Biol 213 Genetics: Wednesday, November 1, 2000
Transcriptional Regulation - Part III

For the first exam you started the analysis of your mystery sequence. For the second exam you found an article of interest and summarized it. For the third exam, you'll combine the two tasks, finding an article that tells you something of interest about the regulation of a gene in your sequence. Bear in mind that you'll almost certainly be the first person on Earth to seriously investigate your gene and so will immediately become the leading authority on the subject.

Stay tuned for hints as to how to go about looking for articles that bear on the regulation of your gene.

Outline

I. Large scale transcriptional regulation in prokaryotes (pp.342-343)

II. Transcriptional regulation in eukaryotes (pp.345-350)

A. EXPERIMENT: Fine structure of b-globin promoter
B. Role of positive-acting regulators on transcription
C. RNA processing

III. Comparison between prokaryotes and eukaryotes

I. Large scale regulation in prokaryotes

OK, we're well versed in regulation of the *lac* operon. One operon down, perhaps a thousand left to go. This may sound daunting for us, but consider the plight of *E. coli*. If every operon were controlled by its own repressor, then something close to half of the genome would be made up of proteins designed to regulate the expression of other protein! Too many chiefs, not enough Indians.

Fortunately, there are ways of regulating many genes at once, and this is a good thing. Suppose *E. coli* finds itself on a garbage heap one hot day. Proteins that worked fine inside a cozy 37° colon start to wilt. This is serious! There are half unfolded protein stuffing the cell that not only don't work any more but threaten to gum up normal operations. To respond to this emergency, *E. coli* needs to induce the expression of a large number of genes, whose products will destroy some aberrant protein and try to refold others back into shape. How can it regulate the expression of all of these genes in the same way?

The solution is to give these genes a special promoter that is active only during heat shock. Conceptually, this is no problem. Promoters are no more than DNA sequences that bind RNA polymerase. All you need to do is precede heat shock-inducible genes by a different DNA sequence and then when heat strikes, call out a special RNA polymerase that recognizes it.

Bacteria are actually more clever than that. RNA polymerase constitutes a major investment for the cell, which needs lots of the protein in order to satisfy its transcription needs. It makes no more sense to manufacture wholly new RNA polymerase during times of heat shock than it does to handcraft a special door for every house. It's much simpler to use the same doors for everyone but slip in a different keyhole. Similarly, RNA polymerase is modular, composed of a constant core polymerase plus a sigma factor hat can be popped in and out. Different sigma factors recognize different promoters. Fig. 13-5 (p.343) shows a cartoon of how a sigma factor recognizes a promoter, enabling RNA polymerase to bind. Once binding is accomplished, the sigma factor is released and the rest of RNA polymerase initiates transcription.

Bacteria can also control large numbers of genes through positive acting factors like CRP. In this way the expression of genes encoding various sugar metabolizing enzymes can be turned on or off in response to the presence or absence of glucose.

SQ1. What kind of mutation do you think would affect the expression of more genes: a mutation in a gene encoding a repressor or a mutation in a gene encoding a sigma factor?

SQ2. What kind of activities do you think would be regulated by the expression of an alternative sigma factor rather than a repressor?

a. Expression of a high efficiency phosphate transporter in response to a phosphate-poor environment

b. Expression of ribosomal proteins

c. Expression of different groups of enzymes required at different stages of development (bacteria also differentiate!)

II. Transcriptional regulation in eukaryotes

II.A. EXPERIMENT: Fine structure of b-globin promoter

We'll leave *E.coli* to put out its fires and turn to a study of ourselves. Do eukaryotes have operons with inducible repressors? If so, they have to solve a horrific problem. Repressors aren't magic. They find their operators by looking for them. We have a thousand times more DNA than *E. coli*. A repressor in a eukaryote figures to take a thousand times longer to find its operator. It is perhaps because of such logistical problems that eukaryotes have devised a different strategy.

