

Study Guide for MICA 8006 Week 3

This study guide covers week three lectures and assigned readings from Tramontano, Problem 1, pg 13-29. 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: BLAST, low complexity region, E-value, bit score, Z score, ROC curve, sensitivity, specificity, positive predictive value.
2. What does “mask low-complexity regions” mean? Why is this done?
3. What are the steps in a BLAST search?
4. What are the steps in a Needleman-Wunsch or Smith-Waterman alignment?
5. What is a position-specific scoring matrix? a Hidden Markov Model of a protein sequence family?
6. What are the steps in creating a position-specific scoring matrix? List three parameters that could be used to determine the weight given to a position.
7. Given a multiple sequence alignment of 100 related sequences: What might indels predict? Conserved residues? Conserved PG residue pairs? Conserved hydrophobic i, i+2, i+4 pattern?
8. Define: primary database; secondary database; UNIPROT; NRL-3D; HMMPfam; BLOCKS; PRODOM; multiple sequence alignments; PDB.
9. Compare BLOCKS and HMMPfam databases. How are they similar and how do they differ?