20.320 Problem Set #4

Due on October 21st, 2011 at 11:59am. No extensions will be granted.

General Instructions:

1. You are expected to state all of your assumptions, and provide step-by-step solutions to the numerical problems. Unless indicated otherwise, the computational problems may be solved using Python/MATLAB or hand-solved showing all calculations. The results of any calculations must be printed and attached to the solutions, and the corresponding code should be submitted on Course website. For ease of grading (and in order to receive partial credit), your code must be well organized and thoroughly commented with meaningful variable names.

2. You will need to submit the solutions to **each problem to a separate mail box**, so please prepare your answers appropriately. Staple the pages for each question separately and make sure **your name** appears on each set of pages. (The problems will be sent to different graders, which should allow us to get graded problem sets back to you more quickly).

3. Submit your completed problem set to the marked box mounted on the wall of the fourth floor hallway between buildings 8 and 16. Codes when relevant should be submitted on Course website.

4. The problem sets are due at noon on Friday the week after they were issued. There will be no extensions of deadlines for any problem sets in 20.320. **Late submissions will not be accepted.**

5. Please review the information about acceptable forms of collaboration, which is available on the course website and follow the guidelines carefully. Especially review the guidelines for collaboration on code. NO sharing of code is permitted.'

This assignment is meant to be an introduction to MATLAB coding. Please review the materials posted online, as an introduction to MATLAB. The basics of will additionally be covered in recitation, so if you do not have experience, please be sure to review the online information or attend recitation. To complete this assignment, you will need access to MATLAB (follow directions to download from http://web.mit.edu/student-matlab/)

Question 1 – Plotting with MATLAB

On homework one, you explored the secondary structure of P-glycoprotein using a ramachandran plot. For that question, you were asked to use pyRosetta to find the phi and psi angles of the protein and plot them.   MATLAB can also make similar plots. To learn how to import files to MATLAB and plot, you'll make the same plot, this time using MATLAB.

Download the excel file containing the phi and psi coordinates from the course website. To import an excel file into MATLAB, try familiarizing yourself with the function "xlsread". To plot a scatter plot, consider the tool "scatter."

If you're not sure how to use these commands, try "help xlsread" or "help scatter"

Please attach your MATLAB ramachandran plot, including axis labels and title generated with MATLAB. Be sure to upload the .m file you used to do this onto course website!

To make your life easier with figures in MALAB, "file publish" from your .m file window will save all the figures generated into your current directory, in the "html" folder.