In order to study with as little bias as possible the regulation of a mammalian gene, Myers et al. (1986) systematically altered each base upstream from the transcriptional start site of mouse b-globin (a component of hemoglobin) and asked what effect the mutations had on transcription. In this way, they could identify regions of DNA that were important in the regulation of transcription of b-globin.

The strategy of Myers et al (Fig 1) was to obtain as many mutations near the transcriptional start site of b-globin and test whether transcription was affected by measuring b-globin transcript from human HeLa cells into which the mutant DNA had been introduced. HeLa cells are cells that were isolated decades ago from a cancer patient and continue to grow in culture. Notice that Myers et al cloned the region of interest just

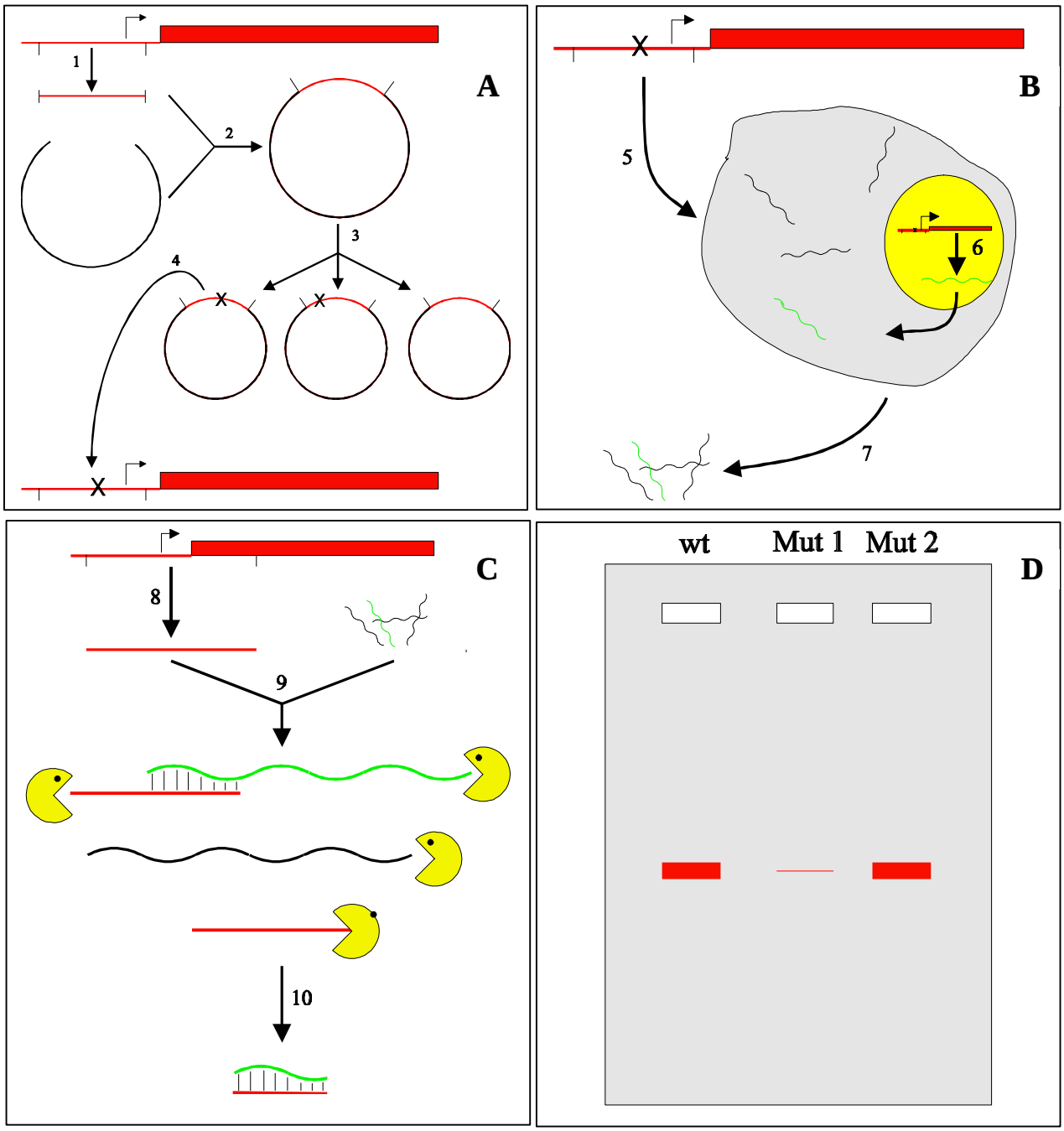

Figure 1. Experiment by Myers et al. (1986) to define regions of DNA important in the regulation of expression of mouse b-globin. (**A**) Mutagenesis of region upstream from b-globin gene. The region upstream from the gene and including the transcriptional start site (arrow) was cut out with restriction enzymes (step 1) and combined with a plasmid also cut with restriction enzymes to form a recombinant plasmid carrying the mouse b-globin upstream region (step 2). The plasmid was propagated in *E. coli* to produce a large quantity, isolated, and exposed to the mutagens nitrous acid. The result was random mutagenesis, sometimes of the mouse DNA, as desired, sometimes of the original plasmid DNA (step 3). Mouse DNA was cut out of the plasmid, separated from unmutagenized DNA, and returned to the mouse b-globin gene, using the same restriction enzymes as before (step 4). (**B**) Expression of gene with mutant upstream region. The mutagenized DNA was introduced into cultured human cells (HeLa cells) (step 5) and incubated to allow transcription (step 6). The cells were broken and RNA was isolated from the suspension (step 7). (**C**) Quantitation of b-globin RNA. A DNA fragment from the mouse b-globin gene was cut out with restriction enzymes and made radioactive (step 8). This radioactive probe was mixed with RNA from the transfected HeLa cells (step 9) and incubated with an enzyme, S1 nuclease, that destroys single stranded RNA and DNA. The only surviving nucleic acid was the 5' end of b-globin message hybridized to the DNA probe (step 10). The DNA/RNA mixture was run out on a gel, blotted, and exposed to X-ray film to visualize the position and quantity of the message hybridizing to the radioactive probe. (from Myers et al. (1986) Science 232:613-618)

