Potential non-covalent SARS-CoV-2 3C-like protease inhibitors designed using generative deep learning approaches and reviewed by human medicinal chemist in virtual reality

Alex Zhavoronkov[1], Bogdan Zagribelnyy[1], Alexander Zhebrak[1], Vladimir Aladinskiy[1], Victor Terentiev[1], Quentin Vanhaelen[1], Dmitry S. Bezrukov[1], Daniil Polykovskiy[1], Rim Shayakhmetov[1], Andrey Filimonov[1], Michael Bishop[2], Steve McCloskey[2], Edgardo Leija[2], Deborah Bright[2], Keita Funakawa[2], Yen-Chu Lin[3], Shih-Hsien Huang[3], Hsuan-Jen Liao[3], Alex Aliper[1], Yan Ivanenkov[1]

[1] Insilico Medicine Hong Kong Ltd, Pak Shek Kok, New Territories, Hong Kong
[2] Nanome Inc. 7310 Miramar Rd #410, San Diego, CA 92126, USA
[3] Insilico Medicine Taiwan Ltd., 17F, No.3, Park Street, Nankang Dist., Taipei, Taiwan

Corresponding author: Alex Zhavoronkov, email: alex@insilico.com

One of the most important SARS-CoV-2 protein targets for therapeutics is the 3C-like protease (main protease, Mpro). In our previous work[1] we used the first Mpro crystal structure to become available, 6LU7. On February 4, 2020 Insilico Medicine released the first potential novel protease inhibitors designed using a *de novo*, AI-driven generative chemistry approach. Nearly 100 X-ray structures of Mpro co-crystallized both with covalent and non-covalent ligands have been published since then. Here we utilize the recently published 6W63 crystal structure of Mpro complexed with a non-covalent inhibitor and combined two approaches used in our previous study: ligand-based and crystal structure-based. We published 10 representative structures for potential development with 3D representation in PDB format and welcome medicinal chemists for broad discussion and generated output analysis. The molecules in SDF format and PDB-models for generated protein-ligand complexes are available here and at https://insilico.com/ncov-sprint/. Medicinal chemistry VR analysis was provided by Nanome team and the video of VR session is available at https://bit.ly/ncov-vr.

Introduction

Coronaviruses (CoVs) are a large family of viruses belonging to the family *Coronaviridae* that can infect the respiratory, gastrointestinal, hepatic and central nervous systems of humans, livestock, avians, bats, mice and many other wild animals [2] [3]. The limited number of coronaviruses known to be circulating in humans cause mild respiratory infections and were regarded as relatively harmless respiratory pathogens [4]. The emergence of the severe acute respiratory syndrome coronavirus (SARS-CoV) and the Middle East Respiratory Syndrome (MERS) coronavirus revealed that coronaviruses can cause severe and sometimes fatal respiratory tract infections in humans [5]. In December 2019, atypical pneumonia cases emerged in Wuhan, Hubei, China, with clinical presentations consistent with viral pneumonia. The cause was quickly identified as a novel CoV, which was eventually named SARS-CoV-2. Investigations of the epidemiological, clinical, laboratory and radiological characteristics, treatment, and outcomes of patients infected by SARS-CoV-2 demonstrated that the infection caused clusters of severe respiratory illness similar to SARS-CoV [6]. Like SARS-CoV, SARS-CoV-2 enters target cells through an endosomal pathway and also uses the same cell entry receptor, angiotensin-converting enzyme II (ACE2) [7] [8]. The coronavirus genome (ranging from 26 to 32 kb) is the largest among all RNA viruses — almost two times larger than that of the second largest RNA viruses. The viral particle is about 125 nm in diameter [9] [10]. The shape of the viral particle is either pleomorphic or spherical, and it is characterized by club-shaped projections of glycoproteins on its surface (diameter 80–120 nm). The genome of CoVs is a single-stranded positive-sense RNA packed in the nucleocapsid protein and further covered with an envelope. The genomic RNA is used as template to directly translate polyprotein (pp) 1a/1ab, which encodes non-structural proteins (nsps) to form the replication-transcription complex (RTC) in double-membrane vesicles (DMVs) [3]. The genes for non-structural proteins, such as the main protease, constitute two-thirds of the CoV genome. Among the structural proteins,four are of special interest: spike (S), envelope (E), membrane (M), and nucleocapsid (N). The S, E, and M proteins are contained within the viral membrane. The M and E proteins are involved in viral assembly, while the N protein is required for RNA genome assembly.

The aim of this paper is to present a set of novel molecular structures of SARS-CoV-2 3C-like protease inhibitors designed using the Insilico Medicine generative chemistry pipeline. This article is organized as follows: The next section discusses the rationale for the design of new SARS-CoV-2 3C-like protease inhibitors. Next, we briefly describe the motivation and recent progress made with use of generative chemistry approaches in drug discovery. The two following sections provide an overview of the methodology used for the design of Mpro inhibitors together with a brief presentation of the Insilico Medicine generative chemistry platform. The two next sections provide a description of the datasets used as input data to train and validate our models and an overview of the training procedures. The last section includes a general

presentation of the results and a structural analysis of the molecular structures generated using the Insilico Medicine pipeline.

Finding new SARS-CoV-2 3C-like protease inhibitors

While initial investigations for finding therapeutics to treat COVID-19 patients relied mostly on the identification of drug repurposing candidates [11], new attempts have been made to develop novel drug-like molecules active against SARS-CoV-2. Indeed, it is estimated that the design of novel drug-like molecules could be a good starting point for long-term coronavirus-related antivirals research, while repurposing already known drugs is the easiest and fastest way to find an effective treatment on a short term basis [12].

Previous studies on SARS-CoV[13] and MERS-CoV focused on the development of small molecule therapeutics using Mpro inhibitors [12–14]. Since the beginning of the COVID-19 outbreak, only a few studies on SARS-CoV-2 Mpro inhibitors have been published. Those publications are mostly based on materials from institutions that have been involved in research on coronaviruses since the first SARS outbreak [15–17]. Currently, nearly 100 crystal structures of ligand-bound and ligand-free SARS-CoV-2 Mpro have been made available in the Protein Data Bank[18], and only a few of them contain drug-like molecules while the remaining PDB 3C-like protease records contain small co-crystallized fragments derived by the UK national synchrotron[19].

