1ª PROVA DE BIOLOGIA MOLECULAR COMPUTACIONAL – FCI0760

Aluno: _____ N.o USP: _____

Data: 14/10/2016 Nota: _____

1) Alinhe, **manualmente**, as sequências **HGSAQVKGHG** e **KTEAEMKASEDLKKHGT** utilizando o algoritmo de programação dinâmica e a matriz de substituição BLOSUM 40. Utilize, para esta finalidade, as matrizes abaixo. Preencha a matriz de escores, determine o *traceback* e o alinhamento final obtido.

		H	G	S	A	Q	V	K	G	H	G
	0	-8	-16	-24	-32	-40	-48	-56	-64	-72	-80
K	-8										
T	-16										
E	-24										
A	-32										
E	-40										
M	-48										
K	-56										
A	-64										
S	-72										
E	-80										
D	-88										
L	-96										
K	-104										
K	-112										
H	-120										
G	-128										
T	-136										

	A	R	N	D	C	Q	E	G	H	I	L	K	M	F	P	S	T	W	Y	V	B	Z	X
A	5																						
R	-2	9																					
N	-1	0	8																				
D	-1	-1	2	9																			
C	-2	-3	-2	-2	16																		
Q	0	2	1	-1	-4	8																	
E	-1	-1	-1	2	-2	2	7																
G	1	-3	0	-2	-3	-2	-3	8															
H	-2	0	1	0	-4	0	0	-2	13														
I	-1	-3	-2	-4	-4	-3	-4	-4	-3	6													
L	-2	-2	-3	-3	-2	-2	-2	-4	-2	2	6												
K	-1	3	0	0	-3	1	1	-2	-1	-3	-2	6											
M	-1	-1	-2	-3	-3	-1	-2	-2	1	1	3	-1	7										
F	-3	-2	-3	-4	-2	-4	-3	-3	-2	1	2	-3	0	9									
P	-2	-3	-2	-2	-5	-2	0	-1	-2	-2	-4	-1	-2	-4	11								
S	1	-1	1	0	-1	1	0	0	-1	-2	-3	0	-2	-2	-1	5							
T	0	-2	0	-1	-1	-1	-1	-2	-2	-1	-1	0	-1	-1	0	2	6						
W	-3	-2	-4	-5	-6	-1	-2	-2	-5	-3	-1	-2	-2	1	-4	-5	-4	19					
Y	-2	-1	-2	-3	-4	-1	-2	-3	2	0	0	-1	1	4	-3	-2	-1	3	9				
V	0	-2	-3	-3	-2	-3	-3	-4	-4	4	2	-2	1	0	-3	-1	1	-3	-1	5			
B	-1	-1	4	6	-2	0	1	-1	0	-3	-3	0	-3	-3	-2	0	0	-4	-3	-3	5		
Z	-1	0	0	1	-3	4	5	-2	0	-4	-2	1	-2	-4	-1	0	-1	-2	-2	-3	2	5	
X	0	-1	-1	-1	-2	-1	-1	-1	-1	-1	-1	-1	0	-1	-2	0	0	-2	-1	-1	-1	-1	-1

Figura 1. Matriz de substituição BLOSUM40.

2) Analise a sequência proteica AMV79340.1 quanto à existência de hélices transmembrana (TM). Para esta finalidade, calcule o índice de hidrofobicidade em janelas de 20 aminoácidos. Para esta sequência. Há hélices TM? Se sim, quantas podem ser inicialmente identificadas?

3) Duas sequências foram alinhadas pelo mesmo programa, porém usando parâmetros distintos. Os alinhamentos obtidos estão mostrados abaixo. Com base nos alinhamentos, responda:

