20.320 — Problem Set # 1

September 17th, 2010

Due on September 24th, 2010 at 11:59am. No extensions will be granted.

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. Staples 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 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. 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 was provided on the first day of class and follow the guidelines carefully.

107 points + 2 EC for problem set 1.

1 Surface Plasmon Resonance (SPR)

You are in charge of characterizing the pharmacokinetics of a set of 20 candidate drugs targeting the same cell surface receptor. In this problem, you will design experiments to characterize their binding kinetics by SPR and analyze the results.

a) Under what circumstances is SPR a suitable experimental method for your purpose? List 3 advantages over other methods for measuring molecular binding parameters and 3 conditions for its applicability.

Solution:

Advantages

- Get kinetic on and off rates as well as K_D
- Typically fairly fast (less than 1 h)
- Can be high-throughput, e.g. in 384-well format (but expensive)
- Requires much less material than ITC
- Lots of redundancy built in → get an idea for how precise the results can be taken to be (K_D from every run; can perform multiple runs at different $[L]_0$ to validate K_D)
- Simple, computer-controlled experiment (but can be deceptive — must still understand principles and analysis to manually check and recognize artifacts)
- can recover binding partners for mass spectrometry (e.g. for ligand fishing)
- multiple types of experiments and chips available (different immobilization chemistries; direct assay, competition experiments, inhibitor or buffer effects, ...)
- can measure weak affinities $\sim 100\mu M$

Conditions for applicability

- k_{on} must be between $\sim 10^3 - 10^7 M^{-1}s^{-1}$
- k_{off} must be between $\sim 10^{-5} - 0.5s^{-1}$; higher-affinity ligands do not fall off over the course of the experiment
- Must be able to immobilize one of the binding partners
- General consideration: Must have purified sample (and enough of it)

6 points overall: 1 point per advantage up to 3 total, and 1 point per condition up to 3 total.

b) In SPR, running just 1 experiment with your sample is *never* enough. List 3 controls to include and explain what possible artifacts they control for. Can you think of a fourth control which is useful to include if your ligand is itself a protein, but cannot be included if the ligand is a small molecule? Explain.

Solution:

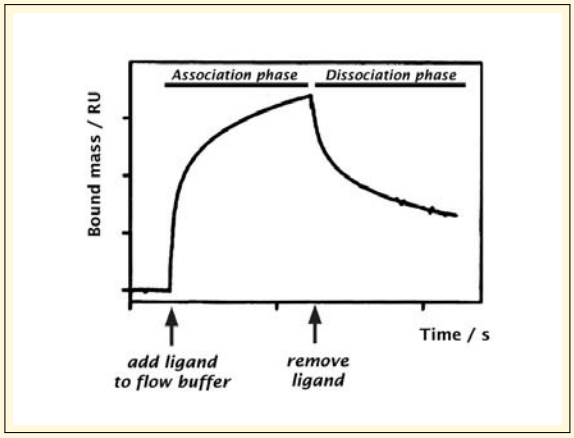
- Run buffer only without ligand → can change teh refractive index of the solution and produce a change in SPR angle
- Run at several (at least 5) different ligand concentrations
- At least one concentration in duplicate to control for drift (machine or regeneration losses due to dirty samples or imperfect regeneration of the chip)
- Different flow rates to control for ligand-rebinding because of insufficient mass transport
- Different densities of the immobilized species to control for avidity effects

One control which should always be included is the "inverse experiment", in which the other of the two binding partners is immobilized. However, for an interaction between a small molecule and a protein, it is not possible to run the experiment with the protein attached to the surface because the change in surface-bound mass caused by capturing a small molecule during the experiment is too small to be reliably measured.
8 points overall: 1 point for each control and 1 point for competent explanation of what it controls for. Liberally give points for mentioning controls, but be very strict about precise, unambiguous, and cogent explanations.

c) What physical quantity does SPR measure? What unit is it commonly given in, and how can this unit be converted into the relevant SI unit? Provide a clearly labelled diagram of the expected data output, indicating where external interventions (e.g. changes in buffer composition) take place. What steps are required to extract the kinetic rate constants kon and koff from the resulting data *without curve-fitting*? Provide an equation and an explanation for each step.

