Homework 1 CS426 K. Kedem

9/12/03-9/19/03

1. Describe briefly how we can find the location of an expressed gene in the genome.

2. In class we gave the algorithm for global alignment (transperancy 9 in week 2). Copy these 6 lines and add to the algorithm the commands needed to allow finding the backtrack path. (No need to show how to find the path itself.)

3. For each case (below) separately, describe how you would modify the global alignment algorithm learnt in class to find the best local alignment, which includes
   - Both the first base (character) of S and the last base of T.
   - Both the first base of S and the last base of S.
   - Both the first base of T and the last base of T.

   For each case draw a sample alignment path in a filled DP matrix.

4. For the following cases, perform the pairwise alignment algorithm by hand, filling the dynamic programming table with scores and paths. Construct sequences S and T of length 6 over the alphabet A,C,G,T. Use the following scores: +2 match, -1 substitution, -2 space.
   - Global alignment: Construct S and T such that the global alignment between them is not unique. Show the paths on your DP table.
   - Local alignment: Construct S and T such that more than one entry in the table has the maximum V value. Show your matrix (paths not needed).

5. Affine gap penalty. Given the following score matrix, M(x,x) = 3; M(A,T)=M(G,A)=2; M(G,T)=M(C,T)=-2; M(C,A)=-3; M(C,G)=-4; and the affine gap penalty by Wopen =-4, Wextend=-1. Fill in the 4 DP matrices and show a path in V for the sequences S=ACAGTG; T=ATCG