20.320 — Problem Set # 2

September 24th, 2010

Due on October 1st, 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.

1 Enzyme Kinetics

An enzyme E catalyzes the reaction of a substrate S into a product P. The product of this reaction is a powerful drug for headaches. An engineer at the production facility is interested in optimizing its production to make it more cost-effective. The enzyme has already been fully characterized previously with $k_{cat} = 50\ [s^{-1}]$, $k_1 = 10^7[M^{-1}s^{-1}]$, $k_{-1} = 50[s^{-1}]$. Answer the following questions:

a) If the reactor contains the enzyme at a concentration of 50nM, what is the maximal turnover rate of the reactor?

b) Using , plot the turnover over rate as a function of substrate concentration.

c) How can the reactor be engineered so that maximal product turnover rate is guaranteed over an extended period of time (i.e. days or weeks)?

d) Using , determine the enzyme concentration range for which the QSSA is valid when i) [S] = 100 $\cdot$K$_M$, ii) [S] = 0.01 $\cdot$K$_M$

e) We want to investigate how long it would take to deplete 90% of substrate if $[S]_0 = K_M$. To do so first obtain an analytical expression by making any assumptions you find necessary. Using that expression, obtain a confidence interval for the 90% depletion time. Then solve the problem using the ODE solver in . Discuss your results.

2 Protein Array

EGFR family receptors regulate cell processes such as proliferation, migration, apoptosis and differentiation, and are implicated in development and cancer. This problem refers to the microarrays in Jones *et al.*, but all the data given here has been artificially generated.

a) In microarrays, the fluorescent signal depends on binding between labeled protein and the tethered protein. Give a control that would allow you to take into account the variation in the number of tethered protein per spot.

b) Is that a useful control? Explain why or why not.

c) Jones *et al.*, performed a titration using 8 different concentrations. The data generated is given in Jones_Nat_2006.mat. The matrix is of size 10x10x8. The first dimension represents the phosphorylated peptides, the second dimension represents the various proteins and the third dimension represents each individual experiment. Use this data to extract the equilibrium dissociation constant for each of these interactions (there are 100, so you need to come up with an automated system).

Index	**Phosphopetides**	**Adaptor proteins**	**Concentrations [nM]**
1	EGFR pY998	ABL2	10
2	EGFR pY1069	GRB2	100
3	EGFR pY1110	PIK3R2-N	200
4	EGFR pY1125	SH3BP2	500
5	ErbB2 pY1023	BMX	1,000
6	ErbB2 pY1221	TENC1	2,000
7	ErbB3 pY1197	TENS1	3,000
8	ErbB3 pY1262	GRAP2	5,000
9	ErbB3 pY1328	PIK3R3-C	
10	ErbB4 pY1188	PLCG1-C	

d) Were you able to estimate a relevant K_D for each interaction? If not why?

e) A typical cell has on the order of 10^4 EGFR receptors. Let us imagine that upon EGF stimulation, there is a phosphopeptide local concentration of 500nM. Draw the interaction network generated. You should have on the x-axis the phosphopeptides, on the y-axis the interaction protein and you should connect a protein to a phosphosite if the probability that they are bound is higher than 50%. *Do not consider competition.*

f) Some cancer cells upregulate the expression of EGFR and have on the order of 10^5 EGFR receptors on their surface. Let us consider 3 different scenarios where the concentration of phosphopeptides is i) 1μM, ii) 1.5μM and iii) 2μM. How does the system evolve?

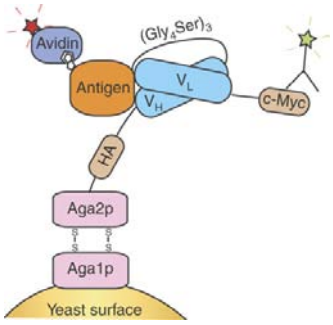
g) Let us consider one specific phosphosite on EGFR and two adaptor proteins A and B. The concentration of the adaptor proteins are 1nM and 500nM respectively. If the equilibrium dissociation constant for each of these adaptor proteins is 0.5nM and 250nM, plot the percentage occupation of the phosphosite by each of these adaptor proteins as a function of the phosphopeptide concentration.

h) What do the results obtained in e) and f) tell you about the importance of EGFR overexpression in cancer cells?

