Inhibitors for Novel Coronavirus Protease Identified by Virtual Screening of 687 Million Compounds

André Fischer, Manuel Sellner, Santhosh Neranjan, Markus A. Lill,* and Martin Smieško*

Computational Pharmacy, Departement of Pharmaceutical Sciences, University of Basel

E-mail: markus.lill@unibas.ch; martin.smiesko@unibas.ch

Abstract

The rapid outbreak of the novel severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) in China followed by its spread around the world poses a serious global concern for public health with almost 90000 people infected and thousands of fatalities. To conquer viral infections, the inhibition of proteases essential for proteolytic processing of viral polyproteins is a conventional therapeutic strategy. To this date, no specific drugs or vaccines are available to treat SARS-CoV-2 despite its close relation to the SARS-CoV-1 virus that caused a similar epidemic in 2003. Thus, there remains an urgent need for the development of specific antiviral therapeutics to conquer SARS-CoV-2. In order to find novel inhibitors, we computationally screened a compound library of over 687 million compounds for binding at the recently solved crystal structure of the main protease of SARS-CoV-2. A screening of such a vast chemical space for SARS-CoV-2 protease inhibitors has not been reported before. After shape screening, two docking protocols were applied followed by the determination of pharmacokinetically relevant molecular descriptors to narrow down the initial hits. Next, molecular

dynamics simulations were conducted to validate the stability of docked binding modes and comprehensively quantify ligand binding energies. After evaluation of off-target binding, we report a list of 11 drug-like compounds with improved binding free energy to the target protease in contrast to the cocrystallized peptidomimetic lead compound that suffers from poor pharmacokinetic properties. Furthermore, we identified one potent binder with comparable properties from the natural compound library.

Introduction

In late 2019, a novel coronavirus, termed severe acute respiratory syndrome coronavirus 2 (SARS-CoV-2) was determined as a cause for several cases of respiratory disease in China. Even though most infected patients only suffer from mild symptoms such as fever and cough associated with a good prognosis, the disease can progress into fatal cases of pneumonia and acute respiratory failure, especially in older males with comorbidities.[1,2] The virus rapidly spread from China to over 55 countries worldwide and, up to this date, infected almost 90000 people claiming more than 2800 fatalities[3] (as of March 1, 2020), already exceeding the 2003 SARS-CoV-1 epidemic.[4] Most probably, the virus originated from a zoonotic transmission between animals such as bats and humans but progressed to transmit from human to human through common droplet infection.[2,5,6] The World Health Organisation declared the SARS-CoV-2 outbreak a Public Health Emergency of International Concern[7,8] and, based on the 2003 epidemic of SARS-CoV-1, it will lead to the loss of many lives exerting enormous social impact and economic loss in the magnitude of billions of dollars.[9] Currently, there are no specific therapeutic options available for the infection meaning that clinical treatment of the disease is mainly symptomatic combined with repurposing of already marketed antiviral drugs such as ritonavir and antibiotics to treat secondary infections.[1] Thus, there remains an urgent need for the development of specific antiviral therapeutics and vaccines against SARS-CoV-2.[4,10–12]

The inhibition of viral proteases necessary for proteolytic processing of polyproteins has

been a successful strategy in the pharmacological treatment of human immunodeficiency virus (HIV) and hepatitis C respectively,[13,14] proving the potential of protease inhibitors for the treatment of viral infections. Similarly, the main protease of SARS-CoV-2 is thought to be essential for viral replication and, therefore, is regarded as promising target for antiviral pharmacotherapy.[10,15] The crystal structure of the SARS-CoV-2 main protease was recently solved[16] enabling the rational design of specific inhibitory compounds. The close relationship of SARS-CoV-2 to SARS-CoV-1 is reflected by the high sequence identity of 96.1% among their proteases.[10,17] In this regard, it was suggested that drugs developed against SARS-CoV-1 might be effective to treat SARS-CoV-2.[12] However, these compounds remained in the preclinical or early clinical stage not leading to an approved medicine[18] and their effectiveness for the novel virus might suffer due differences in single amino acids.[19] Accordingly, the development of specific inhibitors for SARS-CoV-2 main protease remains an urgent necessity in the scientific community[10–12] which is reflected by the multiple projects focusing on this protein. For example, three studies investigating the repurposing of marketed drugs proposed several candidates for SARS-CoV-2 treatment.[10,15,20] In another recent study, a deep learning approach combined with a homology model of the protease was applied to screen a library of approximately one million compounds including already approved drugs, tripeptides, and natural products.[12] However, the disadvantage of applying therapeutics originally designed for a different target is the risk of undesired pharmacological effects and adverse reactions.[21]

In this study, we screened a large library of over 687 million compounds with the ambition to find novel inhibitors for the SARS-CoV-2 main protease. In a first step, we performed a shape-based screening with known binders for the SARS-CoV-1 main protease and relevant substructures as template molecules. After the initial shape screening round, two different docking protocols were applied followed by the assessment of molecular descriptors relevant for pharmacokinetics to narrow down the number of binders. Clustering based on molecular fingerprints was applied to ensure structural diversity of the compounds that were, in the next step, subjected to molecular dynamics (MD) simulations. Based on the obtained trajectories,

the binding free energy of the ligands was quantified using Molecular Mechanics/Generalized Born Surface Area (MM/GBSA) post-processing. In the last step, we assessed potential toxic effects of the compounds due to the interaction with 16 known drug off-targets to make a final selection of 11 compounds. Such a comprehensive exploration of chemical space intending to discover SARS-CoV-2 protease inhibitors was, to the best of our knowledge, not previously reported.