a) Regenerate the ramachandran plot from homework 1, this time using MATLAB

ramachandran plot of P-glycoprotein
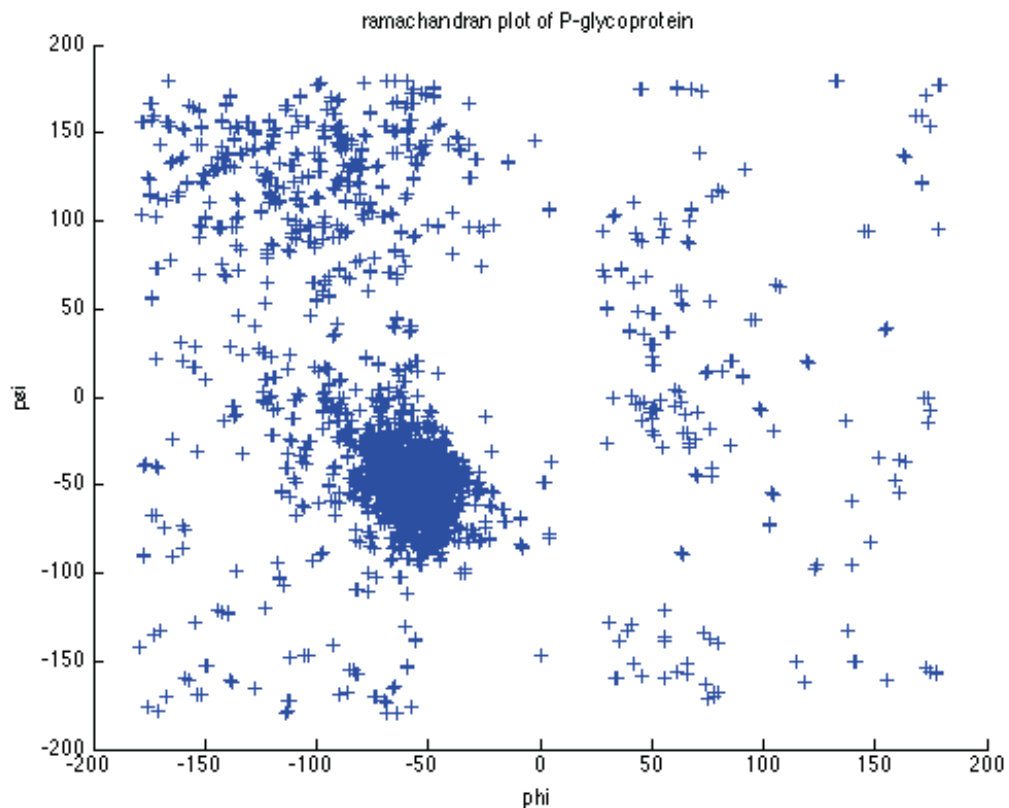
b) Now, suppose you want to look at only the phi or psi angles, not both. Plot phi vs. residue number and psi vs. residue number, each as a subplot on the same plot (Note this plot is not meant to be very informative, but gives you practice at using subplots in MATLAB the function "subplot" will be helpful for you!).
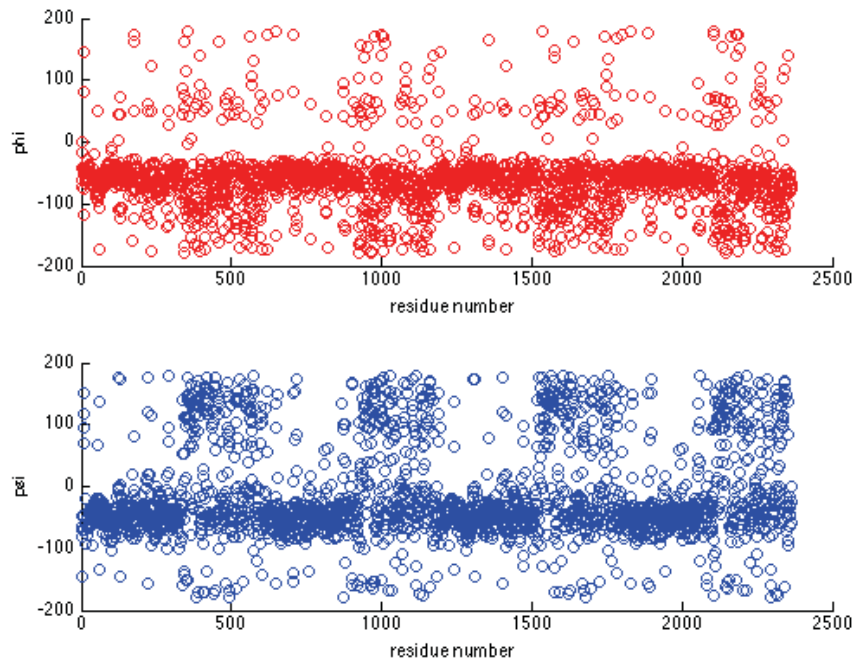
c) Recall in the first homework, we estimated the amount of helical residues in the protein. Now, we'll do the same for helix and for beta sheet. Based on the angles below, what percentage of P-glyprotein is alpha helix? What percentage is beta sheet?
   a. Assume helices have phi from -90 to -10 and psi from -100 to -10
      61.1252% alpha helix
   b. Assume Beta sheets are –phi from -150 to -10 and psi from 60 to 180.
      15.8206% beta sheet

