Potential covalent drugs targeting the main protease of the SARS-CoV-2 coronavirus

Sen Liu1,2,*, Qiang Zheng1,2 and Zhiying Wang1,2

1National “111” Center for Cellular Regulation and Molecular Pharmaceutics, Key Laboratory of Industrial Fermentation (Ministry of Education), Hubei University of Technology, Wuhan 430068, China, 2Institute of Biomedical and Pharmaceutical Sciences, Hubei Key Laboratory of Industrial Microbiology, Hubei University of Technology, Wuhan, 430068, China.

*To whom correspondence should be addressed.

Abstract

Motivation: Since December 2019, the newly identified coronavirus SARS-CoV-2 has caused a massive health crisis worldwide and resulted in over 70,000 COVID-19 infections so far. Clinical drugs targeting SARS-CoV-2 are urgently needed to decrease the high fatality rate of confirmed COVID-19 patients. Traditional de novo drug discovery needs more than 10 years, so drug repurposing seems the best option currently to find potential drugs for treating COVID-19.

Results: Compared with traditional non-covalent drugs, covalent drugs have attracted escalating attention recent years due to their advantages in potential specificity upon careful design, efficiency, and patient burden. We recently developed a computational protocol named as SCAR for discovering covalent drugs. In this work, we used the SCAR protocol to identify possible covalent drugs (approved or clinically tested) targeting the main protease (3CLpro) of SARS-CoV-2. We identified 11 potential hits, among which at least 6 hits were exclusively enriched by the SCAR protocol. Since the preclinical or clinical information of these identified drugs is already available, they might be ready for being clinically tested in the treatment of COVID-19.

Contact: senliu.ctgu@gmail.com

1 Introduction

Starting December 2019, an outbreak of pneumonia of unknown cause took place in Wuhan, Hubei province of China (Chaolin Huang et al., 2020). In a short time, Chinese authorities rapidly isolated and characterized a novel coronavirus closely related to the SARS-CoV that caused the outbreak of a severe acute respiratory syndrome 18 years ago in China (Zhou et al., 2020). The newly identified coronavirus was initially represented by 2019-nCoV, but formally named as SARS-CoV-2 (severe acute respiratory syndrome coronavirus 2) by the International Committee on Taxonomy of Viruses (ICTV) on February 12th, 2020. Meanwhile, the disease caused by this virus was named as COVID-19 by the World Health Organization (WHO). Although China has adopted unprecedented policies to control the spread of the virus including temporarily “shutting down” the Wuhan City on January 23rd, 2020, SARS-CoV-2 has resulted in over 70,000 COVID-19 patients and more than 2,000 fatalities worldwide by February 20th, 2020. With an estimated case fatality rate of 2%-3% and the growing patient numbers, SARS-CoV-2 poses a serious health threat to the whole world (https://www.who.int/emergencies/diseases/novel-coronavirus-2019/situation-reports/).

Unfortunately, specific clinical treatments for the rapidly escalating international crisis caused by SARS-CoV-2 are very limited, which is one of the reasons of the high mortality rate. Due to the deficiency of targeted drugs, the current clinical treatment of COVID-19 focuses on supportive care and symptom relief. Remdesivir, an experimental drug developed to treat the Middle East respiratory syndrome coronavirus (MERS-CoV), has been reported to be effective in treating several COVID-19 cases, but a systematic clinical trial of this drug is still ongoing in Wuhan, China (ClinicalTrials.gov Identifier: NCT04257656). Similarly, chloroquine, lopinavir/ritonavir, and many other drugs have also been reported to be potentially effective, but supportive clinical data are not available for all (Maxmen, 2020). Therefore, it is of great value to identify more potential drugs targeting SARS-CoV-2 to save COVID-19 patients.
In the traditional process, the development of a drug will need over 15 years from target identification, target validation, hit discovery, lead optimization, and preclinical and clinical trials (Kaitin, 2010). Therefore, it is virtually not possible to develop de novo drugs in the time frame needed to impact the current SARS-CoV-2 crisis. Hence, the most feasible approach is to find potential cures from clinical drugs by drug repurposing (also known as drug repositioning, repurposing, redirecting, or rediscovering). Drug repurposing can rapidly expand target disease indications of an existing drug while saving time and money, since the data for human pharmacokinetics, safety and the preclinical results are already available (Cha et al., 2018). Successful examples of drug repurposing include the use of sildenafil in treating erectile dysfunction and the anti-cancer uses of thalidomide (Pushpakom et al., 2019).