The drug design studies recently published (alpha-ketoamide[16], aldehyde inhibitors[15] and Michael acceptors[17]) reported covalent inhibitors with micromolar and submicromolar activities, which elicit promising pharmacokinetics. However, these pharmacokinetic studies were performed on mice models, and covalent inhibitors are not necessarily suitable for treating patients due to the high probability of side effects. With their higher therapeutic index, non-covalent inhibitors offer an alternative despite their reduced potency. One application of computational chemistry to find active molecules against SARS-CoV-2 3C-like protease was recently proposed by the Artem Cherkasov team. In order to find potential main protease ligands, they used a technology called "deep docking", which can provide fast prediction of Glide docking scores. They applied this technology to screen more than 1 billion compounds from the ZINC15 library and then published the 1,000 structures with the highest *in silico* scores [20].

Generative chemistry approaches for Mpro inhibitor design

Considering the virtually unlimited number of chemical structures that can be generated *de novo*, conventional computational drug design approaches tend to include a limited number of fragments and/or employ sophisticated search strategies to sample hit compounds from a predefined area of the chemical space. To enable scientists to exploit the whole drug-like chemical space, a new type of computational method for drug discovery has been developed

using the recent advances in deep learning (DL) and artificial intelligence (AI). Such techniques can automatically extract high dimensional abstract information without the need for manual feature design and learn nonlinear mappings between molecular structures and their biological and pharmacological properties. Deep generative models can utilize large datasets for training and perform *in silico* design of *de novo* molecular structures with predefined properties [21]. The first model of this type, a molecular generator using an adversarial auto-encoder (AAE) to generate molecular fingerprints, was released in early 2017[22]. Since then, many architectures have been proposed to generate not just valid chemical structures, but also molecules matching certain bioactivity and novelty profiles, as well as other features of interest. Several milestones were recently accomplished with the use of generative chemistry in drug discovery, demonstrating that it is possible to generate molecules that can be synthesized, are active *in vitro*, metabolically stable, and elicit *in vivo* activity in disease-relevant models. The first example of an *in vitro* active molecule obtained through generative chemistry was the JAK3 inhibitor [23]. Another generative model, Generative Tensorial Reinforcement Learning (GENTRL), generated discoidin domain receptor DDR1 and DDR2 inhibitors with different property and selectivity profiles, which were assayed *in vitro*, followed by *in vivo* mouse experiments that validated the pharmacokinetics of DDR1 inhibitors [24]. This experiment, performed in 2018, demonstrated that this new generative chemistry approach is capable of finding novel molecular structures with optimized properties. Multiple improvements were made to this generative chemistry approach to address the novelty, diversity, and synthetic accessibility concerns.

During the COVID-19 outbreak, generative chemistry naturally appeared as a promising approach to the fast design of new high-potential chemotypes that could be developed as potent antivirals. In this context, Insilico Medicine launched a program to generate the structures of inhibitors targeting the SARS-CoV-2 main protease at the end of January 2020. Insilico Medicine quickly published representative examples of structures obtained with its internal generative chemistry pipeline to become a pioneer in applying generative chemistry for the development of treatments for COVID-19 [1].

Since then, other groups have proposed their own structures. In March 2020, the scientific group from University of Missouri and Xiamen University used its reinforcement learning (RL) generative models to generate potential SARS-CoV-2 Mpro inhibitors based on the 6LU7 PDB crystal structure, and published 47 representative structures from their experiments [25]. In April 2020, IBM applied its generative model based on VAE (variational autoencoder) to design novel inhibitors with predefined properties for SARS-CoV-2 Mpro, the Spike/ACE2 protein–protein interaction and NSP9. The company released over 3,000 structures as representative examples [26]. Docking studies were carried out to confirm whether the generated molecules possessed the ability to bind the targets, as the binding optimization was not included directly in the generative process.

Insilico Medicine generative chemistry pipeline for COVID-19

Insilico Medicine generative chemistry automatized platform (Figure 1) is designed to perform the key steps of early stage drug discovery in a cost- and time-efficient manner. The platform has already been successfully applied to design small drug molecules for a large range of human diseases. Insilico Medicine has published proofs of concepts, descriptions and validations of the key algorithms integrated within the pipeline as well as the technical description of the fundamental concepts and approaches used in generative chemistry [22–24,27–32]. In this work, our small molecule drug discovery pipeline has been applied to generate inhibitors of the SARS-CoV-2 Mpro.
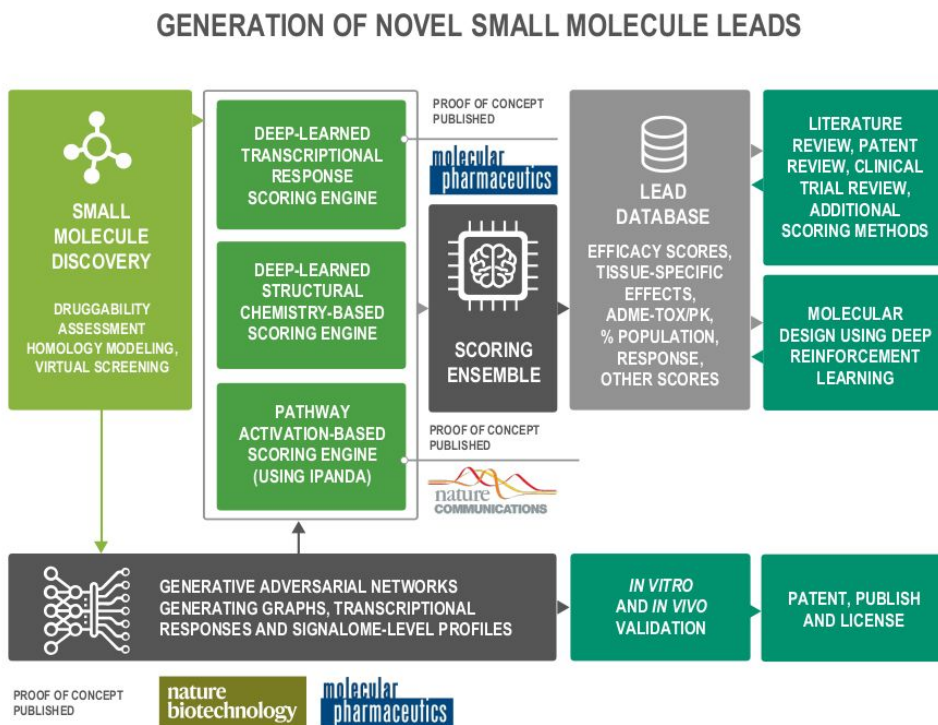
Figure 1: Insilico Medicine generative chemistry pipeline. The generative modules utilizing crystal structure and ligand-based features were used to generate the molecules for the 3C-like protease.

