Data Driven Parameterization of Force Fields

Jessica Maat, Irvine, California, US

Molecular modeling aids diverse fields in chemistry, including high-profile applications to predict the dynamics and properties of macromolecular molecular systems. These models also predict different physical properties such as a molecule's solubility, preferred conformation, or binding. The underlying technology are force fields, which provide the potential energy of a system as a function of atomic positions. Force fields are capable of modeling very large chemical systems, however there are often issues with accuracy. In particular, force fields do not predict the correct energetics and geometries around trivalent atom centers, which is a very prevalent chemical feature of drug molecules. In today's force fields, the improper and angle parameters are responsible for falsely pushing trivalent atom centers into planar geometries with minimal flexibility. Historically improper parameters have been given little attention and have even received arbitrary force constants and phase angles during force field development. Over half of commercially available molecules in the 5.9M eMolecules Database contain trivalent atom centers and will employ overlooked improper parameters when modeled.

In order to utilize the accuracy from quantum mechanical calculations and the low computational cost of force field technologies, my research project in the Open Force Field Initiative focuses on developing a data-driven method for parameterizing impropers and angle potential energy functions in force fields. We calculated the target data for parameterization by performing 1-D and 2-D improper and valence quantum mechanical grid optimizations. The grid optimizations provide energetic landscapes of low energy regions that are ideal for parameter fitting. The data is then fit to a minimally expanded improper and valence parameter set using the optimization tool ForceBalance. This method results in more accurate energetics and allows for a more physically accurate range of planar to tetrahedral geometries around a trivalent atom center. The framework of this method has been applied to a limited set of trivalent nitrogen centered molecules, and in order to expand this data-driven method to a larger set of chemistry a larger set of quantum mechanical data is in need. In preliminary work, we have developed the filtering infrastructure for generating a diverse set of trivalent nitrogen molecules through a fingerprinting method that relies on atom type, Wiberg bond order, aromaticity, and molecule size.

My proposal to the Open Molecular Science Cloud (OMSC) is to compute 1-D and 2-D quantum mechanical grid optimizations for a more expansive and diverse set of molecules containing trivalent nitrogens and other trivalent centered atoms, such as carbon. Having the grid optimizations of a chemically diverse set of molecules in the OMSC database would not only provide the necessary data to extend my

ISSN: 2279-8773

parameterization method but also provide the force field development community with an extensive data set to continue to improve force field accuracy.

ISSN: 2279-8773