Drug repurposing approach targeted against main protease of SARS-CoV-2 exploiting 'neighbourhood behaviour' in 3D protein structural space and 2D chemical space of small molecules

Sohini Chakraborti[1], Narayanaswamy Srinivasan[1*]

[1]*Molecular Biophysics Unit, Indian Institute of Science, Bengaluru 560012, India*

**Corresponding author*

E-mail: ns@iisc.ac.in

Abstract: The current global crisis due to COVID-19 has almost brought normal life to standstill in most parts of the world. With our research interest on repurposing known drugs/drug candidates targeting various diseases, we decided to analyse the available data on the deadly pathogen that has already taken thousands of lives since its outbreak in China in December 2019. Our host institute is now shutdown and we are confining ourselves to our homes with limited access to computational resources. Using a simple *in silico* approach based on the principle of 'neighbourhood behaviour' in three-dimensional (3D) space and two-dimensional (2D) space of protein and small molecules respectively, we have identified potential drugs/drug candidates which can be repurposed against protein targets encoded by the SARS-CoV-2 genome. Based on our preliminary analysis, we have so far prioritized more than 20 known drugs/drug candidates which might elucidate anti-coronavirus properties by binding to main protease of the pathogen. These drugs belong to diverse therapeutic areas such as antiviral, anticancer, antibacterial agents etc. Notably, apart from many synthetic molecules, our analysis also hints that phytochemicals obtained from vinca plant (vinca alkaloids) and camptotheca tree (camptothecin and its derivatives) have the potential to bind to main protease of SARS-CoV-2. In-depth investigation on our findings are currently on-going. Here we are presenting the results we obtained so far. The sole purpose of making these preliminary findings openly available to the community is for the experimental

biologists and biomedical researchers to investigate our predictions in experimental set ups and for the clinicians to evaluate the potential of these findings for anti-COVID-19 treatment. **Our findings should only be used for research purposes and we strongly urge that no individual should interpret these findings for any self-diagnosis or self-medication without the prior approval from competent international health/medical regulatory agencies.**

Keywords: drug repurposing, SARS-CoV-2, main protease, neighbourhood behaviour, structural neighbours, Tanimoto coefficient, docking simulation, antiviral agents, anticancer agents, antibacterial agents, natural products.

Introduction

The purpose of this report is to convey the basic science involved in our analysis and make our interim findings available to the scientific community with the hope that this effort might contribute toward accelerating the research to find potential solutions to combat COVID-19. Consistent with our motive behind writing this interim report, we would restrict ourselves to discuss only our approach and the findings obtained so far. We are therefore purposefully refraining from giving any overview on COVID-19, pathogenicity, statistics, or any other details which would not be directly pertinent in comprehending our results. Interested readers can access such details through other comprehensive resources like COVID-19 resources (supported by Indian Academy of Science; http://confluence.ias.ac.in/covid-19-resources/), LitCovid (supported by US National Institutes of Health; https://www.ncbi.nlm.nih.gov/research/coronavirus/) etc.

In the recent times, drug repurposing approaches have gained importance as these methods are expected to be faster and require less economic investment[1]. Application of *in silico* techniques in early stages of drug discovery either through conventional or repurposing

approaches aid in minimizing the chances of the failures. The last few decades have witnessed tremendous advancement in the field of pharmaceutical sciences. Despite such advancements, the allocation of resources for *de novo* drug discovery to tackle infectious diseases is not very encouraging. Therefore, drug repurposing remains one of the most popular alternate technique in dealing with the unmet medical needs[2–5]. Our previous work on repurposing drugs against multiple infectious diseases have demonstrated the application of protein evolutionary information in identification of potential molecules which could be repurposed for the treatment of tuberculosis, malaria, candidiasis etc[6–9]. Recently we have also employed ligand based screening technique to identify approved drugs which could be targeted against Aurora kinases (a promising anti-cancer target) but are currently known to target other proteins[10].

In this interim report, we are presenting our analyses on potential molecules which could be repurposed against SARS-CoV-2 main protease (SARS-CoV-2 M^{pro}), also referred to as Non-Structural polyprotein 1ab or 3C-Like proteinase ($3CL^{pro}$). At the time when we initiated this work, 13 crystal structures of SARS-CoV-2 M^{pro} were available in the Protein Data Bank (PDB)[11], both in unliganded (PDB codes: 6M03[12],6Y84[13], 6Y2E[14]) and liganded (PDB codes: 6Y2F[14], 6Y2G[14], 6LU7[15], 5R7Y[16], 5R7Z[17], 5R80[18], 5R81[19], 5R82[20], 5R83[21], 5R84[22]) forms. With the latest depositions, at the time of writing this report (30th Mar. 2020), 104 protein structures of SARS-CoV-2 are available in the PDB (http://www.rcsb.org/pdb/results.do?tabtoshow=Current&qrid=FC2F085B). Majority of these structures (~80%) are from main protease of the pathogen bound to different small molecule fragments and one of these structures is bound to a novel inhibitor ('X77' in 6W63[23]). Needless to mention, this statistics indicates that main protease of SARS-CoV-2 is one of the most important drug targets. Earlier reports suggest that no human proteases has a similar cleavage specificity as that of the proteases of coronaviruses, indicating that inhibitors of these proteins are unlikely to cause severe adverse effects[14]. Also, our preliminary attempt to

check the presence of any closely related homologue of SARS-CoV-2 M^{pro} in human using default settings of BLAST search[24] did not yield any significant hit (data not shown). This background data prompted us to start our hunt for potential repurpose-able candidates conidering SARS-CoV-2 M^{pro} as the target of foremost importance.