Results and discussion

Comparison of proteases between SARS-CoV-2 and SARS-CoV-1

The SARS-CoV-2 protease consists of three domains (Figure 1A)[18] and processes polyproteins using a catalytic dyad consisting of histidine and cysteine as catalytic residues. Its active site is located between domains I and II. As SARS-CoV-2 is closely related to SARS-CoV-1, their proteases display a high degree of sequence similarity (96.1%; Figure 1D).[10,17] In the vicinity of the active site, only a single amino acid (Ser46) is different among the two proteases. However, the surface topology of the active site among the two proteins presents distinct differences, especially in the vicinity of the loop centered around Asn142 (Figures 1B and 1C). Additionally, the size and depth of the S1' pocket shows notable differences and, at the center of the S1, S1', and S3, the SARS-CoV-2 protease presents a more distinct subcavity as opposed to the SARS-CoV-1 enzyme. Consequently, inhibitors of the SARS protease might present altered binding affinities for the SARS-CoV-2 protease. Similarities to other viral proteases such as the one of HIV or Middle-East Respiratory Syndrome coronavirus (MERS-CoV) are comparatively low.[22] Although the main protease presumably is the most promising therapeutic target to attenuate viral replication, the inhibition of other functional proteins such as the papain-like protease or the interaction between the viral spike protein and its entry receptor to human cells were considered as well.[8,22]
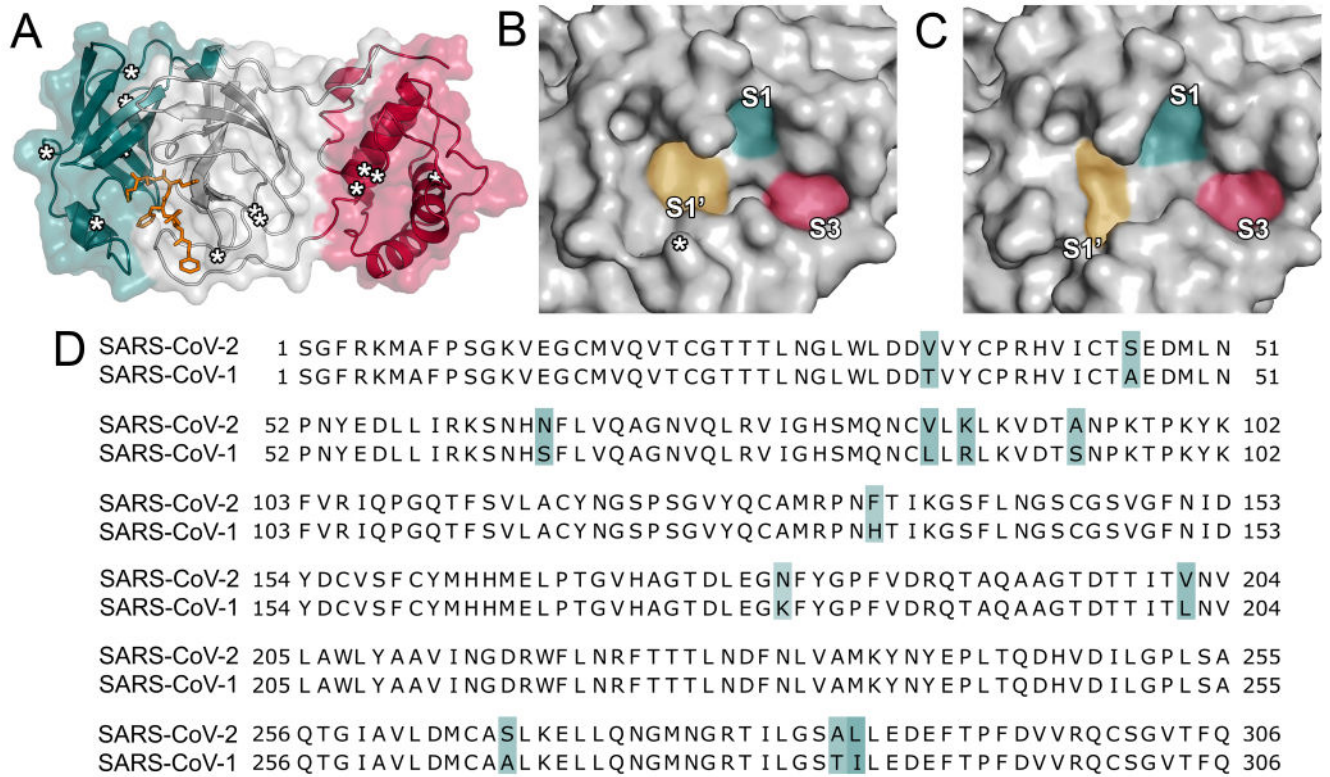
Figure 1: Structural overview and sequence alignment. (A) The three domains (domain I in blue, domain II in grey, domain III in red) of the main protease of SARS-CoV-2 are shown. Amino acid changes between SARS-CoV-1 and SARS-CoV-2 are indicated by asterisks. The cocrystallized ligand (PDB ID 6LU7) is presented in organe. (B) Surface topology of the binding pocket of the SARS-CoV-2 main protease (PDB ID 6LU7). The location of Ser46 is indicated by an asterisk. (C) Surface topology of the binding pocket of the SARS-CoV-1 main protease (PDB ID 2A5I). (D) Sequence alignment of the proteases of SARS-CoV-1 and SARS-CoV-2. Mismatches are marked in blue.

