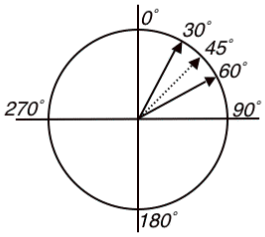
Homework Assignment 6 – Conformational Sampling

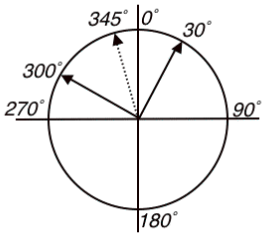
DUE in class, Thursday, Oct 26

In Lecture 4 we learned that α-helices in peptides and proteins show variability in conformational freedom along the length of the helix. Those residues near the ends have larger conformational fluctuations than residues in the middle of the helix. We now address the question of whether or not your simulations agree with this experimental finding. Furthermore, you may remember that we used different values for the hydrogen bond score in various recent simulations. We can ask the question: What value of hydrogen bond score creates an ensemble that best agrees with experimental data?

In this assignment, you will calculate the mean and angular deviation of ϕ and ψ angles at each position along the peptide helix. From a plot of the angular deviation at each position one can answer the question of whether or not the helix shows end fraying in a simulation in agreement with experiment. You will vary the hydrogen bond score and attempt to find the optimal score for simulations; i.e. you will *parameterize the force field*.

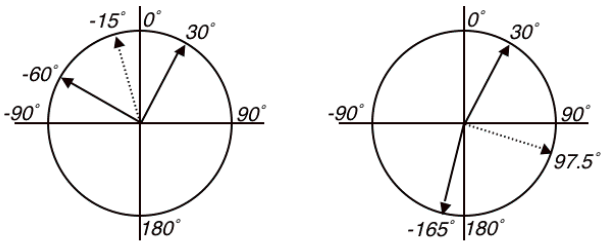
Calculation of statistical properties when the data are circular or periodic presents special problems. For example, in the figure below you can see that if all angles are in the same quadrant the usual definition of the mean gives the correct answer (figure on left); however if the values are spread over the top two quadrants the usual definition gives the wrong answer (figure on right).





$\frac{30 + 60}{2} = 45$ $\frac{300 + 30}{2} = 165$

We obtain the correct answer for the above values on the right (300, 30) if we use the more familiar –180°, +180° range for these angles (-60, 30) (left below). However, the –180°, +180° range for angles will still give the wrong answer if the angles are spread across the circle (right below).



$\frac{-60 + 30}{2} = -15$ $\frac{-165 + 30}{2} = -67.5$

Fortunately algorithms have been worked out to overcome these problems. Here are the steps for calculating the mean angle for circular data.

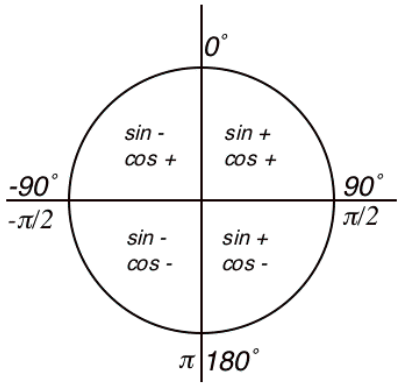
1. Calculate the mean of the unit rectangular coordinates; here angles ($\boldsymbol{a_i}$) are in radians, $\boldsymbol{C_m}$ = mean of cosines, $\boldsymbol{S_m}$ = mean of sines.

$$C_m = \frac{\sum_{i=1}^{n} \cos a_i}{n}, \qquad S_m = \frac{\sum_{i=1}^{n} \sin a_i}{n}$$

2. Calculate mean angle ($\boldsymbol{\theta_m}$); this mean will be in radians

$$\theta_m = \begin{cases} \arctan(S_m/C_m), & \text{if } C_m > 0 \\ \pi + \arctan(S_m/C_m), & \text{if } C_m < 0 \end{cases}$$

The reason we need the conditional outcome is because the inverse tangent (arctan, $\tan^{-1}$) is a multivalued function; it has different values depending on whether the angle data is in the top half or bottom half of the circle. This difference is determined by the sign of the cosine as shown in the figure below.



(Note: For now we will ignore the possibility that the C_m is exactly zero which is very unlikely for our data. Additional outcomes are possible if this is the case.)

Similar problems are presented when one wants to calculate the standard deviation of angular data. There are several ways to do this once you have the mean. We will use a method that builds on the intermediate calculations performed above.

1. Calculate the mean of the unit rectangular coordinates as above, here angles are in radians

$$C_m = \frac{\sum_{i=1}^{n} \cos a_i}{n}, \qquad S_m = \frac{\sum_{i=1}^{n} \sin a_i}{n}$$

2. Calculate the length of the mean vector, *R*

$$R = \sqrt{C_m^2 + S_m^2}$$

3. Calculate the angular deviation in degrees; this deviation will be in radians

$$S_o = \sqrt{2 * (1 - R)}$$

Your task for this assignment calculate the deviation in ϕ, ψ values for each residue along a peptide in at least three different simulation ensembles; each generated with a different value for the hydrogen bond score. **Fortunately, code to calculate angular deviation is provided to you.**

1. Copy the Python module */home/compbio2/Shared/print_phi_psi_mean_SD.py* to your your *linus_mc/* subdirectory on the cluster. The script *print_phi_psi_mean_SD.py* will enable you to read the individual conformations in *sim_ala_prod/ala_traj.pdb* from lab 4b and print out mean and angular deviation of ϕ, ψ values for each residue in the conformations saved during the simulation.
2. Copy */home/compbio2/Shared/normal-python2* to your your *linus_mc/* subdirectory on the cluster. Edit this file to change "yourJHEDID" to your actual JHED_ID. Scroll down and notice the command near the bottom. It should be:

```
$PYTHONBIN ./print_phi_psi_mean_SD.py ala_traj.pdb > phi_psi.stats
```