January 2020 generative chemistry sprint using the 6LU7 crystal structure

At the end of January, the information released on the spread of the COVID-19 demonstrated that the virus is substantially more dangerous than initially thought. While many research institutes around the world started to propose various promising repurposing candidates, we decided to support the ongoing efforts with a different strategy and employed our generative chemistry pipeline to generate novel small molecules specifically designed against SARS-CoV-2. After the selection of the SARS-CoV-2 3C-like protease as a target, we planned out the project timeline (Figure 2). The first important milestone was reached on February 5th with the submission of our first newly generated inhibitors [33]. Three parallel approaches (Figure 3), pocket-based, ligand-based and homology model-based generation, were utilized to

generate novel molecular structures. After a review of the results of the generation process, we published a set of 97 molecular structures whose properties make them suitable as inhibitors of the viral protease [1].
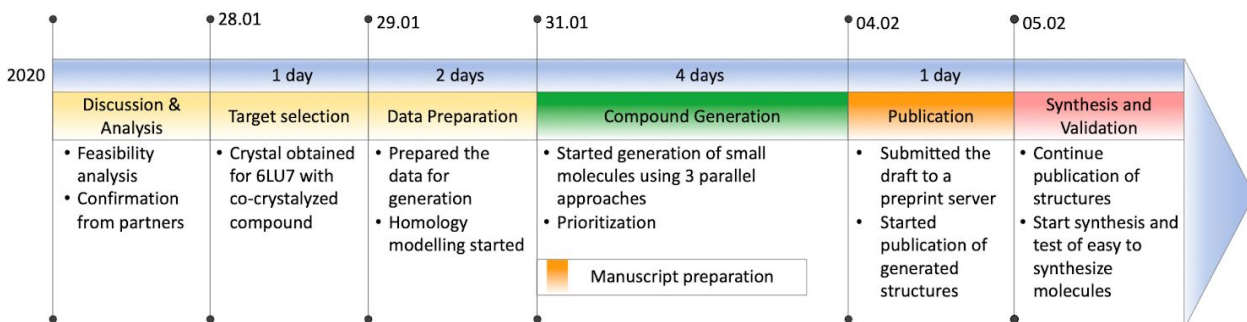

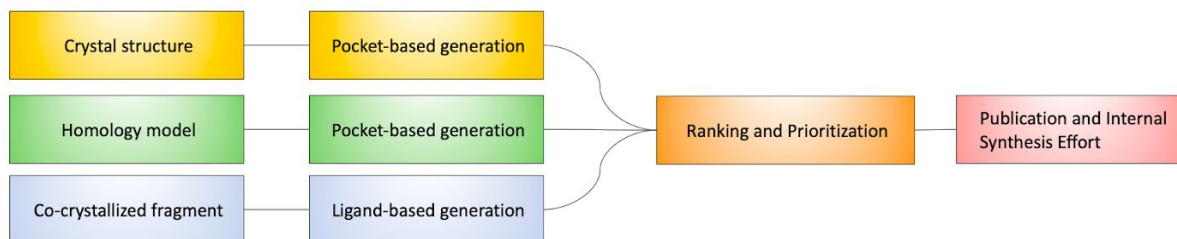Figure 2: Insilico Medicine SARS-CoV-2 small molecule generation sprint timeline [1]

Figure 3: Insilico Medicine SARS-CoV-2 small molecule generation procedure [1]

April–May 2020 generative chemistry sprint using the 6W63 crystal structure

By the end of April 2020, additional Mpro crystal structures had been released on the Protein Data Bank website, but only one of them (PDB: 6W63) contained X77, a non-covalent inhibitor of Mpro [34]. Within the scientific community, most of the current research efforts focus on finding potential SARS-CoV-2 Mpro covalent inhibitors while the possibilities to identify non-covalent inhibitors remain less investigated. For that reason, Insilico Medicine decided to launch a new generation based on the 6W63 crystal structure to find potential non-covalent chemotypes targeting the SARS-CoV-2 3C-like protease. The generation was followed by the selection of representative examples and medicinal chemistry analysis provided by Nanome, Inc. in a virtual reality (VR) environment (Figure 4A, 4B).
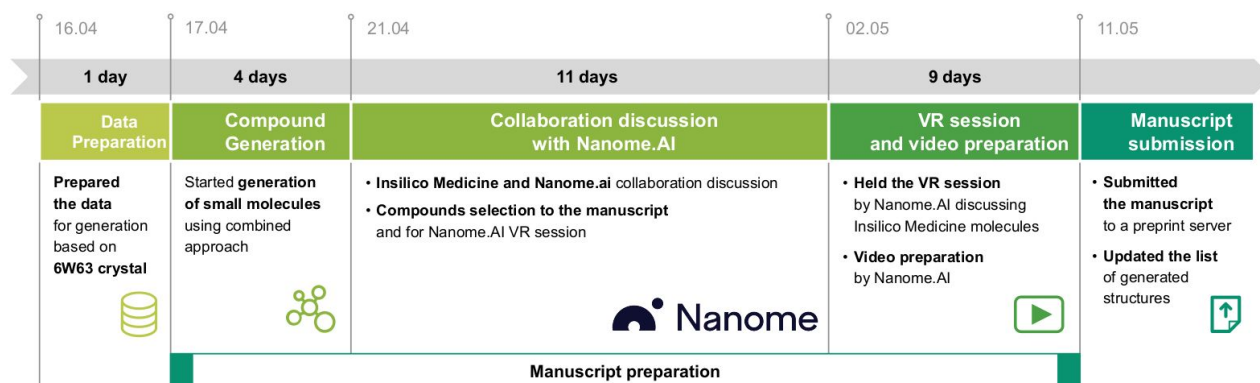
Figure 4A: Sprint timeline for Insilico Medicine SARS-CoV-2 small molecule generation based on 6W63 crystal structure.
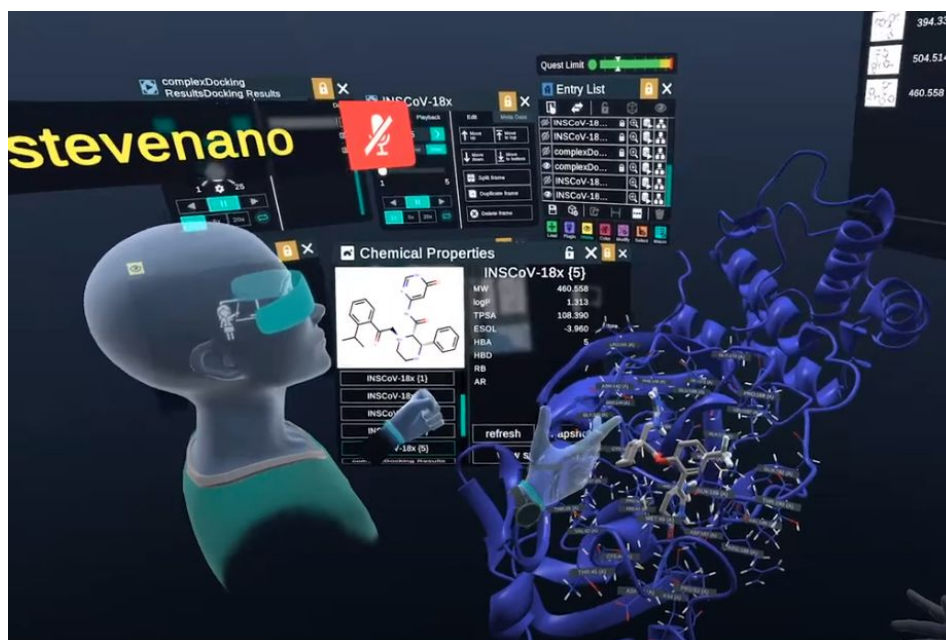