MATLAB Code:

```matlab
% close all figures previously opened and remove old variables - this
% prevents artifacts from prior running of the program
clear all
close all


%Import excel file as "angles"
angles = xlsread('angles.xls');

%plot phi vs. psi
figure(1)
scatter(angles(:,1), angles(:,2), '+')
xlabel('phi')
ylabel('psi')
title('ramachandran plot of P-glycoprotein')

%plot phi and psi vs. residue number
figure(2)
res_num = [1:size(angles,1)]; %make a list of all the residue numbers
from one to total number of residues
subplot(2,1,1)
scatter(res_num, angles(:,1), 'r')
xlabel('residue number')
ylabel('phi')
subplot(2,1,2)
scatter(res_num, angles(:,2), 'b')
xlabel('residue number')
ylabel('psi')


%estimate helices
numres = length(angles); %total number of residues
count_alpha = 0;
count_beta = 0;
for i=1:numres
    phi = angles(i, 1);
    psi = angles(i, 2);

    if (-90<phi) && (phi< -10) && (-100 < psi) && (psi < -10)
        count_alpha = count_alpha + 1;

    elseif (-150 <phi) && (phi <-10) && (60 <psi) && (psi < 180)
            count_beta = count_beta + 1;
    end
end
disp('the percentage residues of alpha helices:')
percent_helix = count_alpha/numres*100
disp('the percentage of residues in beta sheets')
percent_beta = count_beta/numres*100
```
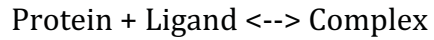
Question 2 – Curve Fitting with MATLAB

Assume a ligand interacts with a protein to form a complex in the basic manor:

Protein + Ligand <--> Complex

You performed an experiment where you were able to measure complex formation, after adding 100M ligand to your system, as a function of time, and want to see if you can determine the parameters k_{on} and k_{off}.

The data you found is as follows:

time(s)	Complex(M)
0	0.023
0.01	0.3887
0.02	0.5934
0.03	0.7252
0.04	0.807
0.05	0.8807
0.06	0.8774
0.07	0.9205
0.08	0.9908
0.09	0.9402
0.1	0.9142
0.11	0.9083

a) Plot this data using MATLAB


Complex formation data
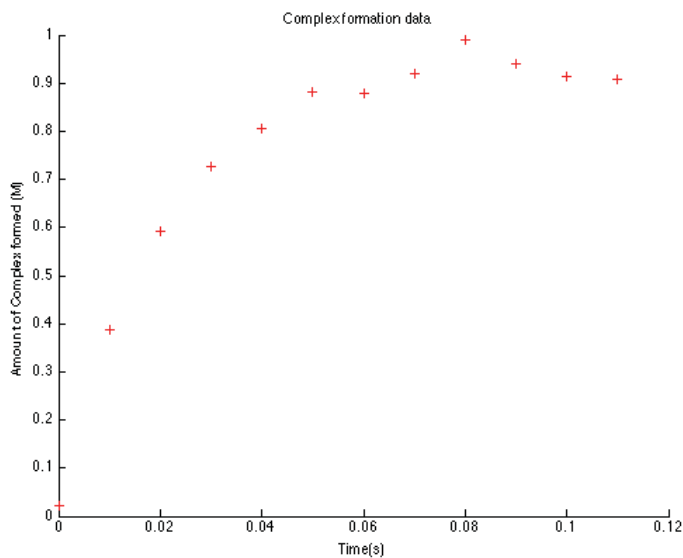
b) You expect your curve to fit the following function:

$$C = \frac{L_0}{L_0 + \dfrac{k_{off}}{k_{on}}} * (1 - e^{-(k_{on}*L_0 + k_{off})*t})$$

Using nonlinear fitting in MATLAB, fit the curve to your data, and extract the parameters koff and kon. Be sure to include your units!