Previously, we described a computer-aided drug discovery protocol named as SCAR (steric-clashes alleviating receptors) (Ai et al., 2016) for the screening of covalent inhibitors of target proteins enlightened by the in silico protein design strategy (Liao et al., 2015). Realizing the potential use of SCAR in discovering both covalent and non-covalent inhibitors, we recently demonstrated that SCAR is also quite efficient in drug repurposing (Y. Zhang et al., 2020). Compared with non-covalent drugs, covalent drugs have the following advantages (Mah et al., 2014): (i) covalent drugs have better biochemical efficiency since they are more competitive than many non-covalent endogenous substrates and cofactors; (ii) covalent drugs cause lower patient burden and delay the emergence of drug resistance due to lower and less frequent dosing; (iii) covalent drugs might have better target specificity by reacting with a non-conserved nucleophilic amino acid with careful designs (Cuesta et al., 2020). Therefore, recent years have witnessed the resurgence of the discovery of covalent drugs. As a result, we set out to use SCAR to identify possible covalent drugs targeting SARS-CoV-2 by drug repurposing in this work.

**Table 1.** The potential drugs that might be repurposed as covalent inhibitors of the SARS-CoV-2 3CLpro from the SCAR strategy. Docking score and conformation rank are listed for the identified pose. Atom distance is the distance between the putative reactive atom of the drug and Cys145-SG. SCAR enriching score is calculated as the top docking score of the drug docked to the wild-type 3CLpro minus the listed docking score.
## Methods

### 2.1 Preparation of the in silico compound library

The structure files of the compounds were downloaded as mol2 files from the ZINC15 database (http://zinc15.docking.org). As described by ZINC15, the 3D conformations were protonated at physiological pH and biologically relevant tautomers were generated for each molecule (Sterling and Irwin, 2015). The “in-trials” catalog (2019-04-22 version) was downloaded, which contains 5811 approved or investigational (clinically tested but not approved) drugs worldwide. MGLTools (version 1.5.6) was used to generate the PDBQT file for docking.

### 2.2 Structure optimization of the protein

The SARS-CoV-2 3CLpro structure was downloaded from PDB (PDB ID: 6LU7). This is a complex structure of the SARS-CoV-2 3CLpro and an inhibitor covalently bonding to Cys145. The structure was energetically minimized in Rosetta (Leaver-Fay et al., 2011) by applying harmonic distance and angle constraints on the bonding atoms in the inhibitor and Cys145. The backbone of the protein was fixed during the minimization. The lowest score model was chosen from 1000 models. For SCAR docking, the Cys145 was mutated to Gly. MGLTools (version 1.5.6) was used to generate the PDBQT file for docking.

### 2.3 In silico docking and analysis

The computational docking process was similar as previously described (Ai et al., 2016; Liao et al., 2015). Briefly, AutoDock Vina (Trott and Olson, 2010) (version 1.1.2) was used to dock the small molecules to the substrate binding pocket of the SARS-CoV-2 3CLpro. The docked poses were then manually evaluated by docking score, ranking, and the distance between the reactive atom and the sulfur atom of Cys145 in the original structure as previously described (Ai et al., 2016).