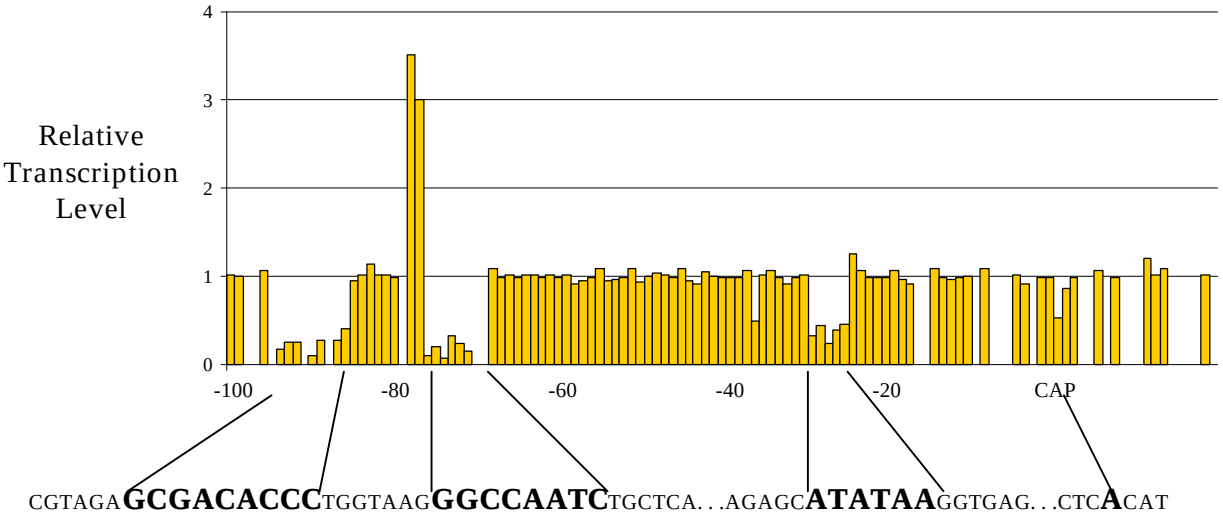

Figure 2. Level of transcription from mouse b-globin gene preceded by mutant 5' region. The numbers under the X-axis represent the distance in base pairs from the transcriptional start site (CAP). The relative transcription level is the amount of b-globin mRNA transcribed in HeLa cells carrying the mutant construct, relative to the amount transcribed in cells carrying the wild type gene (this ratio was normalized to the expression of reference genes to account for variations that might occur from experiment to experiment). No bars are given for the several sites for which mutations were not found. An apparent relative transcription level of zero does NOT imply that the site is important, just that the site was not tested. The sequence of part of the region 5' to the mouse b-globin gene is shown.

like you recently cloned some part of pUR3. This was necessary in order to focus the effect of the mutagen.

It would be wonderful to be able to simply count how many b-globin mRNAs are produced with different constructs, unfortunately, HeLa cells are in business for themselves. They, like all other cells, need to transcribe the genes required to stay alive. For that reason, b-globin message comprises only a small fraction of total mRNA, which, of course, is a small fraction of total RNA. To identify specifically b-globin message, Myers et al used a radioactive DNA probe derived from the cloned b-globin gene (Fig. 1C). Since the probe overlapped the transcriptional start site, part of it was complementary to b-globin mRNA, but the probe is very unlikely to be complementary to any other RNA. This characteristic was used to pick out b-globin mRNA from the crowd. All RNA not basepaired with the radioactive probe was destroyed with an enzyme that chews up single stranded nucleic acids. All that remained, therefore, was the probe paired with the 5' end of b-globin mRNA.

To quantitate the amount of b-globin message, the probe-mRNA hybrid was run out on a gel, which was then blotted onto a piece of filter paper. The paper was placed in contact with X-ray film, and since only the band with the hybrid should have been radioactive, only one region per lane should expose the X-ray film: the more b-globin

mRNA, the more probe-mRNA hybrid, the more radioactivity in the region, the more blackness on the film, which can be quantitated.

The procedure was repeated with all the mutants in the upstream region, leading to the results shown in Fig. 2. Since the position of the mutations were determined by sequencing, it was possible to associate mutations in specific positions with their effects on transcription of the b-globin gene.

SQ3.From Fig. 2, would you say that most mutations within 100 bases of the transcriptional start site adversely affect transcription?

SQ4. If this were a bacterial gene (say the lac operon), where would be the most detrimental mutations?

SQ5. Where were the most detrimental mutations in the case of the b-globin upstream region?

II.B. Role of positive-acting regulators on transcription

If you grew up on the *lac* operon, as did everyone around when the Myers et al experiment was conducted, then the results were truly remarkable. Bacterial genes are immediately preceded by promoters, binding sites for RNA polymerase. If the promoters are destroyed, then the rate of transcription of the associated gene plummets. Take a look at the region immediately upstream of the b-globin gene… nothing really important! You have to go to 26 bp upstream of the transcriptional start site before mutations have any significant effect on b-globin expression.