Figure 4B. INSCoV-188 structure in complex with Mpro visualized using Nanome software.

Input data and datasets

Crystal structure of SARS-CoV-2 3C-like protease with non-covalent inhibitors

The crystal structure of the SARS-CoV-2 3C-like protease was downloaded from the Protein Data Bank (PDB Code: 6W63). The structure was solved with a 2.1-angstrom resolution in complex with X77. It should be noted that currently no activity data for the SARS-CoV-2 3C-like protease has been reported for this compound. However, an IC_{50} value against SARS-CoV Mpro is 2.3 µM [35]. The crystal structure was preprocessed using position restrained minimization with GROMACS[36] with the C_{α} atoms of the protein and all the heavy atoms of the ligand

restrained by harmonic constraints (k_{spring}=100 kJ/mol/nm^2). One protonation state of His41 that forms an H-bond with the ligand pyridyl was considered. The ligand was extracted from the minimized crystal and ligand-based features (pharmacophore hypothesis and shape) were generated exploiting its 3D-conformation. The constructed ligand shape and pharmacophore hypothesis were used for estimating how virtual structures fit these features. Then, the protein binding site was annotated utilizing our proprietary Pocket Module to create amino acid residues mapping as input pocket features for generation (Figure 6).

Co-crystalized ligand features

The 3D structure of the X77 inhibitor was extracted from the solved complex. The conformation was used to build the shape of the ligand and a pharmacophore hypothesis using our proprietary tools. For the hypothesis, 5 pharmacophore points were selected according to the interactions observed within the binding site (Figure 5).
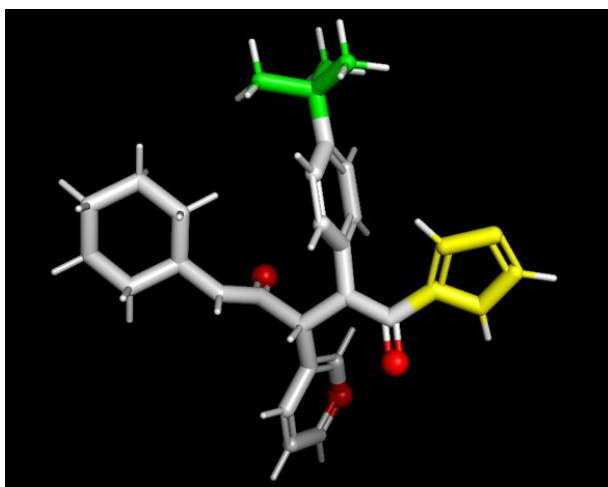

Figure 5: Crucial pharmacophore points (pharmacophore hypothesis) from 6W63 ligand. Red: H-bond acceptor; green: lipophilic area; yellow: aromatic (PyMol graphics).

Protease dataset

The *protease dataset* was assembled with molecules active against various proteases in enzymatic assays extracted from the Integrity database [37], Experimental Pharmacology module and ChEMBL [38,39]. The records from the ChEMBL database were downloaded with the following activity standard types: 'Potency', 'IC$_{50}$', 'K_i', 'EC$_{50}$', 'K_d' (assay confidence score ≥8, assay type: B, F). The activities from the Integrity database were downloaded using the following parameters: 'IC$_{50}$', 'K_i', 'EC$_{50}$', 'K_d', and mass concentrations (e.g., mg/L) were converted to M values by molecular weight. Integrity records were standardized using the pChEMBL value format (logarithmic scale –$\log_{10}$ of a numeric value in M) and merged with the records from

ChEMBL. The resulting records with pChEMBL values less than 5.0 (10 µM in terms of IC_{50}) were then removed.

The structural duplicates were excluded during the standardization procedure, and salt parts were removed. Mild medicinal chemistry filters (MCFs) were applied to exclude non–drug-like molecules (e.g., metals, polycondensed aromatics, chloramines, radicals, hydrazines, isonitriles, nitroso compounds) and structures containing cycles with more than 8 atoms and polypeptides (n≥4). The resulting dataset contained 60,293 unique structures.

To adapt the scoring metrics and the reward functions to the type of structures to be generated, a *protease peptidomimetics dataset* was collected from the *protease dataset* using SMARTS queries for common peptidomimetic substructures. The compounds with pChEMBL value less than 6.0 were filtered and the overrepresented chemotypes were suppressed. The final *protease peptidomimetics dataset* contained 5,891 compounds.

Generative pipeline

Two approaches (ligand-based and pocket-based) were applied for the generation of the novel molecular structures. Pocket and ligand features were obtained from the binding site amino acid environment and from the co-crystallized fragment derived from the same PDB record (6W63), respectively (Figure 6).
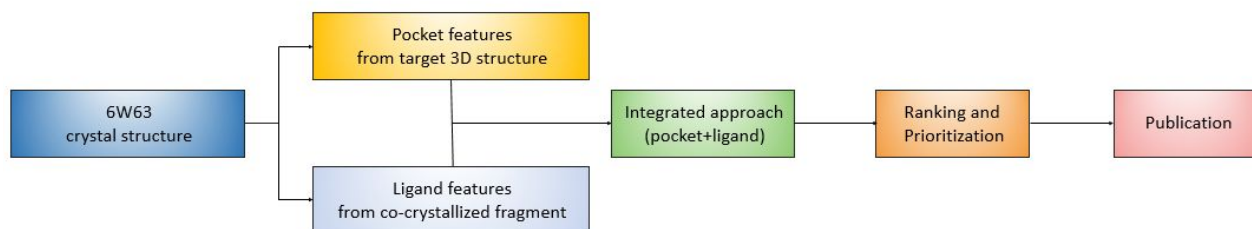

