

**Question 1.** Your lab is studying the MRX complex in *S. pombe*, otherwise known as fission yeast. For your dissertation work, you are studying the function of Rad50, one of the subunits of this complex. RecBCD, an enzyme in *E. coli*, plays a similar role in homologous recombination to the MRX complex.

**1A)** What is different about the resulting resection of ends by these two complexes?

**RecBCD degrades from 5' and 3' ends. MRX only degrades 5' ends.**

When you first joined the lab, you did a genetic screen to isolate temperature-sensitive mutants with defective spore viability. You determined that one of these mutants has a specific missense mutation in the gene encoding Rad50, and you call the mutant *rad50-ts*.

*rad50-ts* mutant spore viability is 0.1% of wild type levels when measured at high temperature. *rec12*<sup>-</sup>, a mutant in the *S. pombe* Spo11 homolog, is also defective in spore viability. *rec12*<sup>-</sup> mutant spore viability is 20% of wild type levels. You make a double mutant strain containing both of these mutations and find that spore viability of *rad50-ts* increases 100-fold.

**1B)** How can you explain this result?

**Rad50 is required to repair/process Spo11-induced DSBs in meiotic recombination. Without Rad50, spores are dead because DSBs are never repaired. In double mutant cells, DSBs are not formed so Rad50 is no longer required, thus increasing the likelihood of a viable spore.**

In further characterization of the *rad50-ts* phenotype, you find that these cells are defective in the removal of the Spo11 protein from DSB ends.

Because you know that homologous recombination is involved in some kinds of DNA repair and the DNA damage response pathway, you test the sensitivity of *rad50-ts* vegetative cells (mitotically dividing) to several different mutagens at the restrictive temperature:

Mutagen	Percentage viability (compared to WT)
Camptothecin*	47%
Gamma-irradiation	95%
Ethyl methane sulfonate (EMS)	98%

\* CPT (Camptothecin) is a topoisomerase I inhibitor. Single-strand breaks persist because CPT prevents completion of the topoI catalyzed reaction.

**1C)** Why does the CPT-sensitivity make sense given what you know about the *rad50-ts* phenotype? What do you hypothesize is a crucial function of Rad50 *in vivo*?

***rad50-ts* cells are unable to remove Spo11 from meiotic DSBs. The DSBs are not processed and lead to inviable spores. In WT CPT-treated cells, topoI remains linked to 3' ends and single stranded breaks persist. Rad50 is recruited to this site of damage and helps remove topoI. In CPT-treated *rad50-ts* cells, however, damage cannot be repaired and viability of cells decreases. One crucial function of Rad50 that is disrupted in the TS mutant is cleavage of Spo11 off of 3' ends to allow for further processing of meiotic DSBs. In addition, Rad50 may be involved in a similar process in DNA damage repair.**

**1D) You might expect that Rad50 mutant cells would be sensitive to gamma irradiation. Why? How can you explain the results shown above for *rad50-ts*?**

**In vegetative cells, DSBs are caused randomly by damaging agents. In meiosis, DSBs are created by the Spo11 protein, which becomes covalently joined to the DNA during cleavage. A good model to explain the results is that the *rad50-ts* mutant protein is only defective in cleaving off Spo11. Repair of DSBs caused by gamma irradiation does not require this cleavage, so this repair process is functional in the mutant.**

**1E) You want to test the activity of purified Rad50 and mutant *rad50-ts* protein to determine whether you are correct about the function that is disrupted in the mutant. You have access to labeled linear dsDNA containing an *S. pombe* meiotic hotspot, purified Spo11/Rec12, purified Rad50 or *rad50-ts*. Describe the assay you would perform.**

**Incubate the ds DNA substrate with Spo11 and Rad50 or *rad50-ts* or a no protein control at the restrictive temperature. Stop the reactions and run on a denaturing gel. You should see the mobility of the band increase after incubation with Rad50 because Spo11 is cleaved off the DNA. With *rad50-ts* and the no protein control, Spo11 will remain bound to the substrate and cause it to run more slowly in the gel.**

**Question 2A.** A fundamental principle of genetic analysis is that physical distance on DNA and genetic distance (e.g. distance on a genetic map) are **generally proportional** over the entire genome. Although some regions exhibit unusually high or low recombination frequencies, all regions of the genome can and do participate in homologous recombination. Based on your knowledge of the mechanism of homologous recombination list two properties of this process that serve to ensure that all DNA sequences can participate in recombination. Explain your answers. (Hint: Think about both how homologous recombination is initiated as well as the mechanism of homologous recombination.)

