Lab1:

- Introduction – brief summary and reason for project
- Procedure – what steps did you actually perform and in what order
- Methods – kinds of calculations, programs used, level of theory, any assumptions or approximation such as quantities varied, restraints and constraints, special program options
- Results – data obtained, usually in tables and plots
- Analysis – explain your result

Chemistry 430 — Simulation in Chemistry & Biochemistry Laboratory #1 — Conformational Analysis of Alanine Dipeptide In this lab you will find the local minima of "Alanine Dipeptide" (i.e., Ace-Ala-NMe) using three different protein force field "molecular mechanics" models, both in the gas phase and using the Generalized Born implicit solvent model.

Protocol (1) If you have not already done so, set up Tinker and FFE on your computer. You can download both packages from the Software Resources section of the course website. FFE comes as the ffe-macosx-8.7.2.dmg installation kit that you can just "double click" on to install. For Tinker, move the gzipped tar file (tinker-8.10.1-macos.tar.gz) to your home directory. Then run gunzip followed by tar xvf on the downloaded file. After all of this, FFE will be under /Tinker-FFE in your home directory, and Tinker will be under /tinker in the same home directory. If it is not already there, drag the FFE executable (~/Tinker-FFE/ffe/Force Field Explorer.app) to your dock. Also, add the line set path = ($path $HOME/tinker/bin) to the .tcshrc file in your home directory, if this line is not already present, or an analogous line if you are using bash as your shell.

(2) Use these programs to run calculations with the **OPLS-AA, Amber ff99SB and CHARMM22-CMAP** force fields. Create a directory to contain the files you will generate in this lab. In this directory, use a text editor to construct three key files, **opls.key, amber.key and charmm.key**. Each file should contain a single line pointing to the appropriate parameter file (i.e., oplsaa.prm, amber99sb.prm and charmm22cmap.prm). The parameter files are in the /params area of your Tinker installation. For example, the line PARAMETERS /user/"yourname"/tinker/params/oplsaa.prm will specify the OPLS-AA parameter file and should go into the file opls.key. In each of the three files, add two more lines. One line should contain the Tinker keyword ENFORCE-CHIRALITY. The other additional line should consist of the keyword ARCHIVE.

(3) In a terminal window, run the Tinker protein program to construct alanine dipeptide for each of the three force fields. The sequence has three "amino acid residues", which should be input as in the order: "ACE", "ALA", "NME".

(4) Open the structures in turn using FFE. Go to the "Keyword Editor" panel and check that the correct force field is being used, and the above keywords are active. From the "Modeling Commands" panel, run the scan program for each structure, using automatic selection of torsional angles (option 0), an energy threshhold of 10 kcal/mol (instead of the default value of 100.0), and an RMS gradient of 0.0001. The minima will be written to a TINKER archive coordinates file (opls.arc, etc.). Open the .arc files in FFE and look through the sequence of minima found by the scan program by playing the .arc file as a "trajectory movie". Alternatively, you can run the scan program in a terminal window, instead of via the FFE GUI interface. To do this, just enter scan in the terminal and answer the questions interactively using the above values.

PSM= Potential Surface Map
NMLS= Normal Mode Local Search

```
scan amber.xyz 0 5 10 0.0001 >& amber.log
```
9 minimums

```
scan charmm22cmap.xyz 0 5 10 0.0001 >& charmm.log
```
34 minimums

```
scan oplsaa.xyz 0 5 10 0.0001 >& oplsaa.log
```
11 minimums

(5) The calculations in step (4) were run on a single, isolated molecule in the gas phase. For each of the three force fields, repeat the scan calculation with the Generalized Born (GB) solvation model activated. The GB model places the dipeptide molecule into a bath of "implicit" water, thereby mimicking the presence of solvent without explicitly including solvent molecule in the computation. To turn on use of the GB solvation model, you should add the keyword phrase SOLVATE GBSA to each of the .key files. Alternatively, this keyword can be activated via the FFE Keyword Editor.