kon = 0.4758 M^{-1}s^{-1} koff = 2.8726 s^{-1}

c) Include a new plot, with the data included as '+' signs, and overlay your best fit line. Be sure to label all axes, and include a legend with your figure. The command "hold on" will likely help you here.
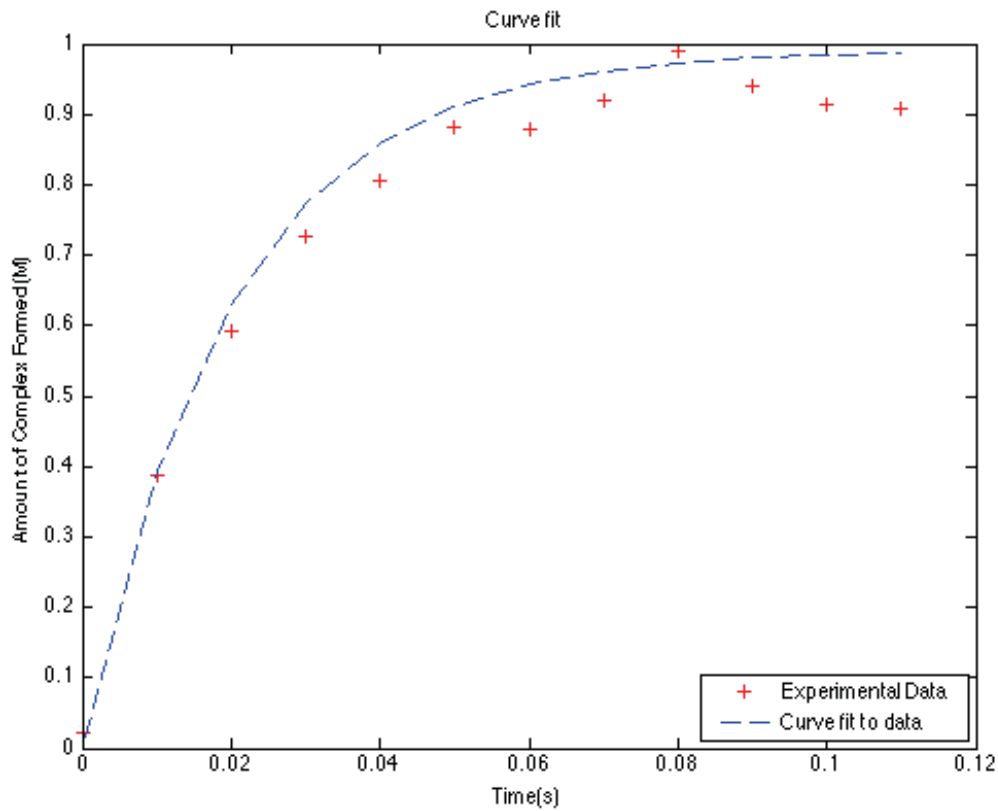
```matlab
function main
clear all
close all


%enter your data
complex = [0.023, 0.3887, 0.5934,0.7252, 0.807, 0.8807, 0.8774, 0.9205,
0.9908, 0.9402, 0.9142, 0.9083];
t = linspace(0,.11,length(complex));

%scatter plot
figure(1)
scatter(t, complex, 'r+')
xlabel('Time(s)')
ylabel('Amount of Complex formed (M)')
title('Complex formation data')

%Parameters
global L0;
L0 = 100; %Initial Molar value of ligand
%k = [kon, koff]
k = [0.5, 0.5]; %initial guess for k

fit_k = nlinfit(t, complex,@C_function, k);

disp('kon = ')
disp(fit_k(1))
disp('koff = ')
disp(fit_k(2))

% get fit values for complex with your optimal k values
fit = C_function(fit_k, t)

% plot fitted data
figure(2)
plot(t, complex, 'r+')
hold on
plot(t, fit, 'b--')
legend('Experimental Data', 'Curve fit to data', 'location',
'southeast')
xlabel('Time(s)')
ylabel('Amount of Complex Formed(M)')
title('Curve fit')
end


function y = C_function(kappa, t)
%a function that returns the value of y for a given t, and k values
global L0
y = L0./(L0 + (kappa(2)./kappa(1))).*(1-exp(-
(kappa(1).*L0+kappa(2)).*t));
end
```