Virtual screening procedures

Virtual Screening is a widely used technique at the early stage of drug discovery that allows to identify potentially bioactive compounds at a high throughput.[23] Due to its speed and cost-effectiveness, using it to identify drug candidates against the globally expanding SARS-CoV-2 offers a promising approach, especially when time is of the essence. The inhibition of proteases as treatment against viral infections has been proven for HIV, rendering the SARS-CoV-2 main protease an attractive drug target in particular since its crystal structure has been recently solved. In our virtual screening workflow (Figure 2), we started with 687 million compounds from the three-dimensional ZINC database in a shape-based screening routine. Such a coarse GPU-accelerated protocol allows for the rapid screening of large databases based on known binders as template. Previous virtual screening efforts were limited to significantly smaller compound libraries.[10,12,15,17,20]
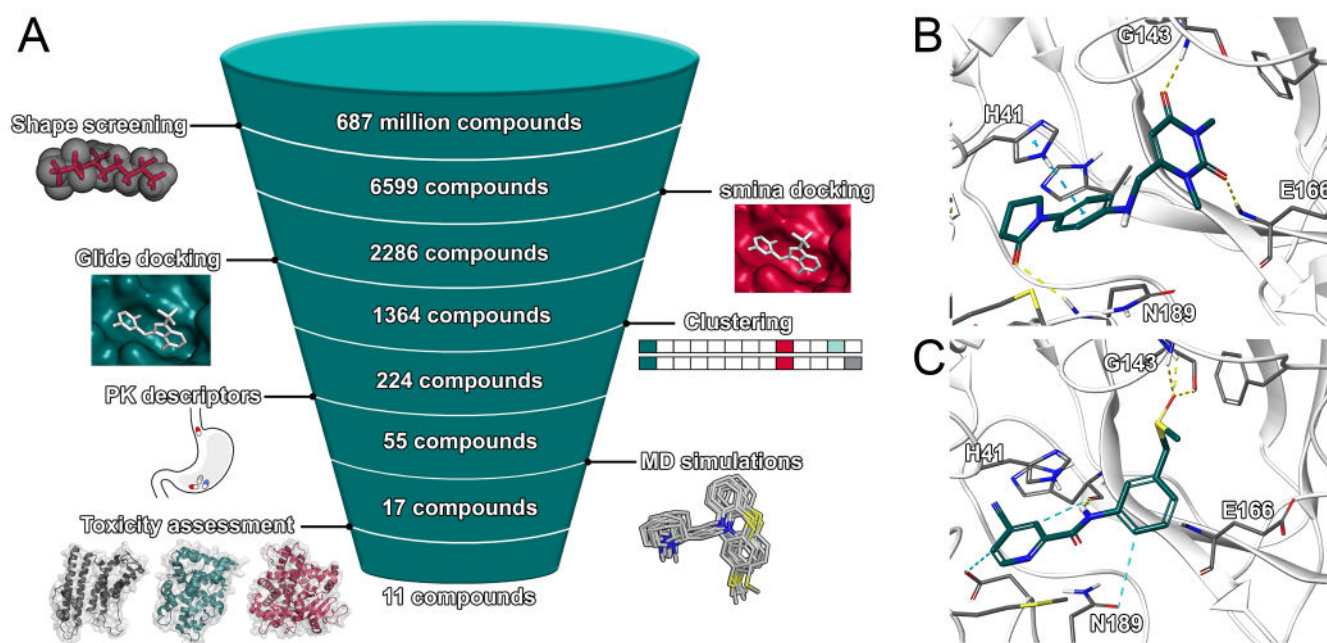

Figure 2: Virtual screening workflow and binding poses of top two ligands. (A) Virtual screening workflow. (B) Binding pose of CP-1. (C) Binding pose of CP-2.

We refined the high number of initial hits (395132 compounds) to 6599 compounds by selecting the best hits regarding shape overlap. The remaining compounds were docked

into the active site of five representative structures of the protease using the smina docking protocol resulting in 2286 hits with a score below a threshold of -7.0 kcal/mol with the majority of compounds scored below -5.0 kcal/mol. While most previous studies regarding SARS-CoV-2 were limited to a homology model of the protease,[10,12,15,17] we could rely on a recently solved crystal structure for our docking procedures. The pharmacokinetic descriptors of these compounds are shown in Figure 3. Due to the origin of the selected compounds in the ZINC database, drug-likeness regarding molecular weight (MW) was to be expected. However, several structures violate commonly accepted criteria distribution coefficient logD and polar surface area (PSA), allowing for their delimitation. Due to the catalytic function of the target enzyme, peptides and peptidomimetics are widely applied in targeting proteases.[18] Nonetheless, disadvantages of peptides or peptidomimetics include limited oral bioavailability due to their large MW, PSA, and high number of rotatable bonds as well as poor metabolic stability and higher production cost.[24] Therefore, the development of small-molecules with balanced and favorable pharmacokinetic properties facilitating oral absorption offers a promising alternative.
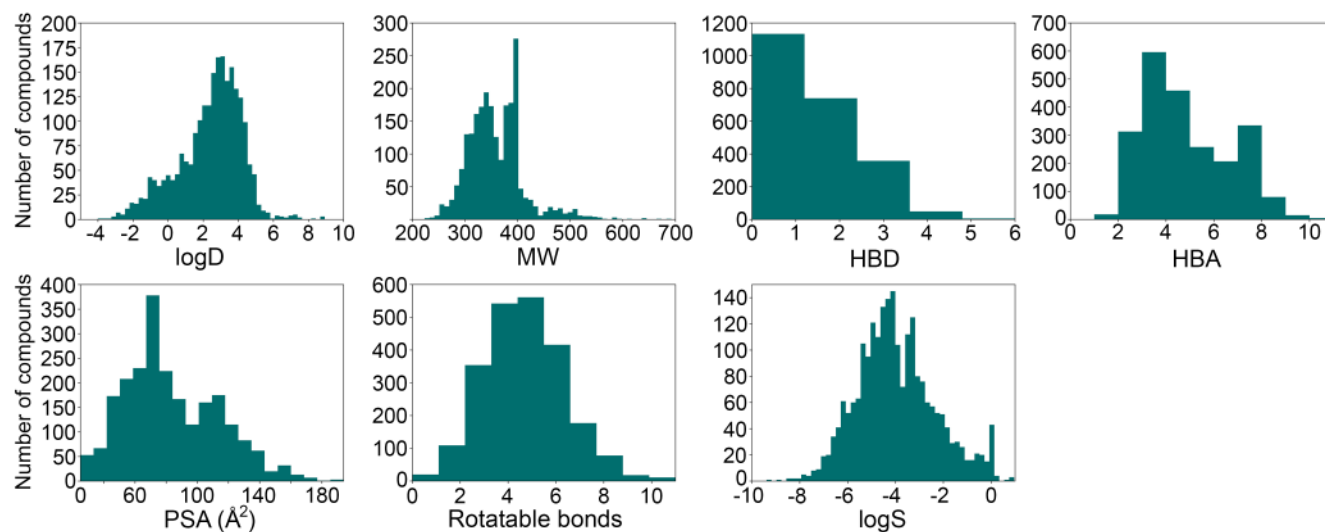

Figure 3: Pharmacokinetically relevant descriptors of the compounds that were subjected to the Glide SP docking protocol.

