Automated Property Predictions in the Open Force Field Initiative and Their Application in Small Organic Molecule Force Field Parameterization

Owen Madin [S]
Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO, U.S.A.

Simon Boothroyd
Open Force Field Initiative, Boulder, CO, U.S.A.

Jeffrey Wagner
Pharmaceutical Sciences, University of California, Irvine, Irvine, CA, U.S.A.

Matthew Thompson
Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO, U.S.A.

David Mobley
Pharmaceutical Sciences, University of California, Irvine, Irvine, CA, U.S.A.

Michael Gilson
Skaggs School of Pharmacy and Pharmaceutical Sciences, University of California San Diego, La Jolla, CA, U.S.A.

John Chodera
Computational & Systems Biology Program, Sloan Kettering Institute, New York, NY, U.S.A.

Lee-Ping Wang
Department of Chemistry, University of California Davis, Davis, CA, U.S.A.

Michael Shirts [C]
Department of Chemical and Biological Engineering, University of Colorado, Boulder, CO, U.S.A.
michael.shirts@colorado.edu

The Open Force Field Initiative (OpenFF) (https://openforcefield.org/) is a multi-group, academic-industry collaboration focused on the development of 1) extensible, open source toolkits for constructing, applying, and evaluating force fields; 2) the curation of public datasets necessary to build high-accuracy biomolecular force fields; and 3) the generation of improved molecular force fields for biomolecular and other soft matter applications. In this presentation I will discuss our recent force field parameterization efforts, and the automatic curation and simulation of physical properties that is crucial in this process. In particular, we use an extensible framework (the OpenFF Evaluator, https://github.com/openforcefield/openff-evaluator) to automatically pull physical property datasets of hundreds of measurements (densities, enthalpies, etc.) from NIST ThermoML, and reproduce these datasets with OpenMM simulations. This process allows for easy benchmarking of new force fields and can be coupled to optimization software in order to re-fit non-bonded interactions. I will also describe how this approach is integrated into the production of our recent small-molecule force fields (OpenFF "Parsley" and "Sage"), as well as how it enables us to systematically answer questions in force field science. In particular, I demonstrate how we have used these tools to investigate the viability of replacing enthalpy of vaporization as a Lennard-Jones fitting target with other targets such as enthalpies of mixing.