

Given the partially completed dynamic programming partial scores table above obtained for global alignment of the two sequences:

- (a) What is the gap penalty (assuming that a linear gap model is used)? What are the match and mismatch scores?
- (b) Fill in the rest of the table.
- (c) Show an optimal global alignment of the sequences. Also, show the traceback path on the partial scores table.
- (d) 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 30000? In other words given two DNA sequences, A and B , where $length(A)=30000$, 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 gaps) 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 100?