Homework # 8

Handed out: 11/9/06
Due: 11/14/06

For this homework you'll have to apply some of the alignment tools to look at biological data. You will need to familiarize yourself with using the resources at the NCBI (http://www.ncbi.nlm.nih.gov), specifically to search for a gene by name and also by using blast.

1. Look for the cytochrome C protein in Genbank using the text search on the front page for NCBI. You'll have to look for a protein record for this protein (i.e. not the gene encoding it). Extract one (the first) instance of the human (*Homo sapiens*) version of the protein.
2. Select the amino-acid sequence of this protein (on the protein page you'll have to select the FASTA display) and paste it into the Blast page for protein-to-protein alignments (blastp). You can reach the Blast server from the top NCBI website.
3. After blasting the human cytochrome C protein against all other proteins in Genbank, identify the top hits from the following organisms:
   a. rhesus macaque (*Macaca mulatta*)
   b. chimpanzee (*Pan troglodytes*)
   c. mouse (*Mus musculus*)
   d. cattle (*Bos taurus*)
   e. dog (*Canis familiaris*)
   f. chicken (*Gallus gallus*)
   g. frog (*Xenopus laevis*)
   Hint: it's easiest to find these organisms by searching for the latin name in the text of the page returned by blast.
4. Save the corresponding entries in a multi-fasta file on your machine (use some cut-and-paste wizardry). Don't forget the human protein as well.
5. Build a multiple alignment of the sequences using
   a. ClustalW (http://www.ebi.ac.uk/clustalw)
   b. Muscle (http://www.bioinformatics.nl/cgi-bin/muscle.pl)
6. Can you identify the regions conserved across these organisms? Are the proteins from closely related species (e.g. human, macaque, chimpanzee) more similar than proteins from distantly related organisms (e.g. human vs. frog) ?
7. Can you identify any differences between the ClustalW and Muscle alignments? Can you guess at the reason for these differences?