CAS 741: Problem Statement

Determination of Conformational Isomers using Evolutionary Computation

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.

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 string into an initial population of possible geometries, as represented by lists of dihedral angles. Then, an evolutionary algorithm will be implemented to search the energy surface and locate conformers.

Environment

The program will have to run on high performance computing clusters (Sharcnet, Compute Canada, etc.), which almost exclusively use Linux OS. The pro-

gram will also need to interface with Vetee, which is another (work-in-progress) software package to which the author is a contributor. Vetee handles databases of molecules and is responsible for setting-up and running calculations using the proprietary software package, Gaussian.

Usage

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. It can also be used 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.).