Problems: Gene analysis
Mic 653 – 4 March 2004

A. Find location and amino acid sequence of mystery gene

1. Go to the web site at Kazusa DNA Research Institute dedicated to the genome of *Anabeana/Nostoc* PCC 7120 (http://www.kazusa.or.jp/cyano/Anabaena)

2. Do a keyword search for the gene all1290

3. Click on all1290 (first entry), note the predicted identify of the gene

4. Download the translated sequence (click on translated sequence and then use your browser to save the page as All1290.FA)

   Download the nucleotide sequence containing all1290 plus 500 nucleotides on either side of the gene (increase the initiation nucleotide by 500 and decrease the termination nucleotide by 500, then press submit)(note that since the gene is on the complementary strand - -see location – the initiation has a higher coordinate than the termination). You should get almost 27 lines of sequence. If you get the message "Requested sequence is too long", then you probably added 500 where you should have subtracted, and vice versa.

5. Submit the ~1600 nucleotides of sequence to GeneMark (select Nostoc PCC 7120 as the species). Click on "Translate predicted genes into protein″.

6. Compare the two amino acid sequences predicted by Kazusa and GeneMark.

7. Do what is necessary to decide which one (if either) is correct.

B. Find motif responsible for binding cyclic AMP

Task
* You have a protein you believe binds cAMP.
* You want to determine region of protein associated with binding

Strategy
B0. Collect sequences of several cAMP-binding protein
    (Done. Available on web in FastA format. Table of contents at end of problem.)

B1. Run *Meme*, a pattern-finding program. Ask it to find any significant motifs

B2. Rerun *Meme*. Demand that every protein has identified motifs

B3. Run Pfam or BlastP (from web) over one of the sequences to check how the motifs you
    found match the opinion of the world at large.
B4. Consider:

    Did you get what you expected? How many of the nine proteins have motifs found by
    Meme? Do any of the motifs found correspond to cAMP-binding motifs? Rationalize your
    results.

SET OF cAMP-binding protein used in Problem B

P07278   Yeast, cAMP-dependent protein kinase regulatory chain
P03020   E. coli, cAMP receptor protein (CRP)
Q64359   rat, Cyclic-nucleotide-gated olfactory channel OCNC2 subunit
P29747   fly, Cyclic-AMP response element binding protein A
P18847   human, Cyclic-AMP-dependent transcription factor ATF-3
P34122   Dictyostelium, Cyclic-AMP-Binding protein CABP1
P27925   cow, cAMP-response element protein 2 CREB2
Q9NP56   human, cAMP-specific phosphodiesterase
Q04758   mouse, cAMP-dependent protein kinase inhibitor

C. Repeat and extend characterization of PEO-related gene

C1. Access the PEO-related sequence on the website

C2. Get better estimate of promoter and polyA site
    (e.g. by TSSW and PolyASH)
    (Is there a TATA box upstream from the predicted promoter? Where?)

C3. Find encoded protein sequence by suitable method
    (e.g. FGeneSH(GC) or comparison with cDNA)

C4. Continue characterization of protein
    a. Contains signal sequence? How do you know? How confident are you?
    b. Contains transmembrane domains? How do you know? How confident are you?