CHEM 498Q / CHEM 630Q: Molecular Modelling of Proteins
TUTORIAL #3a: Empirical force fields

INTRODUCTION

The goal of this tutorial is twofold:

1. Get familiar with the CHARMM empirical force field by calculating the potential energy of a simple molecular structure.

2. Learn the rudiments of the Linux OS (if you have not started yet).

The energy calculation will be performed first with NAMD (http://www.ks.uiuc.edu/Research/namd/), then manually for some of the terms. Finally, in STEP 7, the molecular structure will be optimized with respect to its potential energy.

REQUIRED PRE-LAB READING

None, but you should get ahead in your reading of Leach (see below).

PRE-LAB REPORT

The structure of acetate (CH_3COO^-) is the following:
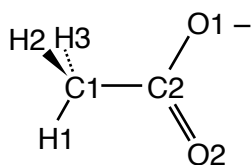
✦ List all pairs of atoms ("A–B") forming a chemical bond. (Note: In the language of empirical force fields, those atoms are called "1–2" pairs.)

✦ List all triples of atoms forming a bond angle. (There are 9 of them.) (Note: The first and last atoms of each triple are called "1–3" pairs.)

✦ List all "quadruples" of atoms ("A–B–C–D") forming a dihedral angle (There are 6 of them.) (Note: The first and last atoms of each quadruple are called "1–4" pairs.)

Note:
Write a triple "A–B–C" following an order in which A is bound to B and B is bound to C. Write a quadruple "A–B–C–D" following an order in which A is bound to B, B is bound to C, and C is bound to D.

READING

Chapter 1 of Leach ("Useful Concepts in Molecular Modelling"): Sections 1.1 to 1.3.
Chapter 3 of Leach ("Empirical Force Field Models: Molecular Mechanics"): Sections 4.1 to 4.10.

REFERENCE MATERIAL

Sections 1 and 2 of "Introduction to Command Line Linux" (by Eric Nodwell)
http://www.phas.ubc.ca/~mbelab/computer/linux-intro/html/

NAMD User's Guide
http://www.ks.uiuc.edu/Research/namd/current/ug/

Instructor: Guillaume Lamoureux

PROCEDURE

STEP 1: Create your directory and download the input files

Start the "Terminal" application and use the command line to create a directory for your files:

```
$ cd
$ mkdir chem498q_yourlastname
$ cd chem498q_yourlastname
$ mkdir tutorial3
$ cd tutorial3
```

You are now in the newly created directory "`tutorial3`", under the directory "`chem498q_yourlastname`".

You will need the following six (6) files in your "`tutorial3`" directory:

- The PDB file "`acetate??.pdb`", containing the atomic coordinates of an acetate molecule. Use the file corresponding to your sequence number (ex: "`acetate02.pdb`" if you have sequence 02). The file is available for download at
    http://faculty.concordia.ca/glamoure/files/acetate??.pdb

- The "Protein Structure File" (PSF) of acetate, that contains the list of atoms, bonds, angles, dihedrals, etc. describing an acetate molecule. This file is available for download at
    http://faculty.concordia.ca/glamoure/files/acetate.psf

- The CHARMM/param27 "topology" file for proteins and nucleic acids:
    http://faculty.concordia.ca/glamoure/files/top_all27_prot_na.rtf

- The CHARMM/param27 "parameter" file for proteins and nucleic acids:
    http://faculty.concordia.ca/glamoure/files/par_all27_prot_na.prm

- The CHARMM "toppar" stream file containing topologies and parameters for additional "model" compounds (acetate, notably):
    http://faculty.concordia.ca/glamoure/files/toppar_all22_prot_model.str

- The NAMD "configuration" file:
    http://faculty.concordia.ca/glamoure/files/acetate_1.conf

Important note:
To download a file from the Linux command line, use the `wget` command from the directory you want to file into:
```
$ wget http://faculty.concordia.ca/glamoure/files/acetate02.pdb
```
You can check that your file was correctly downloaded by using the `ls` ("list") and `more` commands:
```
$ ls
$ more acetate02.pdb
```
(To advance line-by-line, press the space bar. To advance page-by-page, press "return". To get out of the "more" command, type "q" or "control-c".)