Our analysis is based on the fundamental principle of 'neighbourhood behaviour' implemented through two different workflows involving (i) protein 3D space and (ii) 2D chemical space of small molecules. Recognition between biological molecules is governed by complementarity in different fingerprints like shape, volume, electrostatics of binding partners and are likely to be conserved among neighbours in 3D (proteins with similar structures) and 2D (chemically similar compounds) spaces. In general, most protein structural neighbours are known to perform similar functions which are facilitated mainly via conserved/semi-conserved fingerprints of molecular recognition sites between the protein and its binding partner/s (ligands: proteins, small molecules, peptides etc.) [25–27] In favourable cases, structure comparison can reveal distant evolutionary relationships which are otherwise difficult to be captured by sequence comparison. On a similar note, small molecules which share similar chemical scaffolds ('chemical neighbours') generally possess similar topological fingerprints (structural features) and are known to elucidate similar pharmacological responses in many instances. As the number of common features between two small molecules increases, the chances that they will demonstrate similar biological activities increases[28–31].

Materials and Methods

As mentioned earlier, we have implemented two independent workflows in the current work based on the principle of 'neighbourhood behaviour'. The first workflow (workflow-I) involves identifying potential repurpose-able candidates through recognizing structurally similar proteins. The second workflow (workflow-II) involves identification of potential

repurpose-able candidates through analysing chemical similarity of drugs/drug candidates with known inhibitors of SARS-CoV-2 M^{pro}. The workflows we have followed so far are described below.

Workflow-I: This approach can be simplified into five basic steps and has been pictorially represented in Fig.-1.

Step 1: This step involved search for structural neighbours of SARS-CoV-2 M^{pro} (PDB code: 6Y84) using the DALI[32] server (http://ekhidna2.biocenter.helsinki.fi/dali/). DALI is a protein structure comparison server which takes three dimensional (3D) coordinates of a protein as input in PDB format and then searches the PDB for similar structures (structural neighbours) following which it returns a list of structural neighbours (hits), structural alignments and superimposed structures. The hits (indicated by corresponding PDB code, chain code and protein name) are sorted by Z-score in the output file. Similarities with a Z-score lower than 2 are spurious.

Step 2: Here, the reliable hits (Z-score $\geq$ 2) obtained from the previous step are clustered based on their identity. The PDB entries of the hits are mapped on to their corresponding Uniprot[33] accession code using the PDB advanced search tool. Identical Uniprot codes corresponding to multiple PDB entries indicate that the sequence of these proteins are identical (or the parent protein from which the constructs have been derived are identical, in cases when a mutation is introduced in the experimentally determined structure deposited in the PDB).

Step 3: The hits for which Uniprot accession codes could be obtained are then searched in the DrugBank[34] (version 5.1.5) database to check if there is a known molecule associated with the corresponding protein. DrugBank is a curated hub of comprehensive information (description, targets, chemical structure, known use, pharmacodynamic and pharmacokinetic profiles, etc.) on drugs/drug candidates which may fall in any one or multiple following categories: ‘Approved’ (the molecules which have successfully cleared

the clinical trial for an indication and is approved for treatment for the particular indication), 'Investigational' (the molecules which are under clinical trial for at least one indication), 'Experimental' (these are generally the molecules which are in the pre-clinical development stage), 'Vet Approved' (the molecules which are approved for treatment against the indicated veterinary disease), 'Withdrawn' (the molecules which were once approved but have been withdrawn due to toxicity related or commercial reasons).

Step 4: The drug card/s (detailed record of each molecule in DrugBank) of those molecules associated with the proteins filtered from the above step are analysed in this step to extract relevant details. To assess the chemical similarity of the selected small molecules (obtained from DrugBank) with reported SARS-CoV-2 M^{pro} inhibitors, *viz.*, 'O6K' bound to SARS-CoV-2 M^{pro} (PDB codes: 6Y2F and 6Y2G) and 'X77' bound to SARS-CoV-2 M^{pro} (PDB code: 6W63), Tanimoto coefficients (TC1 and TC2, respectively) between each pair of shortlisted DrugBank molecule and 'O6K' followed by 'X77' was calculated using RDKit implemented in an in-house python code. Tanimoto coefficients ranges between 0 and 1. Higher the value, greater is the chemical similarity between two compounds in comparison.

Step 5: Selected compounds are then subjected to docking simulations using Autodock Vina[35] to predict whether those molecules could be favourably accommodated in the binding pocket of SARS-CoV-2 M^{pro} where 'O6K' and 'X77' are shown to bind in the respective crystal structures. In all the docking runs, only the flexibility of the ligands was considered. The binding site residues were considered as rigid. The docking protocols were validated through re-docking experiment to ensure that the docking algorithm is able to reproduce the bound pose of the native ligands ('O6K' and 'X77') present in the respective crystal structures. Dimension of the grid box used for docking was set as 12Å, 12Å and 14Å in x, y and z direction respectively. The grid spacing was set at 1Å and the x, y and z co-ordinates for the centre of the grid boxes were chosen at (-20.396, 18.376 and -27.228). 20 binding modes per ligand were generated with an energy range of 9 kcal/mol.

Workflow-II: Like workflow-I, this workflow is also a five step process and has been shown in Fig.-2. Each step is described below.