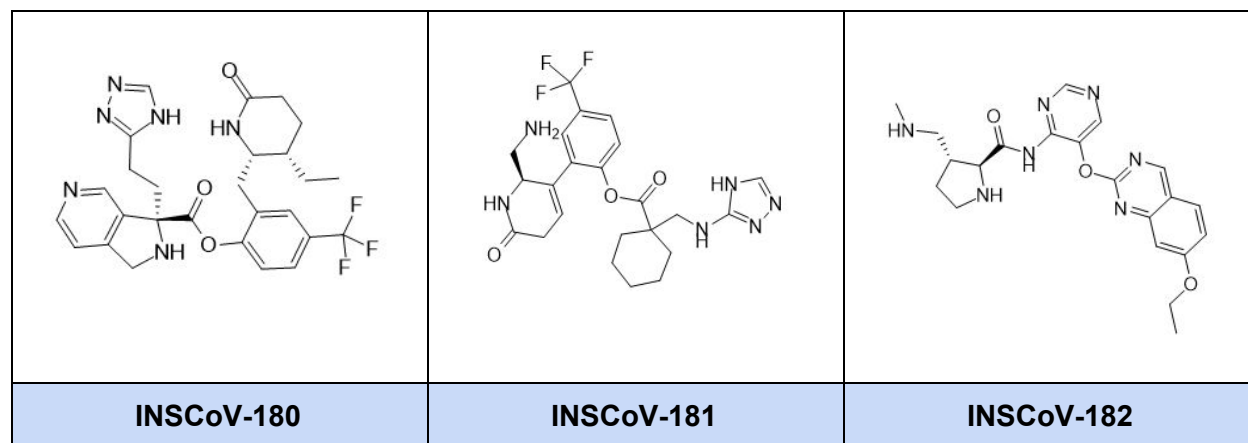Figure 6. Insilico Medicine SARS-CoV-2 small molecule generation procedure.

During the generation step, a total of 28 machine learning (ML) models were used to generate molecular structures. Those structures were optimized with reinforcement learning based on the reward function described below. We used different ML approaches, including generative autoencoders, generative adversarial networks, genetic algorithms, and language models. The models exploited different molecular representations, including fingerprints, string representations, and graphs. Every model optimized the reward function to explore the chemical space, investigate promising clusters, and generate new molecular structures with high scores.

The rewarding function was a weighted sum of multiple intermediate rewards: medicinal chemistry and drug-likeness scoring, active chemistry scoring, structural scoring (fitting to ligand features and/or binding pocket), novelty scoring, diversity scoring and synthetic accessibility scoring (SAScore). Medicinal chemistry scoring assigns low reward for molecules with structural alerts and high reward for molecules with benefit substructures. Drug-likeness scoring directs the generation towards molecules with molecular properties that are representative of *protease peptidomimetics dataset* — logP: -0.50–6.00; Molecular weights (MW): 330–750; Number of hydrogen bond donors (HBD): 0–10; Number of hydrogen bond acceptors (HBA): 2–10; Topological polar surface area (TopoPSA): 40–170; MCE-18[40]: 40–180; Number of stereocenters (nSC): 0–3, Synthetic accessibility score (SAScore): 0–5.00. Target-specific scoring utilizes self-organizing maps trained on the *protease peptidomimetics dataset*. We used novelty and diversity scoring during the optimization procedure to explore the chemical space and output novel and diverse molecular structures. The generated compounds were penalized when they were too similar to the existing molecules or already explored clusters. We performed integral scoring with the provided crystal structure (pocket features) and pharmacophore/shape scoring. We ran the distributed pipeline for 72 hours on the internal computing cluster with 64 NVIDIA Titan V GPUs.

Results

Generated structures

The highest-ranking structures were selected for further analysis. Figure 7 shows representative molecular structures from the chemical space produced by our approach. The generated structures share the structural patterns common to peptidomimetics. We assessed the similarity between our generated structures and the compounds from the ChEMBL database using the ChEMBL search option. The analysis revealed that there is no molecule with the same core structure among the compounds with a similarity coefficient >0.7.
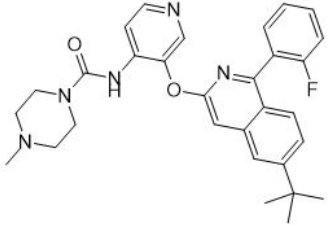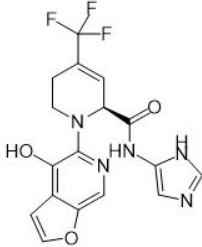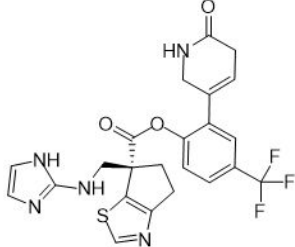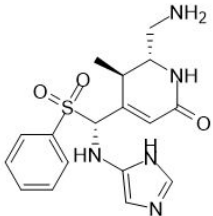

| INSCoV-180 | INSCoV-181 | INSCoV-182 |

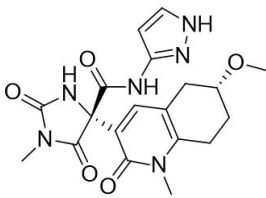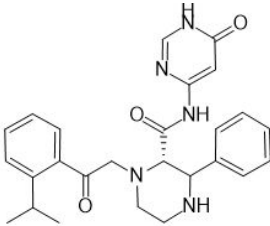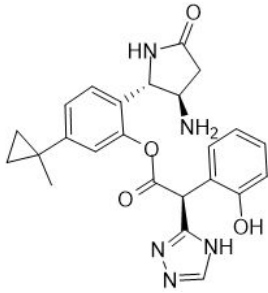|		
INSCoV-183	INSCoV-184	INSCoV-185
INSCoV-186	INSCoV-187	INSCoV-188
	INSCoV-189	

Figure 7. Representative examples of the generated structures targeting the main protease of SARS-CoV-2. The list of generated molecules is available at: https://insilico.com/ncov-sprint/. Interested medicinal chemists are encouraged to provide comments and suggestions.

Table 1. Descriptors of the molecular structures depicted in Figure 7. MW: molecular weight; nRot: number of rotatable bonds; nAR: number of aromatic rings; nSC: number of stereocenters; HBA: number of hydrogen bond acceptors; HBD: number of hydrogen bond donors; MCE-18: medicinal chemistry evolution 2018 descriptor; TopoPSA: Topological polar surface area; SAScore: synthetic accessibility score.