Question 3 - Isothermal Titration Calorimetry

A high-throughput drug screen can identify small molecules with high affinity for a receptor or other target of interest. Small molecule compounds can then be modified to increase affinity or specificity for a target. Determining the thermodynamic properties of the drug-target interaction can guide attempts to increase affinity. For a compound of interest, SM10, the K_d for binding was estimated using the initial screen to be $2x10^{-7}M$ at 25°C.

a) Design an ITC experiment to determine enthalpic vs. entropic contributions to the binding affinity (be sure to include protein and ligand concentrations, and relevant volumes)
From class, Protein concentration should be between 10 and 100 times the Kd value, meaning acceptable protein concentrations are between **$2x10^{-6}$ and $2x10^{-5}$ M**. Ligand concentration should typically be about **10 times whatever protein concentration you chose.** Injection volume is 10uL, and the volume of the cell is 1mL (essentially, you dilute your ligand 100 fold each time, so after 10 injections, it's equal protein and ligand concentration).

The rest of the solutions are based on [Protein] = $2x10^{-6}M$ and [Ligand] = $2x10^{-5}$ M.

b) Based on your experimental design from part a, calculate the heat absorbed or released after the fifth injection of ligand (from injection 4 to injection 5 only!) , given ΔH of binding = -11,000 cal/mol. Be sure to show all of your work!

$$Q_5 = V * \Delta H * ([C]_5 - [C]_4)$$

$$[C] = y * [P]_0 = \frac{(K_d + [L]_0 + [P]_0) - \sqrt{(K_d + L_0 + [P]_0)^2 - 4 * [L]_0 * [P]_0}}{2}$$

The trick here is that L_0 is different at every injection (ligand accumulates), so you need to figure out what this is, and then solve for complex and Injection 5 and injection 4

$$[L]_{0,5} = \frac{[L]_0 * (50 * 10^{-6})}{10^{-3} + 50 * 10^{-6}} = 9.52 * 10^{-7}$$

$$[L]_{0,4} = \frac{[L]_0 * (40 * 10^{-6})}{10^{-3} + 40 * 10^{-6}} = 7.69 * 10^{-7}$$

Plug these values in for $[L]_0$ and you can find $[C]_5 = 8.15*10^{-7}$ and $[C]_4 = 6.69*10^{-7}$ Plug back in for Q, and get:

$$Q_5 = 1.05 * 10^{-3} * 11,000 * (8.15 * 10^{-7} - 6.69 * 10^{-7}) = -\mathbf{1.69 * 10^{-6}}$$

c) Find the change in Entropy (ΔS) of binding given the information above and any necessary information from parts a and b

$$\Delta G = \Delta H - T \Delta S$$

$$\Delta Gbind = RTlnK_d$$

$$\Delta S = \frac{\Delta H - RTln(K_d)}{T} = \frac{-11,000 \frac{cal}{mol} - 1.98 \frac{cal}{mol * K} * 298K * \ln(2 * 10^{-7}M)}{298K}$$

$$\Delta S = -6.3714 \; \frac{cal}{mol * K}$$

d) ITC does not work well for interactions that are extremely weak or extremely strong. Please explain this observation.

If the interaction is too weak, binding is very minimal, heat change is very small, and signal level is low. If Kd is too low (very strong interactions), almost all of the ligand is bound at every injection, so it is difficult to estimate Kd (nothing to plot). Need to have varying levels of fractional saturation to get accurate Kd measurements.

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012