In the attempt to attain a consensus in binding affinity, we used the Glide SP protocol

to evaluate the interaction of the remaining 2286 compounds with the protein active site. Except for two compounds, a valid binding pose was detected in all cases with most compounds scores below -5.0 kcal/mol. The set of compounds with a Glide score below -6.5 kcal/mol presented a high number of congeneric ligands including many structurally similar natural compounds. Therefore, in order to stimulate the structural diversity of the compound set, we computed extended connectivity fingerprints and compared them according to the Tanimoto coefficient for subsequent clustering. For each cluster, the two compounds with the best Glide score were selected and evaluated regarding their pharmacokinetic descriptors. We did not allow the violation of either the Lipinski and Veber criteria resulting in 55 compounds left over for MD simulation. Based on the resulting time-evolved ensemble of ligand-protein complexes, the binding free energies of the ligands were estimated using the MM/GBSA protocol. This protocol was recently applied to predict the interaction energy between nelfinavir and the SARS-CoV-2 main protease.[10] Compounds with a binding free energy improvement in contrast to the cocrystallized ligand were considered regarding their potential toxicity. The VirtualToxLab evaluates the toxic potential of a small molecule on the basis of individual binding affinities to 16 validated off-targets including nuclear receptors, metabolic enzymes, and the human Ether-à-go-go-Related Gene channel (hERG). Such an assessment at an early stage of drug discovery might mitigate the attrition rate of drugs due to toxicity and safety which represent a large share of preclinical and clinical failures of drug development programs.[25] The remaining 17 compounds included several compounds with a toxic potential above 0.5 that could be discarded from the final set. In the ligands subjected to MD simulation, we further selected the natural compound with the lowest binding free energy which was (-)-taxifolin.

Final compound selection

We identified 11 compounds with a lower free energy of binding combined with a higher theoretical potential of absorption after oral administration compared to the cocrystallized
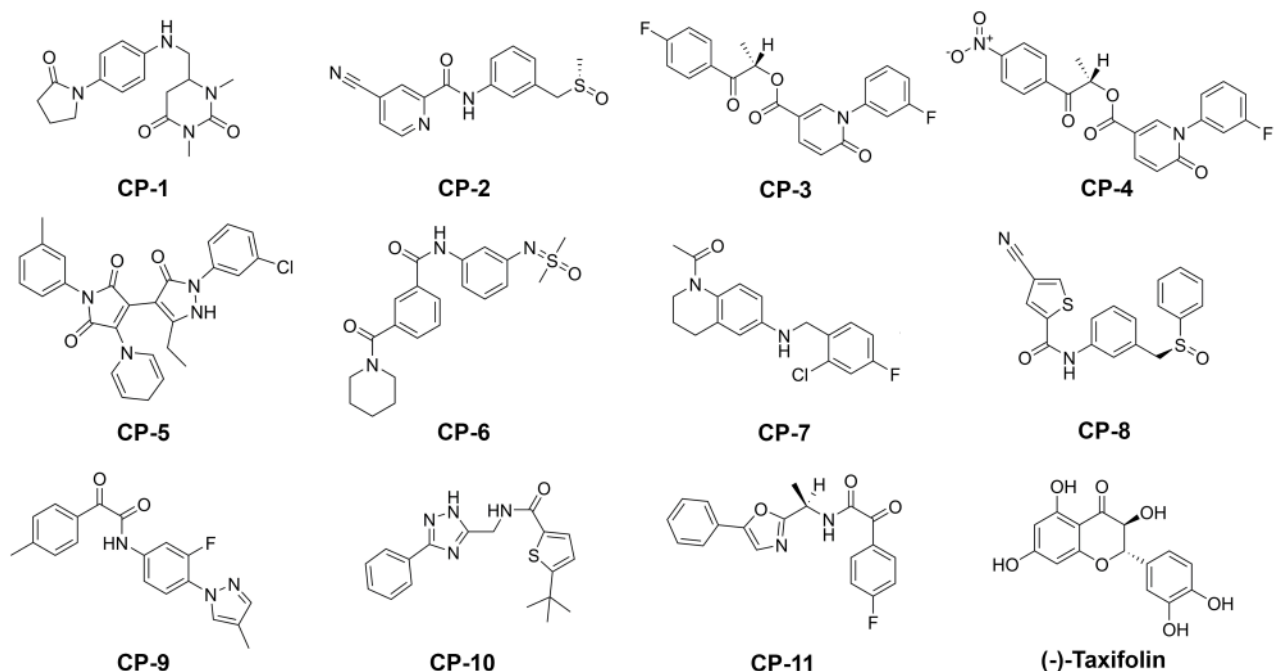
Figure 4: Structures of our final selection of compounds and the highest scored natural compound.

ligand N3 (Table 1 and Figure 4). The reported set of ligands should be regarded as early lead compounds since no experimentally supported optimization was conducted. While some compounds only offered a minor decrease regarding binding energy, CP-1 to CP-4 presented a substantial improvement in comparison to the cocrystallized ligand up to nearly 30%. A study that primarily applied artificial intelligence in the form of a neural network to discovery SARS-CoV-2 inhibitors additionally tested their compound using AutoDock Vina. Their compounds present docking scores ranging from -7.3 to -8.5 kcal/mol,[12] which is comparable to our selection of ligands (Table S1, Supporting Information). Our proposed antivirals interacted with the target with at least one hydrogen bond with an average of over two (Figure 2A and 2B, Table S2). Further, they showed excellent pharmacokinetic descriptors with logD values ranging from 0 to 4.6, MW below 487 g/mol, PSA below 105.5 Å^2 and hydrogen bond acceptors as well as donors in the range described by Lipinski.[26] Consequently, the compounds offer an advantage over commonly described peptidomimetic inhibitors.[18] In contrast to other proposed inhibitors for SARS-CoV-2,[18] the availability of compounds are