Step 1 and 2: These steps involve calculating the Tanimoto coefficient of DrugBank molecules which belong to either or both the category: 'Approved', 'Investigational' with respect to 'O6K' and 'X77'. A cut-off of 0.5 for Tanimoto coefficients was chosen to select molecules for further analysis. In the current report, we have considered the molecules which have TC1 > 0.55 and TC2 > 0.55 for discussion.

Step 3 and 4: The selected molecules were then clustered based on the chemical class as denoted in the 'Taxonomy' field of respective drug card available from DrugBank database (whenever such information is not available in 'Taxonomoy' field, 'Description'/ 'Category' fields were inspected for relevant information). At least one molecule from each chemical class which has a molecular weight < 850 Dalton has been prioritized for docking simulation so far. The docking protocol used here is same as that mentioned in workflow-I.

Step 5: This step involves analysis and compilation of data similar to workflow-I.

Results (so far) and Discussion

Workflow-I

The search for structural neighbours of SARS-CoV-2 M^{pro} yielded 3001 reliable hits which correspond to ~2300 unique PDB entries. These 2300 PDB entries could be mapped to 373 unique Uniprot entries corresponding to 162 different organisms. 101 out of these 373 proteins belong to various groups of viruses (Alphavirus: 2; Flaviviridae: 56; Nidovirales: 32; Norwalk virus: 2; Picornaviridae: 5; Potyvirus: 2; Bacteriophage P42D: 1). Remaining 272 proteins belong to cellular organisms (Bacteria: 41; Eukaryota: 231). Notably, 95 out of the 272 cellular organisms are humans. Out of the 373 proteins, we could find at least one DrugBank molecule associated with ~75 proteins. These ~75 cases have been taken forward further analysis.

The above statistics implies that the SARS-CoV-2 M^{pro} is structurally similar to 373 proteins encoded in the genome 162 different organisms for which at least one 3D structure is available in the PDB. Further, although sequence search using BLAST did not yield any significant human relative of SARS-CoV-2 M^{pro} but structural search against the PDB has yielded some human proteins as hits. This suggests that although the sequence of SARS-CoV-2 M^{pro} is very diverse from any protein encoded by human genome, but its structure is similar to some of the human proteins. Consequently, inhibitors targeted against SARS-CoV-2 M^{pro} might also have the potential to bind to some of the human proteins which have been obtained as structural neighbours of SARS-CoV-2 M^{pro} in our analysis and consequently may pose side-effect related risks. However, in-depth analysis involving comparison of local structures is required to understand the possibilities of such off-target interactions. To minimize the chances of toxicity due to cross-reactivity with host proteins, we are now mainly focussing on analysing the details of the small molecules for which the primary target/s is listed as non-human protein/s in the DrugBank database.

So far, we have analysed data associated with 20 (out of 75 proteins) proteins; 17 among these 20 are non-human proteins *(Table S1)*. Overall, these 20 proteins are known to be the targets for 58 unique DrugBank molecules which are either FDA approved drugs and/investigational and/ experimental molecule. Many of these molecules belong to diverse therapeutic areas, such as anticoagulants, anti-inflammatory and anti-viral agents. However, for few of these molecules, such information is not available. Curiously, seven molecules which possess higher chemical similarity (≥ 0.5) than most other molecules in the current dataset are all known antiviral agents (Table 1). Chemical structures of these seven molecules could be found in the Supplementary Information (Fig.-S1). Three of these 7 compounds were considered for docking simulation. The predicted binding affinity as calculated from the docked poses indicate that these molecules could be favourably accommodated (Table 1; Fig-

3) in SARS-CoV-2 M^{pro} binding pocket which is occupied by the known inhibitors 'O6K' (in 6Y2F, 6Y2G) and 'X77' (in 6W6C).

Interestingly, remdesivir (which was originally developed for treating Ebola infection[36], one of the most discussed antiviral agents which is reported to be effective in treating COVID-19[37,38] (as on 30th March 2020, 17 publications are listed in LitCovid when searched with the keyword "remdesivir"), has been picked up in our analysis too. This emphasizes the strength of basic principles of neighbourhood behaviour exploited in our computational approach of predicting potential candidates for repurposing against the deadly pathogen SARS-CoV-2. Remdesivir is known to inhibit SARS-CoV replicase polyprotein 1ab and RNA-directed RNA polymerase L of Zaire ebolavirus (strain Mayinga-76). The sequence of SARS-CoV replicase polyprotein 1ab is 96% identical to SARS-CoV-2 M^{pro} . Overlay of SARS-CoV replicase polyprotein 1ab structure onto SARS-CoV-2 main protease shows that the binding site of these two proteins are conserved (Fig.-4). Therefore, a molecule (e.g. remdesivir) which inhibits to SARS-CoV replicase polyprotein 1ab will most likely bind to SARS-CoV-2 main protease and subsequently arrest its activity. Our analysis also shows that non-steroidal anti-inflammatory drugs (NSAIDs), like dexibruprofen and nafamostat, might also be potential candidates for inhibiting SARS-CoV-2 main protease. Although there has been some confusion regarding usage of NSAIDs during coronavirus infections may aggravate the diseased conditions, but there is no scientific experimental basis to it and this have been explained in a recent article[39] by Garret A. FitzGerald. Interestingly, benefits of combining antiviral and anti-inflammatory agents to treat COVID-19 has been reported[40] by Justin Stebbing and co-workers. Also, there are publications which argue that ibuprofen can be helpful in lung infections as observed in some bacterial/viral infections by reducing the amount of inflammation, which causes damage to the lungs[41]. Excitingly, Hoffmann and co-workers very recently showed that the human cellular serine protease TMPRSS2 primes SARS-2-S for entry and the serine protease inhibitor camostat (an