ID	MW	nRot	nAR	nSC	HBA	HBD	MCE-18	TopoPSA	SAScore
INSCoV-180	542,23	8	2	3	6	4	102,54	121,89	4,75

INSCoV-181	492,21	7	2	1	5	6	90,24	135,02	4,11
INSCoV-182	423,20	8	1	2	7	5	68,79	123,18	3,76
INSCoV-183	513,25	4	1	0	4	2	66,15	70,59	2,76
INSCoV-184	393,10	3	1	1	5	3	78,86	107,28	4,12
INSCoV-185	503,12	6	2	1	5	3	100,00	137,24	4,25
INSCoV-186	375,14	6	1	3	4	6	60,45	138,35	4,36
INSCoV-187	414,17	4	2	2	6	3	87,63	138,42	4,49
INSCoV-188	459,23	7	1	2	4	5	74,41	107,19	3,64
INSCoV-189	447,19	6	2	3	6	6	97,88	143,22	4,36

Structural analysis of the generated molecular structures

In addition to the 2D structures, we also published 5 examples of the generated molecules in .pdb format as supporting information. The 3D representations of the selected molecules and their interaction interface predicted within the ligand binding site of Mpro are shown in Figure 8. The depicted molecules show multiple H-bonds predominantly with HIS-163, ASN-142 and GLY-143 amino acid residues and fill the key hydrophobic pocket (MET-49, CYS-44, MET-165) occupied with the t-Bu group of the template molecule.

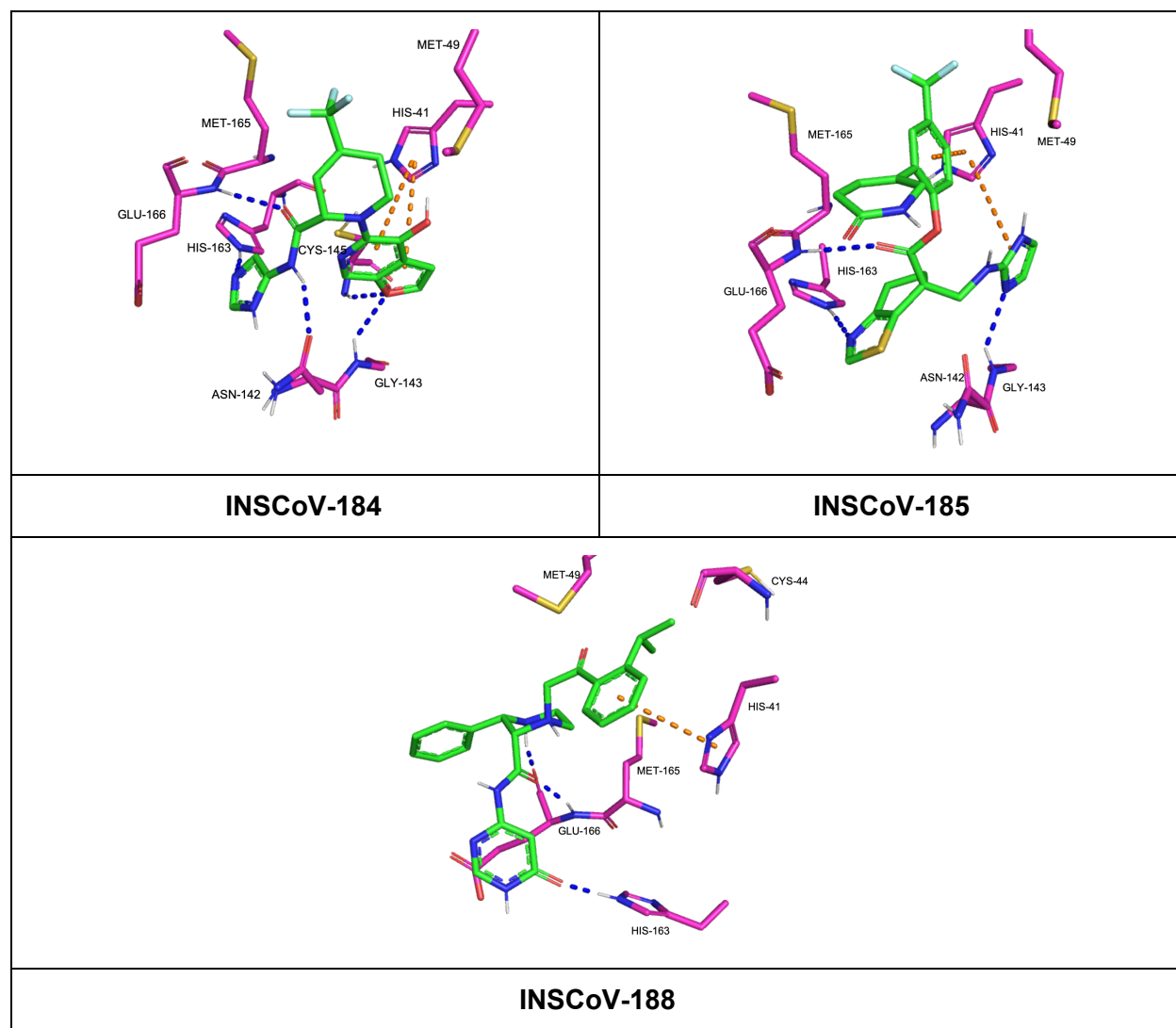| INSCoV-181	INSCoV-182

INSCoV-184

INSCoV-185

INSCoV-188
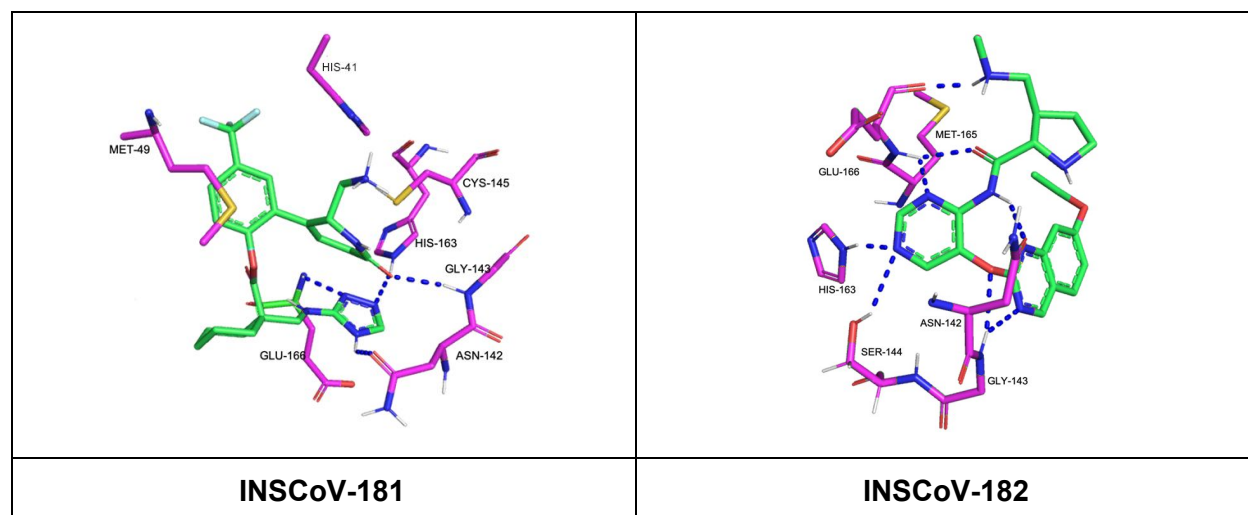
Figure 8. The representative examples illustrate the predicted binding modes for the generated structures (visualization was made in PyMol).

