

20.320 Problem Set 5
Question 1

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?

Binding affinity between a small molecule and its target is an important factor for determining therapeutic effect. In order for a small molecule to affect a target, it must associate with the target long enough to produce the desired change. However, binding affinity is not the sole parameter that must be optimized. In other words, binding tighter to the target does not guarantee a better therapeutic outcome. This stage of the drug discovery process seeks to identify compounds that associate with the target molecule with at least a certain affinity. Choosing between compounds that meet this criteria happens based on results of preclinical and clinical trials.

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.

ITC experiments use two cells filled with solution – a reference cell (where no reaction is taking place), and a sample cell, with target protein in solution. In this case, TAR2 (at a concentration 10x the K_d of the binding reaction) would be placed in a 1 mL sample cell, and 10 μ L of SML1 would be injected 20 times at a concentration 10x that of TAR2. Because the system aims to keep both cells at the same temperature throughout the experiment (and because SML1 binding TAR2 is an exothermic reaction), the amount of heat applied to the sample cell would decrease briefly when SML1 is injected into the sample cell. Integration of these peaks yields points on a binding isotherm, from which binding data can be extracted.

20.320 Problem Set 5
Question 1

As discussed in lecture, ITC experiments typically require a protein concentration roughly 10-fold above the K_d of the drug-protein interaction. Assuming a cell volume of 1 mL:

$$\begin{aligned} [\text{TAR2}] &= 10 \times 290 \text{ nM} = 2.9 \text{ } \mu\text{M} \\ 2.9 \times 10^{-6} \text{ mol/L} \times 10^{-3} \text{ L} &= 2.9 \times 10^{-9} \text{ mol TAR2} \\ 2.9 \times 10^{-9} \text{ mol} \times 98,000 \frac{\text{g}}{\text{mol}} &= 284 \text{ } \mu\text{g TAR2} \end{aligned}$$

For this design, twice as much SML1 is required compared to TAR2. Therefore twice as many moles of SML1 are required.

$$2 \times 2.9 \times 10^{-9} \text{ mol} \times 314 \frac{\text{g}}{\text{mol}} = 182 \text{ } \mu\text{g SML1}$$

Students' answers do not need to match these solutions exactly, but their calculations should be consistent with the values they specified in their design.

- 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 $\Delta H = -10.7 \text{ kcal/mol}$.

Assume that each injection is 10 μL , and that the injected solution consists of SML1 at a concentration of 29 μM (i.e. 10x the concentration of TAR2). Using the equation below,

$$[C] = y[P]_0 = \frac{1}{2} \left[(K_d + [L]_0 + [P]_0) - \sqrt{(K_d + [L]_0 + [P]_0)^2 - 4[L]_0[P]_0} \right]$$

In the above setup, we chose a value of $[P]_0$ (i.e. $[\text{TAR2}]_0$) that was 10x the K_d of TAR2 binding SML1. After the second injection, 20 μL of a 29 μM solution of SML1 has been added to a starting volume of 1 mL. Therefore:

$$[L]_{0,2} = \frac{(29 \times 10^{-6} \text{ mol/L})(20 \times 10^{-6} \text{ L})}{1 \times 10^{-3} \text{ L} + 20 \times 10^{-6} \text{ L}} = 5.69 \times 10^{-7} \text{ M}$$

$$[P]_0 = 2.9 \times 10^{-6} \text{ M}$$

Therefore:

$$[C]_2 = y[P]_0 = \frac{1}{2} \left[(K_d + [L]_0 + [P]_0) - \sqrt{(K_d + [L]_0 + [P]_0)^2 - 4[L]_0[P]_0} \right] = 5.1 \times 10^{-7} \text{ M}$$

Similarly, after the third injection:

$$[L]_{0,3} = \frac{(29 \times 10^{-6} \text{ mol/L})(30 \times 10^{-6} \text{ L})}{1 \times 10^{-3} \text{ L} + 30 \times 10^{-6} \text{ L}} = 8.4 \times 10^{-7} \text{ M}$$

$$[C]_3 = y[P]_0 = \frac{1}{2} \left[(K_d + [L]_0 + [P]_0) - \sqrt{(K_d + [L]_0 + [P]_0)^2 - 4[L]_0[P]_0} \right] = 7.4 \times 10^{-7} \text{ M}$$

20.320 Problem Set 5
Question 1

Therefore:

$$Q_3 = \Delta H([C_3] - [C_2])V_{\text{cell}} = (-10.7 \text{ kcal/mol})(7.4 \times 10^{-7} - 5.1 \times 10^{-7} \text{ mol/L})(1 \times 10^{-3} \text{ L})$$

$$Q_3 = -2.5 \times 10^{-9} \text{ kcal, or } -2.5 \times 10^{-6} \text{ cal}$$

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

The pseudo-first order approximation can be used when ligand is in excess, and the concentration of free ligand in the solution can be approximated as constant as the system approaches equilibrium. In other words, binding protein to excess ligand causes a negligible decrease in the amount of free ligand in solution. Mathematically, this corresponds to setting $[L]_{\text{eq}} = [L]_0$.

