inform. Model.* 50, 742–754. doi: 10.1021/ci100050t
- Scasso, F., Ferrari, G., de Vincentiis, G. C., Arosio, A., Bottero, S., Carretti, M., et al. (2018). Emerging and re-emerging infectious disease in otorhinolaryngology. *Acta Otorhinolaryngol. Ital.* 38, S1–S106. doi: 10.14639/0392-100X-suppl.1-38-2018
- Shiryaev, S. A., Mesci, P., Pinto, A., Fernandes, I., Sheets, N., Shresta, S., et al. (2017). Repurposing of the anti-malaria drug chloroquine for Zika Virus treatment and prophylaxis. *Sci. Rep.* 7:15771. doi: 10.1038/s41598-017-15467-6
- Stumpfe, D., and Bajorath, J. (2011). Similarity searching. *WIREs Comput. Mol. Sci.* 1, 260–282. doi: 10.1002/wcms.23
- Ton, A.-T., Gentile, F., Hsing, M., Ban, F., and Cherkasov, A. (2020). Rapid identification of potential inhibitors of SARS-CoV-2 main protease by deep docking of 1.3 billion compounds. *Mol. Inform.* 39:2000028. doi: 10.1002/minf.202000028
- Turanli, B., Zhang, C., Kim, W., Benfeitas, R., Uhlen, M., Arga, K. Y., et al. (2019). Discovery of therapeutic agents for prostate cancer using genome-scale metabolic modeling and drug repositioning. *EBioMedicine* 42, 386–396. doi: 10.1016/j.ebiom.2019.03.009
- Wang, J. (2020). Fast identification of possible drug treatment of Coronavirus Disease-19 (COVID-19) through computational drug repurposing study. *J. Chem. Inform. Model.* 60, 3277–3286. doi: 10.1021/acs.jcim.0c00179
- Wu, F., Zhao, S., Yu, B., Chen, Y. M., Wang, W., Song, Z. G., et al. (2020). A new coronavirus associated with human respiratory disease in China. *Nature* 579, 265–269. doi: 10.1038/s41586-020-2008-3
- Yang, P., and Wang, X. (2020). COVID-19: a new challenge for human beings. *Cell. Mol. Immunol.* 17, 555–557. doi: 10.1038/s41423-020-0407-x
- Zhang D.-h, Wu K.-l, Zhang X, Deng S.-q, and Peng B. (2020). *In silico* screening of Chinese herbal medicines with the potential to directly inhibit 2019 novel coronavirus. *J. Integr. Med.* 18, 152–158. doi: 10.1016/j.joim.2020.02.005
- Zhang, L., Lin, D., Sun, X., Curth, U., Drosten, C., Sauerhering, L., et al. (2020). Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved a-ketoamide inhibitors. *Science* 368, 409–412. doi: 10.1126/science.abb3405

Conflict of Interest: The authors declare that the research was conducted in the absence of any commercial or financial relationships that could be construed as a potential conflict of interest.

The reviewer KV declared a past co-authorship with one of the authors HP-S to the handling Editor.

Copyright © 2020 Abel, Paredes Ramos, Chen, Pérez-Sánchez, Coluzzi, Rocco, Marchetti, Mura, Simmaco, Bourne, Preissner and Banerjee. This is an open-access article distributed under the terms of the Creative Commons Attribution License (CC BY). The use, distribution or reproduction in other forums is permitted, provided the original author(s) and the copyright owner(s) are credited and that the original publication in this journal is cited, in accordance with accepted academic practice. No use, distribution or reproduction is permitted which does not comply with these terms.



Article

Reynoutria Rhizomes as a Natural Source of SARS-CoV-2 Mpro Inhibitors—Molecular Docking and In Vitro Study

Izabela Nawrot-Hadzik ^{1,*}, Mikolaj Zmudziński ², Adam Matkowski ¹, Robert Preissner ³,
Małgorzata Kęsik-Brodacka ^{4,5}, Jakub Hadzik ⁶, Marcin Drag ² and Renata Abel ^{1,3}

¹ Department of Pharmaceutical Biology and Botany, Wrocław Medical University, 50-556 Wrocław, Poland; bbsekret@umed.wroc.pl (A.M.); renata.abel@charite.de (R.A.)

² Department of Chemical Biology and Bioimaging, Wrocław University of Science and Technology, 50-370 Wrocław, Poland; mikolaj.zmudzinski@pwr.edu.pl (M.Z.); marcin.drag@pwr.edu.pl (M.D.)

³ Structural Bioinformatics Group, Institute for Physiology, Charité—University Medicine Berlin, 10115 Berlin, Germany; robert.preissner@charite.de

⁴ Research Network Łukasiewicz—Institute of Biotechnology and Antibiotics, Starościńska 5, 02-516 Warsaw, Poland; kesikm@iba.waw.pl

⁵ National Medicines Institute, ul. Chełmska 30/34, 00-725 Warszawa, Poland

⁶ Department of Dental Surgery, Wrocław Medical University, 50-425 Wrocław, Poland; jakub.hadzik@umed.wroc.pl

* Correspondence: izabela.nawrot-hadzik@umed.wroc.pl



Citation: Nawrot-Hadzik, I.; Zmudziński, M.; Matkowski, A.; Preissner, R.; Kęsik-Brodacka, M.; Hadzik, J.; Drag, M.; Abel, R. *Reynoutria* Rhizomes as a Natural Source of SARS-CoV-2 Mpro Inhibitors—Molecular Docking and In Vitro Study. *Pharmaceuticals* **2021**, *14*, 742. <https://doi.org/10.3390/ph14080742>

Academic Editor: Osvaldo Andrade Santos-Filho

Received: 12 July 2021

Accepted: 26 July 2021

Published: 29 July 2021

Publisher's Note: MDPI stays neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Copyright: © 2021 by the authors. Licensee MDPI, Basel, Switzerland. This article is an open access article distributed under the terms and conditions of the Creative Commons Attribution (CC BY) license (<https://creativecommons.org/licenses/by/4.0/>).

Abstract: More than a year has passed since the world began to fight the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) responsible for the Coronavirus disease 2019 (COVID-19) pandemic, and still it spreads around the world, mutating at the same time. One of the sources of compounds with potential antiviral activity is Traditional Chinese Medicinal (TCM) plants used in China in the supportive treatment of COVID-19. *Reynoutria japonica* is important part of the Shu Feng Jie Du Granule/Capsule-TCM herbal formula, recommended by China Food and Drug Administration (CFDA) for treatment of patients with H1N1- and H5N9-induced acute lung injury and is also used in China to treat COVID-19, mainly combined with other remedies. In our study, 25 compounds from rhizomes of *R. japonica* and *Reynoutria sachalinensis* (related species), were docked into the binding site of SARS-CoV-2 main protease. Next, 11 of them (vanicoside A, vanicoside B, resveratrol, piceid, emodin, epicatechin, epicatechin gallate, epigallocatechin gallate, procyanidin B2, procyanidin C1, procyanidin B2 3,3'-di-O-gallate) as well as extracts and fractions from rhizomes of *R. japonica* and *R. sachalinensis* were tested in vitro using a fluorescent peptide substrate. Among the tested phytochemicals the best results were achieved for vanicoside A and vanicoside B with moderate inhibition of SARS-CoV-2 Mpro, $IC_{50} = 23.10 \mu\text{M}$ and $43.59 \mu\text{M}$, respectively. The butanol fractions of plants showed the strongest inhibition of SARS-CoV-2 Mpro ($IC_{50} = 4.031 \mu\text{g/mL}$ for *R. sachalinensis* and $IC_{50} = 7.877 \mu\text{g/mL}$ for *R. japonica*). As the main constituents of butanol fractions, besides the phenylpropanoid disaccharide esters (e.g., vanicosides), are highly polymerized procyanidins, we suppose that they could be responsible for their strong inhibitory properties. As inhibition of SARS-CoV-2 main protease could prevent the replication of the virus our research provides data that may explain the beneficial effects of *R. japonica* on COVID-19 and identify the most active compounds worthy of more extensive research.

Keywords: *Polygoni cuspidati rhizoma*; *Reynoutria sachalinensis*; vanicoside; proanthocyanidins; COVID-19 (Coronavirus disease 2019)

1. Introduction

The severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2), responsible for Coronavirus disease 2019 (COVID-19) pandemic, was first reported in December 2019 in Wuhan, China [1,2] Since then, over 180 million confirmed cases and more than 3.8 million

fatalities caused by COVID-19 disease were reported worldwide (Johns Hopkins University database: <https://coronavirus.jhu.edu/data>, accessed on 27 July 2021). SARS-CoV-2 belongs to the *Coronaviridae* family, the same family as previously known coronaviruses such as SARS-CoV, responsible for Severe Acute Respiratory Syndrome and MERS-CoV, which caused Middle East Respiratory Syndrome [3]. The most common symptoms of the infection are: fever, dry cough, shortness of breath, fatigue, diarrhea, rash on skin, or loss of taste and smell. A few of the symptoms usually occurred in combination, depending on the immune system of the individual [4,5]. Not all the infected people are showing symptoms of the disease, which is why extraordinary measures such as social distancing, wearing masks and frequent testing were imposed by authorities around the world.

Task forces from all over the world are working on designing vaccines and suggesting drugs to fight COVID-19. Vaccine development has given us hope to take control of the pandemic, however because of problems with availability and distribution of vaccines, lack of knowledge for how long the vaccine-induced immunity, as well as the virus's ability to mutate, there is an urgent need to find effective drugs for this disease. Different computational methods and drug repurposing techniques such as molecular docking, molecular dynamics, or AI (Artificial Intelligence) approaches [6–9] were used for drugs against SARS-CoV-2 searching. Many of the suggested drugs are currently in clinical trials [10]. There are also many examples of compounds suggested to be potential inhibitors of main coronavirus targets found in different databases such as the ZINC database (ref ZINC 1bilion) or natural product databases or TCM databases [11–13].

In this study, compounds and extracts from underground parts (rhizomes) of two medicinal plants—*Reynoutria japonica* Houtt. (syn. *Fallopia japonica* (Houtt.) Ronse Decr., *Polygonum cuspidatum* Sieb. & Zucc.) and *Reynoutria sachalinensis* (F.Schmidt) Nakai (Polygonaceae) (syn. *Fallopia sachalinensis* (F.Schmidt) Ronse Decr., *Polygonum sachalinense* F.Schmidt) were investigated for inhibition of one of the nonstructural proteins of the virus which is the main protease (Mpro, also called 3CLpro). Inhibition of this enzyme could prevent the replication of the virus. Besides the main protease, SARS-CoV-2 virus encodes other nonstructural proteins such as: papain-like protease (PLpro), RNA-dependent RNA polymerase (RdRp), a helicase–triphosphatase, an exoribonuclease, an endonuclease, and N7- and 2'-O-methyltransferases and four structural proteins: spike, envelope, membrane, and nucleocapsid [14,15]. Many natural products from different medicinal plants have potential antiviral activity against coronaviruses such as SARS-CoV-2, as well as SARS-CoV or other viruses [16–22].