   The *print_phi_psi_mean.py* script was written for python version 2 so that is why you are using a different queue submission script here; the variable $PYTHONBIN is defined in this script as */home/apps/bin/python2.7*. Here the file *ala_traj.pdb* is the trajectory from your polyalanine simulation.
3. Submit your job to the queue with, **`./normal-python2`**. You can monitor the job with **`qstat -u \*`**. The program *print_phi_psi_mean_SD.py* should read individual peptide conformations in the *ala_traj.pdb* trajectory file, calculate the ϕ, ψ values for each residue and print to the *phi_psi.stats* file the mean and angular deviation of ϕ and ψ for each residue. There is relatively little calculation but lots of file reads so it will take a few minutes. The format of

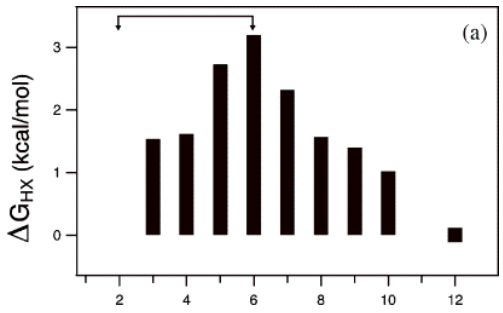
this output is

```
Res       N    Meanphi AngDevphi Meanpsi AngDevpsi
  1   10000      0.000    0.000  176.376   77.706
  2   10000    -83.189   44.287  -36.515   76.614
  3   10000    -82.846   41.612  -20.394   72.527
  .
  .
  .
```

Notice that there is no ϕ for the first residue (values are set to zero).

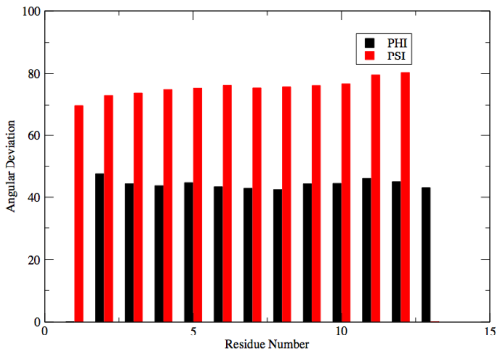
4. Bring the *phi_psi.stats* files back to the Mac.
5. Use **awk** commands to extract the residue number and standard deviation columns to x,y data files for plotting in **xmgrace**. (Don't forget to remove any labels from the x, y data files before importing into xmgrace.)
6. Make a bar chart in **xmgrace** with the following:
   Plot | Graph appearance | Main | Presentation | Type: XY chart
   Plot | Set appearance | Main | Set Presentation | Type: Bar
   Plot | Set appearance | Main | Line Properties | Type: None
   Plot | Set appearance | Symbols | Symbol fill | Pattern: Solid

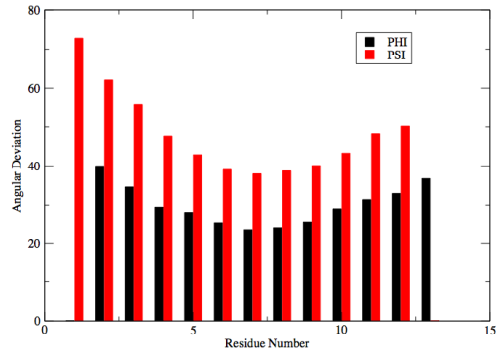
7. Do these results agree with the hydrogen exchange data on helical peptides reported in the paper by Englander et al. (shown below)? Write a short answer to this question explaining why or why not. (Hint: How are free energies of unfolding related to backbone angular variations?)

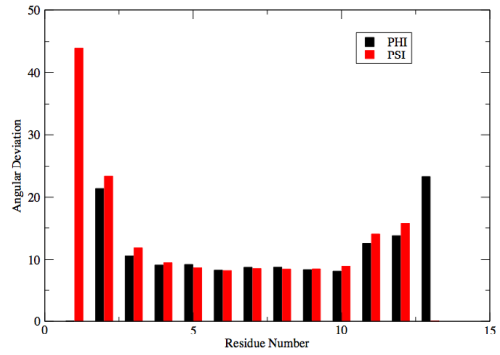


8. Now we come to the question of whether or not the value for the hydrogen bond score in *Linus_prod.py* is the optimal value for agreement with experimental data. Perform at least two other simulations of the alanine peptide with different values for the hydrogen bond score and plot the angular deviation in ϕ, ψ values as above. Try to find the hydrogen bond score that gives approximately the same gradation of stabilities along the peptide as seen for ΔG_{HX} in the figure above (i.e. maximum stability for the center with tapering stabilities toward the ends). Choose values of hbond scores that bracket a good parameter value (i.e. one that mimics the *relative* flexibilities shown by HX).
9. Hand in hard copies of your angular deviation graphs. On the back write your short answer (#7 above) and your best guest for the optimal value of the

hydrogen bond. Your choice of best parameter value should be within those that you tested above. Explain why you chose that value.


PHI
PSI
Angular Deviation
Residue Number
0
20
40
60
80
100
0
5
10
15


PHI
PSI
Angular Deviation
Residue Number
0
20
40
60
80
0
5
10
15


PHI
PSI
Angular Deviation
Residue Number
0
10
20
30
40
50
0
5
10
15