The equation for fractional saturation of protein at equilibrium is:

$$y_{\text{eq}} = \frac{[L]_{\text{eq}}}{[L]_{\text{eq}} + K_d}, \text{ and applying the pseudo-first order approximation, } y_{\text{eq}} \approx \frac{[L]_0}{[L]_0 + K_d}$$

This is not appropriate for use in ITC calculations, because ITC involves injecting small amounts of *ligand* into a solution of *excess free protein* and measuring the heat evolved as the system reaches equilibrium. Therefore ligand is not in excess when heat evolution is measured, and the depletion of ligand must be considered when performing ITC calculations.

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

For very strong interactions, saturation can be reached very rapidly and abruptly without providing data points in the middle to allow us to determine the K_d accurately. For weak binders, the opposite problem occurs: each additional injection may not generate enough new complex formation to register on the detector and give reliable readout. Therefore, the protein-ligand affinity must be such that injections of ligand will result in a full range of fractional saturations.

20.320 Problem Set 5
Question 2

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 $1.25 \text{ cm} \times 1.25 \text{ cm}$. 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?

This configuration was chosen to maximize signal. Immobilizing the small molecule on the surface of the sensor chip results in a large change in mass when a large macromolecule like an enzyme binds. Since the change in resonance angle (and therefore, the RU) is proportional to the change in mass of each immobilized species upon binding, this configuration maximizes the dynamic range, and therefore the sensitivity of the experiment.

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

$$\tau_{1/2} = \frac{\ln 2}{k_{\text{on}}[L_0] + k_{\text{off}}} = \frac{\ln 2}{8.4 \times 10^4 \frac{\text{L}}{\text{mol-s}} \left[15 \times 10^{-9} \frac{\text{mol}}{\text{L}} \right] + 3.1 \times 10^{-3} \frac{1}{\text{s}}} = 159 \text{ sec}$$

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

$$y_{\text{eq}} \approx \frac{[L]_0}{[L]_0 + K_d} = \frac{15 \times 10^{-9} \text{ M}}{15 \times 10^{-9} \text{ M} + 5 \times 10^{-8} \text{ M}} = 0.23$$

$$0.23(450 \text{ fmol}) = 104 \text{ fmol}$$

20.320 Problem Set 5
Question 2

- 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.

The maximum signal (in RU) will result when every immobilized Dasatinib molecule is bound to one molecule of EphB4. Assuming that this is a 1:1 binding reaction, the total amount of material bound to the chip can be calculated.

$$(450 \times 10^{-15} \text{ mol complex}) \left(6.02 \times 10^{23} \frac{\text{molecules}}{\text{mol}} \right) = 2.71 \times 10^{11} \text{ molecules complex}$$

$$(2.71 \times 10^{11} \text{ molecules}) \left(\frac{108488 \text{ Da}}{\text{molecule complex}} \right) \left(\frac{1.66 \times 10^{-24} \text{ g}}{1 \text{ Da}} \right) \left(\frac{10^{12} \text{ pg}}{\text{g}} \right) = 4.88 \times 10^4 \text{ pg complex}$$

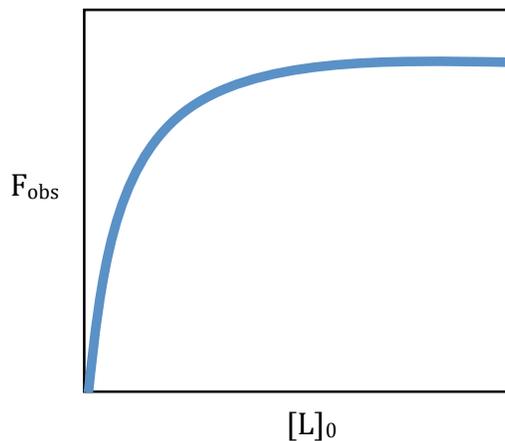
$$RU_{\max} = \frac{4.88 \times 10^4 \text{ pg}}{156.25 \text{ mm}^2} = 312 \text{ pg/mm}^2$$

20.320 Problem Set 5
Question 3

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?

For this experiment, SH2 domains of the 20 proteins of interest will be expressed recombinantly and immobilized on the surface of the microarray. Fluorescently labeled peptides from SUB1, each containing the phosphorylated residue of interest, will be washed over the microarray, and fluorescent spots will be read by an imager. Representative data would resemble a saturation curve for each cell, which allows a K_d to be calculated.