Solution:

- SPR measures the bound mass density on the chip surface. The measurements are commonly given in Resonance Units (RU), where 1000 RU $\hat{=}$ 1 ng mm^{-1}. Note: It is not necessary that the SI *base* units, kilogram and meter be used. Answers using grams or nanograms per unit area are acceptable.
- An SPR sensorgram plots the time-course of bound mass desnity as the ligand flow is turned on, left on until equilibration, and replaced with flow of buffer only until complete dissociation. A schematic drawing of a sensorgram, indicating the begin of ligand flow, the end of ligand flow, the association phase, and the dissociation phase, should be provided. The axis labels should clearly state the quantity and the units.



- To determine k_{off},
  - i) Read the midpoint of the dissocation curve off the diagram and estimate the half-time of dissociation, $\tau_{1/2,off}$, as the time between ligand removal and reaching half-maximal RUs.
  - ii) Calculate k_{off} as
$$k_{off} = \frac{\ln 2}{\tau_{1/2,off}}$$
  - iii) Next, analogously determine k_{obs} from the midpoint of the association curve as
$$k_{obs} = \frac{\ln 2}{\tau_{1/2,on}}$$
  - iv) Then, calculate k_{on} from the definition of k_{obs}, $k_{obs} = k_{on}[L]_0 + k_{off}$, as
$$k_{on} = \frac{k_{obs} - k_{off}}{[L]_0}.$$

8 points overall: 1 for RU, 1 for conversion to pg/mm^2, 2 points for diagram (deduct half point for each missing or incorrect unit or label), 1 point per step i) — iv) in calculation.

d) In one experiment, you flow your soluble species L through at 10 nM and read the following points off your sensorgram:

	Time / s	Signal / RU
Ligand flow begins	5	1
Ligand flow ends after the reaction has equilibrated	130	99
Run ends	300	1
Half-maximal signal is obtained	21.5	50
	147.3	50

Using your equations above, calculate K_D. Show all work.

Solution:

Start with the dissociation phase:

$$k_{off} = \frac{\ln 2}{\tau_{1/2,off}} = \frac{\ln 2}{147.3s - 130s} = 0.0400s^{-1}$$

Then, the association phase:

$$k_{obs} = \frac{\ln 2}{\tau_{1/2,on}} = \frac{\ln 2}{21.5s - 5s} = 0.0420s^{-1}$$

From these two, obtain

$$k_{on} = \frac{k_{obs} - k_{off}}{[L]_0} = \frac{0.0420s^{-1} - 0.0400s^{-1}}{10nM} = \frac{2 \cdot 10^{-3}s^{-1}}{10^{-8}M} = 2 \cdot 10^5 M^{-1}s^{-1}$$

Finally,

$$K_D = \frac{k_{off}}{k_{on}} = \frac{0.04s^{-1}}{2 \cdot 10^5 M^{-1}s^{-1}} = 2 \cdot 10^{-7}M.$$

6 points overall: 2 points for correctly setting up the problem (i.e. first consider dissociation phase, then association phase, then calculate K_D from k_{off} and k_{on}); 1 point for each calculation above. It is also acceptable to first set up a full analytical expression, $K_D = \frac{k_{off}}{k_{on}} = \frac{\frac{\ln 2}{\tau_{1/2,off}}}{\frac{\frac{\ln 2}{\tau_{1/2,on}} - \frac{\ln 2}{\tau_{1/2,off}}}{[L]_0}}$.

e) If the concentration of the soluble species were increased to 50 nM, how long would take for the association phase reaction to reach exactly 99% of equilibrium binding?

Solution:

At 99% equilibrium in the association phase,

$$\begin{aligned}
\frac{[\mathrm{C}]}{[\mathrm{C}]_{\mathrm{eq}}} = 0.99 &= 1 - e^{-\mathrm{k_{obs}}t} \\
&= 1 - e^{-\left(\mathrm{k_{on}}[\mathrm{L}]_0 + \mathrm{k_{off}}\right)t} \\
0.01 &= e^{-\left(2\cdot 10^5 \mathrm{M^{-1}s^{-1}} \times 5\cdot 10^{-8}\mathrm{M} + 0.04\mathrm{s^{-1}}\right)t} \\
&= e^{-\left(0.05\mathrm{s^{-1}}\right)t} \\
t &= 92.1s
\end{aligned}$$

4 points overall: 3 points for correct setup (i.e. first line of the equations; realizing what 99% equilibrium means), and 1 point for numerical answer.

f) Assume your soluble ligand is the protein target of interest with a molecular weight of 450 kDa and that 450 fmol of your drug molecule (MW = 500 Da) have been immobilized on a 1 cm by 1 cm chip. In the scenario just described (i.e. $[\mathrm{L}]_0 = 50$ nM), what is $\mathrm{RU_{eq}}$?

