Chemistry 430 — Simulation in Chemistry & Biochemistry

Laboratory #4 — Global Optimization of "Lennard-Jonesium" and PolyAlanine

In this lab you will attempt to locate the global potential energy minimum for clusters of "Lennard-Jonesium" atoms, and for a capped octapeptide of polyalanine. As mentioned in lecture, global optimization is an "NP-hard" task, so no algorithm running in polynomial-bounded computer time is guaranteed to find the global minimum. A rigorous solution requires brute-force, exhaustive search of all of configurational or conformational space, requiring exponentially increasing time as the system size increases. We will test two different "approximate" global optimization methods: molecular dynamics-based Simulated Annealing (SA), and Monte Carlo Minimization (MCM) on a modified "barrierless" potential surface.

Protocol

(1) Download the **`lj.xyz`** and **`lj.key`** files from the web site for this lab. The **`.xyz`** file contains just a single atom of "Lennard-Jonesium", a hypothetical noble gas-like element with symbol "LJ" that interacts only via a Lennard-Jones van der Waals potential with sigma (σ) = eps (ε) = 1.0.

$$V(r) = 4\epsilon\left[\left(\frac{\sigma}{r}\right)^{12} - \left(\frac{\sigma}{r}\right)^{6}\right]$$

(2) Use the **`xyzedit`** program with the option to "Create and Fill a Periodic Box", and place 13 LJ atoms into a 20 Angstrom cube. Just for clarity, move the resulting **`lj.xyz_2`** file to **`lj13.xyz`**, and then copy the file **`lj.key`** to **`lj13.key`**.

(3) Perform a SA run in a terminal window via the Tinker **`anneal`** program. Use **`lj13.xyz`** as the input file. Choose the default values for all interactive questions, except for following: (a) use a starting temperature of 300 K and an ending temperature of 0 K, (b) the Number of Cooling Protocol Steps should be changed from 2000 to 1000000, and (c) use an Exponential Cooling Protocol, and (d) the Time between Coordinates Saves should be 10.0 ps instead of 0.1 ps. This will finish quickly, so it can be run interactively in a terminal window. From the screen output, what is the energy of the final frame of your SA trajectory? View the resulting 100 frame **`lj13.arc`** file in Force Field Explorer (FFE), VMD or similar. What is the symmetry of the final "global minimum" structure?

(4) Make a second attempt to find the global minimum of the cluster of 13 LJ atoms using the Tinker **`monte`** program. Again, use your original **`lj13.xyz`** file as the input. Default values are reasonable for all questions, except the "temperature" for accepting new Monte Carlo moves which should be raised to 1000 K. The program will report its final estimate of the global energy minimum, and will save all minima it discovers in a new **`.arc`** file, **`lj13.arc_2`**. View the last frame of this file, which corresponds to the global minimum estimate. How does this structure compare to the one found via **`anneal`**?

(5) The global energy values you found can be compared to those tabulated by David Wales at Cambridge and Jon Doye at Oxford (see the web site in Question 2 below). The final structure from your SA trajectory can be extracted into a single numbered file using the Tinker **archive** program, and then minimized without the basin restraint. Note that to find the true global minimum from your calculations, you should comment or remove the helper **BASIN** restraint keyword in the **.key** file, which ensures that all atoms remain in some form of cluster without evaporating into free space during the **anneal** and **monte** calculations. (To comment the **BASIN** restraint, add a **#** symbol before the keyword, i.e., **#BASIN**). Once the basin restraint is turned off, you should be able to exactly replicate the Wales value for the 13-atom global minimum via optimization with the **minimize** program using a tight RMS gradient criterion of 0.0001 kcal/mol/Ang.

(6) Repeat steps 1-3 for clusters of some other sizes. I might suggest clusters of size 55, 26 and 38, but you can reasonably try any small number of L-J "atoms" below about 100.

(7) Build the peptide sequence ACE-ALA-ALA-ALA-ALA-ALA-ALA-ALA-ALA-NME, which is an octomer of alanine capped with an acetyl group at the *N*-terminus and an *N*-methyl amide group at the *C*-terminus. You can build the peptide using Force Field Explorer (FFE), by choosing the **protein** program from the modeling menu and entering the sequence. Or you can run the Tinker **protein** program manually in a terminal window. Choose the Amber 94 force field (*i.e.*, the file **amber94.prm** in the /params directory of your TINKER installation). If you build the peptide manually, create a keyfile (**.key**) containing the key word option **PARAMETERS ~/tinker/params/amber94.prm**. By default, the peptide will be constructed in an extended conformation similar to a strand of β-sheet.