STEP 2: Calculate the potential energy of acetate using NAMD

Examine the contents of the NAMD configuration file:

```
$ more acetate_1.conf
```

Rename your "`acetate??.pdb`" file using the `mv` ("move") command:

```
$ mv acetate??.pdb acetate.pdb
```

Run NAMD and redirect the output into file "`acetate_1.log`":

```
$ namd2 acetate_1.conf > acetate_1.log
```

✦ Examine the log file. How many bonds, angles, dihedrals, and "impropers" were found in the structure?

✦ Report the values of the BOND, ANGLE, ELECT, VDW, and TOTAL energies.

Note:
In NAMD log files, all energies are reported in kcal/mol.

STEP 3: Calculate the BOND energy term by hand

✦ Examine the residue definition of acetate in file "`toppar_all22_prot_model.str`" (look for "`RESI ACET`"). What are the names and atom types of each atom?

You can calculate the lengths or all chemical bonds ("1–2" pairs) from the coordinates in the PDB file, using the 3D version of the Pythagorean theorem:

$$r_{12} = \sqrt{(x_1 - x_2)^2 + (y_1 - y_2)^2 + (z_1 - z_2)^2}$$

but it is much easier to use VMD. Load your structure in VMD

```
$ vmd acetate.pdb
```

activate the "distance" mode by selecting the "Mouse > Label > Bond" menu (or just type "2" from the VMD display window), then click on the two atoms you want to measure. An approximate value of the bond length will be displayed as a label but a more accurate value is printed in the VMD log window. (This is the one you want.) Angles and dihedral angles get calculated the same way—by typing "3" and clicking on three atoms or typing "4" and clicking on four atoms.

✦ Present your data in a table containing the following columns:
   - Name and type of atom #1
   - Name and type of atom #2
   - Bond parameters (k_b and b_0) for the two atom types forming the bond.
   - Bond length b (in angstroms, Å)
   - Bond stretching energy for that pair $= k_b (b - b_0)^2$ (in kcal/mol).

✦ Calculate the total and compare with the BOND value from the NAMD log. (It should be almost exactly the same.)

Note:
Look for the bond parameters first in the '`BONDS`' section of file "`toppar_all22_prot_model.str`", then (if they are not there) in the '`BONDS`' section of file "`par_all27_prot_na.prm`".

Note:
Compute all numbers with as many digits as possible. Your results should closely match those from NAMD.

Note:
If you want to use a spreadsheet application within Linux OS, use the "Calc" program from the OpenOffice suite. (Make sure you save your data in Excel format, though.) Another option is to use Google Spreadsheet.

STEP 4: Calculate the ANGLE energy term by hand

Measure the angles using VMD, as explained above.

✦ Present your data in a table containing the following columns:
   - Name and type of atom #1
   - Name and type of atom #2
   - Name and type of atom #3
   - Angle parameters (k_θ and θ_0) for the three atom types forming the angle.
   - Angle θ (in degrees)
   - Angle difference $\theta - \theta_0$ in radians (1 rad = 57.2957795º)
   - Angle bending energy for that triple $= k_\theta (\theta - \theta_0)^2$ (in kcal/mol).

✦ The ANGLE energy term also contains Urey-Bradley contributions, defined for triples H1-C1-H2, H1-C1-H3, H2-C1-H3, C1-C2-O1, etc. (There are 9 in total.) You should therefore add nine lines to your table, containing the following columns:
   - Name and type of atom #1
   - Name and type of atom #2
   - Name and type of atom #3
   - Urey-Bradley parameters (k_u and u_0, called "Kub" and "S0" in the parameter file).
   - Distance u between atoms #1 and #3 (called "S" in the parameter file).
   - Urey-Bradley energy for that triple $= k_u (u - u_0)^2$ (in kcal/mol).

✦ Calculate the total (angle bending + Urey-Bradley) and compare with the ANGLE value from the NAMD log.

Note:
Look for the angle parameters first in the 'ANGLES' section of file "toppar_all22_prot_model.str", then (if they are not there) in the 'ANGLES' section of file "par_all27_prot_na.prm".

Note:
In CHARMM parameter files, equilibrium angles θ_0 are in degrees, but parameters k_θ are in kcal/mol/rad^2.

STEP 5: Calculate the ELECT energy term by hand

Examine the residue definition of acetate (in file "toppar_all22_prot_model.str"). What are the partial charges of each atom?