analogue of nafamostat) effectively blocks SARS-CoV-2 infection of lung cells[42]. The potential of nafamostat to block MERS-CoV infection has been demonstrated in the past[43]. Both camostat and nafamostat were developed in Japan as treatments for pancreatitis and some other diseases (https://www.u-tokyo.ac.jp/focus/en/articles/z0508_00083.html). Further, ribavirin is also one of the hits identified in our analysis whose anti-coronavirus potential has also been mentioned in some of the recent literatures[37,44]. Our results suggest that besides the indicated targets of remdesivir, nafamostat and ribavirin in the respective literatures, SARS-CoV-2 M^{pro} could also be one of the potential targets of these drugs and probably contributing to multiple levels of viral arrest. This demands further in-depth investigation to have a clear understanding of the molecular mechanism of action of these drugs in SARS-CoV-2 infection. Analysis of rest of the hits are on-going.

Workflow-II

The hunt for chemically similar molecules as those of the known SARS-CoV-2 M^{pro} inhibitors ('O6K' and 'X77') using the protocol as described earlier resulted in identification of 86 approved and/investigational DrugBank molecules (Table S2). Some of these molecules are nutraceuticals and are unlikely to be good drug candidates and hence have not be considered for the current analysis. Information on chemical class could be obtained for 75 molecules which revealed that majority of the hits belong to the class of alkaloids (~37%) followed by tetracyclines (~23%), carboxylic acid derivates/peptide analogue (~12%) and others (~28%) (Fig.-5). Docking simulation of few selected hits indicated that these compounds have the potential to be favourably accommodated within the binding pocket of our interest in SARS-CoV-2 M^{pro} (Table 2; Fig.-3, Fig.-S1). Among the alkaloids, camptothecin derivatives and vinca alkaloids are the major sub-classes of alkaloids which we have been obtained as hits in the search of chemical neighbours of known inhibitors of SARS-CoV-2 M^{pro}. This is particularly exiting to us as the benefits of natural products in several therapeutic areas have been demonstrated through time-tested traditional medicinal

practices and folklore. Moreover, natural products have inspired the development most of the modern day synthetic drugs[45].

Vinca and camptothecin alkaloids have cytotoxic properties and are known for their anticancer properties. They are the active ingredients in many semi-synthetic anticancer drugs[46,47]. Vinca alkaloids are obtained from several species of Vinca genus and periwinkle (*Catharanthus roseus*) plant. While vinca pants are native to Europe, northwest Africa and southwest Asia, periwinkle is native to Madagascar. However, periwinkle is widely cultivated and is naturalised in subtropical and tropical areas of the world like Australia, Malaysia, India, Pakistan and Bangladesh (source of information: Wikipedia). A recent literature on repurposing drugs against interface of S-protein:ACE2 protein identified vidarabine (a vinca alkaloid) as a potential candidate which can be repurposed for treatment of SARS-CoV-2 infections[48]. Camptothecin is isolated from the bark and stem of *Camptotheca acuminata*, a tree native to China used as a cancer treatment in Traditional Chinese Medicine (source of information: Wikipedia). Plants in the genus Ophiorrhiza which grow in South-Western ghats of India show the presence of significant amount of camptothecin. *Ophiorrhiza mungos* is traditionally used in anticancer treatment in Ayurveda[49]. It is interesting to note that potent inhibition of herpes virus by camptothecins could be found in earlier literature[50]. Taken together, our results in the light of earlier reports demand further investigation to probe into the anti-coronavirus properties of these alkaloids obtained naturally from plant sources. The list of compounds which we have been prioritized for extended analysis from workflow-II are mentioned in Table 2 and S2.

Conclusions at the current stage of the project

The approach used in this study is a very general and a simple one which could ideally be applied for any proteins against which repurposed drugs are aimed to be targeted. So far, our analysis has resulted in identification of two sets of drugs/drug candidates. From workflow-I,

we have identified 58 DrugBank molecules which might have the potential to bind to SARS-CoV-2 main protease. Some of these findings need further in-depth analysis and based on the outcome, we might re-think on our priorities. Out of these 58 molecules, 19 molecules are approved by FDA for treating one or multiple therapeutic indications (two of these 19 are not truly FDA approved molecule. One of these two, ximelagatran, have been withdrawn from market post approval and another, nafamostat, is not yet approved by USFDA but is approved in Asian countries and widely prescribed in Japan indicating the availability of safety data (https://www.u-tokyo.ac.jp/focus/en/articles/z0508_00083.html). The 7 most promising candidates obtained form worklow-I are listed in Table 1. All of these have known antiviral properties. From workflow-II, 16 representative compounds are prioritized to be taken forward for further analysis (Table 2). These compounds are either approved and/investigational molecules. As mentioned earlier, the results discussed in this report are from our preliminary findings obtained in the last two weeks solely from *in silico* approaches involving minimum computational resources and need further rigorous evaluation in experimental and clinical set up. Given the necessity to meet the global crisis of combating COVID-19, we believe our preliminary findings could serve as initial reasonable handle to the experimentalists to start with the testing of these molecules. With this vision, we are providing access to the supplementary files containing considerable amount of our raw data and more details (like PubChemID, SMILES code) on each molecule which might be relevant in procuring these compounds from established vendor.