R. japonica is a well-known herb which has rhizomes (Huzhang in Chinese) that are used in China and Japan to treat various inflammatory diseases, infections, skin diseases, and hyperlipidemia [23]. Huzhang is part of the Shu Feng Jie Du Granule/Capsule-TCM herbal formula mainly used in China for the treatment of acute upper respiratory tract infections such as the flu, swelling, and pain in the throat and others [24]. Based on the Shu Feng Jie Du Capsule antiviral effect, the China Food and Drug Administration (CFDA) has recommended use of it for the treatment of patients with H1N1- and H5N9-induced acute lung injury [25]. In 2020, Shu Feng Jie Du Granule/Capsule was used to treat COVID-19, combined with other remedies or alone [26,27]. Rhizomes of *R. japonica* are rich sources of active phytochemicals such as stilbenes, anthraquinones, flavanols, proanthocyanidins, and phenylpropanoid disaccharide esters. The latter are present in greater quantities in the related *R. sachalinensis*, also included in this study. Stilbenes exhibit diverse biological activities such as antioxidative, antitumoral, anti-inflammatory, and antiviral properties [28]. Moreover, recent molecular docking studies showed that stilbenes in general and resveratrol in particular can be promising anti-COVID-19 drug candidates, acting as an inhibitor of the ACE2 receptor and preventing the S1: ACE2 complex formation and entry of the virus into host cells [28]. Emodin, an anthraquinone, blocked the interaction of SARS-CoV S protein and ACE2 in a dose-dependent manner with an IC₅₀ of 200 μM as well as inhibited the infectivity of S protein-pseudotyped retrovirus to Vero E6 cells [29]. Also, simple flavanols like epicatechin or epigallocatechin gallate were

reported to inhibit angiotensin-converting enzyme activity [30]. Molecular docking and dynamics studies carried out by Maroli et al. [31] showed that procyanidins could be a potential inhibitor of SARS-CoV-2 Mpro as well as ACE2 or spike protein. Phenylpropanoid disaccharide esters present in *Reynoutria* species with a predominant amount of vanicoside B and A [32] are still under-studied chemicals in terms of their biological activity. Their antioxidant and cytotoxic activity against some human tumor cell lines [33,34] as well as their activity as acetylcholinesterase and β -glucosidase inhibitors [35] have been revealed. So far, no studies have been conducted to check their antiviral activity.

For this study we pulled 25 compounds belonging to 5 different classes of phytochemicals from *Reynoutria japonica* and/or *R. sachalinensis*. They were: stilbenes: resveratrol, piceatannol, piceatannol glucoside, piceid, resveratrol; anthraquinones: emodin, emodin 8-glucoside, emodin bianthrone, physcion; phenylpropanoid disaccharide esters: vanicoside A, vanicoside B, vanicoside C, hydropiperoside, lapathoside A, lapathoside C, tatariside B; flavan-3-ols and procyanidins: epicatechin, epicatechin gallate, epigallocatechin gallate, procyanidin B2, procyanidin B2 3'-O-gallate, procyanidin B2 3,3'-di-O-gallate, procyanidin C1, procyanidin C1 3',3''-di-O-gallate, cinnamtannin A2. First, molecular docking was performed on those compounds to evaluate them as potential inhibitors against SARS-CoV-2 Mpro. Consequently, the shortlisted compounds as well as the extracts and fractions from the *R. japonica* and *R. sachalinensis* rhizomes were tested in vitro by means of the spectrofluorimetric assay using recombinant enzyme.

2. Results

2.1. Molecular Docking Studies

25 compounds (Figure S1), belonging to five different phytochemical classes that are active compounds of *R. japonica* and *R. sachalinensis*, were docked into the binding site of SARS-CoV-2 main protease. First, poses of re-docked N3 ligand were analyzed and root mean square deviation (RMSD) between re-docked poses and crystallographic N3 ligand was calculated. Also, 3D visualization of the superimposed crystallographic ligand with the best docked pose (RMSD = 1.6369 Å) of N3 ligand is shown below (Figure 1).

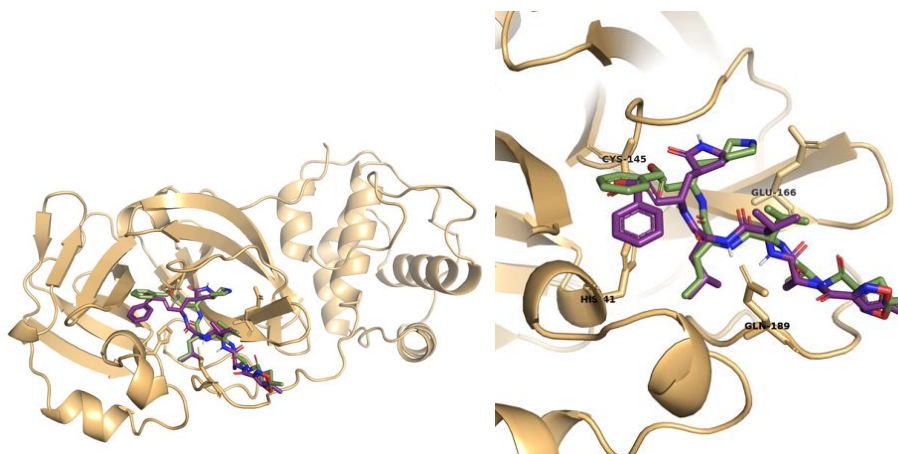


Figure 1. Superimposition of the re-docked (purple) and crystallographic (green) N3 ligand poses.

RMSD analyses and analysis of interactions showed that N3 ligand was successfully re-docked and that hydrogen bonds with such residues as Gly 143, Ser144, Cys145, Glu166, Gln189, Thr190, and Pi-Alkyl interactions with His41 and Ala191 were observed (Figure 2).

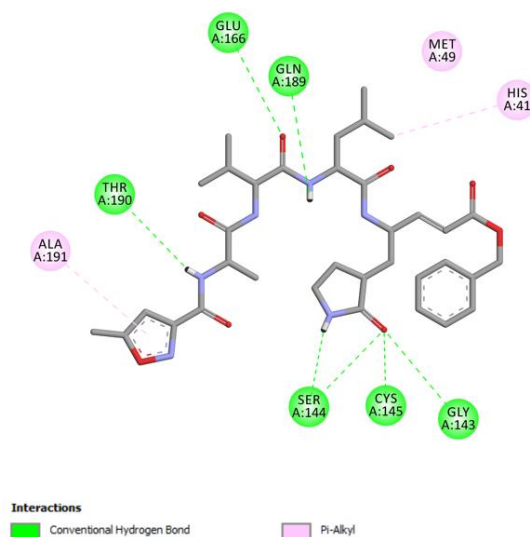


Figure 2. 2D visualization of ligand interactions of re-docked N3 ligand into the binding site of SARS-CoV-2 main protease.

Analysis of results of all docked compounds were based on the visual inspection of interactions with catalytic residues site of Mpro (Cys145 and His41) as well as comparison of interactions of re-docked ligand (N3). Below (Tables 1 and 2), we are presenting 2D interactions diagrams of those 11 compounds, which were also evaluated in vitro as potential inhibitors of Mpro. Conventional hydrogen bonds and also Pi-Pi interactions are listed in the Table 1 below each compound and common interactions with the N3 ligand are marked in bold. Additional docking results are presented in the Supplementary table (Table S1). The choice of compounds for in vitro study was based not only on the best docking results but was also determined by the availability and quantity of isolated compounds. GOLD docking scores of best poses of the compounds are presented in the Supplementary table (Table S2).

Based on the interaction analyses best potential candidates for Mpro inhibitors are presented in Table 1. Table 2 includes compounds which are not supposed to be good inhibitors, but still were evaluated in vitro study to confirm this assumption.

Based on analyses of 11 main compounds we assumed that phenylpropanoid disaccharide esters such as vanicoside A and vanicoside B or procyanidins such as procyanidin B2 3,3'-di-*O*-gallate and procyanidin C1 as well as emodin are potential inhibitors of Mpro. In the case of all those five compounds (Table 1) the interactions with the catalytic residues Cys145 and His41 were observed. Additionally, hydrogen bonds with such residues as Gly143, Glu166, Gln189, His163, or Thr190 were formed. Interactions with those residues were also present in case of re-docked N3 ligand. Epicatechin, (–)-epigallocatechin gallate or epicatechin gallate are probably not as good Mpro inhibitors, because of lack of the interactions with catalytic residues Cys145 or His41. Resveratrol, piceid, and procyanidin B2 were also classified as poor candidates for Mpro inhibitors. In those cases the interactions with Cys145 or His41 were observed, but hydrogen bond interactions with other residues, common with co-crystallized ligand, were absent (Table 2). Further, the analyses of GOLD docking scores (Table S2) shows that in the case of compounds presented in Table 1, their scores are higher (Goldscore.Fitness > 90) than in the case of compounds presented in Table 2, and higher fitness scores indicate better docking results.

Table 1. 2D interactions diagrams of presumably good candidates for Mpro inhibitors.

Phenylpropanoid Disaccharide Esters	
<p>Vanicoside B</p> <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Carbon Hydrogen Bond Pi-Pi Stacked <p>Conventional Hydrogen Bond: Cys145, Gln189, His164, Asn142, Leu141, Tyr54, Cys44; Pi-interactions: His41, Met165</p>	<p>Vanicoside A</p> <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Carbon Hydrogen Bond Pi-Pi Stacked Pi-Alkyl <p>Conventional Hydrogen Bond: Cys145, Glu166, Gln189, Thr190, Thr26; Pi-interactions: His41, Leu27</p>
Procyanidins	
<p>Procyanidin C1</p> <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Pi-Pi T-shaped Amide-Pi Stacked Pi-Alkyl <p>Conventional Hydrogen Bond: Cys145, Met49, Glu166, Gly143, His163; Pi-interactions: His41, Arg188, Met49</p>	<p>Procyanidin B2 3,3'-di-O-gallate</p> <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Pi-Sigma Pi-Alkyl <p>Conventional Hydrogen Bond: His41, Cys145, His163, His164, Cys44, Met49, Gln189; Pi-interactions: Pro52, Met165</p>

Table 1. Cont.

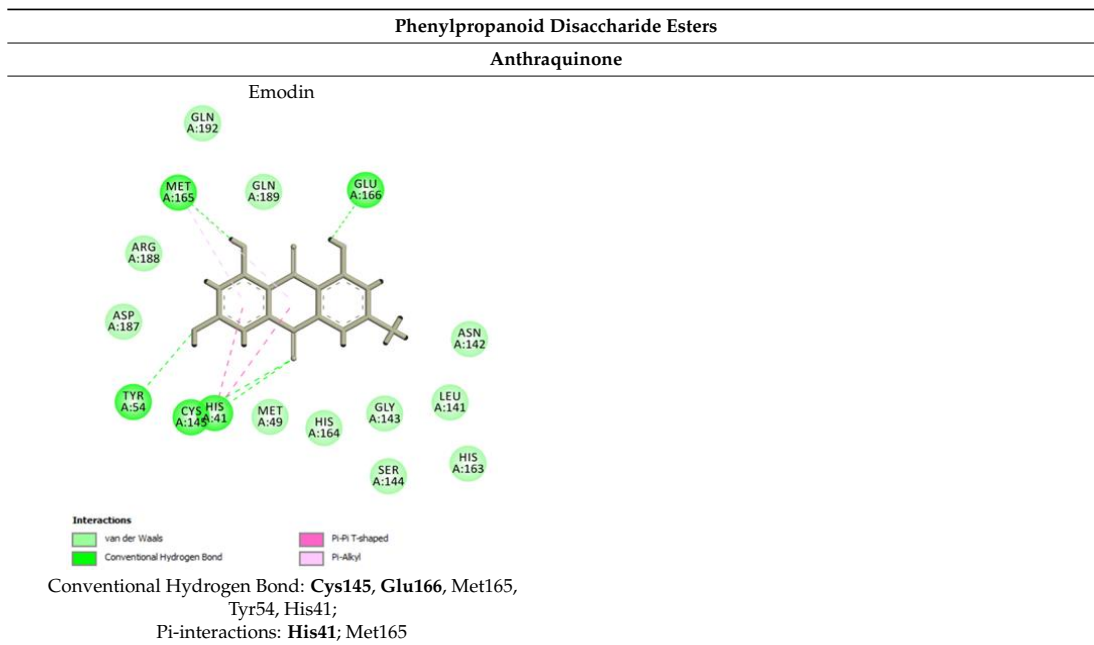


Table 2. 2D interactions diagrams of presumably not good candidates for Mpro inhibitors.

