Problem 1 (40 Points):

Get the protein sequences for mouse AMPA-4 (Accession #: NP_062665) and the worm glr-4 (Accession #: NP_495639). You may use the Entrez Protein Search at NCBI website to fetch the sequences. At the results page you may use the “Display” drop-down menu to get the sequences in the FASTA format and copy the sequences. Compare the local alignment of these sequences under dynamic programming. You may use the SIM tool by Expasy (http://www.expasy.org/tools/sim-prot.html) to align the sequences.

(a) Note the effect of varying gap open and extend penalties on the alignment. Report the top-5 local alignment scores obtained using BLOSUM62 on the following pair of (open,extend) penalties: (12,4), (12,3), (12,2), (12,1). Just write down the scores, you do not need to show the alignment.

(b) Which pair of (open, extend) penalties given in part (a) corresponds to the linear gap model?

(c) Why are the scores of the top match decreasing as the gap extend penalty goes up?

(d) Now, obtain the results by keeping the gap extend penalty fixed at 1 and changing the gap open penalty over [9,10,11,12]. Do you see a pattern? Explain the reason behind the pattern.

Problem 2 (30 Points):

Given the dynamic programming partial scores table above obtained from global alignment of the two sequences:

(a) Show an optimal global alignment of the sequences. Also, show the traceback path on the partial scores table.

(b) How many different optimal alignments exist? Why?
Problem 3 (30 Points):

(a) What is the maximum length of the DNA sequence that is expected to occur at least once in its entirety in another DNA sequence of length 2000? In other words, given two DNA sequences, $A$ and $B$, where $\text{length}(A)=2000$, what should be the maximum length of $B$, so that it is expected to observe a perfect semi-global alignment (all matches, no mismatch or internal gap) between $A$ and $B$? Show the steps of your calculation.

(b) What would be the E-value of such a perfect semi-global alignment if the sequence $B$ in part (a) was of length 10?