Solution:

First, we need to determine the fraction of ligand bound to the immobilized species. At 50 nM, the ligand is in excess and so the pseudo-first order approximation may be made such that the fractional saturation becomes

$$\mathrm{y} = \frac{[\mathrm{L}]_0}{[\mathrm{L}]_0 + \mathrm{K_D}} = \frac{50\mathrm{nM}}{50\mathrm{nM} + 200\mathrm{nM}} = 0.2$$

Since we know that 450 fmol of the small molecule have been immobilized and that 20% of these bind to a receptor molecule, 90 fmol of the protein will be bound to the surface at equilibrium. Neglecting the mass of the small molecule, the total mass bound is approximately

$$450\mathrm{kDa} \cdot 90\mathrm{fmol} = 450 \cdot 10^3 \mathrm{g\ mol^{-1}} \cdot 90 \cdot 10^{-15}\mathrm{mol} = 4.05 \cdot 10^{-8}\mathrm{g}$$

SPR measures the surface density of bound mass. It can be obtained from the total mass by division by the surface area,

$$\frac{4.05 \cdot 10^{-8}\mathrm{g}}{10 \times 10\mathrm{mm}} = \frac{40500\mathrm{pg}}{100\mathrm{mm^2}} = 405\mathrm{RU}.$$

6 points overall: 2 points for each step of the calculation.

38 points overall for problem 1.

2 Isothermal Titration Calorimetry (ITC)

Your drug candidate performed well in kinetic testing by SPR — congratulations. Before commencing cell-based activity assays, you decide to control for artifacts of immobilization in your SPR data by also performing an ITC experiment. First, you will choose reasonable experimental parameters based on your prior knowledge. Then, you will simulate the experiment to ensure that the output is likely to be informative over a range of reasonable values for the key parameters.

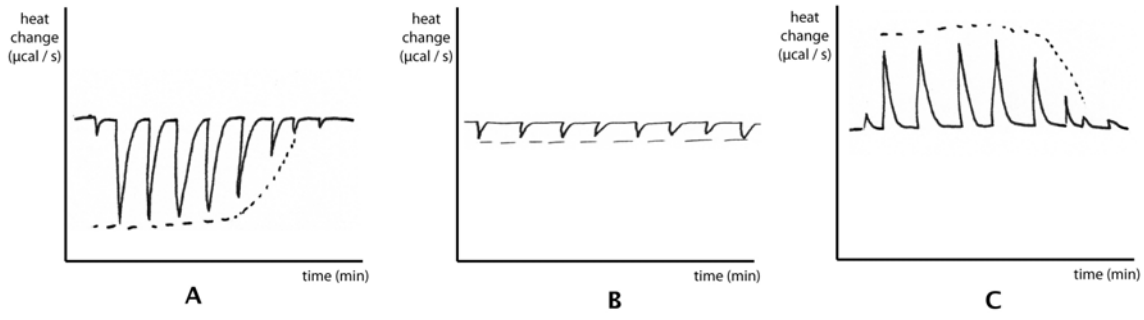
a) ITC is a powerful tool to extract a wide range of thermodynamic parameters without any need for modification or immobilization of the binding partners. Why did you not use it in initial screening? State two reasons.

Solution:

- Require much more sample than for SPR
- Experiment takes longer (order of hours)
- No simple way to parallelize for higher throughput
- Equilibrium data, hence no information about kinetic parameters

2 points.

b) You may assume that the ΔH for your reaction is $\mathcal{O}$(–10 kcal mol^{-1}). Which of the following graphs provides the best qualitative approximation to the expected raw data readout (dQ/dt vs. time) from your experiment? What do the other graphs represent? Extra credit for suggesting two qualitatively distinct interpretations of the peaks in (B).



Solution:

- (A) is the output from an exothermic reaction — this is what is expected here.
- (B) shows small peaks and no saturation. This could be a too low concentration of ligand, which does not achieve saturation over the time of the experiment. It could also be the readout from an ITC experiment with no binding at all – the peaks could represent the heat of solvation of the ligand being introduced (quite noticeable e.g. for titration of DNA, which is a polyelectrolyte, into an ITC chamber with no binding partners present).
- (C) shows the readout from an endothermic reaction.