   a. Qual a principal diferença entre os alinhamentos?
   b. Para este caso, qual alinhamento deve ser mais apropriado?
   c. Explique as diferenças conceituais entre os parâmetros utilizados nestes alinhamentos.

```
                10        20        30        40        50        60
cellob MVDSPSIYKTALLLSMLATSNAQQVGTYTAETHPSLTWQTCSGSGSCTTTSGSVVIDANW
       : .:    . :  :...: .:::                        :: :
PQQ-de MRSS----SLAWALGLVALANAQ---------------------GSPT------------
                  10                              20

                70        80        90       100       110       120
cellob RWVHEVGGYTNCYSGNTWDSSICSTDTTCASECALEGATYESTYGVTTSGSSLRLNFVTT
       .:                                          :.:  :::
PQQ-de QW-------------------------------------YDSITGVT-------------
                                                 30

               130       140       150       160       170       180
cellob ASQKNIGSRLYLLADDSTYETFKLFNREFTFDVDVSNLPCGLNGALYFVSMDADGGVSRF
                                 :.: .  :.:.:                  : .  :
PQQ-de ------------------------FSRFYQQDTDASW-----------------GYI--F
                                    40                            50

               190       200       210       220       230
cellob PTNKAGAKYGTGYCDSQCPRD-LKFIDGQANIEGWEPSSTDVNAGTGNHGSCCPEMDIWE
       :. ..:          : : . . ...: :.  ::
PQQ-de PSASGG----------QAPDEFIGLFQGPASA-GW-------------------------
                         60        70

     240       250       260       270       280       290
cellob ANSISSAFTAHPCDSVQQTMCTGDTCGGTYSDTTDRYSGTCDPDGCDFNPYRFGNTNFYG
                             :.. ::.. .                 ::   :
PQQ-de ---------------------IGNSLGGSMRN----------------NPLLVG------
                               80                         90

     300       310       320       330       340       350
cellob PGKTVDNSKPFTVVTQFITHDGTDTGTLTEIRRLYVQNGVVIGNGPSTYTAASGNSITES
           ::.: :     .. .. .:: .                   :: :..
```

```
PQQ-de ---WVDGSTP-----RISARWATDYAP------------------PSIYSGPR-------
                     100       110

     360       370       380       390       400       410
cellob FCKAEKTLFGDTNVFETHGGLSAMGDALGDGMVLVLSLWDDHAADMLWLDSDYPTTSCAS
                            :. .:..  .: .  .                 :   .:.
PQQ-de --------------------LTILGSSGTNGNIQRIV---------------YRCQNCTR
                           120       130                       140

     420       430       440       450       460       470
cellob SPGVARGTCPTTTGNATYVEANYPNSYVTYSNIKFGTLNSTYSGTSSGGSSSSSTTLTTK
         : : :    :::.:..                                : .  :::
PQQ-de WTGGAGGI--PTTGSAVF-----------------------------GWAFHSTT----
           150                                       160

     480       490       500       510       520       530
cellob ASTSTTSSKTTTTTSKTSTTSSSSTNVAQLYGQCGGQGWTGPTTCASGTCTKQNDYYSQC
                  : :  :.     ...:: ::   :.. :
PQQ-de ---------KPLTPSDPSSGLYRHSHAAQ-YGFDIGNART--------------------
                170       180       190

     540
cellob L

PQQ-de TLYDYYLQQLTNAPPLSGGAPTQPPTQQPPTTTAPPPPPPSSTFVSCPGAPQPRYQMNVA
       200       210       220       230       240       250
```

Figura 2. Alinhamento 1.

```
         480       490       500       510       520       530
cellob LTTKASTSTTSSKTTTTTSKTSTTSSSSTNVAQLYGQCGGQGWTGPTTC-ASGTCTKQND
       .:    :..  . ::  . .  :: .       :.:::::::::::: : :...: ..:.
PQQ-de ITQPPITTSPPTPTTPPVVQPPTTVAPPQASQTLWGQCGGQGWTGPTLCPANSVCRESNQ
      660       670       680       690       700       710

         540
cellob YYSQCL
       .:::::.
PQQ-de WYSQCV
      720
```

Figura 3. Alinhamento 2.

4) A cristalina é um produto gênico presente em diversas espécies, incluindo em humanos. Usando a ferramenta **Entrez** do **NCBI**:
    a. Identifique quantos genes diferentes existem para a cristalina (*crystallin*, em inglês).
    b. Analise a composição de domínios para estes genes. Há diferenças entre eles? Com base na composição de domínios e em análises complementares, é possível associar diferenças funcionais entre estes genes.

5) Dado o arquivo *prova.fasta*, disponível no site http://nascimento.ifsc.usp.br e contendo múltiplas sequências de cristalinas:
    a. Faça um alinhamento múltiplo das sequências usando a ferramenta **ClustalX**.
    b. Utilizando este alinhamento, construa uma árvore filogenética e avalie sua árvore através do método de *bootstraping*.
    c. A árvore filogenética obtida está de acordo com o que seria esperado? Para responder a esta pergunta, consulte os códigos de entrada das sequências na base **Entrez**.