19.E: Transcriptional regulation in eukaryotes (Exercises)

19.1 (POB) Specific DNA binding by regulatory proteins.

A typical prokaryotic repressor protein discriminates between its specific DNA binding site (operator) and nonspecific DNA by a factor of $10^5$ to $10^6$. About ten molecules of the repressor per cell are sufficient to ensure a high level of repression. Assume that a very similar repressor existed in a human cell and had a similar specificity for its binding site. How many copies of the repressor would be required per cell to elicit a level of repression similar to that seen in the prokaryotic cell? (Hint: The *E. coli* genome contains about 4.7 million base pairs and the human haploid genome contains about 2.4 billion base pairs).

Use the following information for the next 3 problems. Let's imagine that part of the regulation of expression of the OB gene is mediated by a protein we will call OBF1. There is one binding site for OBF1 in the OB gene, and let's assume that is the only specific binding site in the haploid genome, or 2 specific sites in a diploid genome. The haploid human genome has about $3 \times 10^9$ bp, or $6 \times 10^9$ bp in a diploid genome. If we assume that about 33.3% of the nuclear DNA is in an accessible chromatin conformation, that means that about $2 \times 10^9$ bp of DNA are available to bind OBF1 nonspecifically.

19.2 The diameter of a mammalian nucleus is about 10 mm. If you model a nucleus as a sphere, what is its volume? What is the molar concentration of specific and nonspecific binding sites in the nucleus?

Binding of OBF1 to a specific site and to nonspecific sites is described by the following equations.

Let $P = \text{OBF1}$

$D_s = \text{a specific binding site in DNA}$
Dns = a nonspecific binding site in the genomic DNA

P + Ds

Ks = = 1011 M⁻¹ (eqn 2)

Knns = = 105 M⁻¹ (eqn 3)

19.3 What fraction of the OBF1 (or P in the equations) is not bound to either specific or nonspecific sites in the DNA?

19.4 How many molecules of OBF1 are needed per nucleus to maintain 90% occupancy of the specific sites? This condition means

\[ = 9 \]

Use the following information for the next seven questions.

The *agouti* gene in mice controls the amount and distribution of pigments within coat hairs. Some mutations of this gene also lead to adult-onset obesity, a mild diabetes-like syndrome, tumor susceptibility and recessive embryonic lethality. The gene encodes a predicted protein of 131 amino acids that has the structural features of a secreted protein, but no striking homology to other known proteins has been recognized. This protein is likely to be a regulator of melanin pigment synthesis, and it may also be a more general metabolic regulator.

Let's suppose that you are investigating the regulation of the *agouti* gene, and have the capacity to transfect a melanocyte cell line, which transcribes the wild-type *agouti* gene, and an adipocyte cell line, which transcribes the wild-type *agouti* gene only at a very low level. Further, you already know that the basal promoter is in a DNA segment located between -100 and +50. You make progressive 5’ deletions of a fragment that includes -300 to +50, link it to a luciferase reporter gene, and transfect the constructs into melanocyte and adipocyte cells, with the following results.

19.5 What do you conclude about the region between -250 and -200?

19.6 What do you conclude about the region between -200 and -150?

19.7 What do you conclude about the region between -150 and -100?

You also investigate the binding of nuclear proteins to these DNA segments located upstream of the *agouti* gene. Extracts containing nuclear proteins from melanocytes were tested for the ability to bind to the fragments delineated...
in the deletion series above.

The fragment from -150 to -100 was used as the labeled probe in a mobility shift assay. The mobility of the free probe is shown in lane 1, and the pattern after binding to melanocyte nuclear extract is shown in lane 2. Lanes 3-14 show the mobility shifts after addition of the competitors to the binding reaction; the triangle above the lanes indicates that an increasing amount of competitor is used in successive lanes. "Self" is the same -150 to -100 fragment that is used as a probe, but it is unlabeled and present in an excess over the labeled probe (lanes 3-5). A completely different DNA (sheared E. coli DNA) was used as a nonspecific competitor (lanes 6-8). Two different duplex oligonucleotides, one containing the binding site for AP1 (lanes 9-11) and the other containing the binding site for Sp1 (lanes 12-14) were also tested. Thinner, less densely filled boxes denote bands of less intensity than the darker, thicker bands. Use these results to answer the next two questions.

**19.8.** What do you conclude from these data?

**19.9.** What sequence within the -150 to -100 segment might you expect to be bound in melanocyte nuclei?

**19.10.** The fragment from -200 to -150 was also used as a labeled probe in a mobility shift assay similar to that described for the -150 to -100 segment, as shown below.

What do you conclude from these data?

**19.11.** Some mutant alleles of the *agouti* gene are expressed ectopically (i.e. in the wrong tissue). Just using the information on the 5' deletions above, what region is a likely candidate for the position of a loss-of-function mutation that leads to ectopic expression in adipose tissue?

**19.12** (POB) Functional domains in regulatory proteins.

A biochemist replaces the DNA-binding domain of the yeast GAL4 protein with the DNA-binding domain from the lambda repressor (CI) and finds that the engineered protein no longer functions as a transcriptional activator (it no longer regulates transcription of the GAL operon in yeast). What might be done to the GAL4 binding site in...
19.13 What is the DNA-binding domain of the transcription factor Sp1?

19.14 What is the dimerization domain of the transcription factor AP1?

19.15 (ASC) Describe three mechanisms for regulating the activity of transcription factors.

19.16 (ASC) You have constructed a plasmid set containing a series of nucleotide insertions spaced along the length of the glucocorticoid-receptor gene. Each insertion encodes three or four amino acids. The map positions of the various insertions in the coding sequence of the receptor gene is as follows:

| Insertion: | A | B | C | D | E | F | G | H | I | J | K | L | M | N | O | P | Q | R | S |
| 0 Glucocorticoid-receptor coding sequence 783 |

The plasmids containing the receptor gene can be functionally expressed in CV-1 and COS cells, which contain a steroid-responsive gene. Using these cells, you determine the effect of each of these insertions in the receptor on the induction of the steroid-responsive gene and on binding of the synthetic steroid dexamethasone. The results of these analyses are summarized in the table below.

<table>
<thead>
<tr>
<th>Insertion</th>
<th>Induction</th>
<th>Dexamethasone binding</th>
</tr>
</thead>
<tbody>
<tr>
<td>A</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td>B</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td>C</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td>D</td>
<td>0</td>
<td>++++</td>
</tr>
<tr>
<td>E</td>
<td>0</td>
<td>++++</td>
</tr>
<tr>
<td>F</td>
<td>0</td>
<td>++++</td>
</tr>
<tr>
<td>G</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td>H</td>
<td>++++</td>
<td>++++</td>
</tr>
<tr>
<td>I</td>
<td>+</td>
<td>++++</td>
</tr>
<tr>
<td>J</td>
<td>++++</td>
<td>++++</td>
</tr>
</tbody>
</table>

The DNA to make the engineered protein functional in activating GALoperon transcription?
a) From this analysis, how many different functional domains does the glucocorticoid receptor have? Indicate the position of these domains relative to the insertion map.

b) Which domain is the steroid-binding domain?

c) How could you determine which of the domains is the DNA-binding domain?