6+2 points total: 2 per explanation.

c) The K_D you determined by SPR is $2 \cdot 10^{-7}$M, and you will use an ITC cell with $V_{cell} = 1$ mL and an injection volume of 10 uL per injection. What would be reasonable values for the protein concentration in the cell, $[P]_0$, and for the ligand concentration in the injected buffer solution, $[L]_0$? Justify your answer.

Solution:

- $[P]_0$ should be about 10-fold higher than the K_D, i.e. $[P]_0 \sim 2 \cdot 10^{-6}$M.
- $[L]_0$ should be markedly higher than $[P]_0$ so saturation can be achieved without significantly diluting P. Let us use $[L]_0 = 10[P]_0 = 2 \cdot 10^{-5}$M.

2 points total: 1 for each.

d) To calculate the heat released or absorbed after each injection, you will require an expression for the fractional saturation of protein molecules with ligand at equilibrium,

$$[y]_{eq} \equiv \frac{[C]_{eq}}{[C]_{eq} + [P]_{eq}} = \frac{[C]_{eq}}{[P]_0} = \frac{[L]_{eq}}{[L]_{eq} + K_D}.$$

State and explain the pseudo-first order approximation. Does it apply for ITC? Justify why or why not. Write down an appropriate mathematical expression for the fractional saturation in terms of only the quantities for which values have been given or estimated in part c) of this problem.

Solution:

- 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 $[\mathrm{L}]_{\mathrm{eq}} \approx [\mathrm{L}]_0$ and hence

$$[\mathrm{y}]_{\mathrm{eq}} \approx \frac{[\mathrm{L}]_0}{[\mathrm{L}]_0 + \mathrm{K_D}}.$$

- In ITC, the ligand is slowly titrated into the reaction cell in small increments – for most of the experiment, it is not present in excess over its binding partner. Therefore, the pseudo-first order approximation cannot be applied.

- Fractional saturation at equlibrium must thus be calculated as

$$[\mathrm{y}]_{\mathrm{eq}} = \frac{(\mathrm{K_D} + [\mathrm{L}]_0 + [\mathrm{P}]_0) - \sqrt{(\mathrm{K_D} + [\mathrm{L}]_0 + [\mathrm{P}]_0)^2 - 4[\mathrm{L}]_0[\mathrm{P}]_0}}{2[\mathrm{P}]_0}$$

8 points total: 3 for stating and explaining the PFOA, 3 for explanation why it cannot be applied in ITC, and 2 for writing down the correct equation. For the equation, deduct half points for minor inaccuracies.

e) The heat absorbed or released after the i^{th} injection (i.e. the area under the i^{th} peak in the graphs provided for Problem 2b)) can be calculated as

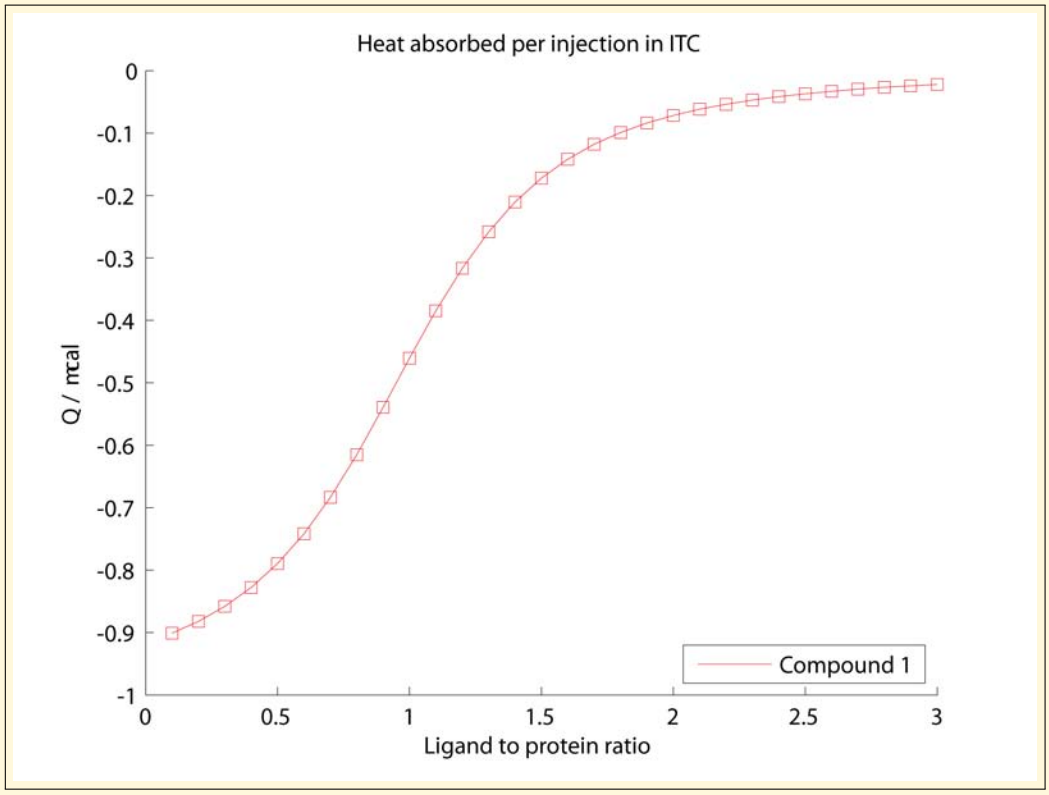
$$\mathrm{Q_i} = \Delta\mathrm{H} \cdot \mathrm{V_{cell}} \cdot ([\mathrm{C}]_{\mathrm{i}} - [\mathrm{C}]_{\mathrm{i\text{-}1}}),$$

