Parsimonious Comparison of Non-Bonded Models Using Bayesian Inference

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

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

Richard Messerly
Theoretical Division, Los Alamos National Laboratory, Los Alamos, NM, U.S.A.

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

Non-covalent interactions play critical roles in molecular biology. Molecular dynamics is the most common in silico method to probe these interactions, due to the larger system sizes and longer timescales enabled by the use of relatively simple force fields. The speed and power of these models comes at the cost of explicit electronic interactions, which are replaced (generally) by point charges and interactions; this forces the modeler to make many decisions about these simpler representations, including atom typing, combination rules, polarizability, and charge modeling.

In the Open Force Field Initiative, we aim to develop force fields, including models for non-covalent interaction, using data-driven techniques. To this end, we explore the use of Bayesian inference to make data-driven choices between Lennard Jones-(LJ) dispersion-repulsion parameters and functional forms, by calculating Bayes factors, which are essentially 'odds' between different models. Bayes factors are advantageous in the way they incorporate parsimony into their predictions, penalizing unnecessary complexity and rewarding generalizability. In this study, we test this strategy on the 2-center Lennard Jones plus Quadrupole (2CLJQ) model for simple fluids, as its simple functional form is easily modified and analytical 'surrogate models' exist in the literature, allowing for the fast, repeated evaluation of parameter sets required for Bayes factor computation.

Through this strategy, we are able to evaluate whether the model's quadrupole parameter is useful in reproducing temperature-dependent density, saturation pressure, and surface tension data for simple molecules. Additionally, this process produces parameter probability distributions for each compound, valuable information about the parameter uncertainty and sensitivity. This work demonstrates the utility of Bayesian inference as a tool for model selection and informs our future application of this technique to more complex decisions required in fitting biomolecular force fields.