9

marked as either for sale, in-stock or on-demand in the ZINC database (Table S1). Regarding potential toxicity through off-target binding, the compounds regularly presented toxic potentials close to 0.5. The top two compounds CP-1 and CP-2 presented comparatively low values and therefore represent excellent candidates for further investigation in vitro. Similarly, the natural compound (-)-taxifolin only exhibited a low potential for binding to the 16 off-targets. Natural compounds such as flavonoids were previously considered to inhibit the main protease of SARS-CoV-1 and their effectiveness was demonstrated by fluorescence resonance energy transfer (FRET).[18,27] Further, in vivo experiments with (-)-taxifolin evidenced activity against other viruses such as coxsackievirus B4.[28] The sibirian larch (Larix sibirica) is known to produce (-)-taxofolin offering a natural resource for its extraction. Furthermore, food preparations containing (-)-taxifolin are readily available from pharmacies and other companies allowing direct and fast access to the potential antiviral. Remarkably, (-)-taxofolin underwent seven hydrogen bonds which represents the highest count in our selection of compounds. Since hydrogen bonds are a key determinant for drug specificity,[29] (-)-taxofolin might offer a naturally occuring alternative to the proposed inhibitors CP-1 to CP-11.

Conclusions

The outbreak of SARS-CoV-2 around the globe and the resulting consequences are an impressive reminder of the menace of zoonotic diseases for public health and the economy of many countries.[30] The lack of specific therapeutics against the novel virus urges for the discovery of new drug compounds for which computational methods offer a fast and cost-efficient approach. Here, we employed a virtual screening workflow consisting of seven individual steps to ultimately determine 11 potential binders. For all reported compounds, we estimated low to moderate to toxicity caused by off-target binding, improved drug-likeness, as well as as enhanced binding to the therapeutic target compared to the cocrystallized ligand. In addition,

Table 1: Final selection of compounds compared to the cocrystallized ligand and the highest scored natural compound.

Compound	ΔG^a (kcal/mol)	Score^b (kcal/mol)	MW^c (g/mol)	logD	logS	PSA^d (Å)	ToxPot^e
CP-1	-70.6 ± 3.9	-14.5	328.4	0	-2.3	73.0	0.352
CP-2	-69.9 ± 4.0	-14.6	299.4	0.6	-2.5	82.9	0.245
CP-3	-66.7 ± 4.7	-15.8	383.4	3.8	-5.4	63.7	0.492
CP-4	-65.3 ± 3.6	-15.2	391.4	3.5	-6.0	105.5	0.494
CP-5	-60.3 ± 4.8	-15.5	487.0	3.5	-5.9	73.0	0.291
CP-6	-58.2 ± 6.0	-15.6	399.5	2.3	-5.2	78.8	0.257
CP-7	-56.3 ± 6.2	-14.6	332.8	3.5	-4.9	32.3	0.482
CP-8	-56.3 ± 5.7	-15.3	366.5	3.4	-5.2	70.0	0.154
CP-9	-56.1 ± 4.6	-14.7	337.4	4.2	-5.1	64.0	0.418
CP-10	-56.0 ± 4.5	-14.9	340.5	4.6	-5.2	70.7	0.499
CP-11	-55.2 ± 3.7	-14.9	338.3	2.9	-4.9	72.2	0.482
N3	-55.0 ± 6.9	-16.0	587.8	-3.1	-0.6	205.3	n/a^f
(-)-taxifolin	-53.3 ± 5.1	-16.8	304.3	1.7	-2.1	127.5	0.289

^{a} Ligand free binding energy predicted by MM/GBSA approach (excluding entropic contributions) with standard deviation; ^{b} Consensus docking score based on sum of highest smina and Glide scores; ^{c} Molecular weight; ^{d} Polar surface area; ^{e} Toxic potential predicted by VirtualToxLab. ^{f} Value not available.

we report a natural compound occurring in the sibirian larch that offers a promising alternative with commercially available plant extracts and derived supplements. To this date, our study marks the most extensive computational screening for the discovery of SARS-CoV-2 main protease inhibitors. Experimental validation and subsequent optimization of our proposed early lead compounds might offer a valuable strategy to conquer SARS-CoV-2.

Materials and Methods

System preparation and ensemble generation

The protein structure of the SARS-CoV-2 main protease was retrieved from the Protein Data Bank (PDB ID 6LU7) and processed using the Protein Preparation Wizard that comes with the Maestro Small-Molecule Drug Discovery Suite (v2019-4)[31] by adding hydrogen atoms, predicting protonation states at physiological pH (pH = 7.4), and reorienting the hydrogen bonding network. Sequence alignment to the protease of SARS-CoV-1 was performed using

the Ugene suite of tools combined to the ClustalW algorithm.[32,33] Next, the system was subjected to a restrained minimization with an root-mean square deviation (RMSD) convergence threshold of 0.3 Å using the OPLS_2005 force field. To obtain an ensemble of structures for following procedures, an MD simulation with a duration of 20 ns was performed using the Desmond (v2019-1) simulation engine.[34] To solvate the orthorhombic periodic boundary systems extending 10 Å from the complex in every dimension, we selected the TIP3P water model. First, the default relaxation protocol of Desmond was executed, followed by the production phase at 310 K in an NPT ensemble. For all production simulations in this study, we applied the Martyna–Tobias–Klein barostat, with a relaxation time of 2.0 ps at 300 K and the Nose–Hoover thermostat, with a relaxation time of 1.0 ps. We used the u-series algorithm to treat long-range interactions with a cutoff of 9 Å for short-range interactions[35] and the M-SHAKE algorithm to constrain bonds to hydrogen atoms. The timestep of the RESPA integrator was left at 2.0 fs and snapshots with atomic coordinates were saved at an interval of 10 ps. Using the resulting simulation trajectory, we used the trj_cluster.py script provided in Maestro, which relies on an affinity propagation algorithm, to determine five representative structures of the protease.

Pharmacokinetic descriptors

Chemical descriptors according to Lipinski and Veber,[26] as well as the aqueous solubility (logS), known to influence the pharmacokinetics of drug candidates, were determined using the cxcalc module of the Marvin (v20.4.0) suite.[36] Instead of the commonly used logP value, we decided on using the more representative logD value to predict the partitioning of the ligands with respect to ionization states.[37] Properties dependent on the ionization of the compound were computed at a physiological pH value of 7.4. The descriptors were computed for all ligands that were supplied to the smina docking protocol.