Comparing the K_d of SH2 domains between different pairs of proteins and phosphorylation sites will determine which protein-phosphosite interactions are most favorable. Proteins whose SH2 domain has the lowest K_d value for a specific phosphorylation site are likely to be involved with signaling pathways that are affected by the phosphorylation of a specific residue.

20.320 Problem Set 5
Question 3

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.

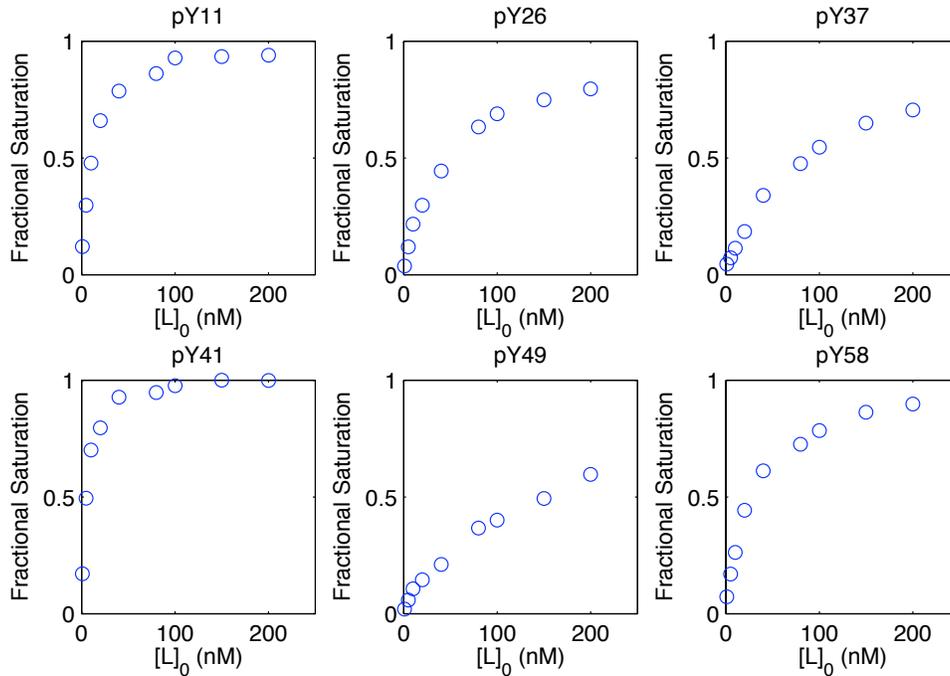
Note: `phosphoSites` is a cell array and does not require quotation marks or apostrophes when called from the `ylabel`, `xlabel`, or `title` commands in MATLAB.

```
load PS5Q3data.mat

% Transform data into fractional saturations by dividing by the
absolute maximum
y = Data./(max(max(Data)));

% Plot fractional saturations in individual subplots
figure(1)
for i = [1:6]
    subplot(2,3,i)
    plot(L0, y(i,:), 'o')
    ylabel('Fractional Saturation')
    xlabel('[L]_0 (nM)')
    title(phosphoSites(i))
    axis([0 250 0 1])
end
```

20.320 Problem Set 5
Question 3



- 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. Be sure to follow the coding guidelines established in Part B).

For each phosphorylation site (pY11, pY26, pY37, pY41, pY49, and pY58):

```
Kd =
    10.8103    46.6429    82.9007    4.6511   141.6944    26.1938
```

See Part D) for code.

- D) In a new figure, overlay the binding curves based on the K_d values you calculated with 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.

