Discovering Energy Functions by Machine Learning

Prof. David van der Spoel, Department of Cell- and Molecular Biology.
Email: David.vanderspoel@icm.uu.se

The energy for forming a molecule out of its elements is called the enthalpy of formation. For instance, the energy released in the reaction to build a water molecule from its elements, $(1/2)\text{O}_2 + \text{H}_2 \rightarrow \text{H}_2\text{O}$, is $-241.8 \text{ kJ/mole}$. Such energies can be measured using, for instance, combustion experiments or they can be computed by the tools of quantum chemistry[1]. However, quantum chemistry calculations are very time consuming and therefore it would be very useful to have a mathematically simple energy function (traditionally called force field) to describe molecular energies.

In this project we will use machine learning to derive a new energy function from quantum chemistry data. A sensitive test for such models is to compute vibrational frequencies and derived quantities such as the standard entropy[2].

Project plan

- Install the GROMACS software to evaluate frequencies and infrared spectra (see figure) for supported compounds for the dataset in the Alexandria library[3].
- Extract data (molecular structures and energies) from the Alexandria library version 2 for compounds that are out of equilibrium. Process the data using the Alexandria Chemistry Toolkit to get lists of bonds and angles for each compound in many conformations.
- Implement Python code to do a linear regression analysis to fit a tabular energy function. This means, we assume the A-B bond energy is a read from a table $V_{AB}[r_{AB}]$, the A-B-C angle energy is read from a table $V_{ABC}[\theta_{ABC}]$, etc. The total energy for a compound is then the sum of these tabular functions and the tables can be derived using a linear regression analysis.
- Refine the table lookup using linear or cubic spline interpolation and derive analytical functions to match the tables.

Time line

- Week 1-2 reading and installing software
- Week 3-6 working with GROMACS and Alexandria
- Week 7-15 derive refined models based on the Alexandria library data and compare results to experimental data.
- Week 16-20 finalize and write report.

References
