Metabolic modelling, Spring 07, Exercise 1, Friday, 23.3.2007

1. Summarize the operational principles of the BLAST sequence alignment algorithm. How the BLAST can be used to detect homologous protein sequences? How are the sequence alignments scored in the BLAST? How the BLAST can help in metabolic reconstruction? (Hint: Almost any textbook in bioinformatics, *http://www.ncbi.nml.nih.gov/Education*)

   Your group has just finished sequencing few new proteins of currently unknown organism, that seems to be a close relative to yeast *Saccharomyces Cerevisiae*. There is a clear evidence that these new proteins catalyze metabolic reactions. You have already discovered that enzymes with EC number 4.2.1.2, 1.3.5.1, 1.2.7.3 and 6.2.1.4 exist in your organism. Also, gene expression studies indicate the recently sequenced proteins might catalyze reactions that reside in the same metabolic pathway than reactions catalyzed by the known enzymes. Your job is now to assign metabolic functions to new protein sequences 1 – 5 found from the www-page of the course.

2. By using KEGG LIGAND database (*http://www.genome.jp/kegg/ligand.html*), find a metabolic pathway (connected bipartite graph of reactions and metabolites) corresponding known enzymes 4.2.1.2, 1.3.5.1, 1.2.7.3 and 6.2.1.4. Draw a schematic picture of the pathway.

3. By using BLAST service found from the home page of KEGG database (*http://blast.genome.jp/*), find homologous metabolic enzymes to new proteins 1 –4. In the BLAST search, limit the query database to the protein sequences of *S. Cerevisiae* (type "SCE" to the field "Favorite organisms"). Try to select such set of homologous enzymes that they form a biologically meaningful pathway together with already known enzymes – do not blindly select the best hit. Record the BLAST scores (bits, E-val) of the enzymes you selected as well as ranks of the selected enzymes in the search results. Judging from the scores and ranks, does our hypothesis, that all the enzymes reside in the same pathway, seem justified? Compare the scores to ones obtained with random protein sequences of the same length (*http:au.expasy.org/tools/randseq.html*). Draw a schematic picture of the reconstructed pathway. (Hint: you don't have to change the default parameter values of KEGG BLAST. Also, it is enough to check few top hits in each search.)

4. The reconstruction of the previous exercise should contain a hole in the network. While you were doing exercises, your lab partly sequenced another protein (sequence 5 on the www page of the course) of the unknown organism. Run BLAST to the new sequence to test whether it can catalyze a reaction filling the hole. Compare the BLAST scores to the scores of previous exercise and to scores obtained with random protein sequences of the same length (*http:au.expasy.org/tools/randseq.html*) How would you interpret the results?

   ExPASy Proteomics Server offers a service where you can try to find functional motifs stored into PROSITE database from protein sequences (*http://au.expasy.org/prosite/*, quick scan). Try the service with sequence 5. How would you interpret the results?