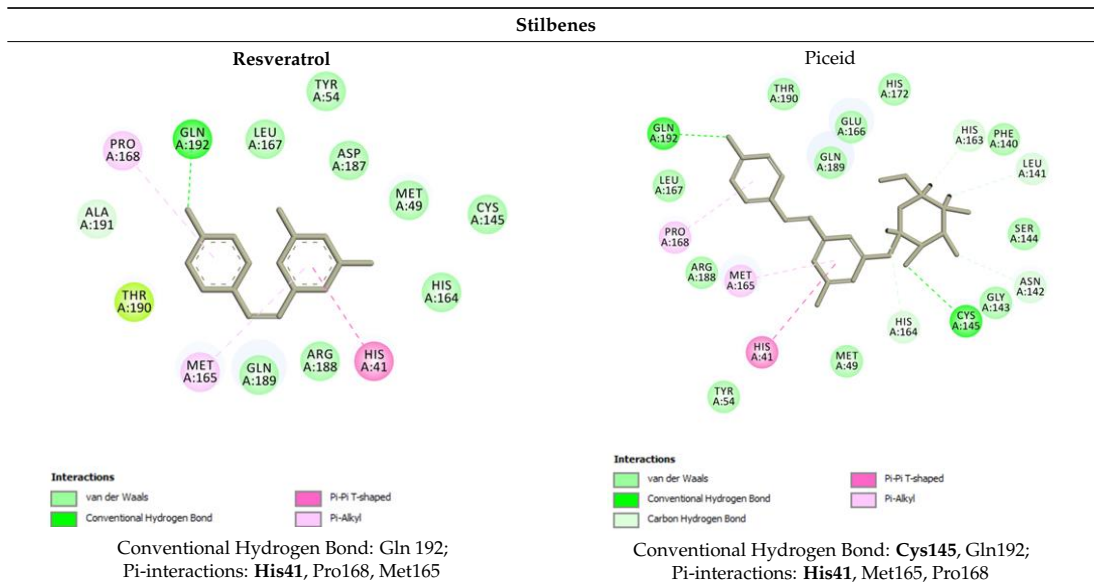
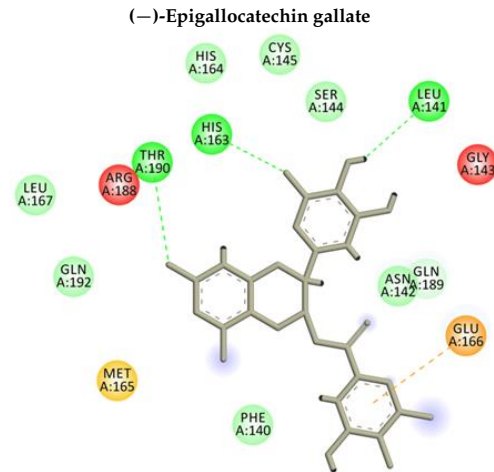
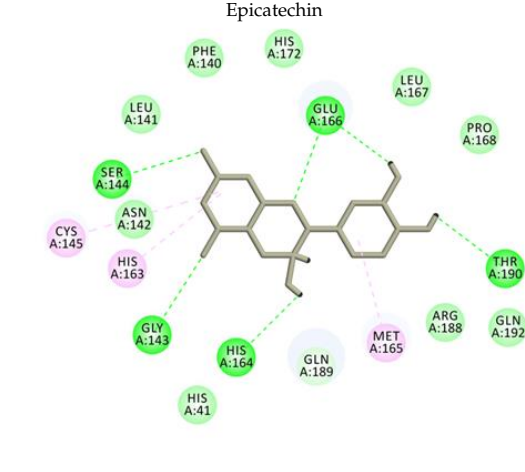
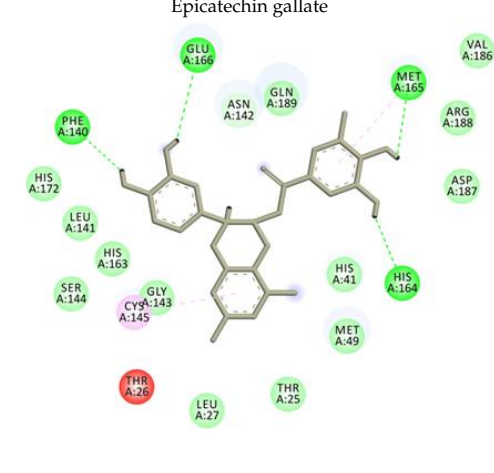
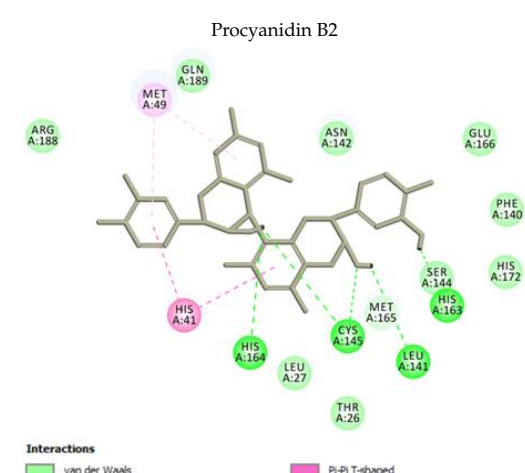


Table 2. Cont.

Stilbenes	
Flavanols and Procyanidins	
<p>(-)Epigallocatechin gallate</p>  <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Pi-Anion <p>Conventional Hydrogen Bond: Thr190, Leu141, His163; Pi-interactions: Glu166</p>	<p>Epicatechin</p>  <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Alkyl Pi-Alkyl <p>Conventional Hydrogen Bond: Ser144, Glu166, Gly143, His164, Thr190; Pi-interactions: Met165, His163, Cys145</p>
<p>Epicatechin gallate</p>  <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Pi-Alkyl <p>Conventional Hydrogen Bond: Met165, His164, Glu166, Phe140; Pi-interactions: Cys145, Met165</p>	<p>Procyanidin B2</p>  <p>Interactions</p> <ul style="list-style-type: none"> van der Waals Conventional Hydrogen Bond Pi-Pi T-shaped Pi-Alkyl <p>Conventional Hydrogen Bond: Cys145, Leu141, His163, His164; Pi-interactions: His41, Met49</p>

Analyses of interactions with Mpro residues of additional 14 compounds, which are included in supplementary table (Table S1), but were not tested *in vitro*, shows that procyanidins such as procyanidin C1 3',3''-di-*O*-gallate and cinnamtannin A2 as well as phenylpropanoid disaccharide esters such as hydropiperoside, tatariside B, lapathoside C or vanicoside C, and emodin bioanthrone could all be promising candidates for Mpro inhibitors and considered for further *in vitro* testing. However, these compounds occur in the studied plant material in minor amounts only and as such, they were unavailable in amounts sufficient for thorough pharmacological investigations. GOLD docking scores (Table S2) in case of those compounds are also higher than in case of the remaining seven compounds presented in Table S1, with the exception of emodin bioanthrone, where docking score is much lower. Compounds such as resveratroliside and piceatannol, piceatannol glucoside, procyanidin B23'-*O*-gallate, physcion, emodin-8-glucoside, and lapathoside A are assumed to be worse candidates either because of low number of hydrogen bonds, which could cause the instability of complexes with protease or because of lack of interaction with both catalytic residues Cys154 and His41. Moreover, GOLD fitness scores are also lower (<80) in all those cases. In addition to 2D interaction visualizations presented above, the 3D visualizations of two highest scored compounds: Vanicosides (A and B) in complex with receptor were generated and shown in Figure 3.

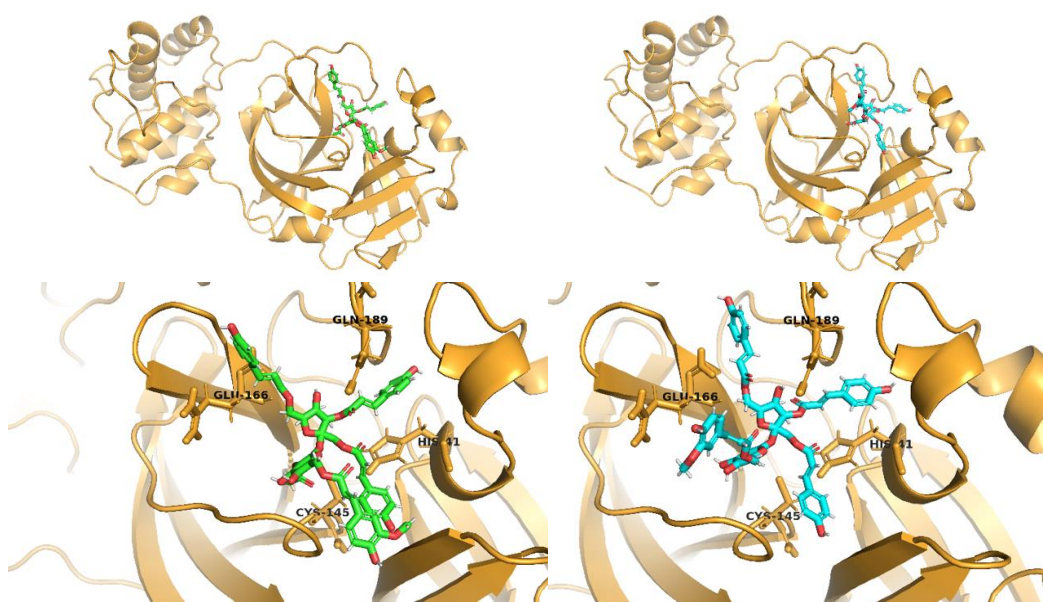


Figure 3. 3D visualization of best scored compounds: Vanicoside A (images on the left—green), Vanicoside B (images on the right—cyan).

2.2. Inhibition of SARS-CoV-2 Mpro Enzyme-*In Vitro* Study

Eleven compounds (Figure 4A, Table S3) and 12 plant extracts and fractions (Figure 4B, Table S4) were studied *in vitro* against recombinant SARS-CoV-2 Mpro in spectrofluorimetric assay. During the experiment we used a novel fluorescent peptide substrate (QS1, Ac-Abu-Tle-Leu-Gln-ACC) [36].

