

Algoritmos para Bioinformática / Bioinformática 2021/2022

Aula 7 – Sequence Alignments

Task 1 – Alignment parameters

Considere as seguintes sequências:

S1: CTC

S2: CGTC

S3: ACTA

S4: AACT

Usando o algoritmo Needleman-Wunsch para calcular o alinhamento global entre S1 e S2 obtivemos as seguintes matriz S e T:

	-	C	G	T	C
-	0	-3	-6	-9	-12
C	-3	2	-1	-4	-7
T	-6	-1	1	1	-2
C	-9	-4	-2	0	3

	-	C	G	T	C
-	0	3	3	3	3
C	2	1	3	3	3
T	2	2	1	1	3
C	2	2	2	1	1

- Determine os valores dos parâmetros que foram usados no alinhamento: *match*, *mismatch* e *gap*.
- Indique um alinhamento ótimo e o respectivo score ótimo?
- Usando código desenvolvido nas aulas escreva uma função de teste que implemente as respostas às alíneas a) e b).

Task 2 – Alignment Information

- Consider the Needleman-Wunsch algorithm. Adopt the provided code to detect in which cells may occur the evidence of possible multiple optimal alignments. Print the index (i,j) of the cell and the possible directions of the alignment.
- Calculate the total number of possible optimal multiple alignments.

Task 3 – Independent gap and Affine gap score

In the gap independent scoring system each gap contributes independently to the final score, i.e. the contribution of n gaps in the alignment is given by $n * gap_penalty$.

The affine gap score determines that the penalty for opening a gap is different than that of extending it. Create a test function that compares the scores from the two modes.

- i) start by producing a global alignment of two sequences;
- ii) calculate the score with the two-scoring system and compares values;
- iii) repeat the procedure with local alignments.
- iv) From the produced alignment calculate the total number of gaps in both sequences
- v) Determine the number of gaps that are new openings and gaps that are extensions
- vi) Determine the largest continuous gaps in one of the sequences.