### 3 Results

In SARS-CoV, the 3C-like proteinase (3CLpro) is the main protease, which cleaves the large replicase polyprotein 1a (pp1a) and pp1ab to produce non-structural proteins (NSPs) for the transcription and replication of the virus (Zumla et al., 2016). Therefore, 3CLpro is a key drug target in inhibiting SARS-CoV. The recent data showed that the

| ZINC ID | Atom | Docking Score | Pose rank | Warhead | Drug name | CAS Number | Drugbank ID | Approved or Investigational treatment | SCAR enriching score |
|---------|------|---------------|-----------|---------|-----------|------------|------------|--------------------------------------|---------------------|
| ZINC0001 | 2.0 | -9.0 | 7 | -CN (Furber et al., 2014) | Itacitinib | 1334298-90-6 | DB12154 | Melanoma, endometrial cancer, B-cell malignancies, etc. | -0.6 |
| 18795962 | 2.5 | -8.9 | 1 | -CN | Oberadilol | 114856-44-9 | Not available | Heart failure; hypertension | 0.5 |
| ZINC0000 | 1.2 | -8.8 | 2 | -Ph-F (Shannon et al., 2014) | Telcagepant | 781649-09-0 | DB12228 | Migraine | 0.1 |
| 28827350 | 1.6 | -8.7 | 1 | -Ph-C1 (Shannon et al., 2014) | Vidaliprant | 1169483-24-2 | DB12272 | Asthma | 0.2 |
| ZINC0001 | 1.3 | -8.5 | 6 | -Ph-Cl | Pilaralisib | 934526-89-3 | DB11772 | Cancer | -0.4 |
| 00472223 | 1.3 | -8.4 | 1 | -Ph-F | Poziotinib | 1092364-38-9 | DB12114 | Breast cancer; adenocarcinoma of lung | 0.8 |
| 95930125 | 1.2 | -8.2 | 4 | -Ph-F | Fostamatinib | 901119-35-5 | DB12010 | Rheumatoid arthritis; Immune | 0.1 |
| ZINC0000 | 1.2 | -8.1 | 9 | -CN | CL-275838 | 115931-65-2 | Not available | Cognition enhancer | -0.9 |
| 22442861 | 1.1 | -8.0 | 1 | -Ph-Cl | Ziprasidone | 146939-27-7 | DB00246 | Schizophrenia; bipolar disorder | 0.0 |
| 00538550 | 2.2 | -8.0 | 1 | -C=O (L. Zhang et al., 2016) | Leucal/Folinic acid | 58-05-9 | DB00650 | Toxic effects of methotrexate and pyrimethamine | 0.6 |
| ZINC0000 | 2.7 | -8.0 | 1 | -N-CO-CO- (Barrett et al., 2004) | ITX5061 | 1252679-52-9 | Not available | Rheumatoid arthritis; hepatitis C | -0.1 |
| 58540931 | 0.6 | -8.1 | 9 | -CN | CL-275838 | 115931-65-2 | Not available | Cognition enhancer | -0.9 |

2 Methods

2.1 Preparation of the in silico compound library

The structure files of the compounds were downloaded as mol2 files from the ZINC15 database (http://zinc15.docking.org). As described by ZINC15, the 3D conformations were protonated at physiological pH and biologically relevant tautomers were generated for each molecule (Sterling and Irwin, 2015). The “in-trials” catalog (2019-04-22 version) was downloaded, which contains 5811 approved or investigational (clinically tested but not approved) drugs worldwide. MGLTools (version 1.5.6) was used to generate the PDBQT file for docking.

2.2 Structure optimization of the protein

The SARS-CoV-2 3CLpro structure was downloaded from PDB (PDB ID: 6LU7). This is a complex structure of the SARS-CoV-2 3CLpro and an inhibitor covalently bonding to Cys145. The structure was energetically minimized in Rosetta (Leaver-Fay et al., 2011) by applying harmonic distance and angle constraints on the bonding atoms in the inhibitor and Cys145. The backbone of the protein was fixed during the minimization. The lowest score model was chosen from 1000 models. For SCAR docking, the Cys145 was mutated to Gly. MGLTools (version 1.5.6) was used to generate the PDBQT file for docking.

