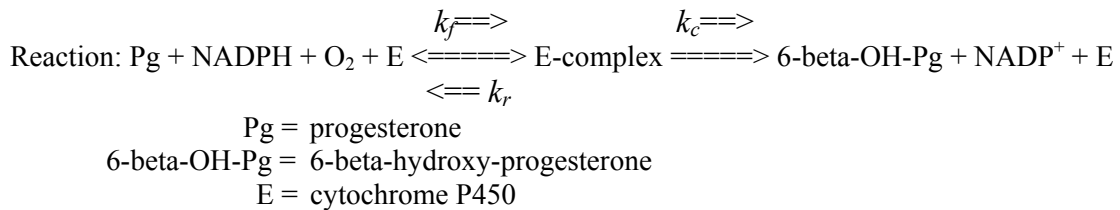# Biol 591 Introduction to Bioinformatics (Fall 2003)
Problem Set 4 – Modeling

PS4.1. Write a subroutine called Divide that takes as input a number and an array and returns an array in which each element is divided by the number.

PS4.2. Write a subroutine called Average that takes as input any number of arrays and returns an array in which each element contains the average of all the elements of the input arrays (at that position). *[Hint: If you think of each input array as constituting a row of a table, then the output array will be the bottom row of that table, each element being the average value of all the values above it]*

PS4.3. The liver enzyme cytochrome P450 is involved in the metabolism a large number of compounds, particularly those of foreign origin, like acetaminophen, caffeine, codeine, and AZT. Since and the effect of a drug is closely related to its lifetime in the body, understanding of the action of cytochrome P450 on a particular drug is often critical for predicting the drug's efficacy. You want to model the reaction below, catalyzed by cytochrome P450. Determine whether each of the proposed rate equations that follow is correct. If it isn't, make it correct.

$$\text{Reaction: Pg + NADPH + O}_2\text{ + E} \underset{\underset{k_r}{\Longleftarrow}}{\overset{\overset{k_f \Longrightarrow}{\Longleftrightarrow}}{}} \text{E-complex} \overset{k_c \Longrightarrow}{=\!=\!=\!=\!=\!>} \text{6-beta-OH-Pg + NADP}^+\text{ + E}$$

Pg = progesterone
6-beta-OH-Pg = 6-beta-hydroxy-progesterone
E = cytochrome P450

3a. d[6-beta-OH-Pg]/dt = [E-complex] k_c - [6-beta-OH-Pg] [NADP] [E] k_c

3b. d[E]/dt = [Pg] [NADPH] [O$_2$] [E] k_f

3c. d[E-complex]/dt = - [Pg] [NADPH] [O$_2$] [E] + [E-complex] k_c

PS4.4. Alter the program *Glycolysis.pl* in the following ways:

4a. Get rid of the irritating lines of values printed on the monitor.

4b. Have the program print instead the current time step each iteration through the loop, so you can tell where it is.

4c. Have the program print instead every 100th time step (to cut down on printing), so you can tell about where it is.

Hint: Use the Perl function int( ) (which gives you the greatest integer less than a given number) in conjunction with division. Only every 100th time step $time_step/100 will be equal to int($time_step/100). If you aren't sure how int( ) works, try writing a one-line program to test it's function.

PS4.5. *Glycolysis.pl* has in the subroutine Model the following line:

$$my \ \$dGbisP\_dt = +\$v\_Rxn7 \ - \$v\_Rxn8;$$

5a. What is the physical significance of this equation?

5b. Give the equation that produced the value of $v_Rxn7.

PS4.6. Alter the program *Glycolysis.pl* so as to test what effect inhibiting hexokinase with a mythical compound that exhibits uncompetitive inhibition with respect to glucose will have on the level of metabolites. Suggested strategy:

- Note that Eisenthal and Cornish-Bowden (1998) have already done the enzymological derivation for you.

- Write a subroutine (perhaps called v_of_rxn2 that calculates the velocity of the reaction.

- Note that Eisenthal and Cornish-Bowdin defines *i* as the concentration of the inhibitor divided by the constant of inhibition. This constant is equal to the concentration of inhibitor required to lead to half-maximal binding to the enzyme. Thus the ratio *i* is related (nonlinearly) to the occupancy of the enzyme by the inhibitor and hence the degree of inhibition. You don't have any value for the constant, but you can vary *i* over reasonable values. I suggest you run the program multiple times, each time with different values of *i*.

PS4.7. The concentrations of the metabolites in *Glycolysis.pl* are given using units of mM. Suppose that's all right with you, but in the end you want to **print** the results in units of M. Add <u>one line</u> to the program and modify another to accomplish this task.

