

Dynamic Programming Exercise

Written by Mingchao Xie for Bio 4342; Revised by Wilson Leung

- Using the following matrix, describe how one could use dynamic programming to calculate the optimal global alignment score $S(i, j)$ for the cell at position (i, j) . Be sure to define the parameters in your scoring system (i.e. for matches, mismatches, and gaps).

	X_{i-1}	X_i
Y_{j-1}	$S_{i-1, j-1}$	$S_{i-1, j}$
Y_j	$S_{i, j-1}$	$S_{i, j}$

- What are the major differences between local and global alignments? Provide an example where you would use a global alignment. Provide an example where you would use a local alignment.
- a) Complete the following dynamic programming matrix using **global** alignment with the following scoring system: 5 for match, -2 for mismatch, and -6 for gap.

Global Alignment		A	G	T
A				
A				
G				
C				

b) Complete the following dynamic programming matrix using **local** alignment with the same scoring system (i.e., 5 for match, -2 for mismatch, and -6 for gap)

Local Alignment		A	G	T
A				
A				
G				
C				

c) Based on the calculations of the optimal alignment score above, what is the optimal global alignment between AGT and AAGC? What is the optimal local alignment between AGT and AAGC?

- The questions above used fixed values for matches, mismatches, and gaps in the alignment. How can you improve this scoring system so that it more accurately reflects differences in real DNA and protein sequences?

For the next two questions, we will use two worksheets in an Excel workbook (Dynamic_Programming_1_0.xlsx) to explore the impact of the scoring system on the global and local alignments. (See the user manual for instructions on how to use this workbook.)

- If we change the scoring system by multiplying all the scores (i.e., match bonus, mismatch penalty, gap penalty) with a positive constant (such as 5), will this change the optimal global alignment? Why? What about the optimal local alignment?
- If we increase the match bonus and mismatch penalty by 5 and increase the gap penalty by 2.5 (i.e. match = 10, mismatch = -7, gap = -8.5), does this change the optimal global alignment between the query and subject sequences in the worksheet? Why? What about the optimal local alignment? (Include screenshots of the traceback matrix before and after you have changed the scoring system.)