Docking and shape screening

In a first step, the GPU-accelerated shape screening protocol of Maestro was employed to rapidly compare the drug-like portion of the three-dimensional ZINC[38] library. Additionally, due to the comparatively large size of the cocrystallized peptidomimetic inhibitor, compounds with a molecular weight above 500 g/mol were considered. As templates, a substructure of the cocrystallized ligand of SARS-CoV-2 protease (compound 4), a cocrystallized ligand (compound 9) and three substructures of the equivalent SARS-CoV-1 protease inhibitors (compounds 1, 3, and 7), as well as four known SARS-CoV-1 inhibitors with binding affinity below 10 μM (compounds 2, 5, 6, and 8) derived from the PubChem database[39] were selected (cf. Figure 2). The experimentally derived inhibitors were selected based on their structural diversity assessed with molecular extended connectivity fingerprints followed by a Tanimoto comparison and subsequent clustering in the Discovery Informatics panel within Maestro. For the shape comparison a threshold of 45% shape overlap was selected. The best 5000 hits regarding shape similarity, as well as the top 500 compounds of each input template were selected for further steps.
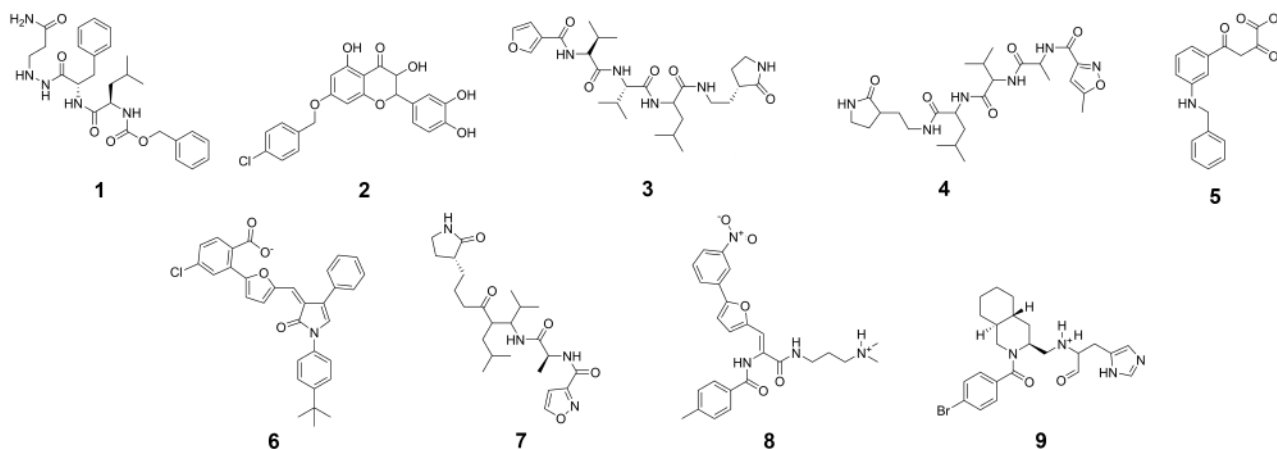

Figure 5: Input structures for shape screening derived from crystal structures and the PubChem database.

Next, we used the smina[40] docking protocol, a fork of AutoDock Vina (v1.1.2),[41] on the selected compounds to quantify their interaction with each of the representatives structures

from the generated ensemble. The ligand PDBQT files for docking were generated using the prepare_ligand4.py script that comes with AutoDockTools (v1.5.4).[42] Docking was performed at an exhaustiveness of 16, the search space was defined using the automatic method with smina, and a random seed of 42 was selected. Hits with a binding free energy below -7.0 kcal/mol were then subjected to the default Glide standard precision (SP) docking protocol in Maestro.[43] Likewise, the interaction with all five ensemble structures was quantified. Compounds with a free binding energy below -6.5 kcal/mol were then subjected to clustering based on the Tanimoto comparison of per-atom molecular fingerprints. For each of the 138 resulting clusters, the best two compounds (one compound in the case of a single structure per cluster) according to the Glide score were selected for further assessment. The commercial availability of the compounds was retrieved from the ZINC website. The final set of compounds that was subjected to MD simulations was selected according to the following criteria: (i.) availability not orphaned, (ii.) no violations of Lipinski and Veber criteria, (ii.) average ensemble score below -6.5 kcal/mol in smina docking, (iv.) average ensemble score below -6.0 kcal/mol in Glide SP docking, and (v.) sum of best docking scores in either protocol below -14.5 kcal/mol. The same docking and ranking procedure was performed with the cocrystallized ligand N3.

MD simulations and post-processing

In order to get insight into the stability and interaction energies of the previously filtered set of ligands, we conducted MD simulations of each compound. As input for the simulations, we chose the ligand-protein complex with the highest Glide SP docking score among the ensemble structures. The duration of the simulations was selected to be 25 ns with atomic coordinates recorded at an interval of 25 ps, while the remaining settings were left as described above.

The stability of the binding modes was quantified using RMSD in the Simulation Interaction Diagram panel in Maestro. The binding free energy was quantified using the ther-

mal_mmgbsa.py routine that comes with Maestro for the last 100 frames of the trajectory with a step size of two. This routine is based on MM/GBSA calculations that essentially compare the apo and holo states of the enzyme with various enthalpic terms. The resulting binding free energies were averaged for the selected frames of the trajectories.

Toxicity assessment and selection of final set of compounds

Potential toxic effects of the compounds that were subjected to MD simulations was assessed using the VirtualToxLab[21,44] that evaluates the binding affinity of a probe compound to 16 known off-targets. The target structures include androgen receptor (AR), aryl hydrocarbon receptor (AhR), cytochrome P450 1A2 (CYP1A2), CYP2C9, CYP2D6, CYP3A4, estrogen receptor α (ERα), ERβ, glucocorticoid receptor (GR), mineralocorticoid receptor (MR), liver X receptor (LXR), hERG channel, peroxisime proliferator-activated receptor γ (PPARγ), progesterone receptor (PR), thyroid receptor α (TRα), and TRβ. The routine quantifies the binding affinities to each target and, based on them, estimates a numerical value of toxicity termed toxic potential ranging from 0.0 to 1.0.

