

# Chemoinformatics

## 1 Chemical Similarity

**Question 1:** What are we measuring with the Tanimoto coefficients in this project?

**Question 2:** Compare the histograms that you generated for drugs that share a target versus drugs that do not share a target. Are they different? If so, what does this suggest?

**Question 3:** We used drug fingerprints to generate the Tanimoto scores which is a simplistic view of a molecule. How could you more accurately compare the structure of two molecules?

## 2 Ligand Set Similarity

In this section you will be looking at the similarity of proteins using ligand sets. We will refer to the following proteins:

SwissProt ID	Name	Description
P00374	DHFR	DHFR dihydrofolate reductase
P00734	F2	Prothrombin
P04818	TYMS	Thymidylate synthase

**Question 4:** Go to the BLAST website,

<http://blast.ncbi.nlm.nih.gov/Blast.cgi>. Click on Protein blast and then click the checkbox for Align two or more sequences. If you enter one SwissProt ID in the top box and one in the bottom box you will get Blast scores for aligning the query to the subject.

- Compare DHFR to F2, what is the total score and E value?
- Compare DHFR to TYMS, what is the total score and E value?
- Compare F2 and TYMS, what is the total score and E value?

**Question 5:** Now go back to Protein blast and un-check Align two or more sequences and instead look at “Choose Search Set” and make sure “Database” is set to Non-redundant protein sequences (nr). Enter the SwissProt ID for DHFR and run the query. Go to Formatting and limit the organism to human. Are there any significant alignments that are not forms of DHFR? Using these results from BLAST could you predict that there are other human proteins that bind some of the same ligands as DHFR?

**Question 6:** Based on your findings in Question 4 and Question 5, are the proteins in Question 4 similar by BLAST? In one sentence, explain how you arrived at this answer.

**Question 7:** Now compare the three proteins above using your program pvalue with the option -n 1000.

- a) Compare DHFR to F2, what is the p-value?
- b) Compare DHFR to TYMS, what is the p-value?
- c) Compare F2 and TYMS, what is the p-value?

**Question 8:** Which of the proteins are similar to each other according to your pvalue program in Question 7? What does this mean with regards to the drugs that bind to the proteins?

### 3 Network Visualization

In this section you will be look at the network you generated with Cytoscape.

**Question 9:** What does it mean if a protein node is highly connected in this network? How would you interpret that biochemically?

**Question 10:** Do the proteins tend to form interconnections based on their annotated indications? Why might a node appear connected with nodes with different colors (indications)?