where the concentration of bound complex is $[\mathrm{C}] = \mathrm{y}[\mathrm{P}]_0$. Implement a function in which calculates $\mathrm{Q_i}$ from these parameters. Implement another function which calculates the fractional saturation from appropriate parameters. Use these functions and the parameter estimates above to plot a graph of the heat released or absorbed, Q, against the molar ratio of ligand and protein, [L]/[P]. You have been provided with skeleton code for this problem; edit it where indicated to implement your solution. Remember to label the axes with the relevant quantities and units.

Solution:

itc.m:

```
%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%
function itc
clc;
close all;

Po = 2e-6;   % 2 x 10^-6 M
Ninj = 30;
Linj = 2e-5;
Vinj = 1e-5;
Vcell = 1e-3;
KD = 2e-7;
ΔH = -10;

%%%%%%%%%%%%%%%%%%%%%

[L,C,Q] = calc_all(Po,Ninj, Linj, Vinj, Vcell,KD,ΔH);
plot_itc_figure(L./Po, Q);
% For Problem 2F, can add these lines:
Linj = 10 * Linj;
[Lnew,Cnew,Qnew] = calc_all(Po,Ninj, Linj, Vinj, Vcell,KD,ΔH);
plot_itc_figure(Lnew./Po, Qnew);

%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%%

function [L,C,Q] = calc_all(Po,Ninj, Linj, Vinj, Vcell,KD,ΔH)
L(1) = ( (Vinj / Vcell)*Linj );
C(1) = fracSat(KD, L(1), Po);
Q(1) = calcQi(ΔH, Vcell, C(1), 0);
for i=2:Ninj
  L(i) = L(i-1) + ( (Vinj / Vcell)*Linj );
  C(i) = fracSat(KD, L(i), Po);
  Q(i) = calcQi(ΔH, Vcell, C(i), C(i-1));
end

function Qi = calcQi(ΔH, Vcell, Ci, Cprev)
   Qi = ΔH * Vcell * (Ci - Cprev);

function yeq = fracSat(KD, Lo, Po)
   KLP = KD + Lo + Po;
   yeq = (KLP - sqrt(KLP^2 -4*Lo*Po)) / (2*Po);

function plot_itc_figure(LP_ratio, Q)
figure;
hold on;
plot(LP_ratio, Q .* 1000, 'rs');
hl1 = plot(LP_ratio, Q .* 1000, 'r-');
hold off;
legend([hl1], 'Compound 1', 'Location', 'SouthEast')
title('Heat absorbed per injection in ITC');
xlabel('Ligand to protein ratio');
ylabel('Q / \mucal');
```

