RESOURCES SHARING PLAN

Aims 1-3 of this work focus on generating reference data tailored for a specific purpose – SAMPL blind prediction challenges. We envision this data going through a life cycle of collection, curation, use in blind challenges, publication, and then post-challenge use as reference data and as part of benchmark sets. Sharing of this data will be particularly vital post-challenge. Additionally, Aim 4 focuses on running blind challenges and reference calculations, which will generate data of a different sort – how different methods perform in these blind prediction challenges, and how they stack up against other methods. Again, dissemination of this data is particularly key.

We plan to make all data collected, including all primary data, available through timely publications. Additionally, we plan to archive all experimental data in easily-accessible machine-readable formats for perpetuity. To ensure broadest dissemination, several different resources will be used. Binding data will be archived in BindingDB, and all data cross-posted or linked to from Alchemistry.org. All data will also be posted on the D3R website because of its connection with the SAMPL challenges.

Some specific types of data or products are worth separate attention:

Software. All computer software developed for this project will be made freely available through free (libre) open source software licenses (such as LGPL) on online collaborative public code repositories such as GitHub [http://github.com], where codes produced by the Mobley and Chodera laboratories are currently hosted [http://github.com/choderalab/ and http://github.com/mobleylab/].

Experimental datasets. All experimental data and results will be shared, when practical, through the D3R website, through BindingDB (for binding data) and through online repositories such as GitHub [http://github.com], Dryad [http://datadryad.org/], FigShare [http://figshare.com], and our group websites [http://www.choderalab.org/data/ and http://www.mobleylab.org/]. Primary data will also be provided to the extent possible.

Simulation datasets. All simulation and model datasets will be shared, when practical, through the online repositories such as GitHub [http://github.com], Dryad [http://datadryad.org/], FigShare [http://figshare.com], and our group websites [http://www.choderalab.org/data/ and http://www.mobleylab.org/].

Data on SAMPL results. Aim 4 involves collection of blind predictions from participants in the SAMPL challenge and analysis of these results. Results will be analyzed and published, with the full set of predictions, method descriptions, and their analysis made available (a) in the supporting information and (b) via a separate GitHub repository; and (c) via the D3R website.

Simulation protocols, best practices, and benchmark sets. We will continue to actively support and help maintain the online repository of simulation protocols, best practices, and references for alchemical free energy calculations at the community site alchemistry.org. SAMPL challenge results will also contribute to the development of future benchmark systems which will be discussed at http://github.com/mobleylab/benchmarksets and linked to from there.

Experimental protocols. In addition to publishing detailed accounts in papers, all experimental protocols for Aims 1 and 3 will be made available online on our group website(s) [http://choderalab.org and http://mobleylab.org].

3D printable laboratory parts. Numerous useful 3D printed parts are fabricated in the Chodera laboratory to aid in our research projects. Electronic printable versions of these parts are made available on both the group website [http://www.choderalab.org/3dparts/] and the NIH 3D Print Exchange [http://3dprint.nih.gov].