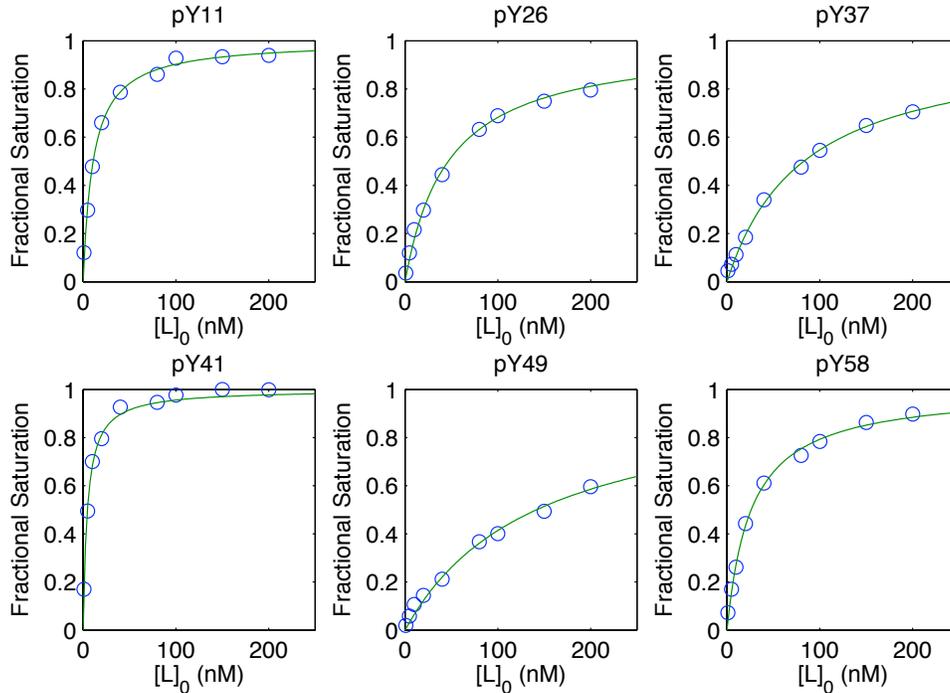
```
% Initialize array of Kd values which will be filled later
Kd = zeros(1,6);

% Plot fractional saturations with binding isotherms overlaid
figure(2)
x = linspace(0, 250); % Basis vector for drawing binding curves
for i = [1:6]
    subplot(2,3,i)
    Kd(1,i) = nlinfit(L0, y(i,:), @isotherm, 1); % Fit data to model
    y_fit = isotherm(Kd(1,i), x); % Calculate points for binding curves
    plot(L0, y(i,:), 'o', x, y_fit); % Plot data and binding curves
    ylabel('Fractional Saturation')
    xlabel('[L]₀ (nM)')
    title(phosphoSites(i))
    axis([0 250 0 1])
end
```

20.320 Problem Set 5
Question 3

Kd

```
function out = isotherm(Kd, L0)
out = L0./(L0 + Kd);
```



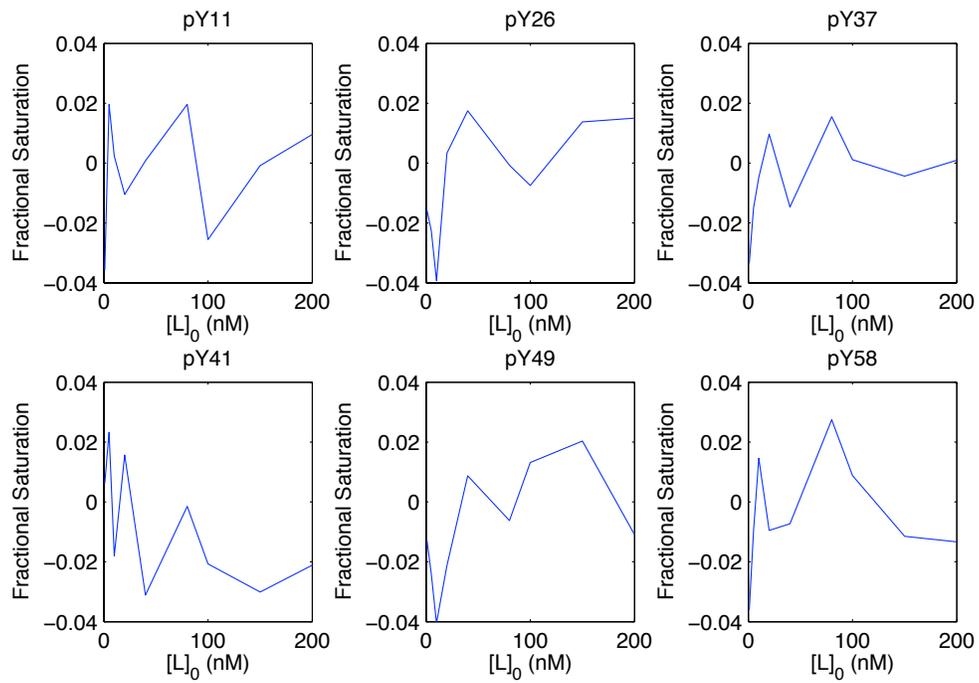
E) Based on your analysis, which phosphorylation site on SUB1 is likely responsible for the most downstream signaling through SUB17?

pY41 on SUB1 has the highest affinity for the SH2 domain of SUB17, and therefore is most likely involved with SUB17-mediated signaling.

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, be sure to follow the coding guidelines established in Part B).

```
% Calculate and plot residuals
figure(3)
for i = [1:6]
    y_resids = isotherm(Kd(1,i), L0) - y(i,:); % Calculate residuals
    subplot(2,3,i)
    plot(L0, y_resids)
    ylabel('Fractional Saturation')
    xlabel('[L]₀ (nM)')
    title(phosphoSites(i))
end
```

20.320 Problem Set 5
Question 3



The model appears to be a good fit for the binding data, as the residuals are small ($\leq 3\%$ variation from the model) and do not show systematic error.

MIT OpenCourseWare
<http://ocw.mit.edu>

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012

For information about citing these materials or our Terms of Use, visit: <http://ocw.mit.edu/terms>.