JCIM JOURNAL OF CHEMICAL INFORMATION AND MODELING

Perspective

A Community Letter Regarding Sharing Bimolecular Simulation Data for COVID-19

Rommie E. Amaro, and Adrian J. Mulholland

J. Chem. Inf. Model., **Just Accepted Manuscript** • DOI: 10.1021/acs.jcim.0c00319 • Publication Date (Web): 07 Apr 2020

is published by the American Chemical Society. 1155 Sixteenth Street N.W., Washington, DC 20036

Molecular simulation and modelling can contribute to combating the current COVID-19 global pandemic, e.g. helping to identify leads for therapies, diagnostics and vaccines for COVID-19, suggesting potential antiviral drugs, analyzing structural effects of genetic variation in the SARS-CoV-2 virus and molecular recognition relevant to protein-protein and protein-inhibitor systems. Such efforts complement experiments, can be linked to emerging artificial intelligence (AI) methods, and add molecular insight to genetic studies and mathematical models for pandemic risk assessment and predicting the impact of different potential non-pharmaceutical interventions (NPI). To maximize the impact and value of molecular simulation methods in this crisis, we as an international community recognize the need to modify our standard practices to maximize the effectiveness of our global response to the pandemic.

There is an urgent need to share our methods, models, and results openly and quickly to test findings, ensure reproducibility, test significance, eliminate dead-ends, and accelerate discovery. Sharing of data for COVID-19 applications will help connect scientists across the global biomolecular simulation community, and also improve connection and communication between simulation and experimental and clinical data and investigators.

We, as a community, commit to the following principles:

- We commit to making results from our work on the SARS-CoV-2 virus available as preprints as quickly as possible, using preprint servers such as arXiv, bioRxiv, and ChemRxiv, and open access data repositories such as Zenodo.

- We commit to making available the input files, model building / processing scripts (e.g. Jupyter notebooks) required to set up, run, and analyze the simulations, and data necessary to repeat analysis *upon deposition* to the preprint sites following the FAIR data principles (findable, accessible, interoperable, reusable) [1]. Doing so will also enable others to test, extend and augment developed models without duplicating efforts, delivering results more rapidly, and developing and testing hypotheses.

- We will make models and trajectories available as soon as possible through open data sharing platforms such as the Molecular Sciences Software Institute (MolSSI) SARS-CoV-2 Biomolecular Simulation Data and Algorithm Store [2], the Open Science Framework [3], and the European Open Science Cloud [4].

- Where appropriate, we will also share algorithms and methods in order to accelerate reuse and innovation. Well-validated and functional machine learning methods and heuristic property calculators would be especially desirable, as are Monte Carlo models of infectious disease spread and prediction of the impact of different NPI strategies. Custom code will be made rapidly available in appropriate repositories (e.g., GitHub).

- We commit to applying thoughtful permissive (and open source) licensing strategies (such as those recommended by Reproducible Research Standard) [5] to ensure that our models and data can be maximally reused, modified, and redistributed to rapidly advance the field

[1] https://www.nature.com/articles/sdata201618
[2] https://github.com/MolSSI/CoVMME
[3] https://osf.io
[4] https://www.eosc-portal.eu
[5] https://web.stanford.edu/~vcs/talks/VictoriaStoddenCommuniaJune2009-2.pdf

in developing new therapies, while appropriately recognizing and acknowledging original authors and contributors.

In support of these efforts, the SARS-CoV-2 Biomolecular Simulation Data and Algorithm Store draws on the expertise and discussion of several recent workshops, as well as ongoing community discussions [6,7]. The NSF MolSSI has recently created a special call for Seed Fellowship applications for students and postdocs that focus on software development, data science, workflows, and machine learning challenges that are especially relevant to the on-going COVID-19 research. Furthermore, the NSF MolSSI is setting up a centralized github and file sharing service to provide a centralized site for community data, and is also working with Zenodo [8] and the Open Science Grid [9] to help store data and data analysis outcomes for this global initiative. Our community should be aware of high performance compute resources made available for COVID-19 research (e.g. https://www.ibm.com/ covid19/hpc-consortium).

We recognize that we represent only a cross section of our community, and encourage others to follow these principles; all are welcome to join this effort. We offer our support to others already working on open data efforts in the hope that others working on COVID-19 in biomolecular simulation and other areas will adopt similar best practices.

Signed:

Rommie E. Amaro*, UC San Diego, USA ramaro@ucsd.edu
Adrian J. Mulholland*, University of Bristol, UK Adrian.Mulholland@bristol.ac.uk
Teresa Head-Gordon, UC Berkeley, USA
John D. Chodera, MSKCC, USA
Zoe Cournia, Academy of Athens, Greece
Vittorio Limongelli, USI Lugano, Switzerland and University of Naples, Italy
Shantenu Jha, MolSSI / Rutgers / Brookhaven National Laboratory, USA
Marco De Vivo, Italian Institute of Technology, Italy
Matteo Dal Peraro, EPFL, Switzerland
Andrea Cavalli, U Bologna, Italian Institute of Technology, Italy
Alexandre Bonvin, Utrecht University, The Netherlands
David Mobley, UC Irvine, USA
Jean-Philip Piquemal, Sorbonne Université, France
Pengyu Ren, UT Austin, USA
Arvind Ramanathan, Argonne National Laboratory/ University of Chicago, USA
Lynn Kamerlin, Uppsala University, Sweden
Julien Michel, University of Edinburgh, UK
Marc van der Kamp, University of Bristol, UK
Shozeb Haider, University College London, UK
Munir Skaf, University of Campinas, Brazil
Adrian Roitberg, U Florida, USA
Andrew McCammon, UC San Diego, USA
Ivet Bahar, U Pittsburgh, USA
Lillian Chong, U Pittsburgh, USA
Xuhui Huang, The Hong Kong University of Science and Technology, Hong Kong, China SAR

[6] https://pubs.acs.org/doi/full/10.1021/acs.jcim.9b00665
[7] http://www.hecbiosim.ac.uk/covid-19
[8] https://zenodo.org/
[9] https://opensciencegrid.org/