Our future work would involve in-depth analysis of the full dataset of ~75 proteins from workflow-I and their associated drugs/drug candidates and extending the study to other SARS-CoV-2 proteins. We would also like to look into the available resources of phytochemicals which are structurally similar to the shortlisted hits from workflow-I and further assess the potential of all these molecules to bind to SAR-CoV-2 proteins using docking simulations. Additionally, we would also conduct similar analyses on bound ligands

of the PDB entries which are structural neighbours of the proteins under study. Further, as we get access to more resources, we would set up large scale virtual screening of database compounds to identify novel inhibitors of SARS-CoV-2 proteins. The docked poses of the molecules shortlisted from both the workflows will also be subjected to rigorous analysis with respect to assessment of biological relevance of the poses in terms of stability of the poses and residence time (using molecular dynamics simulations). The aim of the entire analyses would be to rationally expand the chemical space of the ligands which could potentially bind to the proteins of the pathogen. We also look forward to collaborations for validating our predictions. It is our personal view that experimentalists who would like to test our predictions, should first consider performing cell-based assays to assess the feasibility of success of these molecules in cellular environment. The current crisis demands evaluation of the computational predictions in experimental setups which more closely mimics physiological conditions. Therefore, post cell-based assays, promising candidates can be evaluated through *in vitro* assays to understand the molecular mechanism of action and pave the path for compound optimization.

P.S.: As mentioned earlier, we once again emphasize that the sole purpose of making these preliminary findings openly available to the scientific community is to aid the ongoing research efforts. These results might serve as a reasonable starting point for experimental biologists, biomedical researchers and clinicians to investigate and evaluate our predictions in experimental and clinical set ups for anti-COVID-19 treatment.

WARNING: Our findings should only be used for research purposes and we strongly urge that no individual should interpret these findings for any self-diagnosis or self-medication without the prior approval from competent international health/medical regulatory agencies.

Acknowledgements

Research in our group is generally supported by Mathematical Biology program and FIST program sponsored by the Department of Science and Technology and also by the Department of Biotechnology, Government of India in the form of IISc-DBT partnership programme. Support from UGC, India – Centre for Advanced Studies and Ministry of Human Resource Development, India is gratefully acknowledged. SC acknowledges the financial support by Department of Science & Technology, Government of India towards her research through DST-INSPIRE fellowship. NS is a J. C. Bose National Fellow. The authors are thankful to Ms. Yazhini Arangasamy for providing technical assistance remotely.

References:

1. Pushpakom, S. *et al.* Drug repurposing: progress, challenges and recommendations. *Nat. Rev. Drug Discov.* **18**, 41–58 (2019).
2. Fitchett, J. R., Head, M. G., Cooke, M. K., Wurie, F. B. & Atun, R. Funding infectious disease research: a systematic analysis of UK research investments by funders 1997-2010. *PLoS One* **9**, e105722–e105722 (2014).
3. Furuse, Y. Analysis of research intensity on infectious disease by disease burden reveals which infectious diseases are neglected by researchers. *Proc. Natl. Acad. Sci. U. S. A.* **116**, 478–483 (2019).
4. Park, K. A review of computational drug repurposing. *Transl. Clin. Pharmacol.* **27**, 59–63 (2019).
5. Law, G. L., Tisoncik-Go, J., Korth, M. J. & Katze, M. G. Drug repurposing: a better approach for infectious disease drug discovery? *Curr. Opin. Immunol.* **25**, 588–592 (2013).
6. Ramakrishnan, G., Chandra, N. R. & Srinivasan, N. Recognizing drug targets using evolutionary information: implications for repurposing FDA-approved drugs against

Mycobacterium tuberculosis H37Rv. *Mol. Biosyst.* **11**, 3316–3331 (2015).

7. Ramakrishnan, G., Chandra, N. & Srinivasan, N. Exploring anti-malarial potential of FDA approved drugs: an in silico approach. *Malar. J.* **16**, 290 (2017).
8. Chakraborti, S., Ramakrishnan, G. & Srinivasan, N. Chapter 16 - In Silico Modeling of FDA-Approved Drugs for Discovery of Anticandida Agents: A Drug-Repurposing Approach. in (ed. Roy, K. B. T.-I. S. D. D.) 463–526 (Academic Press, 2019). doi:https://doi.org/10.1016/B978-0-12-816125-8.00016-X.
9. Chakraborti, S., Ramakrishnan, G. & Srinivasan, N. Repurposing Drugs Based on Evolutionary Relationships Between Targets of Approved Drugs and Proteins of Interest BT - Computational Methods for Drug Repurposing. in (ed. Vanhaelen, Q.) 45–59 (Springer New York, 2019). doi:10.1007/978-1-4939-8955-3_3.
10. Chakraborti, Sohini; Chakravarthi, Pushpaveni Srinivasan, N. A ligand-centric approach to identify potential drugs for repurposing: Case-study with Aurora Kinase inhibitors. in *Drug Repurposing in Cancer Therapy: Approaches and Applications* (eds. To, K. & William, C.) (2020).
11. Berman, H. M. *et al.* The Protein Data Bank. *Nucleic Acids Res.* **28**, 235–242 (2000).
12. Zhang, B., Zhao, Y., Jin, Z., Liu, X., Yang, H., Rao, Z. The crystal structure of COVID-19 main protease in apo form. (2020) doi:10.2210/pdb6M03/pdb.
13. Owen, C.D., Lukacik, P., Strain-Damerell, C.M., Douangamath, A., Powell, A.J., Fearon, D., Brandao-Neto, J., Crawshaw, A.D., Aragao, D., Williams, M., Flaig, R., Hall, D., McAauley, K., Stuart, D.I., von Delft, F., Walsh, M. A. COVID-19 main protease with unliganded active site (2019-nCoV, coronavirus disease 2019, SARS-CoV-2). (2020) doi:10.2210/pdb6Y84/pdb.
14. Zhang, L. *et al.* Crystal structure of SARS-CoV-2 main protease provides a basis for design of improved α-ketoamide inhibitors. *Science (80-.).* (2020) doi:10.1126/science.abb3405.

