

### 3 Exercises – Extension of global alignment

#### 1. DP with score matrix

Use the score matrix below with gap penalty  $g = 1$  and answer the following questions.

	C	G	A	T
C	1	0	0	0
G		1	1	0
A			1	0
T				1

(a) Calculate the alignment score.

- Alignment 1

q: ATGCT

d: CA--T

- Alignment 2

q: CAGCT

d: C-A-T

(b) Calculate the score of  $H_{i,j}$ .

- Table A

			C	
		0	-1	
C		-1	$H_{i,j}$	

- Table B

			C	A	
			0	2	
A					
			-1	$H_{i,j}$	
G					

- (c) Fill the empty cells with appropriate scores in the DP table. What is the optimal alignment score?

q\d		C	A	T
	0	-1	-2	-3
C	-1		0	-1
A	-2	0	2	1
G	-3	-1		2
C	-4	-2		1
T	-5	-3	-1	

- (d) There are two different alignments that give the same optimal score in the solution above. Specify both of them.

## 2. Affine gap penalty

Affine gap penalties are often preferable ways to calculate gap scores than linear penalties. A gap with length  $l$  can be calculated as:  $g_l = g_{open} + (l - 1) * g_{extend}$ .

Use the following scoring scheme and gap penalties to answer the questions.

**Scoring scheme:**

$$R_{ab} = 1 \text{ for } a = b$$

$$R_{ab} = 0 \text{ for } a \neq b$$

$$g_{open} = 1, g_{extend} = 0.1$$

- (a) What is the gap penalty when  $l = 2$ .
- (b) Calculate the scores of the alignments.

1.      q: CAGCT  
           d: CT--T

2.     q: CAGCT  
       d: C-T-T
  
3.     q: CCT--  
       d: ---CT

### 3. Initialization for affine gap penalty

Initialize the following tables when  $g_{open} = 10$  and  $g_{extend} = 1$ .

		<b>E</b>		
		T	G	C
A				
	A			
	A			

		<b>F</b>		
		T	G	C
A				
	A			
	A			

		<b>G</b>		
		T	G	C
A				
	A			
	A			

### 4. Affine gap with three DP tables

Use the following scoring scheme and gap penalties to find the optimal alignment score of two sequences  $q = AG$  and  $d = GGGC$ .

**Scoring scheme:**

$$R_{ab} = 1 \text{ for } a = b$$

$$R_{ab} = 0 \text{ for } a \neq b$$

$$g_{open} = 1, g_{extend} = 0.1$$

(a) Fill all blank cells in the DP tables E, F, and G.

**E**

		G	G	G	C
	0	-1			
A	-1		-2.1	-2.2	-2.3
G		-1	-2		-2.2

**F**

		G	G	G	C
	0	-1			
A	-1	-2	-1		-1.2
G			-1	0	-0.1

**G**

		G	G	G	C
	0	-1			
A	-1	0	-1	-1.1	-1.2
G		0	1	0	

(b) What is the optimal score?

### 5. Trackback with affine gap penalty

Perform backtracking on E, F, and G tables to find the optimal alignment. The cells with double border should be visited during backtracking.

**E**

		A	C	G	T
C					
G					

**F**

		A	C	G	T
C					
G					

**G**

		A	C	G	T
C					
G					

(a) Write the optimal alignment.

### 6. Sequence distance with DP DP can be used to calculate the edit distance (Levenshtein distance) between two sequences.

**Scoring scheme:**

$$R_{ab} = 0 \text{ for } a = b$$

$$R_{ab} = -1 \text{ for } a \neq b$$

$$g = 1$$

With the scoring scheme above, the edit distance  $d$  is calculated as  $1 * T$  where  $T$  is the optimal score of the DP.

Find the edit distance between two sequences  $q = AG$  and  $d = ACG$ .

(a) Fill the DP table.

$q \backslash d$		A	C	T
A				
G				

(b) What is the edit distance between  $q$  and  $d$ ?