CAS 741: Problem Statement

Determination of Conformational Isomers

Jen Garner (PeaWagon//garnej2)

Due: September 14, 2018 (Friday)

Table 1: Revision History

Date	Developer(s)	Change
September 13, 2018 (Thursday)	J. Garner	Creation of document.
September 14, 2018 (Friday)	J. Garner	Make changes as per issues in Github.

Motivation for Project

Computational chemists use software to predict the properties of molecules pertinent to their research or profession. Before running calculations, a starting geometry is required. This initial configuration can be random, or it can be based on, for example, a crystal structure. Reasonable input geometries are important, since some computational methods will not converge (or will take much longer to converge) given a poor input structure.

A number of reasonable input structures may exist, especially for molecules with many rotatable bonds. These structures are known as conformational isomers, and they represent the local minima of the potential energy surface. Since conformational isomers do not require the breaking or formation of bonds, their lifetimes are very short; therefore, they are suitable for computational study.

Proposal

This program will construct a set of conformers for each molecule in a list of input molecules. The program will convert a SMILES (simplified molecular-input line-entry system) string into an initial population of possible geometries, as represented by lists of dihedral angles. Then, an optimisation, such as an evolutionary algorithm, will be implemented to search the energy surface and locate conformers.

Environment & Usage

The author will try to package this program using Anaconda, so that it might be installed in a Conda environment. Python is likely going to be the language of choice, such that no compiling step will be needed (for initial stages - future work may be to profile the code and replace slower areas with a compiled language such as C++ or Rust).

Attempts will be made to locate an open-source software package for evaluating the energy of a given structure (to confirm the location of an energy minimum). For future research, the program will be linked to Vetee - a work-in-progress project to which the author is a contributor. Vetee handles databases of molecules and is responsible for setting-up and running calculations using Gaussian. Since Gaussian is proprietary software, the energy calculation step should be left open such that the user is not restricted behind a paywall.

This program will be used by the author to conduct research into databases of floppy organic molecules, polycyclic aromatic hydrocarbons, carbenes, and common drug molecules. The program will have to run on high performance computing (HPC) clusters (Sharcnet, Compute Canada, etc.), which almost exclusively use Linux OS. A serial version will be written first to ensure that the project is executable on personal computers. Gaussian is available on the HPC cluster, Graham, and its use is dependent on having a Compute Canada account and in agreeing to the following terms:

1. I am not a member of a research group developing software competitive to Gaussian.
2. I will not copy the Gaussian software, nor make it available to anyone else.
3. I will properly acknowledge Gaussian Inc. and Compute Canada in publications.
4. I will notify Compute Canada of any change in the above acknowledgement.

Ultimately, this program should be usable by other chemists to determine good initial structures for molecules where the structure is unknown. The benefit of using this program over other geometry optimisation techniques will be the production of a set of possible geometries, rather than a single solution. For example, some drug molecules may adopt certain conformations in active sites that may not be the global minimum of the energy surface. The program should be at least as fast as performing a geometry optimisation with a high level of quantum mechanical theory (coupled-cluster, full configuration interaction, etc.).