SGN-6156 Computational Systems Biology II

Exercise 1, April 9, 2008, at 12:15-13:45, in class TC407

This exercise will familiarize you with the probabilistic methods for pairwise sequence alignment, both in theory and in practise (using Matlab). Exercises can be done in class (during the exercise session).

1. Exercise 2.9 from (Durbin et al., 1998).

2. Exercise 3.2 from (Durbin et al., 1998).

3. Exercise 3.4 from (Durbin et al., 1998).

4. Try to align the two "test sequences" that are extensively used in the course text book (Durbin et al., 1998: Section 2). Try both global alignment using the Needleman-Wunsch (N-W) algorithm as well as local alignment using the Smith-Waterman (S-W) algorithm. Both of these algorithms are implemented in Matlab (type 'help nwalign' and 'help swalign' for details). Try different parameters (e.g. substitution matrices). Tune the parameters of the affine gap penalty score such that it corresponds to the linear gap score (in order to produce the same results).

5. Finding homologous genes is the most common way of generating new annotations for genes (although homologous genes do not need to have the same or similar function). Assume that you are given a database of annotated protein sequences for which you know the function. For simplicity, the database contains only 3 proteins in this case

   P1: KGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKL
   P2: NNPELQAHAGKVFKLVYEAAIQLQVTGVVVTDATLKNLGSVHVSKG
   P3: GSGYLVGDSLTFVDLLVAQHTADLLAANAALLDEFPQFKAHQE

   You are also given an additional, unknown protein sequence X

   PX: GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKL

   Try to estimate the function of the unknown protein (using the given database).