

## IN4030 Introduction to bioinformatics

### Exercises - week 2

#### Exercise 1

Examine the two alignments below:

##### Alignment 1

```
Seq1      AGTGTGAAGGTCCCGGCTAAT---CG
          :  ::  :   :  :   ::  ::  :  :
Seq2      A-TGCG--GCTAATGGC-AATATACG
```

##### Alignment 2

```
Seq1      AGTGTGAAGGTCCCGGCTAAT-----CG
          :   ::           :::::      ::
Seq2      A---TG-----CGGCTAATGGCAATATACG
```

The alignments have been created with the same gap opening penalty, but different gap extension penalties.

- Can you say anything about the relative gap extension penalties for the two alignments?
- What would the alignment have looked like if the gap opening penalty was very large?
- What would the alignment have looked like if both penalties were zero?

#### Exercise 2

Study the two alignments below. Which one do you believe shows the greater biological similarity?

##### DNA alignment

```
Seq1      AGGCTGCCAAAACGCACTGTTTAAT
          :  ::  :   :   ::           ::
Seq2      ACGCA-CGTTATGGCTAAAGCCTAT
```

##### Amino acid alignment

```
Seq3      PVALGLKEKNLYLSCVLKDKGQDIT
          :   ::           :   ::
Seq4      PADLGLMNNYNMIQLRCADLHYIT
```

#### Exercise 3

Calculate the score for the following alignment with the three different scoring systems:

```
--TCATAC--TCATGAACT
GGTAATCCCTC---AA--
```

- Match = 1, mismatch = 0, indel = -1

- b) Match = 1, mismatch = -1, initial gap = -2, additional indel = -1  
 c) Match = 0, mismatch = -1, initial gap = -2, additional indel = -1

#### Exercise 4

Given two sequences  $S$  and  $T$  (not necessarily of the same length), let  $G$ ,  $L$ , and  $H$  be the scores of an optimal global alignment, an optimal local alignment, and an optimal global alignment without counting initial gaps inserted into  $S$  and trailing gaps inserted into  $T$ , respectively.

- a) Give an example of  $S$  and  $T$  so that all three scores  $G$ ,  $L$ , and  $H$  are different.  
 b) Prove or disprove the statement  $L \geq H \geq G$

#### Exercise 5

Consider two DNA sequences of the same length  $n$  and let the scoring system be defined as follows: +1 for match, -1 for mismatch and -2 for each single nucleotide deletion or insertion. Let the score of the optimal global alignment be  $G$  and the score of the optimal local alignment be  $L$ .

- a) Prove that  $L \geq G$  and construct an example such that  $L = G$ .  
 b) What is the maximum value of  $L - G$ ? Construct an example with this maximum value.  
 c) If we want to find a pair of non-overlapping substrings within a given sequence  $S$  with the maximum global alignment score, can we simply compute the optimal local alignment score between  $S$  and itself? Explain!

#### Exercise 6

Given two DNA sequences  $S_1$  and  $S_2$  of length  $n$  and  $m$ , respectively, can you give an efficient algorithm which returns the number of possible global alignments between  $S_1$  and  $S_2$ ? What is the time complexity of your algorithm?

#### Exercise 7

Consider the following alignment

```

A C - G G T T A T -
- C T G G - - A T C

```

and the following score matrix:

	A	C	G	T
A	1			
C	-1	2		
G	-1	-2	1	
T	-2	-1	-2	1

- a) Suppose the gap penalty is  $-5-g$  for a gap of size  $g$ . Compute the alignment score for the above alignment using the above score matrix.  
 b) Is the above alignment an optimal global alignment? If not, what should be the optimal alignment score and the corresponding global alignment?

- c) Give a score matrix and a gap penalty so that the above alignment is an optimal global alignment.