

COMP 555 Bioalgorithms

Fall 2012

Problem Set 3

Assigned: Tue Oct 30

Due: Thu Nov 8 (in class)

1. Problem 8.6 (pg 302) in the text.
2. Problem 8.7 (pg 303) in the text.
3. Problem 8.9 (pg 303) in the text.
4. Use the table of single-letter amino acid codes and corresponding residue masses on slide 4 of the protein sequencing lecture (Lecture 15) for the following question. Use the nearest integer to the residue mass in calculations.
 - a. Determine the theoretical MS/MS spectrum S_1 of "PEPTID". You may assume no mass is lost in MS/MS fragmentation of the molecule.

Given the following two spectra:

$$S_2 = \{ 115, 133, 264, 280, 383, 395, 498, 514, 645, 663, 778 \}$$

$$S_3 = \{ 115, 133, 280, 337, 395, 456, 514, 571, 718, 736, 851 \}$$

- b. Determine the shared peak count between your spectrum in (a) and the two given experimental spectra.
 - c. Use spectral convolution to estimate the number of sequence differences between your theoretical spectrum and the two given spectra.
 - d. Determine the residue substitutions and their positions that might give rise to S_2 and S_3 .
5. **Programming Assignment.** Construct a python program that, given a directed graph, yields a Eulerian path or `None`, if no such path exists. The directed graph is given as an unordered sequence of graph edges. Each edge is a triple of the form (v, w, l) denoting a directed edge from vertex v to vertex w with edge label l . A Eulerian path is the triple (v, s, w) where v is the start vertex of the path, s is the sequence of labels of edges traversed in the path and w is the final vertex of the path.

Example input: `[('a','b',1), ('b','c',2), ('c','a',3)]`

Output (one of several possible): `('a', [1,2,3], 'a')`