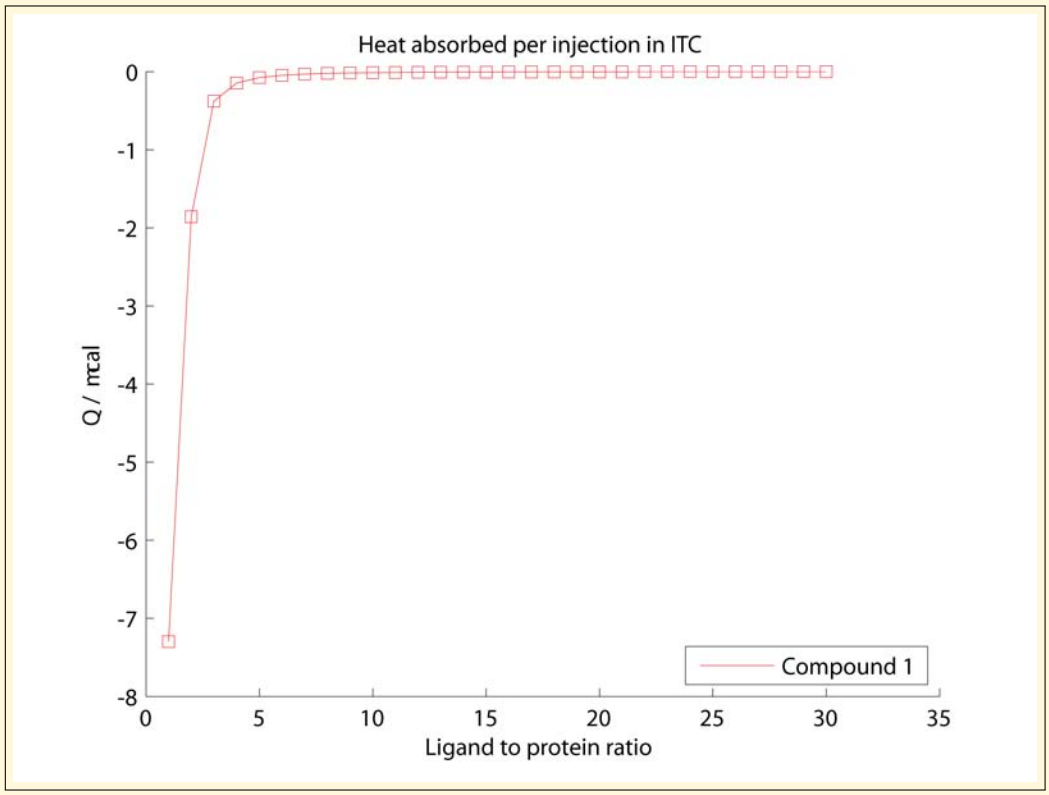


10 points total: 5 for code that runs and performs the right calculations, 5 for correct and fully labelled plot.

f) Assuming that your initial parameter estimates were close and allowing 5 minutes per injection, how long at least should you let your ITC experiment run? Show how you derive your estimate from the simulation output. What would happen if you increased your ligand concentration tenfold? Plot the expected output and comment.

Solution:

- About 30 injections ($[L]/[P]_0 = 3$) until fully equilibrated, so need 150 min total runtime.
- Higher $[L]_0$ quickens approach to equilibrium, but makes urve so steep that only very few data points carry information for fitting, making the result much less accurate.



5 points total: 3 for estimate and explanation, 1 for plot, and 1 for explanation of what happens at higher $[L]_0$.

33 points +2 EC overall for problem 2.

3 Special cases: Radioactive labeling measurement

A scientist is interested in quantifying the interaction between a ligand (L) and a receptor (R). A radioactive label is attached to the ligand and cells are incubated on tissue-culture polystyrene with the therapeutic at 4°C long enough to guarantee equilibrium.

Ligand / nM	**Radioactive signal / AU**
0.001	5.56 E-2
0.01	5.55 E-1
0.1	5.46 E+0
0.5	2.54 E+1
1	4.64 E+1
5	1.44 E+2
10	1.95 E+2
50	2.73 E+2
100	2.89 E+2
300	3.10 E+2
500	3.22 E+2
1000	3.99 E+2