The above 5 ligand–protein complexes were additionally assessed during a VR session kindly carried out by the Nanome team [41] (Figure 9). The binding and fit of the ligands inside the pocket, 3D conformations and pharmacophore features were discussed during the session, and valuable commentaries on metabolic stability and synthetic accessibility were provided by the medicinal chemist from Nanome. A recording of medicinal chemistry analysis of Insilico Medicine AI-generated compounds inside Nanome software is available at https://bit.ly/ncov-vr.
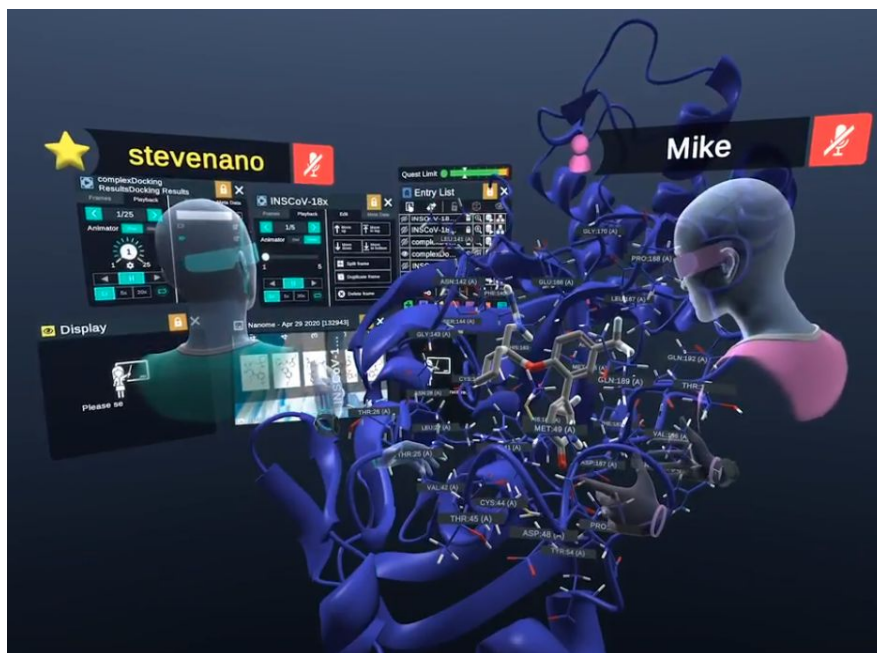
Figure 9. INSCoV-181 structure in complex with Mpro visualized using Nanome software.

Availability of the molecular structures

The most recent data package is available at insilico.com/ncov-sprint. We will continue to update the data package with new compounds from future COVID-19 Insilico Medicine sprints. The data could be used to perform subsequent computer modelling simulations or to synthesize and test the compounds *in vitro* against the SARS-CoV-2 main protease.

Conclusion

Despite the economic and societal impact of CoV infections and the likelihood of future outbreaks of even more serious pathogenic CoVs in humans, there is still a lack of effective antiviral strategies to treat CoVs and few options to prevent CoV infections [42]. Given the high prevalence and wide distribution of CoVs, the novel virus could emerge periodically in humans as a consequence of frequent cross-species infections and occasional spillover events [43]. The development of effective and time-efficient computational methods for designing compounds that can treat CoV infections is critical. In this study, we have used our integrated AI-based drug discovery pipeline to generate novel potential compounds targeting the SARS-CoV-2 main protease. The results demonstrate the time- and cost-effectiveness of this method for the development of novel treatments against CoV infections. We plan to investigate potential compounds targeting other essential SARS-CoV-2 target proteins and PPIs.

Acknowledgments

Conflicts of interest

Insilico Medicine is a company developing an AI-based end-to-end integrated pipeline for drug discovery and development and engaged in aging and cancer research.

Nanome is a company developing VR tools for scientific research and visualisation.

References

1. Zhavoronkov, A. *et al.* Potential COVID-2019 3C-like Protease Inhibitors Designed Using Generative Deep Learning Approaches. *ChemRxiv* (2020) doi:10.26434/chemrxiv.11829102.v2.

2. Chen, Y. & Guo, D. Molecular mechanisms of coronavirus RNA capping and methylation. *Virol. Sin.* **31**, 3–11 (2016).

3. Chen, Y., Liu, Q. & Guo, D. Emerging coronaviruses: Genome structure, replication, and pathogenesis. *J. Med. Virol.* **92**, 418–423 (2020).

4. Song, Z. *et al.* From SARS to MERS, Thrusting Coronaviruses into the Spotlight. *Viruses* **11**, (2019).

5. de Wit, E., van Doremalen, N., Falzarano, D. & Munster, V. J. SARS and MERS: recent insights into emerging coronaviruses. *Nat. Rev. Microbiol.* **14**, 523–534 (2016).

6. Menachery, V. D. *et al.* Corrigendum: A SARS-like cluster of circulating bat coronaviruses shows potential for human emergence. *Nat. Med.* **22**, 446 (2016).

7. Zhou, P. *et al.* Discovery of a novel coronavirus associated with the recent pneumonia outbreak in

humans and its potential bat origin. doi:10.1101/2020.01.22.914952.

8. Letko, M. & Munster, V. Functional assessment of cell entry and receptor usage for lineage B
    β-coronaviruses, including 2019-nCoV. doi:10.1101/2020.01.22.915660.

9. Shereen, M. A., Khan, S., Kazmi, A., Bashir, N. & Siddique, R. COVID-19 infection: Origin,
    transmission, and characteristics of human coronaviruses. *J. Advert. Res.* **24**, 91–98 (2020).

