

Laboratory Worksheet, Tuesday, April 15.

End of Term Procedures:

Today is our last laboratory, and we have to close things out. What is new for today is to incorporate some suggestions of a colleague into the tRNA-synthetase project, see part (c) below. Overall, I would like to indicate what your final assignment is to turn in. I would like each of you to turn in one of the three continuing projects which we have been doing for most of the term. The due date will be **Friday, April 25, 2003**. What follows is a review of what each of those projects means, and what would be expected for each of them. In general, the idea would be to gather – in a coherent and readable form, please! – the data and analyses you have collected and performed over the term, together with a brief cover page or two to explain what you make of it, and why the analyses do or do not support or prove what you are concluding.

(a) Actin/HSB Comparison: Recall that the basic premise was to study the distant homology between the actin proteins and the heat shock proteins. We finished the first part of the project, which was to try to see whether we could draw meaningful homology information from sequence data alone. The second part of the project, needed to complete it, is to go back to the Eisenberg paper distributed in the lab (I still have some hard copies available; the PubMed abstract is linked from the 548 resources page) and follow the lines indicated there to see the homology taking advantage of structural similarities. One can do this either by following (updated) versions of what the authors explain in the paper, or by using SAM alignment tools which take advantage of comparisons with structural databases (see the Lab Worksheet for April 1). In fact, a good way to finish the project would be to try both a “purely” structural version from Eisenberg, and then the SAM (or similar) version, and compare to see whether sequence analysis has improved since the time of the Eisenberg paper, especially hybrid methods which incorporate structure information.

(b) Codon Usage Bias and PHX Genes: This is the one about the Karlin-Mrázek paper linked in the 548 resources page. I draw your attention to the directory <http://www.math.lsa.umich.edu/dburns/548/codusagebiasPHX/> where you will now find (almost – see below) all the data necessary to run these experiments for the three species we chose in February: *E. coli*, *M. tuberculosis* and *P. horikoshii*. Thus, these are the nucleotide sequence data for the reference families of highly expressed genes in those three species: the chaperones (HSP's), ribosomal proteins and protein synthesis proteins (as defined by K-M for these purposes). Here what I would be expecting would be working code which calculates the bias comparison functions which are described on p. 5239 of the KM paper. You may borrow pieces of code from one another (you must reference such borrowing, however), from the BioPerl module or the modules BeginPerlBioinfo.pm or Bfn.pl, all in the 548 resources page. You would demonstrate that your code worked by applying it to several query genes from each of the three reference species.

Please NOTE: We are having trouble, as of this writing (April 15, 2003), straightening out the chaperone data for *Pyrococcus horikoshii*, that is, the heat shock protein data for

the extreme thermophile *P. horikoshii*. I hope to have this settled in the next couple of days. If this remains unresolved, you do not have to complete this part of the exercise.

(c) tRNA Synthetase and TrpB Synthase Project: As noted several times in the laboratory, the point of this project is to examine the hypothesis that there is a temporal hierarchy of amino acid residues according to the epoch at which they were incorporated into the repertory of protein catalyzed life. We have been trying to examine this through a very complex family of proteins, the aminoacyl-tRNA synthetases, the proteins responsible for the specificity of the genetic code. I have consulted with Professor Stephen Freeland of UMBC, and he suggests looking for this temporal hierarchy among the simpler amino acid synthases, the enzymes responsible for amino acid synthesis in the cell. The 548 resource page now contains a new subdirectory on this project:

<http://www.math.lsa.umich.edu/dburns/548/rnasyn/>. There are a collection of materials there now concerning the temporal hierarchy and the evolution of the genetic code, especially by E. Trifonov. Roughly speaking, the synthase version of this project would be to take the amino acid synthase of a protein “recent” on the consensus dating scheme (cf., p.9 of Trifonov’s lecture slides [.../548/rnasyn/trifITP.pdf](http://www.math.lsa.umich.edu/dburns/548/rnasyn/trifITP.pdf) in the subdirectory), and find that the consensus a.a.-synthase for that residue only involved “older” residues. (The Miller-Urey “abiotic” residues are highlighted by a red dot in the reference above.) There is data there which you helped gather on the TrpB-synthase across species (TrpB = the beta subunit of tryptophan synthase, the component directly responsible for the synthesis of tryptophan from serine and indole). It is found under [.../rnasyn/TrpBsynthasedata](http://www.math.lsa.umich.edu/dburns/548/rnasyn/TrpBsynthasedata) in the 548 resources directory.

Specifically, however, for the end of term project, you could carry out the technical analysis already begun on the aminoacyl-tRNA synthetases, it is just that one suspects the conclusions will be less well-defined, for biological reasons. As an exercise to demonstrate your proficiency it will be sufficient, however. I do strongly urge you to try the TrpB-synthase version as well!

Have a Good Summer!