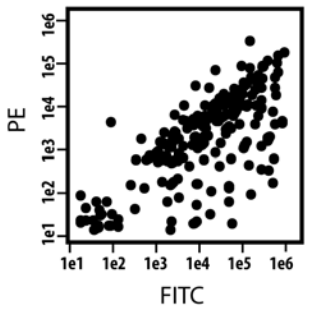
3 Yeast Surface Display (YSD)

In this problem, you will apply the principles for the analysis of binding interactions covered in class to a new experimental context.

Yeast Surface Display is an alternative method for measuring kinetic and thermodynamic parameters for binding interactions. By expressing the protein of interest as a fusion with Aga-2p cell surface protein, you can have it presented on the outside surface of the yeast cell. Furthermore, the protein is flanked by two short peptide tags (HA and cmyc) against which good antibodies are commercially available. In the diagram below, the protein of interest is a singel chain variable fragment (scFv) of an antibody, comprising the light variable domain V_L, the heavy variable domain V_H, and a flexible $(Gly_4Ser)_3$ linker. The ligand of interest is a biotinylated antigen. They will be referred to as P and L in this problem.



You have generated a library of mutated scFvs by using error-prone PCR on the gene for a ligand-binding antibody. You incubate a population of yeast cells expressing the library of P mutants on their surface first with FITC-labelled anti-myc antibody (green; binds to cmyc peptide) and then with phycoerythrein (PE)-labelled streptavidin (red) and biotinylated ligand. You run the library through a flow cytometer and plot red vs. green fluorescence for each cell. In the resulting plot, each dot corresponds to a single cell and its coordinates mark the red and green fluorescence intensities.

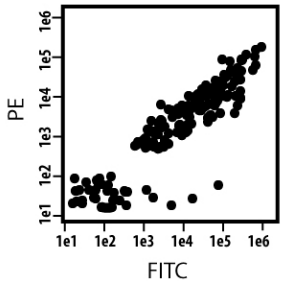


a) You want to sort the best 1% of binders for the next round of directed evolution. On the plot above, circle the part of the cell population which shows significantly above-average protein-ligand binding affinity (a very rough sketch is sufficient).

b) To quantitatively track the improvement of your library through successive rounds of evolution, you run another sample on the flow cytometer, this time containing only the top binders which you selected for the next round. You now want to determine the average K_D for this population. For this purpose, you will incubate a number of separate samples containing on

average a few 100 cells displaying each mutant antibody with a number of different ligand concentrations, label with FITC-anti-cmyc antibody, and perform the measurement.

i) On the FACS plot below, indicate where you can read off a measure of the *average fractional saturation* of protein with ligand for all functional antibody mutants at any one given ligand concentration. Explain in one short sentence how you found that point.



ii) Schematically (no numbers needed on the axes), draw a labelled graph to illustrate how you would plot the quantity you measure for each ligand concentration against [L]. Indicate how to graphically estimate K_D from this plot.

c) The following is data from an equilibrium binding experiment by YSD in triplicate [1]

[L] / nM	Experiment 1	Experiment 2	Experiment 3
3160	47.0	40.9	56.6
1000	42.9	39.9	58.8
316	41.4	41.6	55.0
100	34.3	39.1	53.1
31.6	35.9	33.8	42.4
10	22.1	24.7	33.5
3.16	9.2	16.0	20.0
1	7.2	7.5	10.6
0.316	4.0	6.3	4.6
0.1	2.2	2.5	3.4
0	2.0	1.9	0.0

i) Plot the raw data. Comment.

ii) Extract K_D vakues by regression for each experiment. Using the `subplot` command in , plot the experimental data (normalized) and fitted isotherm in one panel, and the residuals in a second panel of the same figure. Are you satisfied with your results? Explain.

d) In this type of experiment, incubation with ligand is performed in tubes. Why will ligand depletion become a constraint when testing very high-affinity interactions? What can you do to mitigate it?

[1] The interaction being measured was between an engineered antibody and a small molecule chelator of radioisotopes for pre-targeted radioimmunotherapy (PRIT). This antibody has since been tested in a mouse model for treatment of micrometastases; the results were accepted for publication three weeks ago.

20.320 Analysis of Biomolecular and Cellular Systems
Fall 2012