

## HW 4: due Tuesday, December 7, in class

For this assignment, all answers are to be submitted in hardcopy in class.

1. The FASTA files for four sequences appear in file hw5.protseq available through the class webpage. Run each of these protein sequences through several (at least three) of the protein secondary structure prediction programs available on the web. (Make sure what you hand in clearly indicates which programs you used).
  - (a) Output a “consensus” secondary structure prediction for the three-states H/E/other using (1) unweighted voting (assign each position the residue that the majority of predictions assign it; remember to explicitly state what you do about tie votes) (2) another method that you think might do better than unweighted voting (justify what you chose to do and why).
  - (b) For each of these four sequences, which top-level family of the SCOP hierarchy do you think it belongs to (eg. all alpha, all beta, alpha+beta, etc.)?
2. This exercise requires a web browser with a CHIME plugin, which is available for windows and mac OS only (not unix or linux). Follow the straightforward instructions on the web at the address given to download the required software. (Email us if you don't have access to the appropriate machines or can't make this work. If you only have access to unix based machines, you can do most of this with a program called rasmol instead).
  - (a) Go to <http://molvis.sdsc.edu/protexpl/frntdoor.htm> and click on quick start, and do their 1 hour tour of Protein Explorer.
  - (b) Enter 1tie in their window for getting pdb files. View it in space-filling, ball-and-stick, and cartoon modes.

- (c) Using the graphics, and the ability of the program to select pieces of the protein by residue number, annotate the secondary structure “by eye”. Submit your annotations underneath a copy of the sequence in your hardcopy.
  - (d) Compare your annotations with the annotations from 2 of your favorite web-based secondary structure predictors. How did the predictors do?
3. The family described below has been genotyped for three linked markers, A, B, and C. The pair of alleles found (each represented by a digit) is listed below for each marker and each person in the family. All the alleles come from genes located on a single chromosome.

**Father:** I-1  
**Mother:** I-2  
**Offspring:** II-1 through II-10

Person	marker		
	A	B	C
I-1	1,5	2,3	6,8
I-2	1,9	4,7	3,6
II-1	1,5	2,7	6,6
II-2	1,9	3,4	3,8
II-3	1,5	2,7	6,8
II-4	5,9	3,4	3,6
II-5	1,9	3,7	3,8
II-6	5,9	2,4	6,6
II-7	1,1	3,4	6,8
II-8	1,5	2,7	6,6
II-9	5,9	2,4	3,6
II-10	1,9	3,4	6,8

- (a) For each of the offspring, list which alleles were inherited from the mother and which from the father.
- (b) Count the frequency of each paternally- and maternally-derived haplotype. What are the parents’ haplotypes most likely to be?
- (c) For each pair of markers, count the total number (both maternal and paternal ) of crossovers that occurred between just those

markers. (I.e., for markers A and C, ignoring the data for marker B entirely, how many crossovers do you see?) Use this information to build a map of the markers.

- (d) Verify your map by determining the haplotypes corresponding to double crossovers. How often do they occur?