The creation of an open access database of computer-generated poses and scores towards the discovery of SARS-CoV-2 RNA-dependent RNA polymerase inhibitors: A proposed collaboration between the University of California Berkeley, Lawrence Berkeley National Laboratory, and Atomwise.

Benjamin M. Samudio[§], Camille Huang[¥], Muhan Zhang[¥], Tony Drummond[‡¥], and Silvia Crivelli[‡¥]

[§]Atomwise, 717 Market St. #800, San Francisco, CA 94103

[‡]Lawrence Berkeley National Laboratory, 1 Cyclotron Rd., Berkeley, CA 94720

[¥]Masters of Molecular Science and Software Engineering, University of California Berkeley, Berkeley, CA 94720

Introduction

SARS-CoV-2 is an emergent human coronavirus that is the epidemiological agent of the COVID-19 pandemic. Globally and as of this writing, the pandemic has caused over 230 million cases of illness and claimed over 4.7 million lives [1]. As the pandemic continues, there are no SARS-CoV-2-specific small-molecule therapeutics. Though highly effective prophylactic vaccines have been developed against COVID-19, a sizeable portion of the population may not be able to use them. In addition, new variants continue to emerge that can potentially undermine vaccine effectiveness. These challenges motive the search for small-molecule therapeutic options.

The RNA-dependent RNA polymerase (RdRp) of SARS-CoV-2 is critical for viral replication, and inhibiting it has been shown to be a promising antiviral approach [2]. A structure of SARS-CoV-2 RdRp has recently been released [3]. This structure can be used in computer modeling to simulate the interactions of RdRp with small molecules. In this process, each molecule is virtually docked (joined) to RdRp, forming a set of structures called poses. The poses are then scored and used to predict the potential of their associated molecule to inhibit the normal functioning of RdRp. Those compounds with high predicted RdRp inhibition capacity are evaluated further as potential therapeutic candidates.

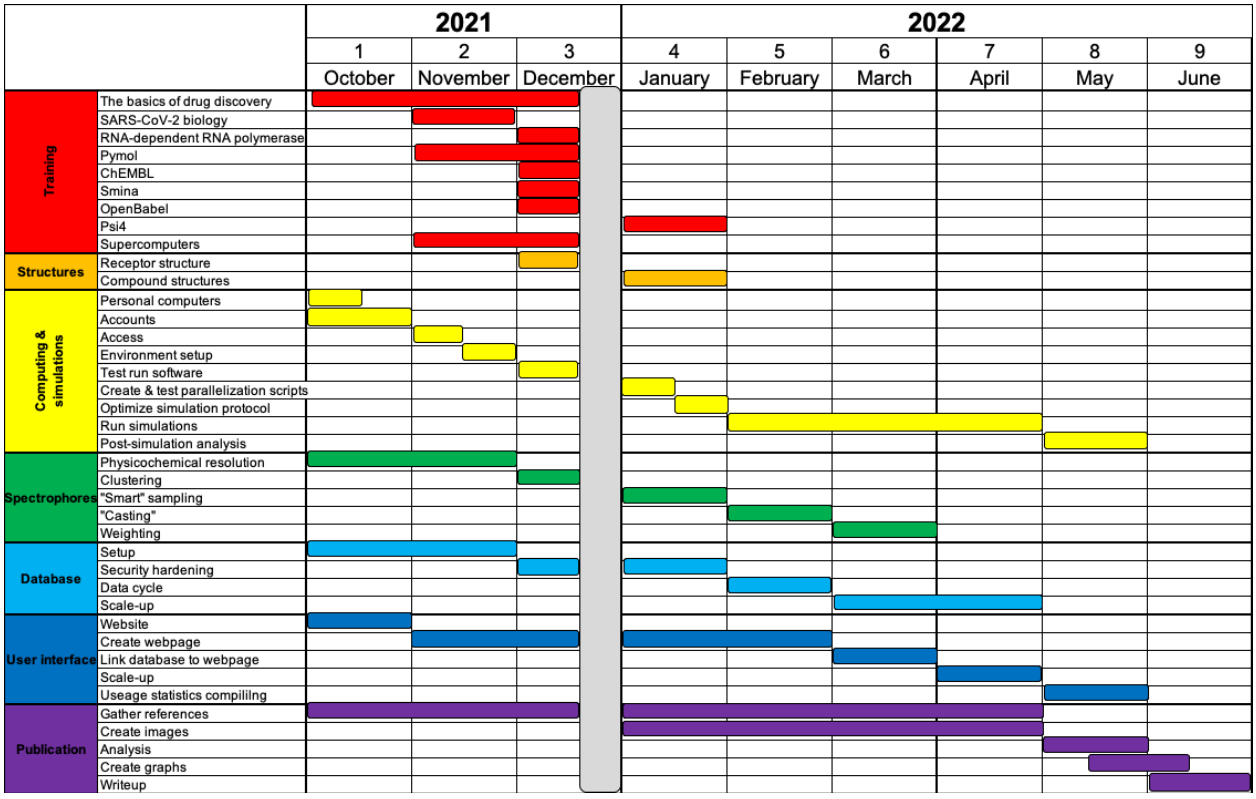
We propose a collaboration between the Masters of Molecular Science and Software Engineering (MSSE) program at the University of California Berkeley (UCB), Lawrence Berkeley National Laboratory (LBNL), and Atomwise to accomplish five main objectives: 1) curate a computational model of RdRp and a set of compounds to target it, 2) virtually dock these compounds to RdRp, generate associated poses, and score them, 3) create a database to organize these data, 4) construct a user interface so that the data may be openly and freely shared, and 5) produce a publication describing this work (Table 1).