The important sites are much further upstream, centered around 75- and 93-bp before the start of transcription. Is it possible that RNA polymerase binds so far away from the transcriptional start site and still works? As it turns out, RNA polymerase does NOT bind to these two sites. Instead, what binds there are positive-acting regulators. The binding sequences, called GC boxes for the sequence at -95 beginning GC… and CAAT boxes for the sequence at -77, are found preceding a large number of genes. In fact genes usually have a varied collection of binding sequences preceding the transcriptional start site, far more than the lone CRP binding site sported by *lac*. The positions of these binding sites are not critical, except in one case. The binding site near the transcriptional start site, called the TATA box is always about 26 bp from the start site, since it determines where transcription will begin.

These positive acting regulators are cut in the same mold as CRP, however. You will recall that RNA polymerase binds only weakly to the *lac* promoter and that CRP stabilizes the interaction. Eukaryotes take this idea to the extreme, having the binding of RNA polymerase totally dependent on positive acting regulators. The complexity of the interaction is hinted at by Fig. 16-13 in the Brooker. The transcriptional factor TFIID ("TF" for transcriptional factor) binds to the TATA box and interacts with positive acting regulators bound far away. This is possible only because DNA can bend. Only when the

collected positive acting factors have made an inviting docking place does RNA polymerase join in, binding to the proteins, not to the DNA.

SQ6. Why is it that the positioning of CAAT boxes is unimportant but TATA boxes are invariably about 26 bp from the transcriptional start site?

SQ7. Given that in eukaryotes, unlike prokaryotes, transcriptional regulation is centered around the binding of RNA polymerase to protein not DNA, speculate on how a eukaryotic protein might serve as a repressor of transcription.

II.C. RNA processing

Please look over Brooker pp.349-350 and pp.357-358

We will discuss the following RNA modifications:

Capping: The 5' end of RNA is altered by adding a modified guanine nucleotide. This unusual base serves as the ribosome binding site, which may be quite distant from the start codon. Translation begins at the first ATG from the 5' end of the message.

Poly-A Tailing: The 3' end of RNA is altered by adding a string of adenine nucleotides, sometimes hundreds. This poly-A tail protects the mRNA from degradation in the cytoplasm. The polyadenylation is not encoded by the gene but instead added after transcription.

Splicing: Genes on RNA transcripts may be interrupted by noncoding regions, which must be removed before translation. Note that the DNA is not affected; it is the RNA transcribed from the gene that is altered. The interruptions are called introns and the pieces that remain are called exons (Fig. 12-7, p.325). The introns are spliced out of a transcript coding, and the exons are knitted together to make the mature message, generally much smaller than the original message.

The end result is that mature mRNA in the cytoplasm may be quite different from the RNA that is initially transcribed in the nucleus.

SQ8. Could eukaryote-type ribosome binding sites (^{7}methyl-G caps) work with the lac operon? Suppose that ribosomes bound to lac mRNA at a ^{7}Me-G cap. What problem would there be in the expression of the lac operon?

SQ9. In bacteria, mRNA is translated even before it is finished getting transcribed (see Fig. 14-16, p.388). Why is this not possible in eukaryotes?

III. Comparison between prokaryotes and eukaryotes

RNA polymerase: Prokaryotic RNA polymerase is modular, with the information on how to recognize a promoter residing within the sigma factors. Eukaryotic RNA polymerase is fixed. Actually, there are three distinct eukaryotic RNA polymerases, each with different purposes as discussed in your text, but they're all fixed -- no sigma factors. As a result, bacteria can alter RNA polymerase to turn on different

classes of genes in response to environmental stimuli, but eukaryotes don't do that: they rely on postivie-acting regulators.

Promoters: In prokaryotes, the promoter sequence is critical in deciding the degree to which a gene will be transcribed: good sequence, good promoter, lots of transcription, lots of protein. In eukaryotes, the promoter sequence itself is of minor importance.

Regulation: Transcriptional regulation in prokaryotes is achieved through a variety of strategies: repressors, activators, different sigma factors. Most of the action takes place within several basepairs of the promoter. In eukaryotes, there is a reliance on a single strategy and multitudinous variations on it: positive-acting regulators acting through their binding sites, often enhancers that act at a considerable distance.

