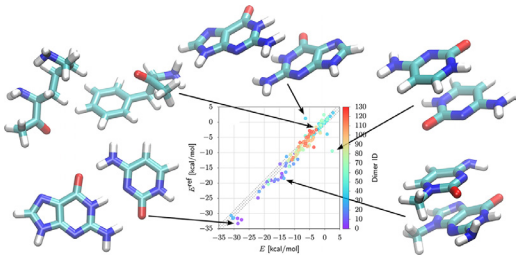
Scilight

15 March 2018

A hybrid physics-based and machine learning approach tackles molecular modeling

Meeri Kim

A new study describes modeling of large molecular systems with an approach that combines classical physics-based potentials with machine learning methods.



The modeling of large molecular systems requires computationally efficient methods to evaluate intra- and intermolecular interactions. Physics-based potentials encode assumptions about the governing physics of the interactions, but new parameters must be determined each time the model is extended to new kinds of molecules and materials. A recent approach involves predicting the potential energy surface of a system using machine learning. While these models have proven extremely accurate, they require prohibitively large training sets to interpolate across conformational space.

A new study published in *The Journal of Chemical Physics* presents a balanced approach to molecular modeling that combines physics-based potentials with machine learning. The approach, called IPML, uses machine learning to make predictions about specific atoms in their local environment, and these parameters are then fed into a physics-based model.

The main drawback of physics-based potentials is the necessity to optimize a number of parameters for any new system. These often require time-consuming reference quantum-chemistry calculations, which must be performed not only for any new molecule, but ideally for any new conformation.

In IPML, these parameters are predicted with machine learning models that rely on kernel-ridge regression, one of the algorithmic approaches to defining machine learning processes, to interpolate reference quantum-chemistry calculations across a diverse set of small organic molecules. The parameters included electrostatic multipole coefficients, the population and decay rate of valence atomic densities, and polarizabilities across conformations and chemical compositions of H, C, N, and O atoms.

The authors validated IPML on various gas-phase dimers at and away from equilibrium separation, where it demonstrated excellent accuracy. They also tested the model on hydrogen-bonded complexes and denser systems like water clusters and supramolecular host-guest complexes. The latter systems proved more challenging, and the authors hope to find strategies for more accurate predictions in the condensed phase.

Source: "Non-covalent interactions across organic and biological subsets of chemical space: Physics-based potentials parametrized from machine learning," by Tristan Bereau, Robert A. DiStasio, Jr., Alexandre Tkatchenko, and O. Anatole Von Lilienfeld, *The Journal of Chemical Physics* (2018). The article can be accessed at https://doi.org/10.1063/1.5009502.