The choice of *R. japonica* and *R. sachalinensis* extracts and fractions was dictated by the promising docking results with compounds present in these plants. The detailed phytochemical composition of the tested extracts and fractions were presented in our earlier studies [33]. Nine out of 11 tested compounds with final concentration of 100 μM inhibited SARS-CoV-2 Mpro enzyme significantly, whereas five of them displayed over 20% inhibition during the screening and were selected for further analysis. Three compounds: vanicoside A, vanicoside B, and emodin revealed over 50% inhibition of the enzyme. All of the studied extracts and fractions with the final concentration of 50 $\mu\text{g}/\text{mL}$ significantly inhibited SARS-CoV-2 Mpro, displaying over 50% inhibition of the enzyme, and were selected for further analysis. In the next step, determination of enzyme inhibition in serial dilutions of the selected compounds and extracts/fractions was defined (Tables S5–S8, Figures 5–8). Three compounds—vanicoside A, vanicoside B, and emodin—showed significant inhibition of SARS-CoV-2 Mpro, also at low concentrations (starting at 13.2 μM), while the remaining two compounds—procyanidin C1, procyanidin B2 3,3'-*O*-gallate—showed significant inhibition only at the highest concentration—100 μM . The Log IC_{50} , IC_{50} , and R^2 were calculated for vanicoside A ($\text{IC}_{50} = 23.10 \mu\text{M}$) and vanicoside B ($\text{IC}_{50} = 43.59 \mu\text{M}$), (Figure 8). Among extracts, stronger inhibition of SARS-CoV-2 Mpro was seen for *R. sachalinensis* acetone extract than for *R. japonica* acetone extract, however both achieved low $\text{IC}_{50} = 9.42 \mu\text{g}/\text{mL}$ and $16.90 \mu\text{g}/\text{mL}$, respectively (Figure 8). All fractions (dichloromethane (CH_2Cl_2), diethyl ether (Et_2O), ethyl acetate (AcOEt), butanol (*n*-BuOH) and water) were obtained during the fractionation process of acetone extracts [33]. Among the fractions, only butanol fractions showed stronger enzyme inhibition than the corresponding acetone extracts (Figures 6 and 7, Table S8). The IC_{50} was 4.031 $\mu\text{g}/\text{mL}$ for *R. sachalinensis* butanol fraction and 7.877 $\mu\text{g}/\text{mL}$ for *R. japonica* butanol fraction (Figure 8). It is supposed that compounds present in the butanol fractions are responsible for observed strong inhibition of main protease by *Reynoutria* extracts.

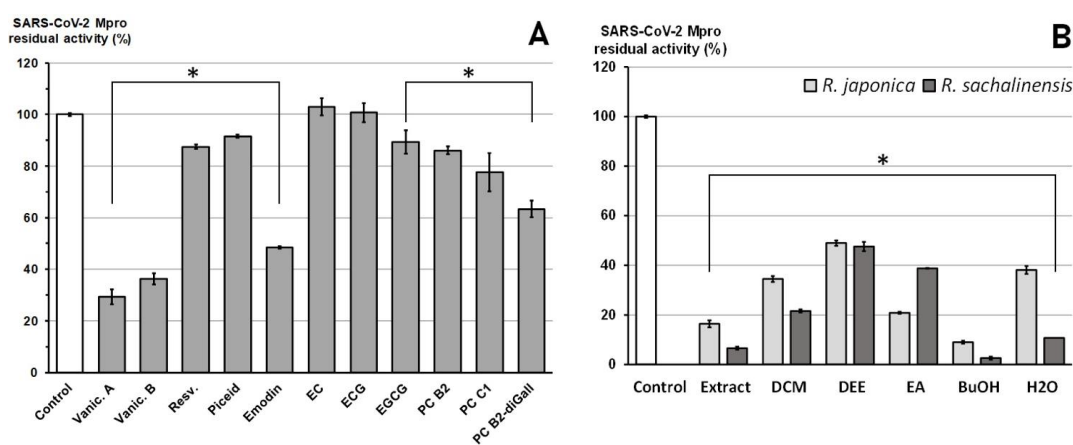


Figure 4. Screening for SARS-CoV-2 Mpro inhibitors. Inhibitors (I), with final concentration equal to 100 μM for individual compounds (A) and 50 $\mu\text{g}/\text{mL}$ for extracts and fractions (B) The results were presented as SARS-CoV-2 Mpro residual activity (%), in relation to control without inhibitor. Error bars shown in this figure are means \pm SD for $n \geq 3$. * Statistically significant at $p \leq 0.05$ compared to control. Abbreviations: Vanic.—vanicoside, Resv.—resveratrol, EC—epicatechin, ECG—epicatechin gallate, EGCG—epigallocatechin gallate, PC—procyanidin, di-Gall—3,3'-*O*-digallate, DCM—dichloromethane, DEE—diethyl ether, EA—ethyl acetate, BuOH—*n*-butanol.

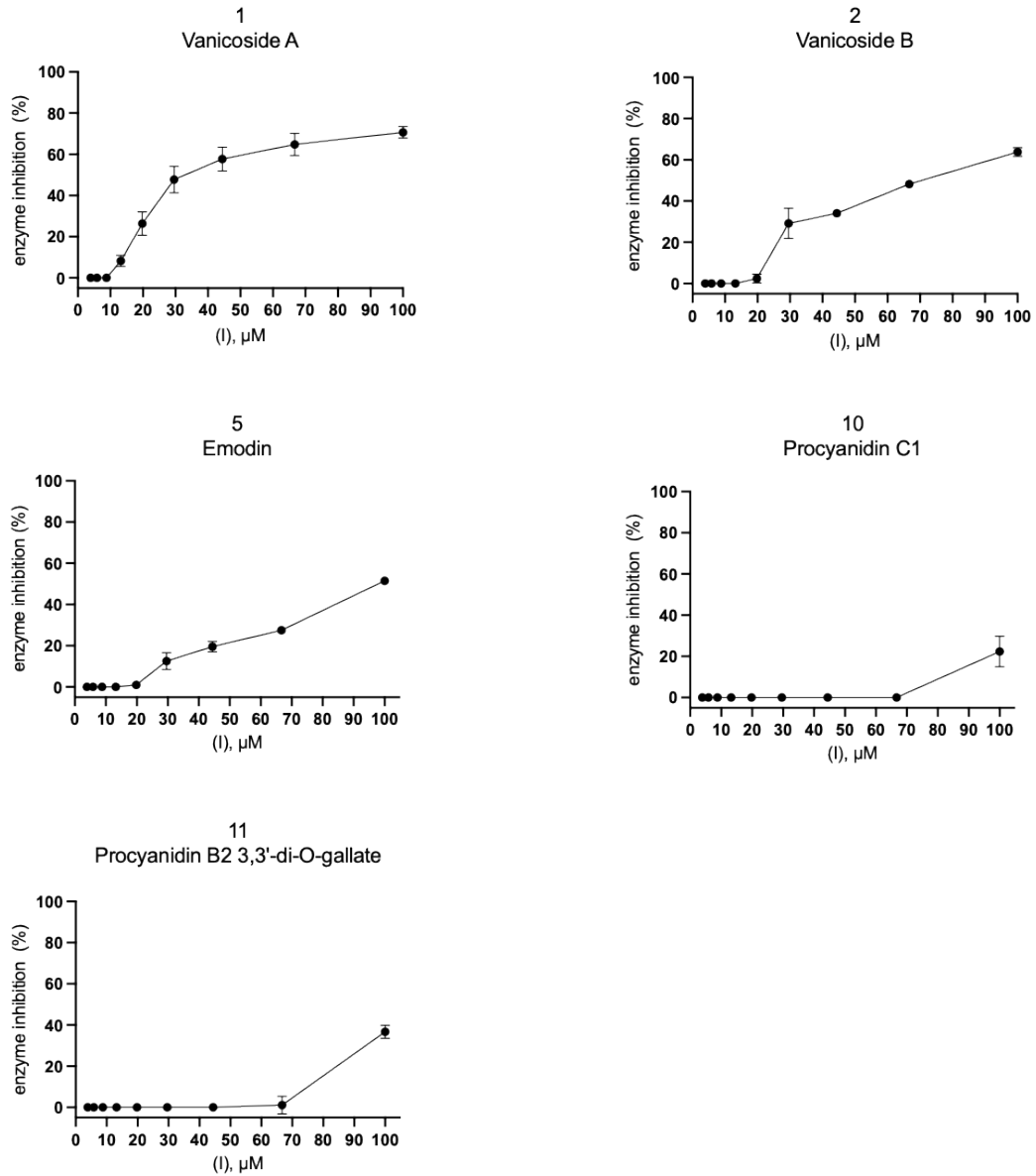


Figure 5. SARS-CoV-2 Mpro activity in serial dilution of compounds. The results were presented as SARS-CoV-2 Mpro inhibition (%).

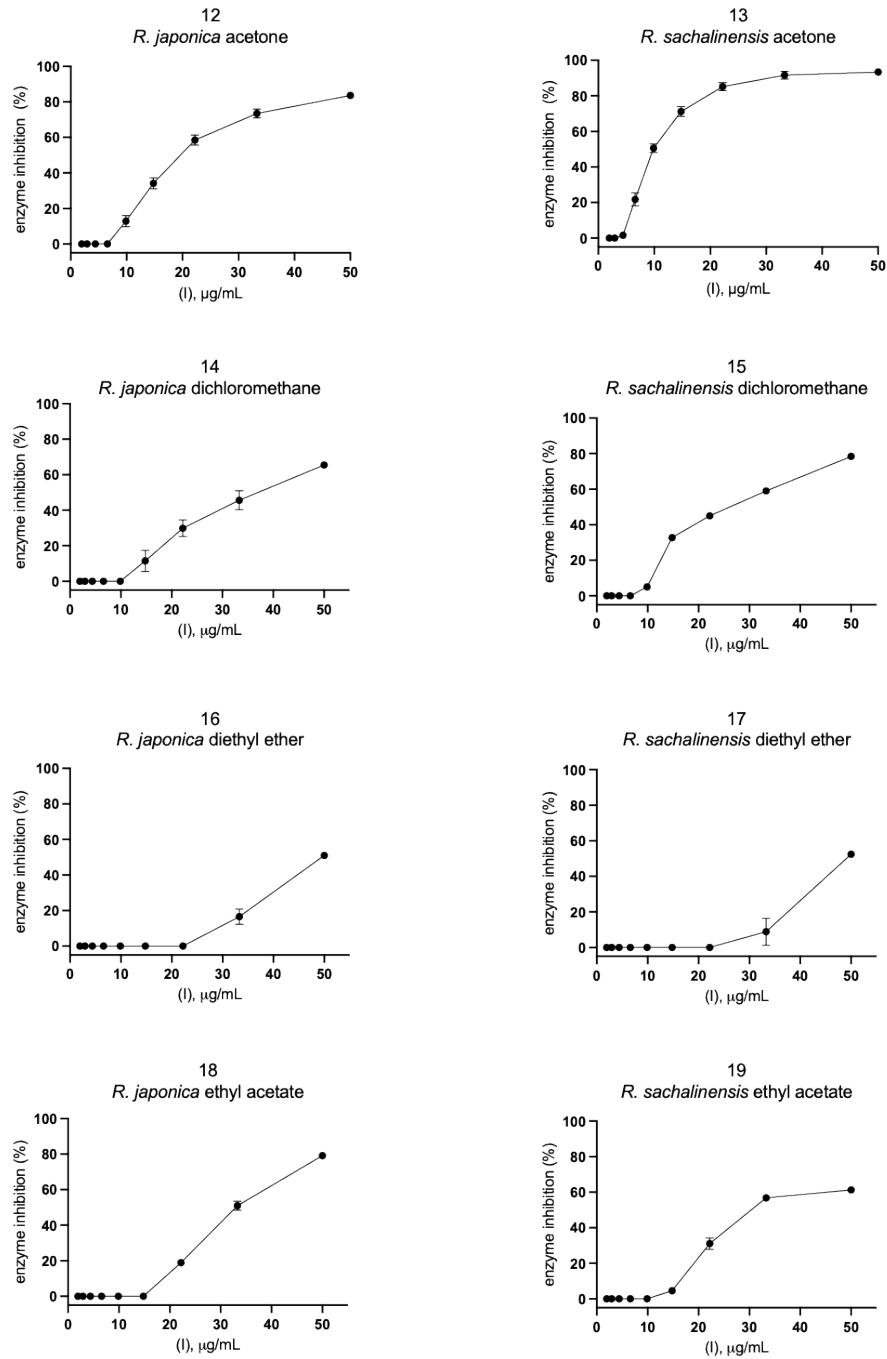


Figure 6. SARS-CoV-2 Mpro activity in serial dilution of extracts and fractions. The results were presented as SARS-CoV-2 Mpro inhibition (%).