15. Zhenming Jin, Xiaoyu Du, Yechun Xu, Yongqiang Deng, Meiqin Liu, Yao Zhao, Bing Zhang, Xiaofeng Li, Leike Zhang, Chao Peng, Yinkai Duan, Jing Yu, Lin Wang, Kailin Yang, Fengjiang Liu, Rendi Jiang, Xinglou Yang, Tian You, Xiaoce Liu, Xiuna Yang, Fang Bai, H, H. Y. Structure of Mpro 1 from COVID-19 virus and discovery of its inhibitors. *bioRxiv* (2020) doi:10.1101/2020.02.26.964882.

16. Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik, P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F. PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z45617795. (2020) doi:10.2210/pdb5R7Y/pdb.

17. Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik, P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F. PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z1220452176. (2020) doi:10.2210/pdb5R7Z/pdb.

18. Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik, P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F. PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z18197050. (2020) doi:10.2210/pdb5R80/pdb.

19. Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik, P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F. PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z1367324110. (2020) doi:10.2210/pdb5R81/pdb.

20. Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik, P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F. PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z219104216. (2020) doi:10.2210/pdb5R82/pdb.

21. Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik,

P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F. PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z44592329. (2020) doi:10.2210/pdb5R83/pdb.

22. Fearon, D., Powell, A.J., Douangamath, A., Owen, C.D., Wild, C., Krojer, T., Lukacik, P., Strain-Damerell, C.M., Walsh, M.A., von Delft, F. PanDDA analysis group deposition -- Crystal Structure of COVID-19 main protease in complex with Z31792168. (2020) doi:10.2210/pdb5R84/pdb.
23. Mesecar, A. D. X77, Structure of COVID-19 main protease bound to potent broad-spectrum non-covalent inhibitor. (2020) doi:10.2210/pdb6W63/pdb.
24. Altschul, S. F., Gish, W., Miller, W., Myers, E. W. & Lipman, D. J. Basic local alignment search tool. *J. Mol. Biol.* **215**, 403–410 (1990).
25. Orengo, C. A., Jones, D. T. & Thornton, J. M. Protein superfamilles and domain superfolds. *Nature* **372**, 631–634 (1994).
26. Wood, T. C. & Pearson, W. R. Evolution of protein sequences and structures11Edited by J. M. Thornton. *J. Mol. Biol.* **291**, 977–995 (1999).
27. Illergård, K., Ardell, D. H. & Elofsson, A. Structure is three to ten times more conserved than sequence—A study of structural response in protein cores. *Proteins Struct. Funct. Bioinforma.* **77**, 499–508 (2009).
28. Maggiora, G., Vogt, M., Stumpfe, D. & Bajorath, J. Molecular Similarity in Medicinal Chemistry. *J. Med. Chem.* **57**, 3186–3204 (2014).
29. Willett, P., Barnard, J. M. & Downs, G. M. Chemical Similarity Searching. *J. Chem. Inf. Comput. Sci.* **38**, 983–996 (1998).
30. Martin, Y. C., Kofron, J. L. & Traphagen, L. M. Do Structurally Similar Molecules Have Similar Biological Activity? *J. Med. Chem.* **45**, 4350–4358 (2002).
31. Patterson, D. E., Cramer, R. D., Ferguson, A. M., Clark, R. D. & Weinberger, L. E. Neighborhood Behavior: A Useful Concept for Validation of "Molecular Diversity"

Descriptors. *J. Med. Chem.* **39**, 3049–3059 (1996).

32. Holm, L. Benchmarking fold detection by DaliLite v.5. *Bioinformatics* **35**, 5326–5327 (2019).

33. Consortium, T. U. UniProt: a worldwide hub of protein knowledge. *Nucleic Acids Res.* **47**, D506–D515 (2018).

34. Wishart, D. S. *et al.* DrugBank: a comprehensive resource for in silico drug discovery and exploration. *Nucleic Acids Res.* **34**, D668–D672 (2006).

35. Trott, O. & Olson, A. J. AutoDock Vina: improving the speed and accuracy of docking with a new scoring function, efficient optimization, and multithreading. *J. Comput. Chem.* **31**, 455–461 (2010).

36. Tchesnokov, E. P., Feng, J. Y., Porter, D. P. & Götte, M. Mechanism of Inhibition of Ebola Virus RNA-Dependent RNA Polymerase by Remdesivir. *Viruses* **11**, (2019).

37. Wang, M. *et al.* Remdesivir and chloroquine effectively inhibit the recently emerged novel coronavirus (2019-nCoV) in vitro. *Cell Res.* **30**, 269–271 (2020).

38. Kupferschmidt, K.;Cohen, J. Race to find COVID-19 treatments accelerates. *Science* 1412–1413 (2020) doi:10.1126/science.367.6485.1412.