The set of ligands subjected to MD simulations was refined based on their toxic potential and average free binding energy to derive the final set of compounds. Compounds with an MM/GBSA binding energy below -55.0 kcal/mol and a toxic potential below 0.5 were considered.

Funding

Acknowledgement

Supporting Information Available

Comprehensive information on all 55 compounds subjected to MD simulations, as well as information on binding poses, pharmacokinetic properties, and hydrogen bonding for the final selection of compounds are available in a separate Supporting Information PDF document.

References

(1) Chen, N.; Zhou, M.; Dong, X.; Qu, J.; Gong, F.; Han, Y.; Qiu, Y.; Wang, J.; Liu, Y.; Wei, Y.; Xia, J.; Yu, T.; Zhang, X.; Zhang, L. Epidemiological and clinical characteristics of 99 cases of 2019 novel coronavirus pneumonia in Wuhan, China: a descriptive study. *The Lancet* **2020**, *395*, 507–513.

(2) Zhou, P. et al. A pneumonia outbreak associated with a new coronavirus of probable bat origin. *Nature* **2020**,

(3) World Health Organisation, Novel Coronavirus (2019-nCoV) situation reports. `https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/`.

(4) Wang, F.; Chen, C.; Tan, W.; Yang, K.; Yang, H. Structure of Main Protease from Human Coronavirus NL63: Insights for Wide Spectrum Anti-Coronavirus Drug Design. *Scientific Reports* **2016**, *6*, 22677.

(5) Chen, Z.-M. et al. Diagnosis and treatment recommendations for pediatric respiratory

infection caused by the 2019 novel coronavirus. *World journal of pediatrics : WJP* **2020**,

(6) Wu, F. et al. A new coronavirus associated with human respiratory disease in China. *Nature* **2020**,

(7) World Health Organisation, International Health Regulations Emergency Committee on novel coronavirus in China. `https://www.who.int/news-room/events/detail/2020/01/30/default-calendar/international-health-regulations-emergency-committee-on-novel-coronavirus-in-china.`

(8) Stoermer, M. J. Homology Models of the Papain-Like Protease PLpro from Coronavirus 2019-nCoV. **2020**,

(9) Keogh-Brown, M. R.; Smith, R. D. The economic impact of SARS: How does the reality match the predictions? *Health Policy* **2008**, *88*, 110–120.

(10) Zhijian Xu, Y. S. Z. Z. K. M. X. W. W. Z., Cheng Peng Nelfinavir was predicted to be a potential inhibitor of 2019-nCov main protease by an integrative approach combining homology modelling, molecular docking and binding free energy calculation. **2020**, *1201*, 1–20.

(11) Zumla, A.; Hui, D. S.; Azhar, E. I.; Memish, Z. A.; Maeurer, M. Reducing mortality from 2019-nCoV: host-directed therapies should be an option. *The Lancet* **2020**, *395*, e35–e36.

(12) Zhang, H.; Saravanan, K. M.; Yang, Y.; Hossain, T. Deep learning based drug screening for novel coronavirus 2019-nCov. **2020**, *19*, 1–17.

(13) Yang, S. et al. Synthesis, Crystal Structure, Structure-Activity Relationships, and Antiviral Activity of a Potent SARS Coronavirus 3CL Protease Inhibitor. *Journal of Medicinal Chemistry* **2006**, *49*, 4971–4980.

(14) Ghosh, A. K.; Osswald, H. L.; Prato, G. Recent Progress in the Development of HIV-1 Protease Inhibitors for the Treatment of HIV/AIDS. *Journal of Medicinal Chemistry* **2016**, *59*, 5172–5208.

(15) Liu, X.; Wang, X.-J. Potential inhibitors against 2019-nCoV coronavirus M protease from clinically approved medicines. *Journal of Genetics and Genomics* **2020**,

(16) Liu, Z.-B. J. Z. Y. H. R. Z., X. The crystal structure of COVID-19 main protease in complex with an inhibitor N3. `http://www.rcsb.org/structure/6LU7`.

(17) Li, Y.; Zhang, J.; Wang, N.; Li, H.; Shi, Y.; Guo, G.; Liu, K.; Zeng, H.; Zou, Q. Therapeutic Drugs Targeting 2019-nCoV Main Protease by High-Throughput Screening. *bioRxiv* **2020**, 2020.01.28.922922.

(18) Pillaiyar, T.; Manickam, M.; Namasivayam, V.; Hayashi, Y.; Jung, S. H. An overview of severe acute respiratory syndrome-coronavirus (SARS-CoV) 3CL protease inhibitors: Peptidomimetics and small molecule chemotherapy. *Journal of Medicinal Chemistry* **2016**, *59*, 6595–6628.

(19) Raugi, D. N.; Smith, R. A.; Gottlieb, G. S. Four Amino Acid Changes in HIV-2 Protease Confer Class-Wide Sensitivity to Protease Inhibitors. *Journal of Virology* **2016**, *90*, 1062–1069.

(20) Chen, Y. W.; Yiu, C.-p.; Wong, K.-y. Prediction of the 2019-nCoV 3C-like protease (3CLpro) structure_ virtual screening reveals velpatasvir, ledipasvir, and other drug repurposing candidates.pdf.crdownload. **2020**,

(21) Vedani, A.; Dobler, M.; Hu, Z.; Smieško, M. OpenVirtualToxLab-A platform for generating and exchanging in silico toxicity data. *Toxicology Letters* **2015**, *232*, 519–532.

(22) Gao, K.; Nguyen, D. D.; Wang, R.; Wei, G.-W. Machine intelligence design of 2019-nCoV drugs. *bioRxiv* **2020**, 2020.01.30.927889.