RNA processing: Bacteria do very little of it, eukaryotes do lots: capping, tailing, splicing.

Translational initiation: Bacteria use a specific sequence close to the start codon to bind the ribosome, while eukaryotes take the first AUG that appears in the message, using the ^{7}methyl G cap to bind the ribosome.

How do we make sense out of these differences? Let's generalize: bacteria use nearby sequences for regulation while eukaryotes rely on signals that may be distant from the site of action.

SQ10. Flesh out that generalization with specific examples.

What purpose is served by distant regulation? We can get at the answer to this question, I think, by observing that compared to the diversity within bacteria, all higher eukaryotes are basically the same (Fig. 3). Bacterial genomes cover a huge range of diversity: from about 20% to 80% G+C content. In contrast, higher eukaryotes cover a much smaller range, and the range from fish to elephants is smaller still. The wide diversity of base composition in bacteria is matched by the diversity of what prokaryotes can do: thrive in boiling water, eat nitrogen gas, and so forth.

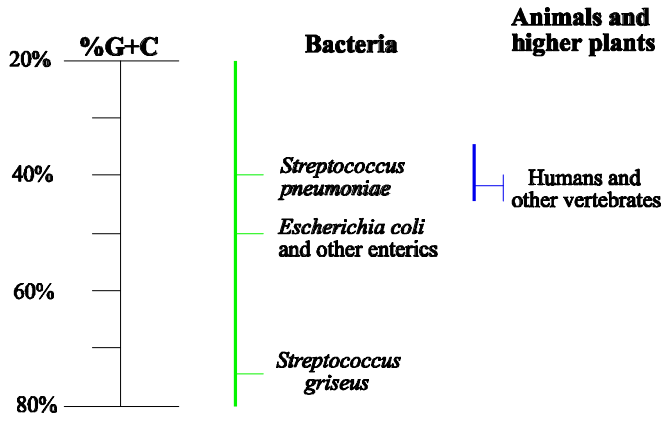

Fig. 3: Diversity amongst living things. The base composition of DNA within bacteria or higher plants and animals was used as a measure of diversity. The thick line represents the range of values within the group. Specific examples are shown.

Eukaryotes are much less diverse,… except in morphology. We eukaryotes all have similar tools, but we are remarkable in the amazing new tricks we come up with them. I think at root, the explanation for the diversity of form amidst the uniformity of sequence amongst eukaryotes is the peculiar

eukaryotic gene structure and strategy of regulation. Prokaryotes must change by basepair mutation, base changes that make the promoter stronger or weaker or that change the function of a protein. This is a very slow process. Bacteria are slow, patient inventors. Eukaryotes, on the other hand, are able to piece together novel combinations of gene components. An regulatory region enhancer from one gene can readily be moved upstream, maybe far upstream, from another as an evolutionary experiment. Exons can be shuffled around, because there is no restriction on how far away one exon need be from another. Eukaryotes are impetuous tinkerers. Suppose you're a eukaryote that needs a new secreted DNA binding protein (who knows why?): grab a leader sequence from an immunoglobulin, a DNA-binding domain from a regulatory protein, other motifs from other proteins, etc. The odds of these different regions of DNA coming together into one contiguous sequence are prohibitive, but if the need merely to get close together, separated by a variable length of garbage intron DNA, the odds are much higher.

Eukaryotes, as opposed to prokaryotes, have evidently opted for the fast track. By permitting coding sequences to be split up, the rate of evolution of new proteins is increased. Likewise, by permitting regulation at a distance by modular enhancers, the rate of evolution of new types of regulation is increased. This may explain why at the DNA level eukaryotes are so homogeneous. Eukaryotes can generate diversity by shuffling around what they already have.

SQ11. Draw a picture illustrating why it might have been easier for the b-globin regulatory region to evolve than the regulatory region of the lac operon.