Biol 591 Introduction to Bioinformatics (Fall 2003)
Problem Set 2M – Regulatory protein and their binding sites

PS2M-1. Most transcriptional factors bind to DNA the same way as CRP. Two of the following sequences are binding sites of actual transcriptional factors. Which do you think they are?

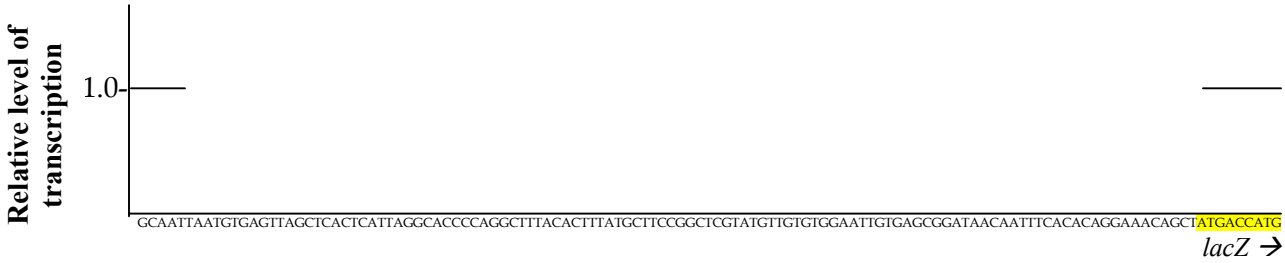
a. AATTGTGAGCGGATAACAATT

b. CACTAGACGGCTGTGATAGT

c. ACCTGTAGGATCGTACAGG

d. AAATTAGAGACTTGTGACATG

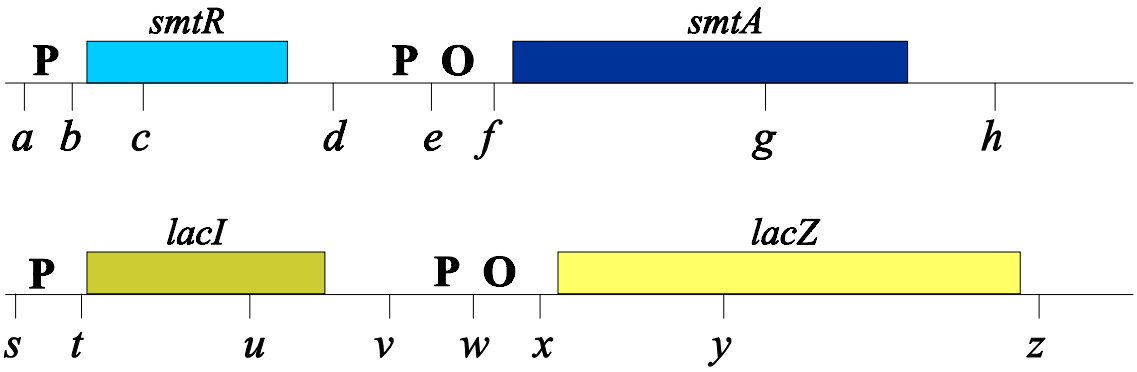
PS2M-2. The chart below shows on its X-axis the DNA sequence from *E. coli* preceding the *lacZ* gene, from 66 bases prior to the start of *lacZ* to 9 bases within *lacZ* (the same sequence is given below the chart in a larger font for those without microscopic vision). Draw two curves relating the predicted level of transcription of *lacZ* in different mutants. The first curve should reflect transcription when *E. coli* is growing in the presence of lactose and absence of glucose. For the second curve, consider *E.* coli to be growing in the presence of glucose and absence of lactose. The height of the curve at each point should indicate the level of transcription you predict if the nucleotide at that position were deleted, relative to the level of *lacZ* transcription if the nucleotide were not deleted. For example, since you wouldn't think that changing the nucleotides within *lacZ* should affect the transcription of the gene, the height of the far right of the curve should be 1.0.



GCAATTAATGTGAGTTAGCTCACTCATTAGGCACCCCAGGCTTTACACTTTATGCTTCCGGC
TCGTATGTTGTGTGGAATTGTGAGCGGATAACAATTTCACACAGGAAACAGCTATGACCATG

PS2M-3. A cell in the heart, a cell in the bicep muscle, and a cell in the pancreas have the same set of genes. All three cells express some of the same genes, the heart cell and bicep muscle cell express some genes that the pancreas cell does not, and each cell type expresses genes not expressed in the other cell types. Explain how all this could happen. What's fundementally ***different*** in these cells?

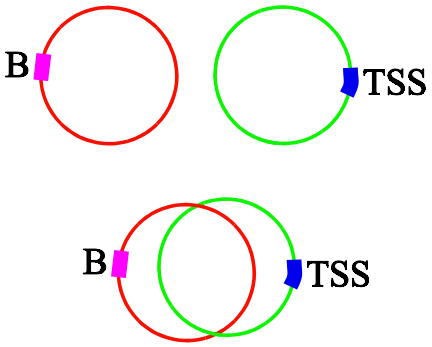
PS2M-4. You want to market a home testing unit that will enable households living nearby mining operations to monitor minute levels of toxic heavy metals. Needless to say, your customers will want to know about heavy metals before they reach levels that are lethal. Your plan is to make use of a gene from the bacterium *Synechococcus* PCC 7942, which is highly resistant to heavy metals. *Synechococcus* achieves its resistance by producing in large quantities a protein, metallotheionein, that binds to heavy metals and prevents them from acting on the cell. The protein is encoded by *smtA*. *smtA* is preceded by *smtR*, which encodes a repressor of the gene. You have cloned the *smt* region and from the sequence and other experiments deduced the following map (shown on the next page with the previously known map of the *lacZ* region):



Maps of *smtA* and *lacZ* regions. Promoters and operators for the genes are indicated by **P** and **O**, respectively. Lower case letters indicate restriction sites.

Your goal is to cut out part of the *smtA* region and part of the *lacZ* region in such a way so that the two pieces can be ligated together, returned to *Synechococcus*, and used to detect heavy metals. The test you envision is that the customer adds suspect water to a test tube containing the genetically modified *Synechococcus*. Then, the customer adds *o*-nitrophenylgalactoside, a colorless chemical that turns yellow when acted upon by ß-galactosidase. The test is positive or negative depending on the resulting color of the liquid in the tube. Describe exactly what pieces you would use from the *smtA* and *lacZ* regions and how you would put them together to achieve the desired end.

PS2M-5. Wedel et al [Science (1990) 248:486-489] studied how positive acting transcription factors work by analyzing the requirements for binding in vitro (in a test tube) of RNA polymerase to the position near the start of transcription. Two situations were examined (shown at right). In the first, a plasmid carrying the binding site (**B**) for a positive acting transcriptional regulator (like CRP) was mixed with a separate plasmid carrying the transcriptional start site (**TSS**) plus 32 bases upstream. In the second, these two plasmids were linked together (like in a chain).



The authors added RNA polymerase, the positive acting transcriptional regulator, plus other necessary components to the two plasmids and measured transcription. They found that transcription was significantly higher when the plasmids were linked than when they were not. What does this result say about what is required for binding sites to enhance transcription?