Solution Set 1

Due 4pm, Friday, September 15

Collaboration is allowed on this homework. You must hand in homework assignments individually. List the names of the people you worked with:

You may not use an alignment program to calculate the alignments on this problem set. Turn in your handwritten answers on alignment template that is are available on the course website.

Homework must be submitted by 4pm in MI646 or electronically to comp-bio@cs.cmu.edu.

1. Semiglobal Alignment

   (a) Using the same scoring function ($M = 4$, $m = -3$, $g = -3$), compute the semiglobal alignment of `SCIONS` with `CONSCIENCES`. Assume no penalty for aligning indels with the leading characters of `CONSCIENCES`; that is, indels can be inserted at the beginning of `SCIONS` with no cost. Show your alignment matrix with scores and traceback on the template provided.

	0	**C**	**O**	**N**	**S**	**C**	**I**	**E**	**N**	**C**	**E**	**S**
0	0	0	0	0	0	0	0	0	0	0	0	0
S	-3	-3	-3	-3	**4**	1	-2	-3	-3	-3	-3	4
C	-6	1	-2	-5	1	**8**	5	2	-1	1	-2	1
I	-9	-2	-2	-5	-2	5	**12**	9	6	3	0	-2
O	-12	-5	2	-1	-4	2	9	**9**	6	3	0	-3
N	-15	-8	-1	6	3	0	6	6	**13**	**10**	**7**	4
S	-18	-11	-4	3	10	7	4	3	10	10	7	**11**

   (b) What is the score of the optimal semiglobal alignment? How many different optimal alignments are there? Show them.

C	**O**	**N**	**S**	**C**	**I**	**E**	**N**	**C**	**E**	**S**
			4	4	4	-3	4	-3	-3	4
_	**_**	**_**	**S**	**C**	**I**	**O**	**N**	**_**	**_**	**S**

There is only one optimal semiglobal alignment. The optimal alignment score is 11.

2. Local alignment

(a) Compute the local alignment of `SCIONS` with `CONSCIENCES`, using the following scoring system: matches = 4, mismatches = -3, indels = -3. Show your alignment matrix with scores and traceback on the attached alignment template.

	0	**C**	**O**	**N**	**S**	**C**	**I**	**E**	**N**	**C**	**E**	**S**
0	0	0	0	0	0	0	0	0	0	0	0	0
S	0	0	0	0	4	1	0	0	0	0	0	4
C	0	4	1	0	1	8	5	2	0	4	1	1
I	0	1	1	0	0	5	12	9	6	3	1	0
O	0	0	5	2	0	2	9	9	6	3	0	0
N	0	0	2	9	6	3	6	6	13	10	7	4
S	0	0	0	6	13	10	7	4	10	10	7	11

(b) What is the score of the optimal local alignment? There is more than one non-overlapping optimal alignment. Show all optimal local alignments.

The score is 13.

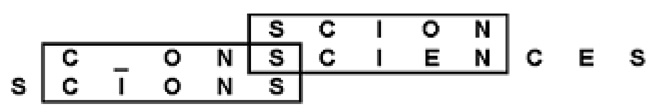
C	**_**	**O**	**N**	**S**
4	-3	4	4	4
C	**I**	**O**	**N**	**S**

S	**C**	**I**	**E**	**N**
4	4	4	-3	4
S	**C**	**I**	**O**	**N**

3. Comparing semiglobal and local alignments:
The semiglobal and local alignment algorithms both have provisions for forgiving indels in specific situations. However, the algorithms are fundamentally different. However, for some inputs and scoring functions, both algorithms will produce the same output. In some cases, both algorithms will produce the same output given the same input, but only when given different scoring functions. There are other cases, where the results obtained with one algorithm cannot be obtained with the other algorithm with *any* scoring function.

In this question, we consider whether we can force the semiglobal alignment algorithm to find the same alignment of `SCIONS` and `CONSCIENCES` as the local alignment algorithm, and vice versa.

(a) The semiglobal alignment in Problem 1 contains an optimal local alignment found in Problem 2, but not the other(s). Is there *any* scoring function that could result in a semiglobal alignment that incorporates two non-overlapping, optimal local alignments found in Problem 2? If so, give it. If not, explain why not.



No. The letters in SCIONS in the first and second local alignments match different parts of the word CONSCIENCES. Only one of them can be part of a global alignment.

(b) Find a scoring function (i.e., give values of M, m, and g) for which the local alignment algorithm yields a *single optimal* alignment of `SCIONS` with `CONSCIENCES` that is the same as the semiglobal alignment you obtained in Problem 1 (disregarding the leading overhang). Your alignment will not necessarily have the same score as the semiglobal alignment in Problem 1.

The scoring function $M = 3, m = g = -1$ *is one example of a scoring function that gives the same alignment as the semiglobal alignment. There are many (see below).*

(c) There is a set of scoring functions for which the local alignment algorithm yields a *single optimal* alignment of `SCIONS` with `CONSCIENCES` that is the same as the semiglobal alignment you obtained in Problem 1 (disregarding the leading overhang). Give a set of inequalities that specify the set of all possible scoring functions for which this is true.

	1	2	3	4	5	6	7	8
	S	**C**	**I**	**E**	**N**	**C**	**E**	**S**
	S	**C**	**I**	**O**	**N**	**_**	**_**	**S**
score	M	M	M	m	M	g	g	M

To obtain an optimal local alignment that is the same as the semiglobal alignment in Problem 1, we need a scoring function that guarantees that the alignment in Problem 1, when scored as a local alignment, has a better score than the two shorter local alignments in Problem 2. For local alignment, we also need to guarantee that the alignment score is positive for all cells on the trace back path.

To achieve these goals, the scoring function must satisfy three criteria:

i. *The cumulative score at site* 8*, which is* $5M + 2g + m$*, must be greater than the cumulative score at site* 5*, which is* $4M + m$*.*
ii. *The cumulative score at site* 8 *must be greater than the score of* *`C_ONS`* *aligned with* *`CIONS`**, which is* $4M + g$*.*
iii. *The cumulative scores at sites* 6 *and* 7 *must be positive to obtain a local alignment that includes the final S in SCIONS.*

Note that the score at site 7, which includes two gap penalties, must be lower than the score at site 6, which only has one gap. So it is sufficient to check that the score at site 7 is greater than zero. The cumulative score at site 7 is $4M + m + 2g$*.*

These criteria yield three inequalities

$$5M + 2g + m > 4M + m$$

$$5M + 2g + m > 4M + g$$

$$4M + 2g + m > 0$$

which reduce to

$$M + 2g > 0$$

$$M + g + m > 0$$

$$4M + 2g + m > 0$$

The scoring function M=3, m = g = -1 is one scoring function that satisfies these inequalities.