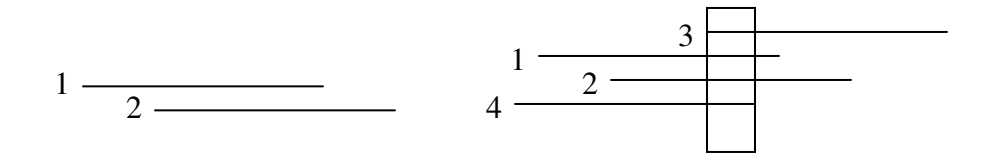


CMSC423 Homework 3  
Handed out: 2/28/07  
Due: 3/5/07

For this assignment you can use the web server at: <http://www.ebi.ac.uk/MPsrch/>

1. Perform the local sequence alignment of the following two sequences:  
>A  
TCAGTTGCCCAAACCCGCT  
>B  
AGGGTTGACATCCGTTTTT
  - a. First set gap open = gap extension = 1000 and perform the alignment with match score = 10, and with mismatch penalty score = 10, 7, 5, 3. Observe the effect of reducing the penalty score by examining the ten highest-scoring alignments for each case. What trend do you observe? Explain it.
  - b. Now with match score = 10 and mismatch penalty score = 10, explore the effect of changing alpha from 15 to 10, then 5 (holding beta at the default 3). What trend do you observe? Explain it.
2. Download sequences contained in accession numbers P21189 and NP\_143776.1 from the NCBI GenBank database (<http://www.ncbi.nlm.nih.gov>).
  - a. Perform global alignment using the “pam 120” as the penalty matrix and the default indel settings (alpha = 15, beta = 3). Examine your result: can you discern a region that likely will produce high-scoring local alignments?
  - b. Perform local alignment on the same sequences. Did the result from local alignment agree with your prediction based on the global alignment?
  - c. Use the contiguous sequence of the alignment found at step b (remove gaps from NP\_143776.1 portion of the alignment) as a query for an NCBI Blast search of the non-redundant databases. Are any putative conserved domains found? Check to see if there are any significant hits to *Danio rerio* (zebrafish) or *Arabidopsis* sequences.
3. Given a set of intervals (pairs of coordinates on a line), describe how you would identify the place on the line (also a pair of coordinates) covered by the largest number of intervals. In the picture below, I’m interested in the region marked by a box.



For all these exercises don't go too overboard on the answers, short answers are more than OK. Give me a quick printout of the alignments with the interesting regions highlighted in pen. Similarly, just print out the significant hits for 3.c.