GBSA doesn't look like an option in the Tinker documentation (page 83) but GB and GBSA give same results at the end.

```
scan amber.xyz 0 5 10 0.0001 >& amber.GBSA.log
scan amber.xyz 0 5 10 0.0001 >& amber.STILL.log
```
18 minimums

```
scan charmm22cmap.xyz 0 5 10 0.0001 >& charmm.GBSA.log
```
48 minimums

```
scan oplsaa.xyz 0 5 10 0.0001 >& oplsaa.GBSA.log
```
22 minimums

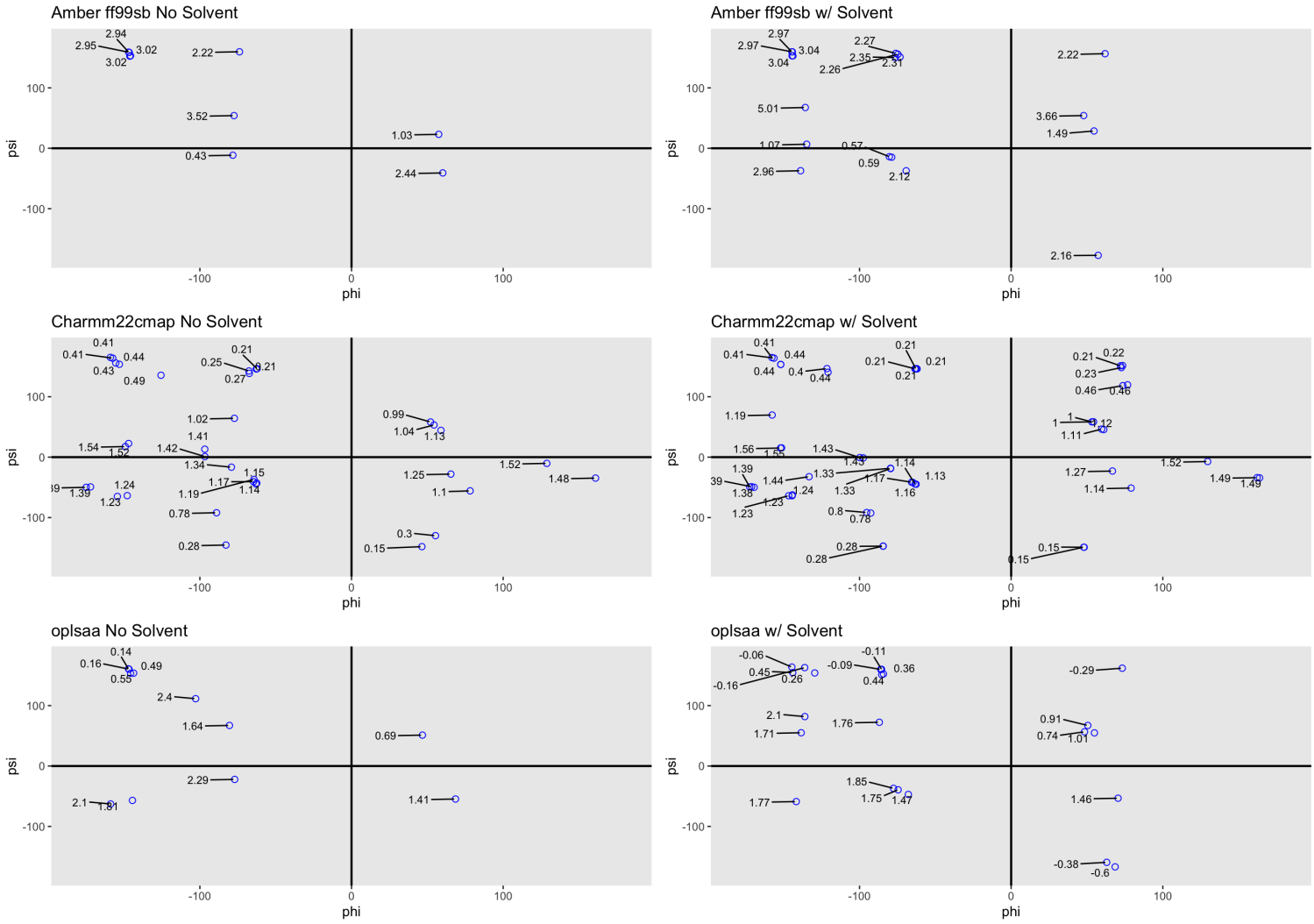
(6) Construct separate “Ramachandran plots” for each force field, both with and without GB solvation, to show the position and energy of each of the minima. You can find the values of the phi and psi angles interactively using FFE, or you can run the Tinker analyze program on the .arc file from each scan calculation. If using analyze, include the D option to get detailed output. Then the phi and psi angle values can be found in the list of torsional angles for each structure.