39. FitzGerald, G. A. Misguided drug advice for COVID-19. *Science* (2020) doi:10.1126/science.abb8034.

40. Justin Stebbing, Anne Phelan, I. G. & Catherine Tucker, Olly Oechsle, Dan Smith, P. R. COVID-19: combining antiviral and anti-inflammatory treatments. *Lancet Infect Dis* (2020) doi:10.1016/ S1473-3099(20)30132-8.

41. Lands, L. C. & Dauletbaev, N. High-Dose Ibuprofen in Cystic Fibrosis. *Pharmaceuticals* vol. 3 (2010).

42. Hoffmann, M. *et al.* SARS-CoV-2 Cell Entry Depends on ACE2 and TMPRSS2 and Is Blocked by a Clinically Proven Protease Inhibitor. *Cell* (2020) doi:https://doi.org/10.1016/j.cell.2020.02.052.

43. Yamamoto, M. *et al.* Identification of Nafamostat as a Potent Inhibitor of Middle East Respiratory Syndrome Coronavirus S Protein-Mediated Membrane Fusion Using the Split-Protein-Based Cell-Cell Fusion Assay. *Antimicrob. Agents Chemother.* **60**, 6532 LP – 6539 (2016).

44. Gordon, D. E. *et al.* A SARS-CoV-2-Human Protein-Protein Interaction Map Reveals Drug Targets and Potential Drug-Repurposing. *bioRxiv* 2020.03.22.002386 (2020) doi:10.1101/2020.03.22.002386.

45. Mushtaq, S., Abbasi, B. H., Uzair, B. & Abbasi, R. Natural products as reservoirs of novel therapeutic agents. *EXCLI J.* **17**, 420–451 (2018).

46. RL, N. The discovery of the vinca alkaloids--chemotherapeutic agents against cancer. *Biochem Cell Biol.* **68**, 1344–51 (1990).

47. Venditto, V. J. & Simanek, E. E. Cancer therapies utilizing the camptothecins: a review of the in vivo literature. *Mol. Pharm.* **7**, 307–349 (2010).

48. Smith, M. & Smith, J. C. Repurposing Therapeutics for COVID-19: Supercomputer-Based Docking to the SARS-CoV-2 Viral Spike Protein and Viral Spike Protein-Human ACE2 Interface. (2020) doi:10.26434/chemrxiv.11871402.v3.

49. Abdul, Jaleel.K;Malarkodi, V. Anti-Cancer Activity of Ophiorrhiza Species Endemic toSouthern Western Ghats: A Review. *J. Pharm. Sci. Res.* **11**, 1156–1159 (2019).

50. Rastogi, R. P. & Dhawan, B. N. Anticancer and antiviral activities in indian medicinal plants: A review. *Drug Dev. Res.* **19**, 1–12 (1990).

Legends to Figures:

Figure 1: Workflow-I. The step 1 to 5 as discussed in the text have been graphically represented depicting the search for structural neighbours of SARS-CoV-2 main protease in Protein Data Bank and thereafter finding if any DrugBank molecule could be associated with those structural neighbours using different computational tools. Once such association is

made, the data is analysed mainly manually (due to availability of limited computational resources at home and occasional remote access to desktop in our laboratory) to prioritize the compounds for testing. Selected compounds are subjected to docking simulation in step 5. The three domains in SARS-CoV-2 main protease protein structure are highlighted in different colours (left topmost protein structure in surface representation: light blue – domain I: residue 10 -99; purple-domain II: residue 100-182; cyan-domain III: residue: 198-303, green-disordered regions; this information is derived from reference 4).

Figure 2: Workflow-II. The step 1 to 5 as discussed in the text have been graphically represented depicting the search for chemical neighbours of reported inhibitors of SARS-CoV-2 main protease in DrugBank. A cut-off of 0.5 for TC (tanimoto coefficient) is chosen to filter the hits which are then clustered based on their chemical class and at least one representative from each chemical class is chosen for docking simulation followed by data analysis as explained in the legend to figure 1.

Figure 3: Docked poses of few molecules in SARS-CoV-2 main protease binding pocket obtained from workflow-I (upper) and workflow-II (lower). The protein binding site is shown in surface representation (grey) and the ligands are shown as ball and stick models (a) Remdesivir (orange) (b) Nafamostat (teal) (c) Ribavirin (red) (d) Ciluprevir (black) (e) Zoliflodacin (yellow) (f) Bictegravir (cyan) (g) Vindesine (pink) (h) Irinotecan (violet).

Figure 4: Comparison of SARS-CoV replicase polyprotein 1ab (PDB code: 1WOF, blue) and SARS-CoV-2 main protease (PDB code: 5R83, green). (a) Overlay of the two proteins in cartoon representation which shows that these two proteins superimpose very well (the structural alignment has been done using TM-align server; TM-Score: 0.97, RMSD: 1.05Å). The region within the box encloses the inhibitor binding site. The inhibitor ‘O6K’ from PDB entry: 6Y2F have been show in stick representation (white carbon). The bound ligand of 5R83 (‘K0G’) and the protein structure from PDB entry 6Y2F have not been displayed for visual clarity. (b) Zoomed-in view of the inhibitor binding site with the binding site residues

in line representation (c) Overlay of ‘I12’, bound ligand of 1WOF and ‘O6K’ in SARS-CoV-2 main protease binding site shown as green surface.

Figure 5: Percentage distribution of different classes of chemical classes of compounds obtained as hits from workflow-II. The respective percentages are show inside the pie-chart.

Figure S1: Chemical structures of molecules listed in table 1 and 2.