The proposed collaboration will leverage the computing power of LBNL, the creativity and drive of students and educators at MSSE, and the drug discovery expertise of Atomwise. The realization of this collaboration is envisioned to have at least the following benefits: 1) students will have the opportunity to learn about the drug discovery process, apply their skills towards fighting the current pandemic, fulfill a capstone project towards their MSSE, and explore career paths, 2) MSSE, UCB, and LBNL will empower open biomedical research for the common good, forge ties with industry, and provide real-world experience and career training to their students, and 3) researchers will have access to freely available data to support their important work. Atomwise, as the leading artificial intelligence-based drug discovery engine in biopharma, understands that people are central to success. Through this collaboration, Atomwise will foster scientific talent, bolster biomedical research, and bring people together to address COVID-19.

The Project

The realization of this proposal, referred to as the Project, is organized into seven main tracks: training, structures, computing & simulations, spectrophores, database, user interface, and publication (Table 1). The completion of the Project is anticipated to take nine months, from October 2021 to July 2022. The final deliverables are a database, to be completed in April 2022 (Table 1, light blue), a website/user interface, to be ready by May 2022 (Table 1, dark blue), and a publication (Table 1, purple), which will begin writeup in June 2022. Spectrophores [4] have been described as a sort of 3D molecular fingerprint and will be used to quantify the physiochemical diversity of the compounds and their poses (Table 1, green). The use of Spectrophores will be tested from October 2021 to March 2022, with the goal of integrating them fully into the Project beginning in April 2022. The main driver of the Project timeline is the computing & simulations track (Table 1, yellow). Setup is anticipated to span October 2021 to the end of January 2022. After this, the simulations will be launched and are estimated to take about 3 months to complete, with post-simulation analysis to be finished in May 2022.

Table 1. Key milestones and deliverables.



Great table! Def going to use for future projects!

[1] "Johns Hopkins University and Medicine COVID-19 Dashboard." https://coronavirus.jhu.edu/map.html

[2] Y. Wang, V. Anirudhan, R. Du, Q. Cui, and L. Rong, "RNA-dependent RNA polymerase of SARS-CoV-2 as a therapeutic target," *Journal of Medical Virology*, vol. 93, no. 1, pp. 300–310, 2021, doi: 10.1002/jmv.26264.

[3] H. S. Hillen, G. Kokic, L. Farnung, C. Dienemann, D. Tegunov, and P. Cramer, "Structure of replicating SARS-CoV-2 polymerase," *Nature*, vol. 584, no. 7819, pp. 154–156, Aug. 2020, doi: 10.1038/s41586-020-2368-8.

[4] R. Gladysz, F. M. Dos Santos, W. Langenaeker, G. Thijs, K. Augustyns, and H. De Winter, "Spectrophores as one-dimensional descriptors calculated from three-dimensional atomic properties: applications ranging from scaffold hopping to multi-target virtual screening," *Journal of Cheminformatics*, vol. 10, no. 1, p. 9, Mar. 2018, doi: 10.1186/s13321-018-0268-9.