

Bioinformática

Exercícios TP4: Algoritmos de alinhamento e matrizes de score

1. Considere as duas sequências:

HEAGAWGHEE
PAWHEAE

- a) Construa a matriz de alinhamento para estas sequências e calcule o alinhamento global óptimo pelo algoritmo de Needleman-Wunsch, considerando uma gap penalty de -8 e uma matriz de score BLOSUM50.
b) Construa um alinhamento semelhante, mas usando o método de alinhamento local de Smith-Waterman.

2. Obtenha as sequências da tripsina humana e de ratazana. Produza o seu alinhamento global com o programa "needle" (www.ebi.ac.uk/emboss/Tools/psa). O software usa por defeito uma matriz BLOSUM62 para fazer este alinhamento, mas qual lhe parece que deveria ser a matriz apropriada para esta situação. Volte a correr "needle" com esta matriz e compare com o resultado anterior. Obtenha a sequência da tripsina de *Litopenaeus vannamei* (white leg shrimp) e faça o alinhamento desta com a humana. Qual seria a matriz de alinhamento mais apropriada nesta situação?... Repita o alinhamento com esta última e compare.

3. Utilize o site DALI para fazer o alinhamento estrutural das duas proteínas tripsina 1 humana e tripsina da bactéria *Streptomyces griseus*. Compare este alinhamento com o obtido com a ferramenta EMBOSS (water). Modifique os valores por defeito das penalidades de abertura e extensão de *gaps* e observe o efeito sobre o score do alinhamento, procurando encontrar os valores que melhor fazem corresponder o alinhamento observado ao alinhado estrutural.

4. Considere o seguinte fragmento de um bloco da base de dados BLOCKS, para a família das globulinas:

myoglobin	G	H	G	Q	D	I	L	I	R	L	F	K	S	H	P	E	T	L	E	K	F	D	R	F
Hb α	E	Y	G	A	E	A	L	E	R	M	F	L	S	F	P	T	T	K	T	Y	F	P	H	F
Hb β	E	V	G	E	A	L	G	R	L	L	V	V	Y	P	Y	T	Q	R	F	F	E	S	F	
erythrocrurin	G	D	P	V	G	I	L	Y	A	V	F	K	A	D	P	S	I	M	A	K	F	T	Q	F
clam Hb	G	N	G	V	A	L	M	T	T	L	F	A	D	N	Q	E	T	I	G	Y	F	K	R	L
worm Hb	T	S	G	V	D	I	L	V	K	Y	F	T	S	T	P	A	A	Q	E	F	F	P	K	F
leghemoglobin	Q	Y	S	V	V	G	Y	T	S	I	L	E	K	A	P	A	A	K	D	L	F	S	F	L
Glycera Hb	E	Y	G	A	E	A	L	E	R	M	F	L	S	F	P	T	T	K	T	Y	F	P	H	F

com base nas frequências de ocorrência dos aminoácidos e frequência de ocorrência dos pares fenilalanina-leucina (FL), obtenha um valor aproximado para o score da substituição Phe-Leu na matriz BLOSUM.

5. Use o programa PRSS para calcular a significância dos alinhamentos da cadeia alfa da hemoglobina com as seguintes proteínas: cadeia beta da hemoglobina, Leghemoglobina do tremço, GST-7 de *C. elegans*. O que pode concluir?