x-axis = phi (φ) (N-terminal direction) carbonyl carbon, the connecting α-carbon, an amide nitrogen and the next carbonyl carbon
Torsional angle 2-C 7-N 8-CT (C-alpha) 9-C

y-axis = psi (ψ) (COO-terminal direction) amide nitrogen, a carbonyl carbon, an α-carbon and a second nitrogen
Torsional angle 7-N 8-CT (C-alpha) 9-C 17-N

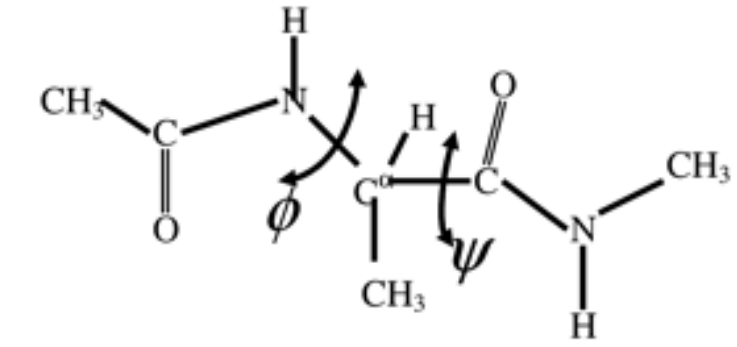
```
for ff in amber oplsaa; do
   analyze $ff.arc $(echo D) -k $ff.key >& $ff.analyze.log
   ## using sum of phi and psi torsion potentials instead of plotting total potential
of molecule
   paste <(ggrep -P "Torsion\\s+2-C\\s+7-N\\s+8-CT\\s+9-C" $ff.analyze.log | awk
'{print $6,$7}') <(ggrep -P "Torsion\\s+7-N\\s+8-CT\\s+9-C\\s+17-N" $ff.analyze.log |
awk '{print $6,$7}') > $ff.rama.coords.txt

   analyze $ff.arc_2 $(echo D) -k $ff.GBSA.key >& $ff.analyze.GBSA.log
   paste <(ggrep -P "Torsion\\s+2-C\\s+7-N\\s+8-CT\\s+9-C" $ff.analyze.GBSA.log | awk
'{print $6,$7}') <(ggrep -P "Torsion\\s+7-N\\s+8-CT\\s+9-C\\s+17-N"
$ff.analyze.GBSA.log | awk '{print $6,$7}') > $ff.GBSA.rama.coords.txt
done

# NH1 instead of N, CT1 instead of CT
for ff in charmm22cmap; do
   analyze $ff.arc $(echo D) -k $ff.key >& $ff.analyze.log
   ## using sum of phi and psi torsion potentials instead of plotting total potential
of molecule
   paste <(ggrep -P "Torsion\\s+2-C\\s+7-NH1\\s+8-CT1\\s+9-C" $ff.analyze.log | awk
'{print $6,$7}') <(ggrep -P "Torsion\\s+7-NH1\\s+8-CT1\\s+9-C\\s+17-NH1"
$ff.analyze.log | awk '{print $6,$7}') > $ff.rama.coords.txt

   analyze $ff.arc_2 $(echo D) -k $ff.GBSA.key >& $ff.analyze.GBSA.log
   paste <(ggrep -P "Torsion\\s+2-C\\s+7-NH1\\s+8-CT1\\s+9-C" $ff.analyze.GBSA.log |
awk '{print $6,$7}') <(ggrep -P "Torsion\\s+7-NH1\\s+8-CT1\\s+9-C\\s+17-NH1"
$ff.analyze.GBSA.log | awk '{print $6,$7}') > $ff.GBSA.rama.coords.txt
done
```