2.3 In silico docking and analysis

The computational docking process was similar as previously described (Ai et al., 2016; Liao et al., 2015). Briefly, AutoDock Vina (Trott and Olson, 2010) (version 1.1.2) was used to dock the small molecules to the substrate binding pocket of the SARS-CoV-2 3CLpro. The docked poses were then manually evaluated by docking score, ranking, and the distance between the reactive atom and the sulfur atom of Cys145 in the original structure as previously described (Ai et al., 2016).

3 Results

In SARS-CoV, the 3C-like proteinase (3CLpro) is the main protease, which cleaves the large replicase polyprotein 1a (pp1a) and pp1ab to produce non-structural proteins (NSPs) for the transcription and replication of the virus (Zumla et al., 2016). Therefore, 3CLpro is a key drug target in inhibiting SARS-CoV. The recent data showed that the
SARS-CoV-2 3CLpro is highly similar with the SARS-CoV 3CLpro both in sequence and structure (PDB ID: 6LU7). Previous studies demonstrated that Cys145 is a key residue in the active site of 3CLpro (Changkang Huang et al., 2004), which makes this residue be an attractive target for covalent binding of covalent 3CLpro inhibitors. Cysteine is also a most popular target of covalent inhibitors because of its high intrinsic reactivity at physiological pH permits the use of relatively unreactive electrophiles (Cuesta et al., 2020). Meanwhile, targeting cysteine renders the high selectivity of covalent inhibitors due to the low prevalence of cysteine in the proteome (Cuesta et al., 2020).

To repurpose potential covalent drugs targeting the SARS-CoV-2 3CLpro, we firstly optimized the X-ray complex structure of this protein with its inhibitor using the macromolecule modeling suite Rosetta (Leaver-Fay et al., 2011) (Figure 1A). As demonstrated in our previous work, this structural optimization could benefit the docking result (Ai et al., 2016). Next, we mutated Cys145 to Gly as required by the SCAR protocol to prepare the docking target (Ai et al., 2016). The “in-trials” dataset (10083 compounds/conformations) obtained from ZINC15 was filtered by the generally used warhead groups targeting cysteine (5010 compounds/conformations) before it was used in the docking process (represented as “SCAR-dock”), among which 1253 were left for manual checking after score filtering (±8.0).

As listed in Table 1, we identified 11 potential covalent inhibitors of the 3CLpro of SARS-CoV-2 following the SCAR protocol described previously (Ai et al., 2016). These hits contain five different covalent warhead groups suitable for targeting cysteine, which represents diverse structural options. As shown in Figure 1B, the putative covalent poses of the identified hits fit in the binding pocket reasonably. In each drug, the distance between the putative nucleophilic atom and the SG atom of Cys145 is shown in Table 1. In 6LU7 (PDB ID), the distance between the covalent carbon atom of the inhibitor and the SG atom of Cys145 is 1.8 Å. Therefore, the proximity between these two atoms indicates high potential for covalent bonding between the drug and Cys145.

To investigate how the SCAR strategy helped enriching these hits, we re-docked these compounds to the SARS-CoV-2 3CLpro without mutating Cys145 (represented as “regular-dock”). For quantitative comparison, we took the lowest score of each compound in the regular-dock results. Then we compared the difference between this lowest score with the score of the same compound in the SCAR-dock results (represented as “SCAR enriching score”). A note is that we used the score of the pose oriented for covalent bonding in SCAR-dock, which is the lowest score only if the rank of the conformation is 1 (Table 1). As shown in Table 1, for 6 out of these 11 compounds (54.5%), the SCAR-dock protocol generated lower scores (higher affinities, corresponding to a lower rank) than the regular-dock protocol (Table 1). As shown in Table 1, for 6 out of these 11 compounds (54.5%), the SCAR-dock protocol generated lower scores (higher affinities, corresponding to a lower rank) than the regular-dock protocol (Table 1). As shown in Table 1, for 6 out of these 11 compounds (54.5%), the SCAR-dock protocol generated lower scores (higher affinities, corresponding to a lower rank) than the regular-dock protocol (Table 1). As shown in Table 1, for 6 out of these 11 compounds (54.5%), the SCAR-dock protocol generated lower scores (higher affinities, corresponding to a lower rank) than the regular-dock protocol (Table 1).

