

2 Exercise solutions – Global pairwise alignment

1. Pairwise alignments

Pairwise alignments are two aligned sequences of DNA, RNA, or protein. DNA and RNA sequences consist of four different nucleotides, whereas protein sequences consist of 20 different amino acids. The “-” sign is used to represent a blank or a gap, which indicates an insertion or a deletion from one sequence to the other.

Use the simple scoring scheme and calculate the score of the following alignments.

Scoring scheme:

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

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

$$g = 1$$

(a) Alignment 1

q: ATGCT

d: CA--T

Solution: -1

(b) Alignment 2

q: CAGCT

d: C-A-T

Solution: 0

2. Brute force approach

A brute force approach can be used to find the optimal alignment. Use the sequences q and d below and answer the questions.

Sequences:

q: CG, d: AC

Scoring scheme:

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

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

$$g = 1$$

(a) Identify all possible alignments.

Solution:

1 = 4	CG--	C-G-	C--G	-CG-	-C-G	--CG
	--AC	-A-C	-AC-	A--C	A-C-	AC--

1 = 3	CG-	CG-	C-G	C-G	-CG	-CG
	A-C	-AC	AC-	-AC	A-C	AC-

1 = 2	CG
	AC

(b) Identify the optimal alignment with its score.

Solution:

CG
AC Score: 0

3. Table representation

An alignment can be represented as a table with arrows. Vertical and horizontal arrows indicate gaps, while diagonal arrows indicate matches and mismatches.

Identify the alignment that corresponds to the arrows in the following tables.

(a) Table 1

q\d		C	A	T
S				
	↘			
C		↘		
		↘		
A			↘	
			↓	
G			↓	
			↓	
C			↓	
			↘	
T				↘
				E

Solution:

q: CAGCT
d: CA--T

(b) Table 2

q\d		C	A	T
S				
	→	→	↓	
C		↓		
		↓		
A		↓		
		↓		
G		↓		
		↓		
C		↓		
		↘		
T			↘	
			→	→
				E

Solution:

q: -CAGCT-
d: C----AT