**EXERCISE 2:**

I predicted that all of the proteins that began with “M” or Methionine, or AUG/ATC (start codon), however you want to refer to it as, as proteins that would be considered “real”. Only frame 4 peptide sequence 2 (or 23rd sequence from the top) was correctly predicted by me. This protein was only a hypothetical protein. The following data shows the results of my blastn and blastp searches.

seqA:

blastn: Drosophila melanogaster chromosome 2R

blastp results:

>seq 1 frame 1 : LAAPSDGGHLPAERSGIKGRGVAESSTLEL length: 30

GE19322 [Drosophila yakuba]

>seq 2 frame 1 : ITRRIPNMHGYRTYNMESHHAHHDASPVDQKPLVVDLLATQYGKPQTPPPSPNGKFKDKAEQT length: 63

seq2 – homeodomain transcription factor, partial [Drosophila melanogaster]

>seq 6 frame 1 : WPPECPRWACPRCPQCRCPDTRDMPAIHRPTDSTATRPTTSPPARRRHIPLVLICIIRT length: 59

seq6 – PREDICTED: dystonin isoform X3 [Condylura cristata]

>seq 7 frame 1 : WDPAQRDRRTPPVPPAFWALCPPPPAIPDWVWVCPRPRRRRWICSRRHRRTPPRCRSRQWDPITPRCSTAVQWLNPLHQFLLPLH length: 85

seq7 – hypothetical protein MIMGU\_mgv1a000031mg [Erythranthe guttata]

>seq 13 frame 2 : LLERQPRLGDSRRPVGTPLSHRLHP length: 25

seq13 – hypothetical protein [Sulfobacillus thermosulfidooxidans]

>seq 15 frame 2 : GVSPRSTHTLSPPPMLPQKIVQTSLSQPHLFIPEDCTDCRVANCSHN length: 47

seq15 – PREDICTED: dystrophin-related protein 2 [Picoides pubescens]

>seq 19 frame 3 : SRANVTSYLHPIFPLSKTECLSSPDNSLNGSRGSEIPADPSVRRYRTAFTRDQLGRLEKEFYKENYVSRPRRCELAAQLNLPESTIKVWFQNRRMKDKRQRIAVAWPYAAVYSDPAFAASILQAAANSVGMPYPPYAPAAAAAAAAAAAVATNPMMATGMPPMGMPQMPTMQMPGHSGHAGHPSPYGQYRYTPYHIPARPAPPHPAGPHMHHPHMMGSSATGSSYSAGAAGLLGALPSATCYTGLGVGVPKTQTPPLDLQSSSSPHSSTLSLSPVGSDHAKVFDRSPVAQSAPSVPAPAPLTTTSPLPAPGLLMPSAKRPASDMSPPPTTTVIAEPKPKLFKPYKTEA length: 348

seq19 – PREDICTED: heparan-alpha-glucosaminide N-acetyltransferase isoform X1 [Bubalus bubalis]

>seq 20 frame 3 : ARDPHTLSPPPPCSPKRLYKLVLVSLIYLFPKIVQIVE length: 38

seq20 – even skipped [Drosophila melanogaster]

>seq 23 frame 4 : GLKSFGLGSAITVVVGGGDMSEAGRLALGIRRPGAGSGLVVVSGAGAGTDGAD length: 53

seq23 – hypothetical protein CYLTODRAFT\_492058 [Cylindrobasidium torrendii FP15055 ss-10]

>seq 32 frame 4 : RRQLVVDCAAVTARVCCSQPPRALLYRCALQANGRRRWERP length: 41

seq32 – hypothetical protein [Catenuloplanes japonicus]

>seq 34 frame 5 : WSGLSHWTAVEHLGVIGSHWRERQRGGVRR length: 30

seq34 –PREDICTED: zinc finger and BTB domain-containing protein 8A [Merops nubicus]

>seq 43 frame 6 : RRCDSGVPTGRRESPSRGCRSRSYPDWIGILFWTEERLGVSNSSRLLGFIFKLTIWRGRRCLGLAVLGGQEVHNQGLLVHGTGVVMGMVALHVVGSVSVHVWYASCDSKLAYSEDINEGS

seq43 – PREDICTED: sister chromatid cohesion protein PDS5 homolog B [Orycteropus afer afer]

>seq 44 frame 6 : LLTVRRSQLECAALSHPAPFYTAALCRQMAAVAGSG length: 36

seq44 –aminoglycoside phosphotransferase [Bacillus sp. J13]

seqB:

blastn: Homo sapiens BIC noncoding mRNA, complete sequence

blastp results:

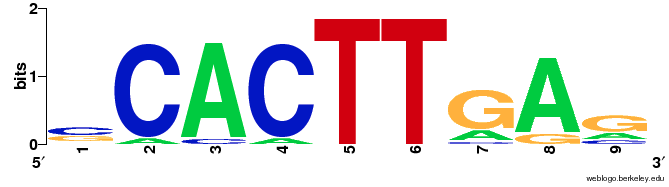
>seq 40 frame 5 : ESNKSINFKLKLPEQKIVTWGKVPVSKSFKKRCQTIHDITGG length: 42

Seq40 – hypothetical protein ACD\_62C00601G0003 [uncultured bacterium]

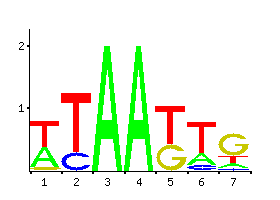
**EXERCISE 3:**

a. consensus sequence:

|  |  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- | --- |
| S | C | A | C | T | T | G | A | V |

b. 

c. They are quite close but the consensus sequence does not display the probability of a particular base as well as the motif logo as each base is represented by its probability.

d. 

This was the logo that I obtained from JASPAR Nkx2-5 search. All the As are Ts and vice versa. The sequence otherwise looks quite conserved.