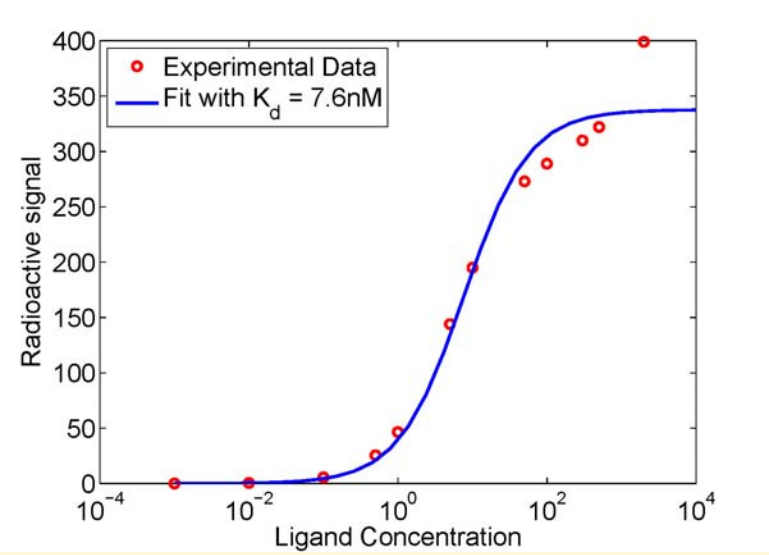
a) The experiment is to be conducted at 4°C. Does this have an impact on the K_D measurement, *i.e.* is it identical to the physiological K_D ?

Solution:

The K_d measured at 4°C differs from that at room temperature or body temperature. The temperature dependence of the binding constant K_D is given by the van't Hoff relationship. *2 points.*

b) The data collected is shown in table 1 above. Using plot this data using the appropriate representation and curve-fit a monovalent binding isotherm to extract K_D. Why does the data look unusual?

Solution:



There does not seem to be saturation of signal. This is typically observed with non-specific binding. The ligand binds non-specifically to the cells, extracellular matrix with zero-th order kinetics. This contribution is small and becomes obvious only at high concentrations. The fit with monovalent binding isotherm yields K_D = 7.6nM. Note: Plotting with linear axes is also acceptable.

The following code was used to answer this question:

```
L = [0.001, 0.01, 0.1, 0.5, 1, 5, 10, 50, 100, 300, 500, 2000];
y = [0.0556 0.555 5.46 25.4 46.4 144 195 273 289 310 322 399];

semilogx(L,y,'ro')
f = inline('C(1)*1./(1+C(2)./L)', 'C', 'L');

C0 = [300, 1];
output = nlinfit(L,y,f,C0);
hold on;
x = logspace(-3, 4, 30);
plot(x, feval(f, output, x))
hold off;
xlabel('Ligand Concentration');
ylabel('Radioactive signal');
legend('Experimental Data', 'Fit with K_d = 7.6nM', 'location', 'NorthWest')
```

11 points total: 2 for plotting the data, 5 for fitting isotherm, 2 for value of K_D, and 2 for observation that data looks unusual because there is no saturation.

c) Perform the necessary modifications to the data set so that a more accurate monovalent binding isotherm can be fitted. Curve-fit the modified data and extract the K_D. Explain

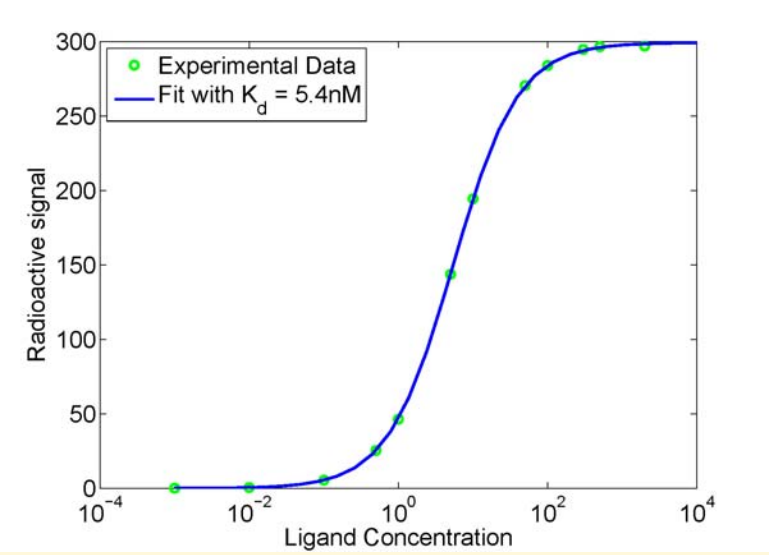
qualitatively why the previously measured K_D was an over- or underestimate.