Caption to tables:

Table 1: Comprehensive details of seven potential molecules for repurposing against SARS-CoV-2 main protease shortlisted from workflow-I

Table 2: Comprehensive details on shortlisted high priority drugs to be considered for further evaluation as obtained from workflow -II

Table S1: Detailed information on analysis of 20 structural neighbours and their associated DrugBank molecules.

Table S2: Detailed information on analysis of 86 chemical neighbours and their associated known targets.

Table 1: Comprehensive details of seven potential molecules for repurposing against SARS-CoV-2 main protease shortlisted from workflow-I

Sl.no.	Name of DrugBank molecule	DrugBank ID	Drug group	TC1, TC2	Known category/description/indication	Pharmacological action	Target's Name (Uniprot Acc. code; *Organism*)	%identity with SARS-CoV-2 main protease (Q. cov%; E-val)	Predicted Binding Affinity (Vina score of top ranked pose in kcal/mol)
1	Remdesivir	DB14761	Investigational	0.54, 0.53	Antiviral agents	inhibitor	Replicase polyprotein 1ab (P0C6X7; Human SARS coronavirus)	96.08 (100; 0)	-7.7
2	Asunaprevir	DB11586	Approved, Investigational, Withdrawn	0.51, 0.50	HCV NS3 protease inhibitor/HIV Protease Inhibitors	inhibitor	Genome polyprotein (P26663; Hepatitis C virus genotype 1b (isolate BK) (HCV))	N/D	N/P
3	Ciluprevir	DB05868	Investigational	0.51, 0.52	HCV NS3 protease inhibitor	N/A	Genome polyprotein (P26664; Hepatitis C virus genotype 1a (isolate 1) (HCV))	N/D	-6.4
4	Simeprevir	DB06290	Approved	0.53, 0.53	Antiviral Agents/Protease Inhibitors	inhibitor	NS3 protease (Q91RS4; Hepacivirus C)	N/D	N/A
5	Danoprevir	DB11779	Investigational	0.52, 0.49	NS3/4A protease inhibitor/Cytochrome P-450 Enzyme Inhibitors	N/A	Genome polyprotein (P26664; Hepatitis C virus genotype 1a (isolate 1) (HCV))	N/D	N/A
6	Glecaprevir	DB13879	Approved, Investigational	0.54, 0.54	Antiviral Agents/NS3/4A Protease Inhibitors	inhibitor	NS3 protease (Q91RS4; Hepacivirus C)	N/D	-5.8
7	3-(1,1-dioxido-4H-1,2,4-benzothiadiazin-3-yl)-4-hydroxy-1-(3-methylbutyl)quinolin-2(1H)-one	DB07275	Experimental	0.51, 0.51	N/A (Targets viral protein)	N/A	Genome polyprotein (Q99AU2; Hepatitis C virus subtype 1b)	N/D	N/A

TC1, TC2: Tanimoto Coefficient1/2 (for details refer Materials and methods); N/A : Not Available; N/D: Not detected in BLAST search; Q-cov: Query coverage; E-val: E-value

Table 2: Comprehensive details on shortlisted high priority drugs to be considered for further evaluation as obtained from workflow-II

Sl. No.	DrugBank ID	Name	Status	Use	Chemical class	Predicted Binding affinity (Vina score of top ranked pose; kcal/mol)
1	DB00696	Ergotamine	Approved	Sympatholytic Agents (antimigraine preparations)	Ergoline and derivatives (Alkaloids)	-9.7
2	DB12817	Zoliflodacin	Investigational	Antibacterial Agents	Quinolines and derivatives	-9.1
3	DB12225	Beclabuvir	Investigational	Antiviral Agents	Indoles and derivatives	-8.9
4	DB01200	Bromocriptine	Approved, Investigational	Anti-Parkinson Agents	Ergoline and derivatives (Alkaloids)	-8.3
5	DB11799	Bictegravir	Approved, Investigational	Antiviral Agents	Pyridines and derivatives	-8.2
6	DB12939	Nikkomycin Z	Investigational	Antifungal Agents	Carboxylic acids (peptides and analogues)	-8.1
7	DB00595	Oxytetracycline	Approved, Investigational, Vet approved	Antibacterial Agents	Tetracyclines	-7.3
8	DB12691	UK-432,097	Investigational	Treatment for Pulmonary Disease, Chronic Obstructive.	Purine nucleosides	-7.3
9	DB00762	Irinotecan	Approved, Investigational	Antineoplastic and Immunomodulating Agents	Camptothecins (Alkaloids and derivatives)	-7.2
10	DB00309	Vindesine	Approved, Investigational	Antineoplastic and Immunomodulating Agents	Vinca alkaloids	-7.1
11	DB00541	Vincristine	Approved, Investigational	Antineoplastic and Immunomodulating Agents	Vinca alkaloids	-7.0
12	DB00305	Mitomycin	Approved	Antibiotics, Antineoplastic	Indoles and derivatives	-6.9
13	DB09050	Ceftolozane	Approved, Investigational	Antibacterial Agents	Lactams	-6.8
14	DB00570	Vinblastine	Approved	Antineoplastic and Immunomodulating Agents	Vinca alkaloids	-6.6
15	DB11641	Vinflunine	Approved, Investigational	Antineoplastic and Immunomodulating Agents	Vinca alkaloids	-6.5
16	DB00361	Vinorelbine	Approved, Investigational	Antineoplastic and Immunomodulating Agents	Vinca alkaloids	-6.3