

The SARS-CoV-2 replication-transcription complex is a priority target for broad-spectrum pan- coronavirus drugs

*Setayesh Yazdani*¹, *Nicola De Maio*², *Yining Ding*¹, *Vijay Shahani*³, *Nick Goldman*², *Matthieu Schapira*^{1,3,4,*}

¹Structural Genomics Consortium, University of Toronto, Toronto, ON M5G 1L7, Canada

²European Molecular Biology Laboratory, European Bioinformatics Institute, Hinxton, United Kingdom

³Cyclica, Toronto, ON M5J 1A7, Canada

⁴Department of Pharmacology and Toxicology, University of Toronto, Toronto, ON M5S 1A8, Canada

KEYWORDS: SARS-CoV-2, druggability, genetic variability

ABSTRACT

In the absence of effective treatment, COVID-19 is likely to remain a global disease burden. Compounding this threat is the near certainty that novel coronaviruses with pandemic potential

will emerge in years to come. Pan-coronavirus drugs – agents active against both SARS-CoV-2 and other coronaviruses – would address both threats. A strategy to develop such broad-spectrum inhibitors is to pharmacologically target binding sites on SARS-CoV-2 proteins that are highly conserved in other known coronaviruses, the assumption being that any selective pressure to keep a site conserved across past viruses will apply to future ones. Here, we systematically mapped druggable binding pockets on the experimental structure of fifteen SARS-CoV-2 proteins and analyzed their variation across twenty-seven α - and β -coronaviruses and across thousands of SARS-CoV-2 samples from COVID-19 patients. We find that the two most conserved druggable sites are a pocket overlapping the RNA binding site of the helicase nsp13, and the catalytic site of the RNA-dependent RNA polymerase nsp12, both components of the viral replication-transcription complex. We present the data on a public web portal (https://www.thesgc.org/SARSCoV2_pocketome/) where users can interactively navigate individual protein structures and view the genetic variability of drug binding pockets in 3D.

The coronavirus SARS-CoV-2 that emerged in late 2019 has so far caused over two million deaths worldwide. While the recent approval of the first vaccines is expected to terminate the pandemic nature of the epidemic, it is believed that COVID-19 will not be eradicated, emphasizing the need for drugs¹. Available vaccines may be less effective against novel and future SARS-CoV-2 variants, a serious global health threat. Additionally, in years to come, novel coronavirus outbreaks are bound to emerge, as did SARS-CoV-1 in 2003, the MERS coronavirus in 2012 and SARS-CoV-2 in 2019. To simultaneously develop anti COVID-19 therapeutics and prepare for the next pandemic threat, SARS-CoV-2 drug discovery efforts should focus on the development of pan-coronavirus agents. A first strategy is to target host proteins that are essential for the replication of past coronaviruses and may therefore be essential

for the replication of coronaviruses emerging in the future². An alternative approach, and the focus of this work, is to identify drug binding sites in the SARS-CoV-2 proteome that are conserved across coronaviruses. An approved anti-COVID-19 drug binding such site could be rapidly deployed against future outbreaks of novel coronaviruses.

Recent reports provided valuable insight into the druggability of SARS-CoV-2 proteins³, on their variation in COVID-19 samples and on their divergence from SARS-CoV-1 and from a specific bat coronavirus⁴. The excellent accompanying COVID-3D web portal allows users to map mutations on SARS-CoV-2 protein structures. A powerful online interface for inspecting variations between the SARS-CoV-2 genome and other viruses was also developed⁵, but to our knowledge, no integrated analysis focusing on the genetic variability of drug binding pockets across the SARS-CoV-2 proteome was reported to date.

We analyzed the structures of all SARS-CoV-2 protein domains represented in the PDB. The proteins analyzed included non-structural proteins (nsp) nsp1, nsp3, nsp5, nsp9, nsp12, nsp13, nsp14 (a SARS-CoV-1 structure was used that shared 100% sequence identity with SARS-CoV-2 at the active site), nsp15, nsp16; structural proteins E (transmembrane domain), Spike; additional factors ORF3a, ORF7a, ORF8, ORF9b. We used the PocketFinder function⁶ in ICM (Molsoft, San Diego) to find potential druggable binding sites and assessed their druggability using SiteMap (Schrodinger, NY)⁷. The druggability evaluated with SiteMap was in general agreement with a previous work using a different method³ and varied widely from poorly druggable, such as the catalytic site of the endoribonuclease nsp15 (Druggability score [Dscore] < 0.8), to highly druggable, such as a functionally uncharacterized ectopic site of the methyltransferase nsp14 (Dscore >1.0) (Figure 1a).