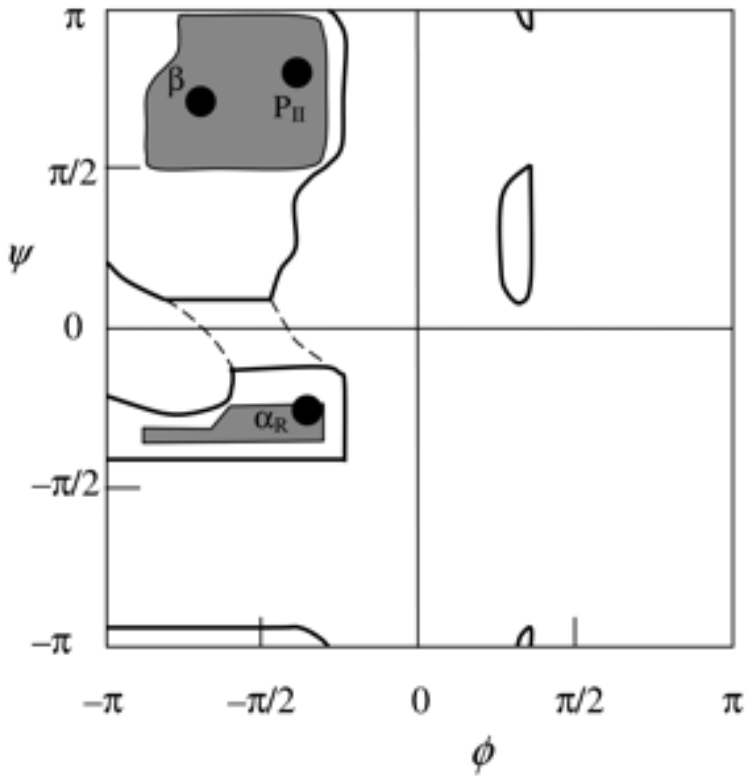
Plot(s):



Energy terms are the some of the sum of the potential energies for φ and ψ torsional angle potential energies per molecule. You should mostly see angles in upper left quadrant for β-sheets and polyproline α-helices followed by lower left for right-handed α-helices, then a few in upper right for left-handed α-helices and not many at all in lower right.

https://www.pnas.org/content/108/8/3095





Questions

(1) How many minima did you find for each of the six scan calculations (three force fields, gas phase and solvated)? Why do the numbers of minima found differ? What is the lowest energy "gas phase" structure for the dipeptide? Why is its energy so low? What about the 2nd through 4th lowest gas phase structures? Rationalize the relative energies of these four lowest energy

conformations. These four structures are not as strongly preferred once GB solvation is included in the force field energy. Why?

A. For amber ff99sb without solvent there were 9 minimums and with solvent 18 minimum
B. For charmm22cmap without solvent there were 34 minimums and with solvent 48 minimum
C. For oplsaa without solvent there were 11 minimums and with solvent 22 minimum

Interesting how it didn't double for charmm22cmap.

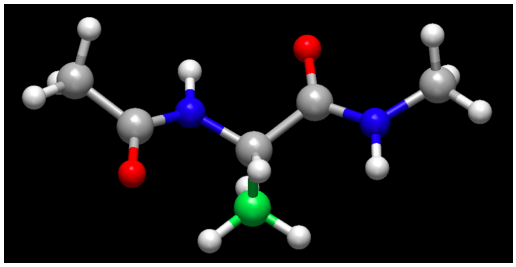
```
for ff in amber charmm oplsaa; do
    grep "Potential Surface Map" $ff.log | sort -k6,6n | head -n4; echo; echo
done
```

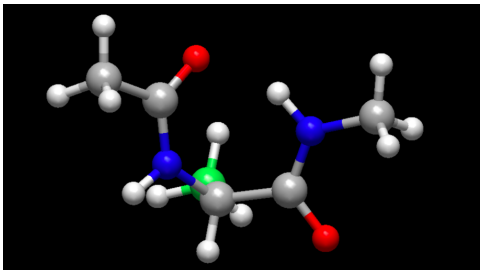
The lowest minima for "gas" phase structures were:

- -21.7347 Kcal/mole for amber ff99sb # 5th of 9 structures
  - Followed by structure 1/9 at -21.1407, structure 4/9 at -20.3137, & structure 3/9 at -18.1346 Kcal/mole
- -17.2417 Kcal/mole for charmm22cmap # 1st of 34 structures
  - Followed by structure 19/34 at -16.2412, structure 14/34 at -15.9106, & structure 5/34 at -15.3390 Kcal/mole
- -41.9374 Kcal/mole for oplsaa # 4th of 11 structures
  - Followed by structure 1/11 at -40.5733, structure 11/11 at -39.4148, & structure 3/11 at -37.1082 Kcal/mole

Interesting how oplsaa forcefield is twice the potential of the other two. It also seems odd that for amber and oplsaa the initial structure, structure 1, being the 2nd most favorable. For example, structure 4 from amber forcefield resulted in the oxygen of the ACE cap and NH of the NME cap in closer conformation for hydrogen bonding that in the first structure from scan but the initial structure. One rational I can see for the minimum structures is the H-bonding between termi-caps of the dipeptide. The reason you don't see these being the most favorable structures in implicit solvent (I am assuming this is H2O) is because it is more favorable to make more intermolecular H-bonds with the solvent (4 in this case, 2 donor NH and 2 accept carbonyls) than with intramolecular H-bonding.

Amber gas phase:
Structure 1 (left; -21.1407) vs Structure 4 (right; -20.3137)





(2) The procedure used by the scan program is sometimes referred to as "Low Mode Search" or LMOD. The original article on this method is Journal of the American Chemical Society, 118, 5011-5019 (1996). Read this paper, which is located in the directory for this lab, and briefly describe how the method works. While this JACS paper is quite old, versions of this basic method are still widely used for conformational search projects. An analysis of more recent methods similar to LMOD is found in Bioorganic Medicinal Chemistry, 21, 7898-7920 (2013). Both of these research papers are provided on the course web site.

JACs (notes)
- bounded by the number of low-frequency modes considered

BioOMC (notes)
Inhibitors of protein–protein interactions also tend to be larger and more flexible since, to gain binding affinity for rather flat and open binding sites, they have to make numerous contacts with the protein.

Uses eigen-vector following algorithm. It searches the entire potential energy hypersurface (n-dimensional surface). Talks a lot about using variation of torsion angles a lot in JACs paper. But based on number of torsional angles and scaling the of peptide size this does not seem like it represents the "meat" of the run time.

(3) Repeat any one of the scan conformational searches using alanine tripeptide (Ace-AlaAla-NMe). Now how many minima do you find? Do you expect the number of minima to grow linearly or exponentially with the length of the peptide? Explain. How large of a peptide structure do you think could be completely searched using this method? In a similar fashion, if you have time, try to use scan to find all the minima for the n-alkanes: methane, ethane, propane, butane, pentane, hexane, etc. Extended n-alkane .xyz and .key files for use with the OPLS-AA force field are on the lab website as alkanes.tar.gz.

oplsaa.prm
```
scan oplsaa.xyz_2 0 5 10 0.0001 >& oplsaa.GBSA.tripep.log
```

There were 43 minimums which was almost twice that for the dipeptide in solvent using opls. I tested a quadpeptide to see if it was around 10*2^(n-1) but with 4 it was double
n=4 ~160 minimums (thought might be ~80) (92 torsion angles)
n=3 ~40 minimums (67 torsion angles)
n=2 ~20 minimums (42 torsion angles)
n=1 ?? can we just have ACE-NME molecule

From a naïve answer with no backup other than reading about N-body systems I would assume this would grow exponentially and just looking at the first 4 up to alanine quad/quadra?-peptide the minimums did not grow linearly.

Will do alkanes later.

(4) If you have the time and/or interest (not required!), figure out how to use Tinker to restrain the phi and psi angles of alanine dipeptide (more keywords!, I can help with this...). Then use the Tinker minimize program to find the minimum energy of the dipeptide on a regular grid of phi/psi values. You can use this data to construct a full Ramachandran map as a 2-D contour plot of the energy as a function of the "phi" and "psi" peptide backbone torsional angles.

PROCHAIN Subroutine "prochain" builds up the internal coordinates for an amino acid sequence from the phi, psi, omega and chi values
PHIPSI Module phi-psi-omega-chi angles for protein

Scripts can be shared if needed. Will probably upload to Github for fun. Set max energy threshold of -18 Kcal/mole