Calculate the distances between all nonbonded atom pairs ("1–4" pairs and above) with the Pythagorean theorem.

✦ Present your data in a table containing the following columns:
   - Name and type of atom #1, and its partial charge q_1 (in electrons, e)
   - Name and type of atom #2, and its partial charge q_2 (in electrons, e)
   - Distance r_{12} between the two atoms (in Å)
   - Coulomb interaction energy for that pair $= q_1 q_2 / r_{12}$ (in e^2 / Å)
   - Coulomb interaction energy <u>including the shift factor</u> (in e^2 / Å)
   - Shifted Coulomb interaction energy, in kcal/mol (multiply the previous number by 332.0636).

✦ Calculate the total and compare with the ELECT value from the NAMD log.

STEP 6: Calculate the VDW energy term by hand

Examine the 'NONBONDED' section of file "par_all27_prot_na.prm" and locate the Lennard-Jones parameters "epsilon" and "Rmin/2" of the atom types involved in "1–4" pairs.

✦ Present your data in a table containing the following columns:
   - Name and type of atom #1, and its ε and Rmin values
   - Name and type of atom #2, and its ε and Rmin values
   - Mixed Lennard-Jones parameters $\varepsilon_{12} = \sqrt{\varepsilon_1 \varepsilon_2}$ and $Rmin_{12} = Rmin_1/2 + Rmin_2/2$
   - Lennard-Jones interaction energy $= \varepsilon_{12} [\,(Rmin_{12} / r_{12})^{12} - 2(Rmin_{12} / r_{12})^6\,]$ (in kcal/mol)
   - Lennard-Jones interaction energy <u>including the switch factor</u> (in kcal/mol).

✦ Calculate the total and compare with the VDW value from the NAMD log.

STEP 7: Optimize the acetate structure

Create a copy of the configuration file:

```
$ cp acetate_1.conf acetate_2.conf
```

Modify this new file such that variable "outputName" is now equal to "./acetate_2" and the last line is "minimize 100" instead of "minimize 0".

Run NAMD for this new file:

```
$ namd2 acetate_2.conf > acetate_2.log
```

Compare the two log files side by side:

```
$ sdiff acetate_1.log acetate_2.log
$ sdiff acetate_1.log acetate_2.log | more
```

Extract the energies of the 100 optimization steps using the grep command and redirect the output into file "acetate_2.ener":

```
$ grep "ENERGY:   " acetate_2.log
$ grep "ENERGY:   " acetate_2.log > acetate_2.ener
```

File "acetate_2.ener" should contain exactly 101 lines. To count the lines, use the wc ("word count") command:

```
$ wc -l acetate_2.ener
101 acetate_2.ener
```

✦ Produce graphs of BOND, ANGLE, DIHED, IMPRP, ELECT, VDW, and TOTAL energy terms as a function of the iteration number (from 0 to 100). What values are each of these terms converging to?

✦ Comment on the deviation of the optimal bond lengths and bond angles relative to their reference values from the "toppar" file (the b_0 and θ_0 values).

Note:

An easy way to produce plots on Linux OS is to open the file in OpenOffice (Calc). Another convenient option is the software `xmgrace`, available from the terminal by typing:

```
$ xmgrace &
```

The more adventurous ones can also try the command-line driven `gnuplot`:

```
$ gnuplot
gnuplot> plot 'acetate_2.ener' using 2:12 with linespoints notitle
gnuplot> set terminal png
gnuplot> set output 'plot_TOTAL.png'
gnuplot> replot
gnuplot> quit
```

INSTRUCTIONS FOR THE LAB REPORT

ACETATE versus ACETIC ACID

Write an additional discussion section to your lab report, in which you explain how you would calculate the potential energy for acetic acid ("`RESI ACEH`" in file "`toppar_all22_prot_model.str`"). Even though you are not provided with any atomic coordinates, be as specific as possible. In particular, explain which energy terms would have the same form as for acetate and which would be different, and describe all additional terms not present in acetate.

The structure of acetic acid is the following:
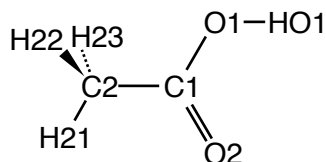
(Notice that atom C1 in residue `ACET` corresponds to atom C2 in residue `ACEH`, and vice versa.)