Since pathogenic coronaviruses responsible for mild or severe respiratory illness all belong to the genera alphacoronavirus (α -CoV) and betacoronavirus (β -CoV), we focused on these families to define the genetic variability of binding pockets found on the SARS-CoV-2 proteome⁸. Over 400 reviewed protein sequences of 58 α - and β -CoV were downloaded from UniProt (<https://www.uniprot.org/>). Automated sequence search retrieved homologs for most SARS-CoV-2 proteins, including all nsp proteins specified above, in 27 α - and β -CoV comprising zoonotic entries (bat and bovine CoVs), clinical and epidemic CoV isolates (SARS-CoV-1, MERS, HCoV-OC43, HCoV-HKU, HCoV-229E, HCoV-NL63). We performed multiple sequence alignments and determined the percent identity and conservation of amino-acids lining each binding pocket across all 27 coronaviruses where protein homologs were found (Figure 1a, Supplementary figure 1, Supplementary Table 1). We find that some binding sites are highly conserved while others are genetically unstable. For instance, binding of the essential fatty acid linoleic acid (LA) to the receptor binding domain of the SARS-CoV-2 Spike protein stabilizes the closed/locked state of Spike, inhibits Spike binding to ACE2 and synergizes with Remdesivir to reduce SARS-CoV-2 replication, suggesting that drugs targeting this site may be of therapeutic interest⁹. While the LA binding site is druggable (Dscore > 1.0), we find that sidechains lining LA are poorly conserved (0% conservation across 27 α - and β -coronaviruses), indicating that this site is not favorable for the discovery of pan-coronavirus inhibitors (Figure 1b). Conversely, a druggable channel partially overlapping with the RNA-binding site of the SARS-CoV-2 helicase (nsp13) is highly conserved (>90% of residues lining the pocket are conserved across 27 α - and β -coronaviruses) (Figure 1b). Interestingly, nsp13 binds to the RNA-dependent RNA polymerase (nsp12) to functionally couple helicase and polymerase activities¹⁰ and we find that the catalytic site of nsp12 is also among the most conserved binding pockets in

Figure 1: Druggability and conservation of drug binding sites in SARS-CoV-2 proteins.

(a) For all binding sites, druggability scores calculated with SiteMap⁷ and conservation of pocket-lining residues across α - and β -coronaviruses are represented in a scatterplot. (b) Structural details and multiple sequence alignments are shown for the linoleic-acid binding site of the protein Spike (left, PDB code 6ZB4) which is poorly conserved, and a channel overlapping the RNA binding site of nsp13 (right, PDB code 6ZSL with RNA from structure 2XZL), which is highly conserved. Both sites are highlighted in the scatterplot (a). The full names of all viral genera are provided in Supplementary Table S1.

Our effort to systematically map the genetic variability of putative drug binding sites onto 3D structures of the SARS-CoV-2 proteome reveals medicinal chemistry strategies for the development of broad-spectrum coronavirus inhibitors. In particular, we find that pharmacologically targeting the viral replication-transcription complex is a promising avenue for the discovery of pan-coronavirus drugs. In this regard, compounds such as EIDD-280, an oral antiviral targeting the catalytic site of nsp12 and currently in phase II-III clinical trials against COVID-19¹¹, represent promising drug candidates for current and future coronavirus pandemic threats. Similarly, targeting the RNA binding site of nsp13 is a priority – though underexplored – avenue that may benefit from recent advances on the chemical inhibition of RNA-binding proteins¹².

ASSOCIATED CONTENT

Supporting Information. The following figures and tables are provided as supporting online information. Percent sequence identity and druggability of drug binding sites in the SARS-CoV-2 proteome represented in the PDB (Figure S1); mutation level of residues lining drug binding

sites found in SARS-CoV-2 proteins in the PDB across >15,000 samples from COVID-19 patients and across 27 α - and β - coronavirus genera (Figure S2); conservation matrix of SARS-CoV-2 proteome represented in the PDB across 27 α - and β - coronaviruses (Table S1). Detailed methods are also provided.

AUTHOR INFORMATION

Corresponding Author

*matthieu.schapira@utoronto.ca

Author Contributions

The manuscript was written through contributions of all authors. All authors have given approval to the final version of the manuscript.

Funding Sources

This work was supported by a grant from the Natural Sciences and Engineering Research Council of Canada awarded to MS (grant ALLRP 555329-20) and by EMBL (NDM and NG). The Structural Genomics Consortium is a registered charity (no: 1097737) that receives funds from AbbVie, Bayer AG, Boehringer Ingelheim, Genentech, Genome Canada through Ontario Genomics Institute [OGI-196], the EU and EFPIA through the Innovative Medicines Initiative 2 Joint Undertaking [EUBOPEN grant 875510], Janssen, Merck KGaA (aka EMD in Canada and US), Pfizer, Takeda and the Wellcome Trust [106169/ZZ14/Z].

ACKNOWLEDGMENT

We are grateful to Aled Edwards, Andrew Leach and Geoff Barton for insightful discussions at the outset of this work and to Dasha Redka for comments throughout the analysis. We are very grateful to GISAID and all the groups who shared their sequencing data, as detailed in <https://github.com/roblanf/sarscov2phylo/tree/master/acknowledgements>.

ABBREVIATIONS

CoV, coronavirus; nsp, non-structural protein; PDB, Protein Data Bank; Dscore: Druggability Score.