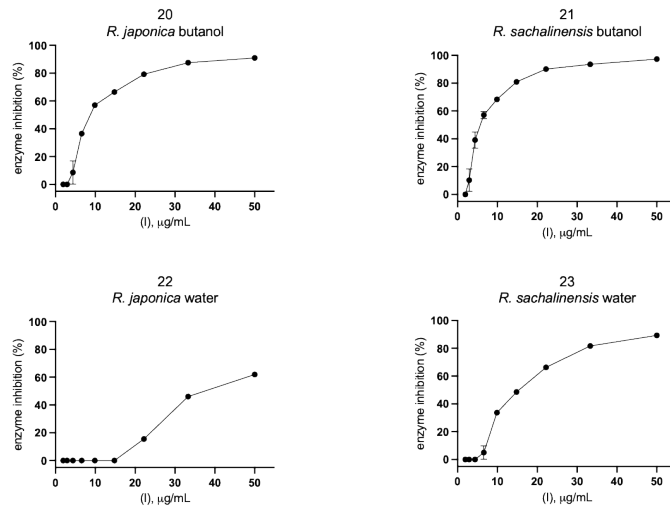


Figure 7. SARS-CoV-2 Mpro activity in serial dilution of extracts and fractions. The results were presented as SARS-CoV-2 Mpro inhibition (%).

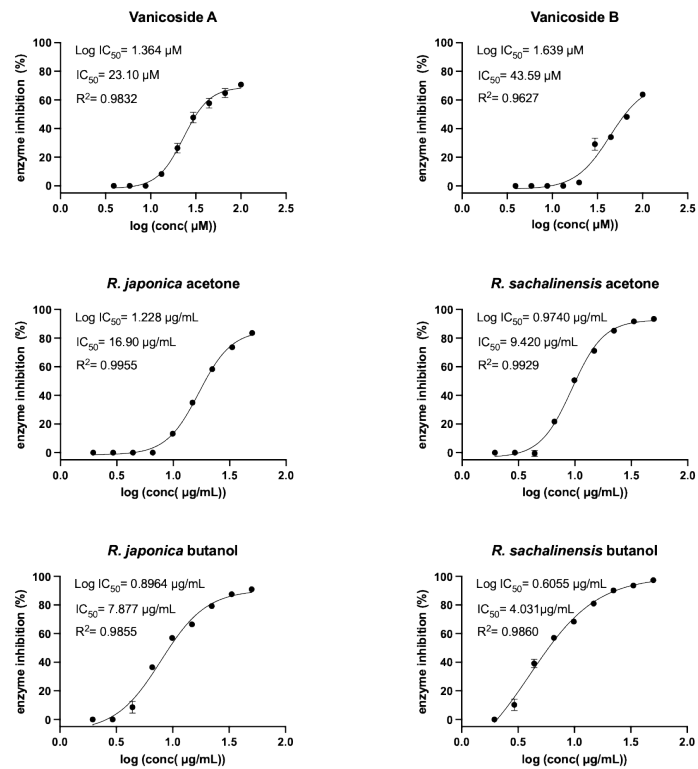


Figure 8. SARS-CoV-2 Mpro activity in serial dilution of the most potent inhibitors. Log IC_{50} , IC_{50} and R^2 were calculated for each sample.

3. Discussion

The 25 compounds, previously identified in *R. japonica* and *R. sachalinensis* extracts, belonging to five different classes of phytochemicals (stilbenes, anthraquinones, phenylpropanoid disaccharide esters, flavan-3-ols, and procyanidins) were evaluated as potential inhibitors against SARS-CoV-2 Mpro in molecular docking study. The most successfully docked were compounds belonging to procyanidins (procyanidin B2 3,3'-di-O-gallate, procyanidin C1, procyanidin C1 3',3''-di-O-gallate, cinnamtannin A2), phenylpropanoid disaccharide esters (vanicoside A, vanicoside B, vanicoside C, hydroxypiperoside, lapathoside C, tatariside B), and anthranoids (emodin, emodin bianthrone) (Table 1 and Table S1). In the case of almost all of those compounds the interactions with the catalytic residues—Cys145 and His41—were observed. Additionally, hydrogen bonds with such residues as Gly143, Ser144, Glu166, Gln189, or Thr190 were formed. Interaction with those residues were also present in case of the N3 ligand. Also, for all those compounds, except emodin bianthrone, the fitness GOLD docking scores are similar or higher than the score for N3 ligand, which could also indicate that those compounds could be classified as good inhibitors. Results from molecular docking study of phenylpropanoid disaccharide esters towards SARS-CoV-2 Mpro are presented for the first time.

Alongside these 11 compounds tested in the in vitro study, we have also tested extracts and fractions from rhizomes of *R. japonica* and *R. sachalinensis*. The results indicated that for strong inhibition of SARS-CoV-2 Mpro by *R. japonica* and *R. sachalinensis* acetone extracts ($IC_{50} = 16.90 \mu\text{g/mL}$ and $9.42 \mu\text{g/mL}$, respectively), the mainly responsible compounds are present in the butanol fractions ($IC_{50} = 4.031 \mu\text{g/mL}$ for *R. sachalinensis* and $IC_{50} = 7.877 \mu\text{g/mL}$ for *R. japonica*). Only these fractions revealed stronger enzyme inhibition than the corresponding acetone extracts (Table S8, Figures 6–8). According to our earlier phytochemical study [33], among all obtained fractions (CH_2Cl_2 , Et_2O , AcOEt , *n*-BuOH, and water), the butanol fractions of *R. japonica* and *R. sachalinensis* contained the highest amount of procyanidins with high degree of polymerization such as procyanidin heptamer or octamer. Next to procyanidins, phenylpropanoid disaccharide esters were another important group of compounds detected in these fractions. The stronger inhibition of SARS-CoV-2 Mpro by *R. sachalinensis* than *R. japonica* could be associated with higher amount of procyanidins and phenylpropanoid disaccharide esters in *R. sachalinensis* rhizomes, which was confirmed in our earlier studies [32,33]. We docked compounds belonging to these phytochemical groups into the binding site of SARS-CoV-2 Mpro. Four of them: vanicoside A, B, procyanidin C1, procyanidin B2 3,3'-di-O-gallate, were selected to in vitro study. However, despite good results in docking study, only vanicoside A ($IC_{50} = 23.10 \mu\text{M}$) and vanicoside B ($IC_{50} = 43.59 \mu\text{M}$) showed moderate inhibition of SARS-CoV-2 Mpro. Therefore, we suggest that other compounds may be responsible for the strong inhibition of SARS-CoV-2 Mpro by butanol fractions, or a phenomenon of synergy between the compounds occurred. However, some of the well scored compounds, belonging to the phenylpropanoid esters (hydroxypiperoside, lapathoside C) and procyanidins (cinnamtannin A2) were not tested because of insufficient amounts obtained from the crude drug. Even so, their scores were comparable but not higher than those of the tested vanicosides/procyanidin C1, respectively. These rare compounds also occur mainly in *R. sachalinensis*, less utilized as a medicinal plant. Hence, future investigation into the anti SARS-CoV-2 potential should focus on this species.

Highly polymerized proanthocyanidins present in butanol fractions, with the degree of polymerization higher than those tested in this in vitro study (dimers, trimers), may have had a significant effect on the strong inhibitory effect of these fraction on SARS-CoV-2 Mpro. Moreover, according to previous studies [33], dimeric and trimeric proanthocyanidins (with weak inhibition activity in our experiment) and simple flavan-3-ols (epicatechin, epicatechin gallate, without inhibitory effect at $100 \mu\text{M}$ in our experiment), apart from phenylpropanoid disaccharide esters, are the main compounds in the Et_2O and AcOEt fractions, which provides a plausible explanation of their weaker inhibitory effect (Figure 6). Additional studies are needed to confirm the inhibition of SARS-CoV-2 Mpro by highly

polymerized proanthocyanidins. We are continuing isolation of these subfractions and single compounds from butanol fractions to fully confirm these assumptions. However, they are already supported by the chemical nature of these compounds as well as by the other studies outlined below. Proanthocyanidins bind proteins also non-specifically due to the numerous phenol (hydroxyl) groups that form cross-linked structures with polypeptides. This seems to depend on the size of the molecule: the more a proanthocyanidin is polymerized, the less specific the protein bonds are. Moreover, a higher degree of polymerization of proanthocyanidins increases their affinity to proteins and enhances the cross-linkages between proteins [36–38]. Some studies have demonstrated antiviral activity of extracts rich in proanthocyanidins. Zhuang et al. [39] found that butanol fraction of *Cinnamomi cortex* containing high amount of proanthocyanidins, inhibit SARS-CoV infection ($IC_{50S} = 7.8 \pm 0.3 \mu\text{g/mL}$, wild-type SARS-CoV). Conzelmann et al. [40] revealed that black chokeberry (*Aronia melanocarpa*) juice, pomegranate (*Punica granatum*) juice and green tea (*Camellia sinensis*), rich in proanthocyanidins, possess high antiviral efficacy against SARS-CoV-2 (BetaCoV/France/IDF0372/2020) and influenza A virus (A/H1N1/Brisbane/59/2007). The 5-min incubation of SARS-CoV-2 with black chokeberry juice resulted in a $\geq 1.52 \log_{10}$ decrease in infectivity (which corresponds to a 96.98% reduction of infectivity). The mean procyanidin polymerization degree (mDP) in *A. melanocarpa* juice is high: 12–52 [41]. Due to general affinity of proanthocyanidins to proteins, in addition to binding the main protease SARS-CoV-2, they can also bind other essential viral structural proteins. So far, an inhibitory effect of extracts rich in proanthocyanidins or isolated compounds on the other enveloped viruses such as influenza and RSV has been observed [42–44]. It was shown that the *Rumex acetosa* extract rich in proanthocyanidins and its main active constituent-procyanidin B2-di-gallate protect cells from influenza A virus infection by inhibiting viral attachment [44]. Similarly, in the case of the herpes simplex virus (HSV-1), inhibition of virus adsorption and penetration by proanthocyanidins was observed [45]. Moreover, proanthocyanidins were proposed as a new class of hepatitis B and D virus entry inhibitors [46]. They directly target the preS1 region of the HBV large surface protein.

