Name: XIONG ZHIPENG__Student ID:_44231536__

11300A   Bioinformatics

Homework 3
Due Date: Nov. 28, 2023

1. Answer the following questions about scoring an alignment.

A. Calculate the score for the DNA sequence alignment shown below, using the scoring matrix below. Use an affine gap penalty to score the gaps, with -11 for opening the gap and -1 for each additional position in the gap. ("Affine gap penalty" refers to a situation when the gap opening and gap extension penalties are not the same. The gap opening penalty should be greater than the gap extension penalty.)

54-5-5-5-5-11-2=21

B. How would the score change if you were to use a nonaffine gap penalty? To answer the question, try a nonaffine penalty of -2, and then -6.

```
GACTACGATCCGTATACGCACA---GGTTCAGAC          A   2
||||||      |||||||||||||   |||||||||       G  -5   2
GACTACAGCTCGTATACGCACACATGGTTCAGAC          T  -7  -7   2
                                            C  -7  -7  -5   2
                                               A   G   T   C
```

Nonaffine penalty of -2:
54-5-5-5-5-2-2-2=28

Nonaffine penalty of -6:
54-5-5-5-5-6-6-6=16

2. Below is part of the BLOSUM62 matrix. Answer the following questions using this matrix.

A. Using this matrix, two aligned cysteines (C) would receive a score of 9 while two aligned threonines (T) would only receive a score of 5. What can you conclude about cysteine relative to threonine?

A score of 9 for two aligned cysteines (C) as compared to a score of 5 for two aligned threonines (T) suggests that the substitution of cysteine is less likely to occur than the substitution of threonine.

B. A serine (S) aligned with a cysteine (C) would receive a negative score (-1) while a serine aligned with a threonine would receive a positive score (1). Offer a possible explanation for this in terms of physicochemical properties of the amino acid side chains.

```
Part of BLOSUM62 matrix:
  C  9
  S -1   4
  T -1   1   5
     C   S   T
```

This could be because serine(S) and threonine(T) are more similar to each other in terms of their physicochemical properties than serine(S) and cysteine (C).

C. The bestng an alignment score. Use the BLOSUM62 matrix alignment of the two amino acid sequences "LDS" and "LNS" is obvious (it's shown below). Given a scoring system, you could easily calculate an alignment score. Set up a matrix and use the dynamic programming algorithm to "prove" that this is the best alignment by calculating x (see Powerpoint notes) to score aligned residues, and use a gap penalty of -1. (You may hand write the matrix in your homework rather than typing it if you like.)

```
seq1    LDS
        | |
seq2    LNS
```

	m	L	D	S
n	0	-1	-2	-3
L	-1	4	3	2
N	-2	3	5	4
S	-3	2	3	9

So, this is the best alignment.

3. Find the optimal global alignment of the two sequences *Seq*1: THISLINE and *Seq*2: ISALIGNED based on the BLOSUM62 matrix with linear gap penalty of -4.

	m	T	H	I	S	L	I	N	E
n	0	-4	-8	-12	-16	-20	-24	-28	-32
I	-4	-1	-5	-4	-8	-12	-16	-20	-24
S	-8	-3	-2	-6	0	-4	-8	-12	-16
A	-12	-7	-5	-3	-4	-1	-5	-9	-13
L	-16	-11	-9	-3	-5	0	1	-3	-7
I	-20	-15	-13	-5	-5	-3	4	0	-4
G	-24	-19	-17	-9	-5	-7	0	4	0
N	-28	-23	-18	-13	-8	-8	-4	6	4
E	-32	-27	-22	-17	-12	-11	-8	2	11
D	-36	-31	-26	-21	-16	-15	-12	-2	7

Alignment Seq1: THIS - LI - NE -
Alignment Seq2: - - ISA LIGNED