

Study Guide for MICA 8006 Week 2

This study guide covers week two lectures and assigned readings from Tramontano, Problem 1, pg 1-12 & Fig. 17. It does NOT include UNIX, Perl or GCG syntax, or other topics discussed only in lab sessions. Some mid-term exam questions may be similar to these:

1. Define: indel, similarity matrix, PAM, BLOSUM, dotplot (dot matrix), global alignment.
2. What is a "gap penalty"? List two types of gap penalty. Which is larger?
3. As two sequences become increasingly identical, how does the difference between a similarity score calculated using PAM256 and one calculated using BLOSUM62 change?
4. What are the steps in a Needleman-Wunsch algorithm?
5. Perform dynamic programming as used in global pairwise sequence alignment.
6. Assume two sequences are $q = \text{CDAA}$ and $d = \text{AEECA}$, and the scoring matrix is:

	A	C	D	E
A	2	-2	-2	-1
C		1	0	0
D			2	-2
E				2

(a) Find the highest score by aligning q and d when the gap penalty is $g_l = 2l$. Then find the best alignment.

(b) Now use the gap penalty $g_l = 1.8 + 0.4l$. The dynamic programming table is partly filled in below, using the general DP procedure:

q/d			A	E	E	C	A
		0	1	2	3	4	5
	0	0.0	-2.2	-2.6	-3.0	-3.4	-3.8
C	1	-2.2	-2.0	-2.2	-2.6	-2.0	-4.2
D	2	-2.6	-4.2	-4.0	-4.2	-2.6	-4.0
A	3	-3.0	-0.6	-2.8	-3.2		
A	4	-3.4	-1.0	-1.6			

Note especially how the value -3.2 in cell 3,3 is found.

The value is cell 3,1 - $g_2 = -0.6 - (1.8 + 0.8)$.

Fill in the rest of the table and find the best alignment(s).

- (c) Compare the alignments found under (a) and (b), and find for each of them the minimum number of mutations which might have occurred, assuming that only one residue is changed in each mutation.
7. Sometimes we want the gap penalty at the ends of an alignment to be scored as 0. In what cases is this reasonable (what is the relation between the two sequences)?
8. (a) How can a dot matrix be used to find substrings of a sequence q that also exists reversed in another sequence d ?
 (b) Show this (manually) on the sequences $d = \text{DABADCE}$ and $q = \text{EDABCBD A}$, when the length of the substrings must be at least 3.