Taking into account the above properties of proanthocyanidins, in the context of COVID-19, it has been hypothesized that proanthocyanidins can inhibit the attachment of SARS-CoV-2 to the oral epithelium [47] and lower viral adsorption and penetration. The lower viral load may lower the risk of developing severe condition [48]. It was suggested that pharmaceutical preparations like gargle and mouthwash solutions as well as lozenges or chewing gums with extracts rich in proanthocyanidins may be useful in the prophylaxis and adjunctive therapy of COVID-19. An important issue in the case SARS-CoV-2 treatment, apart from virus neutralization, is also dealing with overreaction of the immune system and a cytokine storm leading to systemic inflammatory response syndrome (SIRS) and ARDS [48]. Importantly, phytochemicals, including proanthocyanidins, present in the studied extracts and fractions have proven anti-inflammatory effects [49]. However, this issue goes beyond the scope of the present work and requires a separate development.

4. Materials and Methods

4.1. Extracts and Fractions of *Reynoutria Species*

The plant material, extracts, and fractions were obtained according to procedures described in our previous paper [33], stored under -80°C and their composition confirmed to be unchanged until the beginning of current experiments.

4.2. Compounds

Vanicoside A and vanicoside B were isolated earlier from rhizomes of *Reynoutria sachalinensis* (F.Schmidt) Nakai, according procedure described in [32]. The structures of vanicoside B and vanicoside A were identified using ^1H and ^{13}C NMR and HR-MS-qTOF MS analysis and presented in the above article [32]. Emodin, piceid, epigallocatechin gallate, procyanidin B2, procyanidin C1, were purchased in ChemFaces (Wuhan, China),

procyanidin B2 3,3'-di-O-gallate was purchased in Albtchnology (HongKong) and resveratrol, epicatechin, epicatechin gallate in MilliporeSigma (St. Louis, MO, USA) and their purity verified using HPLC and exceeded 98%.

4.3. Molecular Docking

The 25 selected compounds were docked into the binding site of SARS-CoV-2 main protease. Molecular docking was carried out with GOLD software (version 5.7.2), which uses genetic algorithms for generation of ligand conformations [50]. Crystal structure of main SARS-CoV-2protease was obtained from PDB database [51] (PDB code: 6LU7). Binding site was defined based on the position of co-crystallized ligand (N3) from PDB protein structure and all atoms within 10 Å from the ligand were selected. First, the N3 was re-docked into the binding site of the main protease and RMSD with respect to co-crystallized ligand was calculated. Interactions of best pose of re-docked ligand were compared to the interactions of N3 ligand provided in the literature such as: Gly143, Cys145, His163, His164, Glu166, Gln189, Thr190 [52]. Docking protocol, where ligands were flexible, and residues of the binding site were set to rigid, was established and selected compounds were docked into the binding site of the receptor. Even though most of the ligands were large and flexible molecules, none of the rotatable bonds of ligands were fixed and default setting or ligand flexibility were used. Genetic algorithm used in GOLD is able to predict the binding mode of highly flexible molecules [50]. Structures of ligands were obtained from PubChem database in sdf format [53]. In case of lack of 3D structure, 2D structures were uploaded to docking software. Interaction analyses of protein-ligand complexes and 2D interaction diagrams were done with BIOVIA Discovery Studio 2020 [54]. The 3D interaction visualizations of ligands with Mpro were created with PYMOL version 2.3.5 [55]. GoldScore scoring function was chosen for ranking of compounds and for each ligand 10 poses were generated.

Molecular docking analyses and choice of the best candidates for Mpro inhibitors was based on visual inspection of protein-ligand complexes. When interactions with key residues were observed—those compounds were suggested as potential candidates for inhibitors. Key residues were defined based on the literature [52] and by re-docking the co-crystallized ligand. Additionally, GOLD fitness scores of all docked compounds were analyzed.

4.4. Inhibition of SARS-CoV-2 M^{pro} Enzyme-In Vitro Study

The experiment was performed according to the procedure described in the previous article [56] with minor modifications. All experiments were carried out in 96-well assay plates. Buffer solution (pH 7.3) contained 50 mM Tris, 1 mM EDTA, and 1 mM DTT. Inhibitor screenings: to the wells, 1 µL of DMSO inhibitors solutions was added. Then, 79 µL of SARS-CoV-2 Mpro enzyme (E) in buffer was added. Enzyme was incubated with inhibitors (I) for 10' at 37 °C. After incubation, 20 µL of substrate (QS1, Ac-Abu-Tle-Leu-Gln-ACC) [15] in buffer solution was added. Final concentrations were [E] = 100 nM, [QS1] = 50 µM, [I] = 100 µM for compounds 1–11 or [I] = 50 µg/mL for extracts and fractions 12–23. Measurements were carried out in Molecular Devices SpectraMax Gemini XPS spectrofluorometer at 37 °C for 30'. Liberation of ACC fluorophore was measured using $\lambda_{ex} = 355$ nm and $\lambda_{em} = 460$ nm wavelengths. The linear range of progress curves was used for analysis. Measurements were carried out at least in triplicate. The results were presented as mean values with standard deviations. Compounds as well as extracts and fractions displaying >20% inhibition during the screening were selected for further analysis.

Determination of enzyme inhibition % in serial dilutions of inhibitors: serial dilutions in DMSO of selected compounds were prepared (dilution factor 2/3). The experiment was carried out analogically to the screening described above. 1 µL of diluted inhibitors in DMSO were added to the wells. Then, 79 µL of SARS-CoV-2 Mpro enzyme in buffer was added. Enzyme was incubated with inhibitors for 10' at 37 °C. After incubation, 20 µL of substrate (QS1) in buffer solution was added. Final concentrations were [E] = 100 nM and

[QS1] = 50 μ M. Measurements were carried out in Molecular Devices SpectraMax Gemini XPS spectrofluorometer at 37 °C for 30'. Liberation of ACC fluorophore was measured using $\lambda_{\text{ex}} = 355$ nm and $\lambda_{\text{em}} = 460$ nm wavelengths. The linear range of progress curves was used for analysis. Measurements were carried out at least in triplicate. The results were presented as mean values with standard deviations.

4.5. Statistical Analysis

Each assay was performed in at least triplicate and presented as mean \pm SD for $n \geq 3$. Statistical analysis was performed using GraphPad Prism v.7 software (GraphPad Software, San Diego, CA, USA). Initially, the Shapiro-Wilk test was used to assess the distribution of results. Then, in the Student t-test a comparison of the means between the treated and control samples (DMSO instead of the inhibitor) was used. Results with $p \leq 0.05$ was considered statistically significant. Log IC₅₀, IC₅₀ and R² were calculated using GraphPad Prism v. 7.

5. Conclusions

Among the 11 phytochemicals (vanicoside A, vanicoside B, resveratrol, piceid, emodin, epicatechin, epicatechin gallate, epigallocatechin gallate, procyanidin B2, procyanidin C1, procyanidin B2 3,3'-di-O-gallate) selected for in vitro study after docking into the binding site of SARS-CoV-2 Mpro, the best results were achieved with vanicoside A and vanicoside B with moderate inhibition of SARS-CoV-2 Mpro, equal IC₅₀ = 23.10 μ M and 43.59 μ M, respectively. These compounds are important components of the extracts and fractions obtained from the rhizomes of *Reynoutria japonica* and *Reynoutria sachalinensis*. As the first report about an interaction of these compounds with a virus protein, it also indicates a possible biological function in the plant pathogen resistance that should inspire further studies.

Nonetheless, the evident inhibitory activity of the vanicosides does not alone explain a strong inhibition of SARS-CoV-2 Mpro by acetone extracts and mainly butanol fractions. As the main constituents of butanol fractions, besides the phenylpropanoid disaccharide esters (e.g., vanicosides), are highly polymerized procyanidins, we suppose that they could be responsible for strong inhibitory properties of these fractions. Moreover, the less polymerized procyanidins (procyanidin B2, procyanidin C1, procyanidin B2 3,3'-di-O-gallate), abundantly present in the fractions with weaker inhibition of the enzyme, showed no remarkable activity. Further studies are needed to prove the contribution of highly polymerized procyanidins and their potential synergy with vanicosides.

Supplementary Materials: The following are available online at <https://www.mdpi.com/article/10.3390/ph14080742/s1>, Figure S1: Structures of compounds docked into the binding site of SARS-CoV-2 main protease, generated with PubChem Sketcher V2.4 (PubChem Sketcher V2.4), Table S1: Compounds docked to SARS-CoV-2 main protease of (Mpro), Table S2: GOLD docking scores of all compounds tested in vitro, Table S3: Compounds studied in vitro against the proteases SARS-CoV-2 Mpro, Table S4: Extracts and fractions studied in vitro against the proteases SARS-CoV-2 Mpro, Table S5: SARS-CoV-2 Mpro activity in serial dilution of isolated compounds. The results were presented as SARS-CoV-2 Mpro residual activity (%), Table S6: SARS-CoV-2 Mpro activity in serial dilution of extracts and fractions from *R. japonica* (R.j.) and *R. sachalinensis* (R.s.) rhizomes. The results were presented as SARS-CoV-2 Mpro residual activity (%), Table S7: SARS-CoV-2 Mpro activity in serial dilution of isolated compounds. The results were presented as SARS-CoV-2 Mpro inhibition (%), Table S8: SARS-CoV-2 Mpro activity in serial dilution of extracts and fractions from *R. japonica* (R.j.) and *R. sachalinensis* (R.s.) rhizomes. The results were presented as SARS-CoV-2 Mpro inhibition (%).

Author Contributions: Conceptualization: R.A., I.N.-H.; project administration: I.N.-H., R.A.; investigation: R.A., I.N.-H., M.Z., M.K.-B.; methodology: I.N.-H., R.A.; molecular docking study: R.A.; obtainment of extracts, fractions and compounds: I.N.-H.; in vitro study: M.Z., I.N.-H.; statistical analysis: I.N.-H., J.H.; writing—original draft preparation: I.N.-H., R.A.; writing—review and editing: A.M., I.N.-H., R.A.; resources A.M., M.D., R.P.; software R.A., R.P.; supervision: I.N.-H., A.M., R.A.; funding acquisition M.D., R.P., J.H.; visualization: I.N.-H., R.A. All authors have read and agreed to the published version of the manuscript.

Funding: This work was supported by the Medical Research Agency in Poland through its Own Project (grant 2020/ABM/SARS/1), by the National Science Center grant UMO-2020/01/0/NZ1/00063, and the “TEAM/2017-4/32” project, which is carried out within the TEAM program of the Foundation for Polish Science, co-financed by the European Union under the European Regional Development Fund and by Wrocław Medical University subvention No. SUB.D030.21.029 (to A.M.). The provision of plant material and voucher specimen banking is funded by Special Research Facility grant from Ministry of Science and Higher Education—decision No. 96/E-394/SPUB/SP/2019.

Institutional Review Board Statement: Not applicable.

Informed Consent Statement: Not applicable.

Data Availability Statement: Data is contained within the article and supplementary materials.

Acknowledgments: The valuable technical assistance during the plant material extraction and analysis is acknowledged to Marcin Surma.

