APh161: The Physics of Biological Structure and Function
Homework 2
Due Date: Thursday, January 27, 2005

On Exactitude in Science . . . In that Empire, the Art of Cartography attained such Perfection that the map of a single Province occupied the entirety of a City, and the map of the Empire, the entirety of a Province. In time, those Unconscionable Maps no longer satisfied, and the Cartographers Guilds struck a Map of the Empire whose size was that of the Empire, and which coincided point for point with it. The following Generations, who were not so fond of the Study of Cartography as their Forebears had been, saw that that vast Map was Useless, and not without some Pitilessness was it, that they delivered it up to the Inclemencies of Sun and Winters. In the Deserts of the West, still today, there are Tattered Ruins of that Map, inhabited by Animals and Beggars; in all the Land there is no other Relic of the Disciplines of Geography. - Jorge Luis Borges

Reading: read chaps. 7 and 8 of *Essential Cell Biology* and chap. 1 of *Genes and Signals*.

1. A Feeling for the Numbers: The Rates of Things

In the previous homework, we worked hard to get a sense for the physical sizes of various biological entities. Another interesting angle on all of this is to try and get a feel for the *rates* at which things happen. Following in the tradition of the previous problems, here you will try to make some estimates of the rates of some processes. Much of what you will do in this problem I have already done partially in class - your job is to make it your own now.

(a) Consider the division of an *E. coli* cell. Think of such a cell during rapid growth phase where the cell is dividing roughly once every 20 minutes. Make estimates of the number of water molecules being taken on board per second during this phase, the number of lipid molecules that are being added onto the surface membranes, the number of proteins being synthesized per second and how many ribosomes are needed to do so.

(b) In this case, think about the motility of the bacterium *Listeria monocytogenes* and a typical eucaryotic cell. In the case of *Listeria*, the motion of the bacterium is mediated by the formation of actin comet tails which depend in turn upon the linear polymerization of actin filaments. The formation of the actin comet results in a speed for the bacterium of something around 0.1 μm/sec. In the eucaryotic setting, the cell extends arms called filopodia which permit it to crawl, again by virtue of actin polymerization. For *Listeria*, use the measured rate of motion of the bacterium to *estimate* the rate of actin polymerization both in microns/sec and monomers/sec. Make sure you draw a picture of the process and explain your rationale. Now, take that estimate for the rate of actin polymerization and estimate the rate at which a filopodium extends on a eucaryotic cell. Anything you can do to compare these estimates with measurements would be useful - one excellent source is **Cell Movements** by Dennis Bray.

(c) Given that the time scale for the cell cycle of a bacterium is around a half hour, in this part of the problem, we estimate the rate of DNA replication. In particular, compute the number of nucleotides per second that DNA polymerase would have to add as a function of the number of DNA polymerase molecules in the cell. That is, if there were only one DNA polymerase molecule, then this one molecule would have to add $\approx 5 \times 10^6$ nucleotides in 30 minutes. On the other hand, if there are 10 DNA polymerase molecules they can share the burden. Your task is to estimate the mean rate of nucleotide incorporation as a function of the number of DNA polymerase molecules in the cell so as to make sure that the whole genome is replicated. Now let's turn the argument on its side - if I tell you that the rate of nucleotide incorporation is roughly 100 nucleotides per second, then estimate how many DNA polymerase molecules are present in the cell (note that this is all orders of magnitude since we are ignoring subtleties like what fraction of the polymerase molecules are bound, etc.). Also, look at fig. 6-9 of *Essential Cell Biology* and assuming that this is a representative sample of the replication process, estimate the number of DNA polymerase molecules in an *E. coli* cell. Taking our examination of that figure further, note that the fly DNA is about 1.8×10^9 nucleotide pairs in size. Estimate the fraction of the total fly DNA shown in the micrograph. There are eight forks in the micrograph, numbered 1-8. Estimate the lengths of the DNA strands between replication forks 4 and 5 where we count the forks from left to right.

If a replication fork moves at a speed of 100 nucleotides/s, how long will it take for forks 4 and 5 to collide.

2. Cut it Up

In this problem I want you to think about the $\approx 48,000$ bp genome of lambda phage and to work out the lengths of the fragments that you would get if the DNA is cut with both the HindIII and EcoRI restriction enzymes. What are the recognition sequences that these enzymes each cut? Make an estimate for the lengths of the fragments that one would get when only using one of these enzymes - there is a precise mathematical way to do this and it depends upon the length of the recognition sequence - a 5 cutter will have shorter fragments than an 8 cutter - explain that. To do the precise analysis, go to the New England Biolabs website (www.neb.com) and look up the tables that they have for identifying the sites on the lambda genome that get cut by these different enzymes. If you are feeling ambitious, Hernan last year did something very cool which you could imitate. In particular, he went to the New England Biolabs website and basically examined the distribution of cuts for all of their enzymes and examined the extent to which the simple probabilistic model I am advocating here is correct. Take a look.

3. The Frances Arnold Estimate Problem

In a 2001 Bioengineering seminar, Professor Frances Arnold made a startling remark that it is the aim of the present problem to examine. The basic point is to try and generate some intuition for the **HUGE**, **ASTRONOMICAL** number of ways of choosing amino acid sequences. To drive home the point, she noted that if we consider a protein with 300 amino acids, there will be a huge number of different possible sequences.

(a) How many different sequences are there for a 300 amino acid protein?

But that wasn't the provocative remark. The provocative remark was that if we took only one molecule of each of these different possible proteins, it would take a volume equal to five of our universes to contain all of these different *distinct* molecules.

(b) Estimate the size of a protein with 300 amino acids. Justify your result, but remember it is an estimate. Next, find an estimate of the size of the universe and figure out whether Frances was guilty of hyperbole or if her statement was on the money.