General Instructions:

1. You are expected to state all 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. Both the results of any calculations and the corresponding code must be printed and attached to the solutions. **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 get sent to different graders, which should allow us to get the graded problem set 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.

4. This problem set is due at noon on Friday, October 23rd. 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 was provided on the first day of class and follow the guidelines carefully.

As discussed in class, drug discovery often involves screening large libraries of small molecules to identify those that have favorable interactions with a certain "druggable" target. One of the first tests to determine whether a small molecule could potentially have a therapeutic effect is to measure its binding affinity to the target of interest. These are typically performed in high-throughput assays to select for those compounds that bind the target with the highest affinity.

A) Briefly discuss the role of binding affinity in determining the overall therapeutic effect of a small molecule. At this stage in the drug discovery process, given a list of compounds and their affinities to a particular target, what can you conclude about the relative effectiveness of each compound as a drug?

Downstream screening steps reveal a small-molecule lead (code-named SML1, MW = 314 Da) to be a promising candidate. Results from the initial high-throughput screen show that SML1 binds to its target TAR2 (MW = 98 kDa) with K_d = 290 nM. You are interested in improving the affinity of this interaction. In order to do this, you decide to perform an isothermal titration calorimetry (ITC) experiment to deduce more detail about the thermodynamics of the binding reaction.

B) Design the ITC experiment you would use to determine the entropic vs. enthalpic contributions to binding affinity between SML1 and TAR2. Be sure to specify/calculate the following quantities, based on guidelines discussed in class.

   - The concentrations of TAR2 and SML1
   - The number of injections and the injection volume
   - The volume of the sample cell
   - The total amount [i.e. the mass] of purified target protein your experiment will require
   - The total amount [i.e. the mass] of SML1 your experiment will use.

C) For the experimental set-up you described in Part B), calculate the amount of heat absorbed or released after the third injection of ligand, given ΔH = -10.7 kcal/mol.

D) Describe the pseudo-first order approximation (both mathematically and physically) and discuss whether this is a valid assumption for ITC.

E) Why are ITC experiments only accurate within a certain range of binding affinities?

Dasatinib (Sprycell®) is a small molecule tyrosine kinase inhibitor developed by Bristol-Myers Squibb with fairly exquisite specificity for Src family kinases and Abl family kinases, with low affinity for most other kinases. This drug is currently used as second-line therapy for chronic myeloid leukemia patients that have developed resistance to Gleevec, the first-line therapeutic.

Some initial data has emerged that Dasatinib might have an off-target effect in which it is binding to and inhibiting the EphB4 receptor tyrosine kinase. Initial estimates of the K_d for this interaction are 5×10^{-8} M.

To better quantify the affinity and kinetics of this interaction, a Surface Plasmon Resonance (SPR) experiment was performed in which 450 fmoles of Dasatinib was immobilized onto a sensor chip with dimensions 1.25 cm × 1.25 cm. EphB4 at a concentration of 15 nM was then washed over the immobilized Dasatinib, and the reaction was allowed to reach equilibrium.

A) Why was Dasatinib (rather than EphB4) immobilized onto the SPR sensor chip?

B) Given $k_{on} = 8.4 \times 10^4$ L mol^{-1}sec^{-1} and $k_{off} = 3.1 \times 10^{-3}$ sec^{-1}, calculate the time required for this system to reach equilibrium.

C) How much of the Dasatinib on the surface is bound by EphB4 at equilibrium?

D) Determine the maximum Resonance Units (RUs) measured for this binding interaction. Assume a molecular weight of 488 Da for Dasatinib, and 108 kDa for EphB4.

You are interested in investigating the interaction between a protein kinase KIN1 and its substrate SUB1. While it is well established that KIN1 is primarily responsible for phosphorylating SUB1, the mechanism of SUB1-mediated signaling is poorly understood. To date, twenty other proteins have been identified as potential interacting partners for SUB1, which has been shown to possess six accessible phosphorylation sites. You decide that the next step in probing this signaling pathway is to determine which phosphorylation sites on SUB1 are bound by the SH2 domains of these 20 proteins.

A) Design a protein microarray experiment to address this question. In your experimental design, be sure to [briefly] discuss:
   - *Experimental setup*: What will be immobilized on your microarray, and how will you detect binding?
   - *Expected results*: Draw what you anticipate your data will look like for a representative cell in your microarray. What parameter will your data help you calculate?
   - *Interpretation:* How do these calculations help you determine the relative specificity of the SH2 domains for different tyrosine residues on SUB1?

A small subset of the data from this experiment is posted on the Course website. The file `PS5Q3data.mat` contains raw microarray readouts from the SH2 domain of SUB17 (a candidate SUB1 interaction partner) binding to each phosphorylation site on SUB1. The names of the phosphorylation sites are stored in the cell array `phosphoSites`. Each row represents a series of binding experiments using a single phosphorylation site with the columns representing nine initial ligand concentrations. These concentrations (in nM) are stored in the vector `L0`.

B) Begin your analysis by loading these results into MATLAB and transforming each data point into a fractional saturation [Hint: assume that the maximum measured value **in the entire experiment** corresponds to full saturation]. In a single figure, plot the fractional saturation of each phosphorylation site with respect to initial ligand concentration. Each site should be represented by a subplot in your figure, and should have consistent axes to facilitate direct comparison.

To receive full credit for this problem, you must automate the data transformation and plotting processes (i.e. you should only need one line of code to transform all your data, and you should only call the plot function once for this part of the problem). You may need to employ the `max` command and use a `for` loop.

Note: `phosphoSites` is a cell array and does not require quotation marks or apostrophes when called from the `ylabel`, `xlabel`, or `title` commands in MATLAB.

C) Using `nlinfit`, calculate the K_d for the SH2 domain of SUB17 interacting with each phosphorylation site on SUB1. You will need to call `nlinfit` individually for each data set. For maximum credit, be sure to follow the coding guidelines regarding loops established in Part B).

D) In a new figure, overlay the binding curves based on the K_d values you calculated on top of the values you obtained in Part B). Use the `linspace` command to give you an appropriate basis of x-values from which to compute the points for your binding curves. Again, for maximum credit, be sure to follow the coding guidelines regarding loops established in Part B).

E) Based on your analysis, which phosphorylation site on SUB1 is likely responsible for the most downstream signaling through SUB17?

F) A key step in evaluating the fit of a model is examining the residuals between the model and the data. In a new figure, plot these residuals. Was your model a good fit for your binding data? Again, for maximum credit, be sure to follow the coding guidelines regarding loops established in Part B).

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012