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

References

1. Wu, F.; Zhao, S.; Yu, B.; Chen, Y.-M.; Wang, W.; Song, Z.-G.; Hu, Y.; Tao, Z.-W.; Tian, J.-H.; Pei, Y.-Y.; et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**, *579*, 265–269. [CrossRef]
2. Wang, C.; Horby, P.W.; Hayden, F.G.; Gao, G.F. A novel coronavirus outbreak of global health concern. *Lancet* **2020**, *395*, 470–473. [CrossRef]
3. Gordon, D.E.; Hiatt, J.; Bouhaddou, M.; Rezelj, V.V.; Ulferts, S.; Braberg, H.; Jureka, A.S.; Obernier, K.; Guo, J.Z.; Batra, J.; et al. Comparative host-coronavirus protein interaction networks reveal pan-viral disease mechanisms. *Science* **2020**, *370*, eabe9403. [CrossRef] [PubMed]
4. Menni, C.; Valdes, A.M.; Freidin, M.B.; Sudre, C.H.; Nguyen, L.H.; Drew, D.A.; Ganesh, S.; Varsavsky, T.; Cardoso, M.J.; El-Sayed Moustafa, J.S.; et al. Real-time tracking of self-reported symptoms to predict potential COVID-19. *Nat. Med.* **2020**, *26*, 1037–1040. [CrossRef]
5. Pierron, D.; Pereda-Loth, V.; Mantel, M.; Moranges, M.; Bignon, E.; Alva, O.; Kabous, J.; Heiske, M.; Pacalon, J.; David, R.; et al. Smell and taste changes are early indicators of the COVID-19 pandemic and political decision effectiveness. *Nat. Commun.* **2020**, *11*, 5152. [CrossRef]
6. Zhou, Y.; Wang, F.; Tang, J.; Nussinov, R.; Cheng, F. Artificial intelligence in COVID-19 drug repurposing. *Lancet Digit. Health* **2020**, *2*, e667–e676. [CrossRef]
7. Adeoye, A.O.; Oso, B.J.; Olaoye, I.F.; Tijjani, H.; Adebayo, A.I. Repurposing of chloroquine and some clinically approved antiviral drugs as effective therapeutics to prevent cellular entry and replication of coronavirus. *J. Biomol. Struct. Dyn.* **2020**, 1–11. [CrossRef] [PubMed]
8. Jia, Z.; Song, X.; Shi, J.; Wang, W.; He, K. Transcriptome-based drug repositioning for coronavirus disease 2019 (COVID-19). *Pathog. Dis.* **2020**, *78*. [CrossRef]
9. Khan, A.; Ali, S.S.; Khan, M.T.; Saleem, S.; Ali, A.; Suleman, M.; Babar, Z.; Shafiq, A.; Khan, M.; Wei, D.-Q. Combined drug repurposing and virtual screening strategies with molecular dynamics simulation identified potent inhibitors for SARS-CoV-2 main protease (3CLpro). *J. Biomol. Struct. Dyn.* **2020**, 1–12. [CrossRef]
10. Available online: <https://go.drugbank.com/covid-19> (accessed on 10 June 2021).
11. Ton, A.; 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**, *39*, 2000028. [CrossRef]
12. Wu, C.; Liu, Y.; Yang, Y.; Zhang, P.; Zhong, W.; Wang, Y.; Wang, Q.; Xu, Y.; Li, M.; Li, X.; et al. Analysis of therapeutic targets for SARS-CoV-2 and discovery of potential drugs by computational methods. *Acta Pharm. Sin. B* **2020**, *10*, 766–788. [CrossRef]
13. Abel, R.; Paredes Ramos, M.; Chen, Q.; Pérez-Sánchez, H.; Coluzzi, F.; Rocco, M.; Marchetti, P.; Mura, C.; Simmaco, M.; Bourne, P.E.; et al. Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2. *Front. Chem.* **2020**, *8*, 1–19. [CrossRef]
14. Gordon, D.E.; Jang, G.M.; Bouhaddou, M.; Xu, J.; Obernier, K.; White, K.M.; O’Meara, M.J.; Rezelj, V.V.; Guo, J.Z.; Swaney, D.L.; et al. A SARS-CoV-2 protein interaction map reveals targets for drug repurposing. *Nature* **2020**, *583*, 459–468. [CrossRef] [PubMed]

15. Rut, W.; Groborz, K.; Zhang, L.; Sun, X.; Zmudzinski, M.; Pawlik, B.; Wang, X.; Jochmans, D.; Neyts, J.; Mlynarski, W.; et al. SARS-CoV-2 Mpro inhibitors and activity-based probes for patient-sample imaging. *Nat. Chem. Biol.* **2020**, *17*, 222–228. [[CrossRef](#)]
16. Vicidomini, C.; Roviello, V.; Roviello, G.N. Molecular Basis of the Therapeutical Potential of Clove (*Syzygium aromaticum* L.) and Clues to Its Anti-COVID-19 Utility. *Molecules* **2021**, *26*, 1880. [[CrossRef](#)] [[PubMed](#)]
17. 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*, 18–23. [[CrossRef](#)]
18. Vicidomini, C.; Roviello, V.; Roviello, G.N. In Silico Investigation on the Interaction of Chiral Phytochemicals from *Opuntia ficus-indica* with SARS-CoV-2 Mpro. *Symmetry* **2021**, *13*, 1041. [[CrossRef](#)]
19. Dwarka, D.; Agoni, C.; Mellem, J.J.; Soliman, M.E.; Bajjnath, H. Identification of potential SARS-CoV-2 inhibitors from South African medicinal plant extracts using molecular modelling approaches. *S. Afr. J. Bot.* **2020**, *133*, 273–284. [[CrossRef](#)]
20. Verma, S.; Twilley, D.; Esmear, T.; Oosthuizen, C.B.; Reid, A.-M.; Nel, M.; Lall, N. Anti-SARS-CoV Natural Products With the Potential to Inhibit SARS-CoV-2 (COVID-19). *Front. Pharmacol.* **2020**, *11*, 1514. [[CrossRef](#)] [[PubMed](#)]
21. Kulkarni, S.A.; Nagarajan, S.K.; Ramesh, V.; Palaniyandi, V.; Selvam, S.P.; Madhavan, T. Computational evaluation of major components from plant essential oils as potent inhibitors of SARS-CoV-2 spike protein. *J. Mol. Struct.* **2020**, *1221*, 128823. [[CrossRef](#)]
22. Shree, P.; Mishra, P.; Selvaraj, C.; Singh, S.K.; Chaube, R.; Garg, N.; Tripathi, Y.B. Targeting COVID-19 (SARS-CoV-2) main protease through active phytochemicals of ayurvedic medicinal plants—*Withania somnifera* (Ashwagandha), *Tinospora cordifolia* (Giloy) and *Ocimum sanctum* (Tulsi)—A molecular docking study. *J. Biomol. Struct. Dyn.* **2020**, 1–14. [[CrossRef](#)]
23. Peng, W.; Qin, R.; Li, X.; Zhou, H. Botany, phytochemistry, pharmacology, and potential application of *Polygonum cuspidatum* Sieb. et Zucc.: A review. *J. Ethnopharmacol.* **2013**, *148*, 729–745. [[CrossRef](#)]
24. Tao, Z.; Gao, J.; Zhang, G.; Xue, M.; Yang, W.; Tong, C.; Yuan, Y. Shufeng Jiedu Capsule protect against acute lung injury by suppressing the MAPK/NF- κ B pathway. *Biosci. Trends* **2014**, *8*, 45–51. [[CrossRef](#)]
25. Tao, Z.; Meng, X.; Han, Y.Q.; Xue, M.M.; Wu, S.; Wu, P.; Yuan, Y.; Zhu, Q.; Zhang, T.J.; Wong, C.C.L. Therapeutic Mechanistic Studies of ShuFengJieDu Capsule in an Acute Lung Injury Animal Model Using Quantitative Proteomics Technology. *J. Proteome Res.* **2017**, *16*, 4009–4019. [[CrossRef](#)]
26. Luo, L.; Jiang, J.; Wang, C.; Fitzgerald, M.; Hu, W.; Zhou, Y.; Zhang, H.; Chen, S. Analysis on herbal medicines utilized for treatment of COVID-19. *Acta Pharm. Sin. B* **2020**, *10*, 1192–1204. [[CrossRef](#)]
27. Wang, Z.; Chen, X.; Lu, Y.; Chen, F.; Zhang, W. Clinical characteristics and therapeutic procedure for four cases with 2019 novel coronavirus pneumonia receiving combined Chinese and Western medicine treatment. *Biosci. Trends* **2020**, *14*, 64–68. [[CrossRef](#)] [[PubMed](#)]
28. Wahedi, H.M.; Ahmad, S.; Abbasi, S.W. Stilbene-based natural compounds as promising drug candidates against COVID-19. *J. Biomol. Struct. Dyn.* **2021**, *39*, 3225–3234. [[CrossRef](#)]
29. Ho, T.-Y.; Wu, S.-L.; Chen, J.-C.; Li, C.-C.; Hsiang, C.-Y. Emodin blocks the SARS coronavirus spike protein and angiotensin-converting enzyme 2 interaction. *Antiviral Res.* **2007**, *74*, 92–101. [[CrossRef](#)] [[PubMed](#)]
30. Muchtaridi, M.; Fauzi, M.; Ikram, N.K.K.; Gazzali, A.M.; Wahab, H.A. Natural Flavonoids as Potential Angiotensin-Converting Enzyme 2 Inhibitors for Anti-SARS-CoV-2. *Molecules* **2020**, *25*, 3980. [[CrossRef](#)] [[PubMed](#)]
31. Maroli, N.; Bhasuran, B.; Natarajan, J.; Kolandaivel, P. The potential role of procyanidin as a therapeutic agent against SARS-CoV-2: A text mining, molecular docking and molecular dynamics simulation approach. *J. Biomol. Struct. Dyn.* **2020**, 1–16. [[CrossRef](#)]
32. Nawrot-Hadzik, I.; Granica, S.; Domaradzki, K.; Pecio, L.; Matkowski, A. Isolation and Determination of Phenolic Glycosides and Anthraquinones from Rhizomes of Various Reynoutria Species. *Planta Med.* **2018**, *84*, 1118–1126. [[CrossRef](#)]
33. Nawrot-Hadzik, I.; Slusarczyk, S.; Granica, S.; Hadzik, J.; Matkowski, A. Phytochemical diversity in rhizomes of three Reynoutria species and their antioxidant activity correlations elucidated by LC-ESI-MS/MS analysis. *Molecules* **2019**, *24*, 1136. [[CrossRef](#)] [[PubMed](#)]
34. Nawrot-Hadzik, I.; Choromańska, A.; Abel, R.; Preissner, R.; Saczko, J.; Matkowski, A.; Hadzik, J. Cytotoxic effect of vanicosides a and b from *reynoutria sachalinensis* against melanotic and amelanotic melanoma cell lines and in silico evaluation for inhibition of brafv600e and mek1. *Int. J. Mol. Sci.* **2020**, *21*, 4611. [[CrossRef](#)]
35. Chen, X.Y.; Wang, R.F.; Liu, B. An update on oligosaccharides and their esters from traditional Chinese medicines: Chemical structures and biological activities. *Evid. Based Complement. Altern. Med.* **2015**, *2015*, 1–23. [[CrossRef](#)] [[PubMed](#)]
36. Ottaviani, J.I.; Actis-Goretta, L.; Villordo, J.J.; Fraga, C.G. Procyanidin structure defines the extent and specificity of angiotensin I converting enzyme inhibition. *Biochimie* **2006**, *88*, 359–365. [[CrossRef](#)] [[PubMed](#)]
37. Kilmister, R.L.; Faulkner, P.; Downey, M.O.; Darby, S.J.; Falconer, R.J. The complexity of condensed tannin binding to bovine serum albumin—An isothermal titration calorimetry study. *Food Chem.* **2016**, *190*, 173–178. [[CrossRef](#)] [[PubMed](#)]
38. Rauf, A.; Imran, M.; Abu-Izneid, T.; Ihtisham-Ul-Haq; Patel, S.; Pan, X.; Naz, S.; Sanches Silva, A.; Saeed, F.; Rasul Suleria, H.A. Proanthocyanidins: A comprehensive review. *Biomed. Pharmacother.* **2019**, *116*, 108999. [[CrossRef](#)]
39. Zhuang, M.; Jiang, H.; Suzuki, Y.; Li, X.; Xiao, P.; Tanaka, T.; Ling, H.; Yang, B.; Saitoh, H.; Zhang, L.; et al. Procyanidins and butanol extract of *Cinnamomi Cortex* inhibit SARS-CoV infection. *Antiviral Res.* **2009**, *82*, 73–81. [[CrossRef](#)]
40. Conzelmann, C.; Weil, T.; Gross, R.; Jungke, P.; Frank, R.; Eggers, M.; Mueller, J.A.; Muench, J. Antiviral activity of plant juices and green tea against SARS-CoV-2 and influenza virus in vitro. *bioRxiv* **2020**. [[CrossRef](#)]

