Sequence comparison
Exercises (2)

1. Given two DNA sequences $S$ and $iS$ that are each other’s inverse (for example $5’$-GATCAT-3’ and $5’$-TACTAG-3’), what does their dotplot look like?

![Diagram of dotplot options]

A) B) C) D)

2. Which of the following statements on the Needleman and Wunsch algorithm for pairwise sequence alignment is true?

A) The Needleman-Wunsch algorithm is always better in aligning homologous sequences than a multiple sequence alignment
B) The Needleman-Wunsch algorithm provides the single best alignment.
C) The alignment is optimal (as measured by the alignment score), but there might be many equally optimal pathways/traces through the scoring table.
D) The Needleman and Wunsch algorithm computes the best local alignment
E) None of the above

3. The figure below shows a small fragment of a protein. From this figure, is it possible to define which extremity is the N-terminal, and which extremity is the C-terminal?

![Diagram of protein fragment]

a) Yes: 1 is Nter, 2 is Cter
b) No: there is not enough information
c) Yes, 1 is Cter and 2 is Nter
d) No: Nter and Cter are only defined for nucleic acids
e) No: we would need to know the sequence of this protein fragment
4) Which combination of program / substitution matrix will most likely give you the best alignment between two sequences that are highly similar?

A) BLAST / Blosum45  
B) Needleman-Wunsch / Blosum45  
C) BLAST / Blosum90  
D) Needleman-Wunsch / Blosum90

5) We want to find the best alignment(s) between the protein sequences WWYCTY and WCYTY. The scoring scheme S is defined as follows: \( S(i,i) = 10 \), \( S(i,j) = 0 \) otherwise. There is a gap penalty of -2-N, where \( N \) is the length of the gap (gaps at the beginning are considered, see below). The score \( S_{best} \) and the number \( N \) of optimal alignments are (show your final dynamic programming matrix for full credit):

<table>
<thead>
<tr>
<th></th>
<th>W</th>
<th>W</th>
<th>Y</th>
<th>C</th>
<th>T</th>
<th>Y</th>
</tr>
</thead>
<tbody>
<tr>
<td>W</td>
<td>10</td>
<td>7</td>
<td>-4</td>
<td>-5</td>
<td>-6</td>
<td>-7</td>
</tr>
<tr>
<td>C</td>
<td>-3</td>
<td>10</td>
<td>7</td>
<td>6</td>
<td>5</td>
<td>4</td>
</tr>
<tr>
<td>Y</td>
<td>-4</td>
<td>7</td>
<td>7</td>
<td>6</td>
<td>6</td>
<td>15</td>
</tr>
<tr>
<td>T</td>
<td>-5</td>
<td>6</td>
<td>7</td>
<td>7</td>
<td>27</td>
<td>16</td>
</tr>
<tr>
<td>Y</td>
<td>-6</td>
<td>5</td>
<td>16</td>
<td>17</td>
<td>20</td>
<td>37</td>
</tr>
</tbody>
</table>

A) \( S_{best} = 40, N = 1 \)  
B) \( S_{best} = 35, N = 1 \)  
C) \( S_{best} = 35, N = 2 \)  
D) \( S_{best} = 36, N = 2 \)  
E) \( S_{best} = 37, N = 1 \)

6. How many possible alignments of length \( M \), with no gaps, can you form when you compare two sequences of length \( N \) and \( M \), with \( N > M \)?

A) 1  
B) N-M  
C) N-M+1  
D) M  
E) N

7. We want to find the best alignment(s) between the DNA sequences AGTATCT and AGATGc. The scoring scheme S is defined as follows: \( S(i,j) = 2 \) if \( i = j \), and \( S(i,j) = 0 \) otherwise. There is a gap penalty of -2-2N, where \( N \) is the length of the gap (penalty for the first position counts; see table below). What is the value at the position marked (?)

<table>
<thead>
<tr>
<th></th>
<th>A</th>
<th>G</th>
<th>T</th>
<th>A</th>
<th>T</th>
<th>C</th>
<th>T</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>2</td>
<td>-4</td>
<td>-6</td>
<td>-6</td>
<td>-10</td>
<td>-12</td>
<td>-14</td>
</tr>
<tr>
<td>G</td>
<td>-4</td>
<td>4</td>
<td>-4</td>
<td>-6</td>
<td>-6</td>
<td>-10</td>
<td>-12</td>
</tr>
<tr>
<td>A</td>
<td>-4</td>
<td>-4</td>
<td>4</td>
<td>2</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>T</td>
<td>-8</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>G</td>
<td>-10</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>C</td>
<td>-12</td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

A) -2  
B) 0
C) 2
D) -1
E) 1