In summary, we have identified eleven approved or investigational compounds/conformations) before it was used in the docking process (represented as “SCAR-dock”), among which 1253 were left for manual checking after score filtering (±8.0).

References

Ai,Y. et al. (2016) Discovery of Covalent Ligands via Noncovalent Docking by Dissecting Covalent Docking Based on a ‘Steric-Clashes Alleviating Receptor (SCAR)’ Strategy. J. Chem. Inf. Model., 56, 1563–1575. Barrett,D.G. et al. (2004) Potent and selective P2-P3 ketoamide inhibitors of cathepsin K with good pharmacokinetic properties via favorable P1’, P1, and/or P3 substitutions. Bioorganic & Medicinal Chemistry Letters, 14, 4897–4902. Chua,Y. et al. (2018) Drug repurposing from the perspective of pharmaceutical companies. Br. J. Pharmacol., 175, 168–180. Cuesta,A. et al. (2020) Ligand Conformational Bias Drives Enantioselective Modification of a Surface-Exposed Lysine on Hsp90. J. Am. Chem. Soc., 142, 3392–3400. Furber,M. et al. (2014) Cathepsin C inhibitors: property optimization and identification of a clinical candidate. J. Med. Chem., 57, 2357–2367. Huang,Changkang et al. (2004) 3C-like proteinase from SARS coronavirus catalyzes substrate hydrolysis by a general base mechanism. Biochemistry, 43, 4568–4574. Huang,Chaolin et al. (2020) Clinical features of patients infected with 2019 novel coronavirus in Wuhan, China. Lancet, 395, 497–506. Kaitin,K.I. (2010) Deconstructing the drug development process: the new face of innovation. Clin. Pharmaco. Ther., 87, 356–361. Leaver-Fay,A. et al. (2011) ROSETTA3: an object-oriented software suite for the simulation and design of macromolecules. Methods in Enzymology, 487, 545–574. Liao,C. et al. (2015) Discovery of novel inhibitors of human S-adenosylmethionine decarboxylase based on silicon high-throughput screening and a non-radioactive enzymatic assay. Sci. Rep., 5, 10754. Mah,R. et al. (2014) Drug discovery considerations in the development of covalent inhibitors. Bioorganic & Medicinal Chemistry Letters, 24, 33–39. Maxmen,A. (2020) More than 80 clinical trials launch to test coronavirus treatments. Nature, 578, 347–348. Pushpakom,S. et al. (2019) Drug repurposing: progress, challenges and recommendations. Nat Rev Drug Discov, 18, 41–58. Shannon,D.A. et al. (2014) Investigating the proteome reactivity and selectivity of aryl halides. J. Am. Chem. Soc., 136, 3330–3333. Sterling,T. and Irwin,J.J. (2015) ZINC 15 – Ligand Discovery for Everyone. J. Chem. Inf. Model., 55, 2324–2337. Trott,O. and Olson,A.J. (2010) AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. J. Comput. Chem., 31, 455–461. Zhang,L. et al. (2016) Optimize the interactions at S4 with efficient inhibitors targeting 3C proteinase from enterovirus 71. J. Mol. Recognit., 29, 520–527. Zhang,Y. et al. (2020) Repurposing Clinical Drugs as AdoMetDC Inhibitors Using the SCAR Strategy. Front. Pharmacol., 11, 248. Zhou,P. et al. (2020) A pneumonia outbreak associated with a new coronavirus of probable bat origin. Nature, 579, 270–273. Zumla,A. et al. (2016) Coronavirus - drug discovery and therapeutic options. Nat Rev Drug Discov, 15, 327–347.