PS4.8. What happens when you decrease the time interval used in stepping through *Glycolysis.pl*? What happens when you **increase** the time interval? Be sure you try enough values to get a picture of the general trend.

PS4.9. Write an equation to describe the number of bacteria in a culture where the bacteria double every 20 minutes (the maximal growth rate of *E. coli*). Presume that there is no limitation by food or other resources. Now write a **rate** equation for the process.

PS4.10. Suppose that you are interested in studying the dynamics of how sexually transmitted diseases spread. From a study of the general population, it has been found:

- The disease is spread solely by heterosexual contact.

- The frequency of new infection is proportional to both the number of carriers in the population and the number of susceptible partners.

- Males and females are infected at different rates.

- Carrier males are treated and cured at a higher rate than carrier females (becuase the symptoms are more severe in males, hence they are more likely to seek treatment).

- They never learn. Cured males and females return to the susceptible population just as likely as before to contract the disease.

10a. Write rate equations describing the rates of change of the two carrier populations (Cm - carrier males, and Cf - carrier females).

You would like to test your equations on a test population. Accordingly you advertise for volunteers and select 600 healthy females and 899 healthy males for the experiment. You place the volunteers on a south sea island along with one infected male and wish them luck. You will return in 12 years to tabulate the results, but you'd like to know right now what to expect.

10b. Devise a model for the experiment, one that will predict the number of infected males and females over the 12-year duration of the experiment. Use the following constants:

$k_1 = 0.000032$ infected males per day per appropriate male per appropriate female

$k_2 = 0.2$ successfully treated males per day per appropriate male

$k_3 = 0.00033$ infected females per day per appropriate female per appropriate male

$k_4 = 0.025$ successfully treated females per day per appropriate female

10c. Plot and interpret the results.

PS4.11. (As a public service, I have to include this one) You are faced with the all too common problem of dividing a doughnut evenly amongst three people. The easy way is to slice it by three radial cuts (see picture to right).

But this would be WRONG. Doughnuts have holes, yes? This way, no holes. NOBODY gets a doughnut.
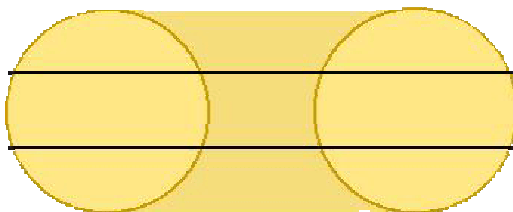
The correct approach is to slice the doughnut twice, parallel to its plane (see left). The problem is, exactly where do you position those two slices? A classical problem of calculus, but who knows calculus when you've got a knife in your hands and eyes staring at you? No, modeling is much simpler.

Here's a strategy. Consider just one of the two cross sectional circles of the doughnut. If you can slice that circle into three equal pieces by slices along the plane, that should do it. Where should the cuts go?

The solution is diagrammed to the right. I'll model the cutting of the circle from the top, little by little, adding up the area of those slices, until the sum of the areas equals one-third of the area of the full circle. The point when that happens is the point where I should slice the doughnut to give one of the three equal pieces. The other cut should be at the equivalent position on the other side of the circle.
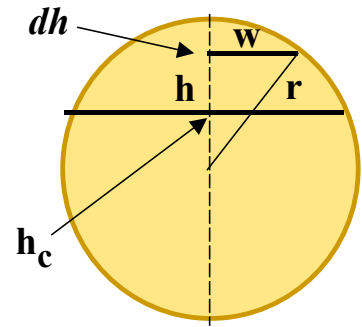
Some terms:

- **r**   The radius of the circle
- **h**   The distance from the center to the site of the slice
- **w**   Half the width of the slice (the full width is the extended line to the other side of the circle). From the Pythagorean Theorum, $w = \text{sqrt}(r^2 - h^2)$.
- **dh**   The thickness of the slices
- **h$_c$**   The position of the last slice, the one that finally gives one-third of the circle.

So, I'll start at the top (where **w** will be very small), adding up each slice (twice the width times the thickness of the slice), and I'll let **h** go from **r** down towards zero. When the sum is equal to one-third the area of the circle, I'll stop and note the value of **h**. That value must be what I'm looking for, **h$_c$**, the place I should cut the doughnut.

Write/steal a program that models the process described above and finds the value for **h$_c$**.

(This is an example of numerical integration, common enough in computational biology, where the functions are often too complicated to integrate analytically)