10. Yang, H., Bartlam, M. & Rao, Z. Drug design targeting the main protease, the Achilles' heel of
    coronaviruses. *Curr. Pharm. Des.* **12**, 4573–4590 (2006).

11. Harrison, C. Coronavirus puts drug repurposing on the fast track. *Nat. Biotechnol.* **38**, 379–381
    (2020).

12. Liu, C. *et al.* Research and Development on Therapeutic Agents and Vaccines for COVID-19 and
    Related Human Coronavirus Diseases. *ACS Cent Sci* **6**, 315–331 (2020).

13. 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. *J. Med. Chem.* **59**, 6595–6628 (2016).

14. Ghosh, A. K., Brindisi, M., Shahabi, D., Chapman, M. E. & Mesecar, A. D. Drug Development and
    Medicinal Chemistry Efforts Toward SARS-Coronavirus and Covid-19 Therapeutics. *ChemMedChem*
    (2020) doi:10.1002/cmdc.202000223.

15. Dai, W. *et al.* Structure-based design of antiviral drug candidates targeting the SARS-CoV-2 main
    protease. *Science* (2020) doi:10.1126/science.abb4489.

16. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of
    improved α-ketoamide inhibitors. *Science* **368**, 409–412 (2020).

17. Jin, Z. *et al.* Structure of Mpro from COVID-19 virus and discovery of its inhibitors. *Nature* (2020)
    doi:10.1038/s41586-020-2223-y.

18. RCSB PDB: Homepage. *Protein Data Bank* https://www.rcsb.org/.

19. Main protease structure and XChem fragment screen. *Diamond Light Source*
    https://www.diamond.ac.uk/covid-19/for-scientists/Main-protease-structure-and-XChem.html.

20. Ton, A.-T., Gentile, F., Hsing, M., Ban, F. & Cherkasov, A. Rapid Identification of Potential Inhibitors

of SARS-CoV-2 Main Protease by Deep Docking of 1.3 Billion Compounds. *Mol. Inform.* (2020) doi:10.1002/minf.202000028.

21. Zhavoronkov, A., Vanhaelen, Q. & Oprea, T. I. Will Artificial Intelligence for Drug Discovery Impact Clinical Pharmacology? *Clin. Pharmacol. Ther.* (2020) doi:10.1002/cpt.1795.

22. Kadurin, A. *et al.* The cornucopia of meaningful leads: Applying deep adversarial autoencoders for new molecule development in oncology. *Oncotarget* vol. 8 (2016).

23. Polykovskiy, D. *et al.* Entangled Conditional Adversarial Autoencoder for de Novo Drug Discovery. *Mol. Pharm.* **15**, 4398–4405 (2018).

24. Zhavoronkov, A. *et al.* Deep learning enables rapid identification of potent DDR1 kinase inhibitors. *Nat. Biotechnol.* **37**, 1038–1040 (2019).

25. Tang, B. *et al.* AI-aided design of novel targeted covalent inhibitors against SARS-CoV-2. *bioRxiv* 2020.03.03.972133 (2020) doi:10.1101/2020.03.03.972133.

26. Chenthamarakshan, V. *et al.* Target-Specific and Selective Drug Design for COVID-19 Using Deep Generative Models. *arXiv [cs.LG]* (2020).

27. Kadurin, A., Nikolenko, S., Khrabrov, K., Aliper, A. & Zhavoronkov, A. druGAN: An Advanced Generative Adversarial Autoencoder Model for de Novo Generation of New Molecules with Desired Molecular Properties in Silico. *Mol. Pharm.* **14**, 3098–3104 (2017).

28. Putin, E. *et al.* Adversarial Threshold Neural Computer for Molecular de Novo Design. *Mol. Pharm.* **15**, 4386–4397 (2018).

29. Putin, E. *et al.* Reinforced Adversarial Neural Computer for de Novo Molecular Design. *J. Chem. Inf. Model.* **58**, 1194–1204 (2018).

30. Kuzminykh, D. *et al.* 3D Molecular Representations Based on the Wave Transform for Convolutional Neural Networks. *Mol. Pharm.* **15**, 4378–4385 (2018).

31. Aliper, A. *et al.* Deep Learning Applications for Predicting Pharmacological Properties of Drugs and Drug Repurposing Using Transcriptomic Data. *Mol. Pharm.* **13**, 2524–2530 (2016).

32. Zhavoronkov, A. Artificial Intelligence for Drug Discovery, Biomarker Development, and Generation of Novel Chemistry. *Mol. Pharm.* **15**, 4311–4313 (2018).

33. Zhavoronkov, A., Aladinskiy, V. A., Zhebrak, A., Zagribelnyy, B. A. & Ivanenkov, Y. A. Potential 2019-nCoV 3C-like protease inhibitors designed using generative deep learning approaches. *ResearchGate* (2020) doi:10.13140/RG.2.2.29899.54569.

34. Mesecar, A. D. & Center for Structural Genomics of Infectious Diseases (CSGID). Structure of COVID-19 main protease bound to potent broad-spectrum non-covalent inhibitor X77. (2020) doi:10.2210/pdb6w63/pdb.

35. St. John, S. E. & Mesecar, A. D. Broad-spectrum non-covalent coronavirus protease inhibitors. *US Patent* (2017).

36. Abraham, M. J. *et al.* GROMACS: High performance molecular simulations through multi-level parallelism from laptops to supercomputers. *SoftwareX* **1-2**, 19–25 (2015).

37. Clarivate Analytics Integrity. https://integrity.clarivate.com/integrity/.

38. Gaulton, A. *et al.* ChEMBL: a large-scale bioactivity database for drug discovery. *Nucleic Acids Res.* **40**, D1100–7 (2012).

39. *CHEMBL database release 25*. http://ftp.ebi.ac.uk/pub/databases/chembl/ChEMBLdb/releases/chembl_25 (2019) doi:10.6019/CHEMBL.database.25.

40. Ivanenkov, Y. A., Zagribelnyy, B. A. & Aladinskiy, V. A. Are We Opening the Door to a New Era of Medicinal Chemistry or Being Collapsed to a Chemical Singularity? *J. Med. Chem.* **62**, 10026–10043 (2019).

41. Nanome. https://nanome.ai/.

42. de Wilde, A. H., Snijder, E. J., Kikkert, M. & van Hemert, M. J. Host Factors in Coronavirus Replication. *Curr. Top. Microbiol. Immunol.* **419**, 1–42 (2018).

43. Zhu, N. *et al.* A Novel Coronavirus from Patients with Pneumonia in China, 2019. *N. Engl. J. Med.* (2020) doi:10.1056/NEJMoa2001017.