(23) Lionta, E.; Spyrou, G.; Vassilatis, D.; Cournia, Z. Structure-Based Virtual Screening for Drug Discovery: Principles, Applications and Recent Advances. *Current Topics in Medicinal Chemistry* **2014**, *14*, 1923–1938.

(24) Recio, C.; Maione, F.; Iqbal, A. J.; Mascolo, N.; De Feo, V. The potential therapeutic application of peptides and peptidomimetics in cardiovascular disease. *Frontiers in Pharmacology* **2017**, *7*, 1–11.

(25) Waring, M. J.; Arrowsmith, J.; Leach, A. R.; Leeson, P. D.; Mandrell, S.; Owen, R. M.; Pairaudeau, G.; Pennie, W. D.; Pickett, S. D.; Wang, J.; Wallace, O.; Weir, A. An analysis of the attrition of drug candidates from four major pharmaceutical companies. *Nature Reviews Drug Discovery* **2015**, *14*, 475–486.

(26) Walters, W. P. Going further than Lipinski's rule in drug design. *Expert Opinion on Drug Discovery* **2012**, *7*, 99–107.

(27) Nguyen, T. T. H.; Woo, H. J.; Kang, H. K.; Nguyen, V. D.; Kim, Y. M.; Kim, D. W.; Ahn, S. A.; Xia, Y.; Kim, D. Flavonoid-mediated inhibition of SARS coronavirus 3C-like protease expressed in Pichia pastoris. *Biotechnology Letters* **2012**, *34*, 831–838.

(28) Galochkina, A. V.; Anikin, V. B.; Babkin, V. A.; Ostrouhova, L. A.; Zarubaev, V. V. Virus-inhibiting activity of dihydroquercetin, a flavonoid from Larix sibirica, against coxsackievirus B4 in a model of viral pancreatitis. *Archives of Virology* **2016**, *161*, 929–938.

(29) Wade, R. C.; Goodford, P. J. The role of hydrogen-bonds in drug binding. *Progress in clinical and biological research* **1989**, *289*, 433–444.

(30) Hui, D. S.; I Azhar, E.; Madani, T. A.; Ntoumi, F.; Kock, R.; Dar, O.; Ippolito, G.; Mchugh, T. D.; Memish, Z. A.; Drosten, C.; Zumla, A.; Petersen, E. The continuing 2019-nCoV epidemic threat of novel coronaviruses to global health — The latest

2019 novel coronavirus outbreak in Wuhan, China. *International Journal of Infectious Diseases* **2020**, *91*, 264–266.

(31) Schrodinger LCC, Maestro Small-Molecular Drug Discovery Suite 2019-4. **2019**,

(32) Okonechnikov, K.; Golosova, O.; Fursov, M.; Varlamov, A.; Vaskin, Y.; Efremov, I.; German Grehov, O. G.; Kandrov, D.; Rasputin, K.; Syabro, M.; Tleukenov, T. Unipro UGENE: A unified bioinformatics toolkit. *Bioinformatics* **2012**, *28*, 1166–1167.

(33) Thompson, J. D.; Higgins, D. G.; Gibson, T. J. CLUSTAL W: Improving the sensitivity of progressive multiple sequence alignment through sequence weighting, position-specific gap penalties and weight matrix choice. *Nucleic Acids Research* **1994**, *22*, 4673–4680.

(34) Bowers, K.; Chow, E.; Xu, H.; Dror, R.; Eastwood, M.; Gregersen, B.; Klepeis, J.; Kolossvary, I.; Moraes, M.; Sacerdoti, F.; Salmon, J.; Shan, Y.; Shaw, D. Scalable Algorithms for Molecular Dynamics Simulations on Commodity Clusters. *ACM/IEEE SC 2006 Conference (SC'06)* **2006**, 43.

(35) Shaw, D. E. et al. Anton 2: Raising the Bar for Performance and Programmability in a Special-Purpose Molecular Dynamics Supercomputer. *International Conference for High Performance Computing, Networking, Storage and Analysis, SC* **2014**, *2015-Janua*, 41–53.

(36) ChemAxon, Marvin (v.20.4.0). 2020.

(37) Bhal, S. K.; Kassam, K.; Peirson, I. G.; Pearl, G. M. The rule of five revisited: Applying log D in place of log P in drug-likeness filters. *Molecular Pharmaceutics* **2007**, *4*, 556–560.

(38) Sterling, T.; Irwin, J. J. ZINC 15 - Ligand Discovery for Everyone. *Journal of Chemical Information and Modeling* **2015**, *55*, 2324–2337.

(39) Kim, S.; Chen, J.; Cheng, T.; Gindulyte, A.; He, J.; He, S.; Li, Q.; Shoemaker, B. A.; Thiessen, P. A.; Yu, B.; Zaslavsky, L.; Zhang, J.; Bolton, E. E. PubChem 2019 update: improved access to chemical data. *Nucleic acids research* **2019**, *47*, D1102–D1109.

(40) Koes, D. R.; Baumgartner, M. P.; Camacho, C. J. Lessons learned in empirical scoring with smina from the CSAR 2011 benchmarking exercise. *Journal of Chemical Information and Modeling* **2013**, *53*, 1893–1904.

(41) Brooks, B. R. et al. AutoDock Vina: Improving the Speed and Accuracy of Docking with a New Scoring Function, Efficient Optimization, and Multithreading. *Journal of computational chemistry* **2009**, *30*, 1545–1614.

(42) Morris, G.; Huey, R. AutoDock4 and AutoDockTools4: Automated docking with selective receptor flexibility. *J Comput Chem.* **2009**, *30*, 2785–2791.

(43) Halgren, T. A.; Murphy, R. B.; Friesner, R. A.; Beard, H. S.; Frye, L. L.; Pollard, W. T.; Banks, J. L. Glide: A New Approach for Rapid, Accurate Docking and Scoring. 2. Enrichment Factors in Database Screening. *Journal of Medicinal Chemistry* **2004**, *47*, 1750–1759.

(44) Vedani, A.; Dobler, M.; Smieško, M. VirtualToxLab - A platform for estimating the toxic potential of drugs, chemicals and natural products. *Toxicology and Applied Pharmacology* **2012**, *261*, 142–153.