REFERENCES

- (1) Phillips, N. The Coronavirus Is Here to Stay — Here’s What That Means. *Nature* **2021**, *590* (7846), 382–384. <https://doi.org/10.1038/d41586-021-00396-2>.
- (2) Schneider, W. M.; Luna, J. M.; Hoffmann, H.-H.; Sánchez-Rivera, F. J.; Leal, A. A.; Ashbrook, A. W.; Le Pen, J.; Ricardo-Lax, I.; Michailidis, E.; Peace, A.; Stenzel, A. F.; Lowe, S. W.; MacDonald, M. R.; Rice, C. M.; Poirier, J. T. Genome-Scale Identification of SARS-CoV-2 and Pan-Coronavirus Host Factor Networks. *Cell* **2021**, *184* (1), 120-132.e14. <https://doi.org/10.1016/j.cell.2020.12.006>.
- (3) Cavasotto, C. N.; Lamas, M. S.; Maggini, J. Functional and Druggability Analysis of the SARS-CoV-2 Proteome. *European Journal of Pharmacology* **2021**, *890*, 173705. <https://doi.org/10.1016/j.ejphar.2020.173705>.
- (4) Portelli, S.; Olshansky, M.; Rodrigues, C. H. M.; D’Souza, E. N.; Myung, Y.; Silk, M.; Alavi, A.; Pires, D. E. V.; Ascher, D. B. Exploring the Structural Distribution of Genetic Variation in SARS-CoV-2 with the COVID-3D Online Resource. *Nature Genetics* **2020**, *52* (10), 999–1001. <https://doi.org/10.1038/s41588-020-0693-3>.

- (5) Fernandes, J. D.; Hinrichs, A. S.; Clawson, H.; Gonzalez, J. N.; Lee, B. T.; Nassar, L. R.; Raney, B. J.; Rosenbloom, K. R.; Nerli, S.; Rao, A. A.; Schmelter, D.; Fyfe, A.; Maulding, N.; Zweig, A. S.; Lowe, T. M.; Ares, M.; Corbet-Detig, R.; Kent, W. J.; Haussler, D.; Haeussler, M. The UCSC SARS-CoV-2 Genome Browser. *Nature Genetics* **2020**, *52* (10), 991–998. <https://doi.org/10.1038/s41588-020-0700-8>.
- (6) An, J.; Totrov, M.; Abagyan, R. Pocketome via Comprehensive Identification and Classification of Ligand Binding Envelopes. *Mol Cell Proteomics* **2005**, *4* (6), 752–761. <https://doi.org/10.1074/mcp.M400159-MCP200>.
- (7) Halgren, T. A. Identifying and Characterizing Binding Sites and Assessing Druggability. *J Chem Inf Model* **2009**, *49* (2), 377–389. <https://doi.org/10.1021/ci800324m>.
- (8) Totura, A. L.; Bavari, S. Broad-Spectrum Coronavirus Antiviral Drug Discovery. *Expert Opinion on Drug Discovery* **2019**, *14* (4), 397–412. <https://doi.org/10.1080/17460441.2019.1581171>.
- (9) Toelzer, C.; Gupta, K.; Yadav, S. K. N.; Borucu, U.; Davidson, A. D.; Williamson, M. K.; Shoemark, D. K.; Garzoni, F.; Staufer, O.; Milligan, R.; Capin, J.; Mulholland, A. J.; Spatz, J.; Fitzgerald, D.; Berger, I.; Schaffitzel, C. Free Fatty Acid Binding Pocket in the Locked Structure of SARS-CoV-2 Spike Protein. *Science* **2020**, *370* (6517), 725–730. <https://doi.org/10.1126/science.abd3255>.
- (10) Chen, J.; Malone, B.; Llewellyn, E.; Grasso, M.; Shelton, P. M. M.; Olinares, P. D. B.; Maruthi, K.; Eng, E. T.; Vatandaslar, H.; Chait, B. T.; Kapoor, T. M.; Darst, S. A.; Campbell, E. A. Structural Basis for Helicase-Polymerase Coupling in the SARS-CoV-2 Replication-Transcription Complex. *Cell* **2020**, *182* (6), 1560-1573.e13. <https://doi.org/10.1016/j.cell.2020.07.033>.

- (11) Wahl, A.; Gralinski, L. E.; Johnson, C. E.; Yao, W.; Kovarova, M.; Dinnon, K. H.; Liu, H.; Madden, V. J.; Krzystek, H. M.; De, C.; White, K. K.; Gully, K.; Schäfer, A.; Zaman, T.; Leist, S. R.; Grant, P. O.; Bluemling, G. R.; Kolykhalov, A. A.; Natchus, M. G.; Askin, F. B.; Painter, G.; Browne, E. P.; Jones, C. D.; Pickles, R. J.; Baric, R. S.; Garcia, J. V. SARS-CoV-2 Infection Is Effectively Treated and Prevented by EIDD-2801. *Nature* **2021**, 1–9. <https://doi.org/10.1038/s41586-021-03312-w>.
- (12) Wu, P. Inhibition of RNA-Binding Proteins with Small Molecules. *Nature Reviews Chemistry* **2020**, 4 (9), 441–458. <https://doi.org/10.1038/s41570-020-0201-4>.

TOC graphic