41. Sidor, A.; Gramza-Michałowska, A. Black Chokeberry *Aronia Melanocarpa* L.—A Qualitative Composition, Phenolic Profile and Antioxidant Potential. *Molecules* **2019**, *24*, 3710. [CrossRef]
42. Derksen, A.; Kühn, J.; Hafezi, W.; Sendker, J.; Ehrhardt, C.; Ludwig, S.; Hensel, A. Antiviral activity of hydroalcoholic extract from *Eupatorium perfoliatum* L. Against the attachment of influenza A virus. *J. Ethnopharmacol.* **2016**, *188*, 144–152. [CrossRef]
43. Kim, S.J.; Lee, J.W.; Eun, Y.G.; Lee, K.H.; Yeo, S.G.; Kim, S.W. Pretreatment with a grape seed proanthocyanidin extract downregulates proinflammatory cytokine expression in airway epithelial cells infected with respiratory syncytial virus. *Mol. Med. Rep.* **2019**, *19*, 3330–3336. [CrossRef]
44. Derksen, A.; Hensel, A.; Hafezi, W.; Herrmann, F.; Schmidt, T.J.; Ehrhardt, C.; Ludwig, S.; Kühn, J. 3-O-galloylated procyanidins from *Rumex acetosa* L. inhibit the attachment of influenza A virus. *PLoS ONE* **2014**, *9*, e110089. [CrossRef]
45. Gescher, K.; Hensel, A.; Hafezi, W.; Derksen, A.; Kühn, J. Oligomeric proanthocyanidins from *Rumex acetosa* L. inhibit the attachment of herpes simplex virus type-1. *Antiviral Res.* **2011**, *89*, 9–18. [CrossRef]
46. Tsukuda, S.; Watashi, K.; Hojima, T.; Isogawa, M.; Iwamoto, M.; Omagari, K.; Suzuki, R.; Aizaki, H.; Kojima, S.; Sugiyama, M.; et al. A new class of hepatitis B and D virus entry inhibitors, proanthocyanidin and its analogs, that directly act on the viral large surface proteins. *Hepatology* **2017**, *65*, 1104–1116. [CrossRef]
47. Hensel, A.; Bauer, R.; Heinrich, M.; Spiegler, V.; Kayser, O.; Hempel, G.; Kraft, K. Challenges at the Time of COVID-19: Opportunities and Innovations in Antivirals from Nature. *Planta Med.* **2020**, *86*, 659–664. [CrossRef]
48. Zhang, Y.; Chen, Y.; Meng, Z. Immunomodulation for Severe COVID-19 Pneumonia: The State of the Art. *Front. Immunol.* **2020**, *11*, 577442. [CrossRef]
49. Huang, Y.-F.; Bai, C.; He, F.; Xie, Y.; Zhou, H. Review on the potential action mechanisms of Chinese medicines in treating Coronavirus Disease 2019 (COVID-19). *Pharmacol. Res.* **2020**, *158*, 104939. [CrossRef]
50. Jones, G.; Willett, P.; Glen, R.C.; Leach, A.R.; Taylor, R. Development and validation of a genetic algorithm for flexible docking 1 Edited by F. E. Cohen. *J. Mol. Biol.* **1997**, *267*, 727–748. [CrossRef]
51. 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. [CrossRef]
52. Jin, Z.; Du, X.; Xu, Y.; Deng, Y.; Liu, M.; Zhao, Y.; Zhang, B.; Li, X.; Zhang, L.; Peng, C.; et al. Structure of M^{pro} from SARS-CoV-2 and discovery of its inhibitors. *Nature* **2020**, *582*, 289–293. [CrossRef]
53. Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B.A.; Thiessen, P.A.; Yu, B.; et al. PubChem in 2021: New data content and improved web interfaces. *Nucleic Acids Res.* **2019**, *47*, D1388–D1395. [CrossRef]
54. Dassault Systèmes BIOVIA, Discovery Studio Modeling Environment, Release 2017, San Diego- Dassault Systèmes. Available online: <https://discover.3ds.com/discovery-studio-visualizer-download> (accessed on 15 January 2020).
55. The PyMOL Molecular Graphics System. Available online: [Citeulike-article-id:240061%5Chttp://www.pymol.org](http://www.pymol.org) (accessed on 31 July 2020).
56. Zmudzinski, M.; Rut, W.; Olech, K.; Granda, J.; Giurg, M.; Burda-Grabowska, M.; Zhang, L.; Sun, X.; Lv, Z.; Nayak, D.; et al. Ebselen derivatives are very potent dual inhibitors of SARS-CoV-2 proteases—PL^{pro} and M^{pro} in vitro studies. *bioRxiv* **2020**. [CrossRef]

Curriculum Vitae

My curriculum vitae does not appear in the electronic version of my paper for reasons of data protection.

List of all publications:

1. Nawrot-Hadzik I., Granica S., **Abel R.**, Czapor-Irزابek H. & Matkowski A.: *Analysis of Antioxidant Polyphenols in Loquat Leaves using HPLC-based Activity Profiling*. Nat Prod Commun. 2017;12(2):163-166.
IF: 0.809
2. Zielińska S., Dąbrowska M., Kozłowska W., Kalemba D., **Abel R.**, Dryś A., Szumny A. & Matkowski A.: *Ontogenetic and trans-generational variation of essential oil composition in Agastache rugosa*. Ind Crop Prod. 2017; Volume 97(March 2017):612-619.
IF: 3.849
3. **Abel R.**, Paredes Ramos M., Chen Q., Pérez-Sánchez H., Coluzzi F., Rocco M., Marchetti P., Mura C., Simmaco M., Bourne P.E., Preissner R. & Banerjee P.: *Computational Prediction of Potential Inhibitors of the Main Protease of SARS-CoV-2*. Front Chem. 2020;8:590263.
IF: 5.221
4. Ślusarczyk S., Senol Deniz F.S., **Abel R.**, Pecio Ł., Pérez-Sánchez H., Cerón-Carrasco J.P., den-Haan H., Banerjee P., Preissner R., Krzyżak E., Oleszek W., E Orhan I. & Matkowski A.: *Norditerpenoids with Selective Anti-Cholinesterase Activity from the Roots of Perovskia atriplicifolia Benth*. Int J Mol Sci. 2020;21(12)
IF: 5.923
5. Nawrot-Hadzik I., Choromańska A., **Abel R.**, Preissner R., Saczko J., Matkowski A. & Hadzik J.: *Cytotoxic Effect of Vanicosides A and B from Reynoutria sachalinensis Against Melanotic and Amelanotic Melanoma Cell Lines and in silico Evaluation for Inhibition of BRAFV600E and MEK1*. Int J Mol Sci. 2020;21(13)
IF: 5.923
6. Erukainure O.L., Atolani O., Banerjee P., **Abel R.**, Pooe O.J., Adeyemi O.S., Preissner R., Chukwuma C.I., Koorbanally N.A. & Islam M.S.: *Oxidative testicular injury: effect of L-leucine on redox, cholinergic and purinergic dysfunctions, and dysregulated metabolic pathways*. Amino Acids. 2021;53(3):359-380.
IF: 3.52
7. Jakowiecki J., **Abel R.**, Orzeł U., Pasznik P., Preissner R. & Filipek S.: *Allosteric Modulation of the CB1 Cannabinoid Receptor by Cannabidiol-A Molecular Modeling Study of the N-Terminal Domain and the Allosteric-Orthosteric Coupling*. Molecules. 2021;26(9)
IF: 4.411

8. Nawrot-Hadzik, I., Zmudzinski, M., Matkowski, A., Preissner, R., Kęsik-Brodacka, M., Hadzik, J., Drag, M., **Abel, R.**: *Reynoutria Rhizomes as a Natural Source of SARS-CoV-2 Mpro Inhibitors-Molecular Docking and In Vitro Study*. *Pharmaceuticals*. 2021;14(8).

IF: 5.863

9. Chen, Q., Springer, L., Gohlke, B.O., Goede, A., Dunkel, M., **Abel, R.**, Gallo, K., Preissner, S., Eckert, A., Seshadri, L., Preissner, R.: *SuperTCM: A biocultural database combining biological pathways and historical linguistic data of Chinese Materia Medica for drug development*. *Biomed Pharmacother*. 2021;144:112315.

IF: 6.529

10. Gallo, K., Goede, A., Mura, C., **Abel, R.**, Moahamed, B., Preissner, S., Nahles, S., Heiland, M., Bourne, P.E., Preissner, R., Mallach, M.: *A Comparative Analysis of COVID-19 Vaccines Based on over 580,000 Cases from the Vaccination Adverse Event Reporting System*. *Vaccines* 2022, 10, 408

IF: 4.422

11. Antonyova, V., Kejik, Z., Brogyanyi, T., Kaplanek, R., Vesela, K., Abramenko, N., Ocelka, T., Masarik, M., Matkowski, A., Gburek, J., **Abel, R.**, Goede, A., Preissner, R., Novotný, P., Jakubek, M.: *Non-psychotropic cannabinoids as inhibitors of TET1 protein*. *Bioorg.Chem*. 2022 Vol.124 art.105793

IF: 5.275

12. Szlasa, W., Ślusarczyk, S., Nawrot-Hadzik, I., **Abel, R.**, Zalesińska, A., Szewczyk, A., Sauer, N., Preissner, R., Saczko, J., Drag, M., Poręba, M., Daczewska, M., Kulbacka, J., Drag-Zalesińska, M.: *Betulin and Its Derivatives Reduce Inflammation and COX-2 Activity in Macrophages*. *Inflammation* (2022)

IF: 4.657

Total IF for 12 publications: 56.402

Acknowledgments

Firstly, I would like to thank PD Dr. Robert Preissner for the opportunity to join the research group at Charité – Universitätsmedizin Berlin, and for the supervision of my doctoral project. I would also like to thank Dr. Mathias Dunkel for the supervision and help with the German translation of my dissertation's abstract.

I am grateful to my colleagues of the Structural Bioinformatics group, especially to Dr. Priyanka Banerjee, for the valuable scientific discussions and suggestions. I am also grateful to Prof. Adam Matkowski, Dr. Izabela Nawrot-Hadzik and my colleagues from Wrocław Medical University in Poland, as well as to all the other collaborators and co-authors of my publications, for the fruitful cooperation.

Furthermore, I would like to thank my family, especially my parents, for their support and encouragement, as well as my friends from Berlin and Wrocław. Special appreciation to John for his support and help with proofreading.