(8) Check the peptide structure you built by looking at it in FFE. Does the structure form any hydrogen bonds between amide hydrogens and carbonyl oxygens? The lab web site provides pre-built idealized α- and 3/10-helices as **helix-alpha.xyz** and **helix-3-10.xyz**. Look at these helix structures in FFE, and count the number of hydrogen bonds contained in each of them. Approximately how many amino acid residues are there per full turn (*i.e.*, 360°) of each kind of helix? Use the **analyze** program to evaluate the energy of your extended peptide structure and each of the helical structures. Explain the energy ordering of the extended and helical structures.

(9) Using the **anneal** program from within a terminal window, perform a simulated annealing run on the peptide. The program uses an MD-based procedure that starts at a high temperature, and slowly lowers the temperature over the course of the simulation. The SA protocol first equilibrates at the high temperature, and begins a longer "cooling" phase. Before starting the run, add the **ARCHIVE** option to your key file. Use 100000 steps at 1000K for the equilibration, and then 1000000 steps to cool from 1000K to 0K. Accept the default 1.0 fs time step, save a trajectory snapshot every 1.0 ps in place of the 0.1 ps default, and use an exponential (**E**) cooling schedule instead of the linear default. The program will save the trajectory frames to a **.arc** file.

(10) Next, perform a Monte Carlo Minimization (MCM) calculation using the Tinker **monte** program. Start from your original extended peptide conformation, and run for

perhaps 5000 Monte Carlo steps, using torsional moves (**T**, instead of Cartesian) and the default temperature of 500K. Before starting, you can optionally add the **SAVECYCLE** keyword to your keyfile. This keyword forces **monte** to save intermediate structures as individual numbered files (**.001**, **.002**, *etc*.) instead of to a single **.arc** file, and the lowest energy structure found will be in the file with the highest number as its file name extension.

Questions

(1) Search online and find a concise definition of "NP-hard". In intuitive terms, why is finding the absolute lowest energy structure of a molecular system formally categorized as an NP-hard problem? Is this a problem in practice? Or in nature? Obviously protein molecules are able to fold relatively quickly *in vivo*. How is this possible?

(2) The global optimum structures for Lennard-Jonesium clusters containing up to a couple of hundred atoms are known to high certainty. A catalog of the global minima as available at **http://doye.chem.ox.ac.uk/jon/structures/LJ/tables.html** as provided by Jon Doye's group at Oxford University, and is also on the lab site as **doye-results.pdf**. Compare the best minima you found with those cited on the above web site. How did you do? Was the **anneal** or the **monte** program more effective?

(3) Based on the relative energies of the extended and helical structures computed in protocol step 7, what is a rough estimate of the stability afforded by an amide hydrogen bond. What kind of energy is this number? *(HINT: It is **not** a free energy. Why?)*

(4) Use FFE to view the trajectory generated by your peptide SA run. What kind of structure is generated at the end of the computation? How much lower in energy is the final SA structure compared to your original extended structure? You can use the **analyze** program to get the energy of individual structures. Why do you think the final SA structure is lower in energy? You run the SA calculation in the "gas phase". How would use of a solvation model affect your results? If you have the time, you can add the "**SOLVATE GB**" keyword to your **.key** file to turn on a so-called generalized Born (GB) implicit solvent model, and rerun the calculations.

(5) The method used by the **monte** program is very similar to that described in a journal article provided on the web site for this lab, *J. Phys. Chem. A*, ***101***, 5111-5116 (1997). Read this article and briefly describe how the algorithm works.

(6) The **monte** program saves a new structure every time it finds one lower in energy than the current lowest energy structure. So the highest numbered saved peptide file will be the one with the lowest energy. View a few of the lower energy MCM structures in FFE. Are they higher or lower in energy than the final SA structure? Why?

(7) Do you think the results you have found for the octapeptide will scale to larger structures? Why? What structure do you expect to be the "gas phase" global minimum as the peptide sequence becomes longer?