Solution:

A brief look at the data suggest a single-digit Kd. Therefore at a ligand concentration hundred-fold higher than the Kd, we would expect 99% saturation. From the table, we see that the signal increases from 322 to 399 between 500 and 2000 nM, given the high concentrations, this rise is probably strictly due to non-specific interactions. If we model non-specific interaction as zero-th order kinetics, then the signal S is directly proportional to the ligand concentration (S = α*L). Given the two data points mentioned above, α can be estimated as (399-322)/(2000-500) = 0.051. Thus, we can now remove the non-specific component from the data. To do so, you can use the following code:

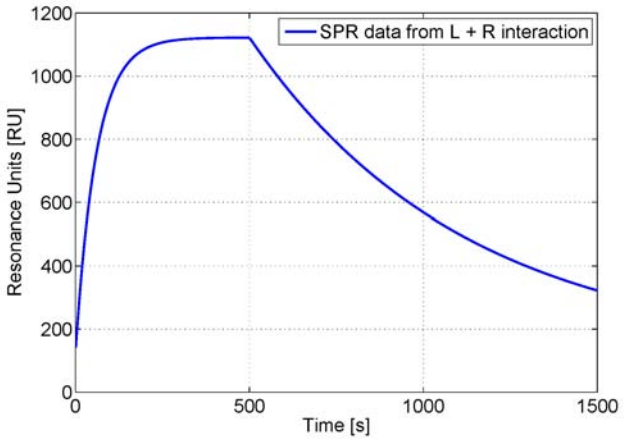
```
y2 = y - 0.051*L;
semilogx(L,y2, 'go')
f = inline('C(1)*1./(1+C(2)./L)', 'C', 'L');
C02 = [300, 1];
output2 = nlinfit(L,y2,f,C02);
hold on;
x = logspace(-3, 4, 30);
plot(x, feval(f, output2, x))
hold off;
xlabel('Ligand Concentration');
ylabel('Radioactive signal');
legend('Experimental Data', 'Fit with \KD = 5.4nM', 'location', 'NorthWest')
```

We now obtain K_D = 5.4nM as shown in figure 2. Non-specific binding results in an over-estimate of the K_D since it cause the fit to estimate the plateau at higher concentrations.



15 points total: 5 for either modification of data or fit to to modified expression for isotherm, 5 for coming up with and plotting and explaining sensible functional form of isotherm with nonspecific binding, 2 for value of K_D, and 3 for explanations.

d) The scientist would like to extract interaction kinetic constants (k_{on} and k_{off}) and performs an SPR experiment with 50nM ligand concentration. The data is shown in figure 1. Extract the values of k_{on} and k_{off}.



Solution:

The dissociation rate constant k_{off} can be extracted from the dissociation phase. In this phase the signal is governed by:

$$RU = RU_0 + (RU_{eq} - RU_0) * e^{-k_{off} * t}$$

Solving for k_{off}:

$$k_{off} = \frac{\ln\left(\frac{RU_{eq} - RU_0}{RU - RU_0}\right)}{\Delta t}$$

From the graph given, we can say that $RU_0 = 100$, $RU_{eq} = 1125$, and Δ t for the dissociation phase is 1000s, with RU = 325 at the end. Therefore $k_{off} = 1.52{*}10^{-3}$ $[s^{-1}]$.
The association phase is governed by the following equation:

$$RU = RU_0 + (RU_{eq} - RU_0) * (1 - e^{(k_{on} * [L] + k_{off}) * -t})$$

Solving for k_{on}:

$$k_{on} = \frac{\frac{\ln\left(\frac{RU_{eq} - RU_0}{RU_{eq} - RU}\right)}{\Delta t} - k_{off}}{[L]}$$

Looking at the graph, we see that RU = 1086, after 200s. Pluging in we obtain $k_{on} = 3{*}10^5 [M^{-1}s^{-1}]$.
6 points total: 2 for k_{off} calculation, 2 for k_{obs}, 2 for k_{on}.

e) Do these result corroborate the estimation obtained by radioactive labeling?

Solution:

By definition, $K_D = \frac{k_{off}}{k_{on}}$, plug in in the values obtained by SPR analysis, we get $K_D = 5.3$ nM. The approaches yielded approximately the same equilibrium dissociation constant. There can be some variance depending on the values you extracted from the graph.

2 points.

36 points overall for problem 3.

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012