Property 1: **Initiation of homologous recombination is sequence independent. Double-strand breaks can occur anywhere in the genome. (However, it is true that the Spo11 proteins that initiate double strand breaks during meiosis in eukaryotes do show some site specificity. This results in some hot and cold regions for homologous recombination. )**

Property 2: **The components that initially act to repair a double strand break are structure specific and sequence independent. RecBCD will load on any flush dsDNA end and RecA will load on a single stranded region. (The MRX complex degrades the DNA strand that is covalently attached to the Spo11 protein.)**

**Note: The overrepresentation of Chi sites is not an acceptable answer in this case because recombination frequency still varies as you move away from a Chi site.**

**Read the next experimental section, and use these data to answer parts b-e below**

Although the frequency of recombination is generally quite uniform throughout the chromosome, detailed analysis of specific chromosomal regions reveals that some regions show a higher-than-average recombination frequency, whereas other regions show a lower frequency. You decide to study this phenomenon in *E. coli* to try to understand the molecular mechanisms responsible for these “hot” and “cold” regions.

The experimental set-up is as follows. You determine the frequency of homologous recombination within 8 regions of the chromosome by generalized phage-mediated transduction. The recipient strain has multiple auxotrophic markers, meaning it has a defective gene in each of the 8 regions. The donor strain has wild-type alleles for each of these genes.

Recipient: *his<sup>-</sup>, trp<sup>-</sup>, lac<sup>-</sup>, ara<sup>-</sup>, val<sup>-</sup>, leu<sup>-</sup>, thi<sup>-</sup>, ura<sup>-</sup>*  
Donor: *his<sup>+</sup>, trp<sup>+</sup>, lac<sup>+</sup>, ara<sup>+</sup>, val<sup>+</sup>, leu<sup>+</sup>, thi<sup>+</sup>, ura<sup>+</sup>*

The transducing phage is grown on the donor cells, and the phage packages, at random, 50kb segments of the donor cells chromosome. The phage are then used to infect the recipient cells, thereby introducing segments of the donor chromosomes. If recombination between the incoming DNA and the recipient host chromosome is successful, than the recipient cells may gain a wild-type allele.

Below is a table of recombination frequencies observed in each of the marked chromosomal locations.

**Table 1: Frequency of Wild-type Recombinants:**

Region 1	His+	0.2 %
Region 2	Trp+	0.2 %
Region 3	Lac+	0.02 %
Region 4	Ara+	0.2 %

Region 5	Val+	0.02 %
Region 6	Leu+	0.6 %
Region 7	Thi+	0.2 %
Region 8	Ura+	0.2 %

To explore the mechanisms responsible for the higher- and lower-than-average recombination frequencies, you decide to study regions 1, 3, 5, and 6 in more detail. To do this, you repeat the same type of *in vivo* recombination assay, except this time the recipient strain carries a mutation in one of the following recombination genes: *recB*, *recD*, or *ruvC*. Below is a table of these results; in each case the frequency of wild-type recombinants is shown (e.g. His<sup>+</sup> for region 1) as in table 1.

**Table 2: Frequency of wild-type recombinants:**

<b>Mutation in Recipient Cells</b>			
	<i>RecB</i>	<i>RecD</i>	<i>ruvC</i>
Region 1	<0.001%	0.6%	<0.001%
Region 3	<0.001%	0.12%	<0.001%
Region 5	<0.001%	0.6%	<0.001%
Region 6	<0.001%	0.6%	<0.001%

**2B.** Based on these findings, suggest a hypothesis for the mechanistic basis of the aberrant recombination frequency observed in table 1 for region 5. Explain the experimental support for your model.

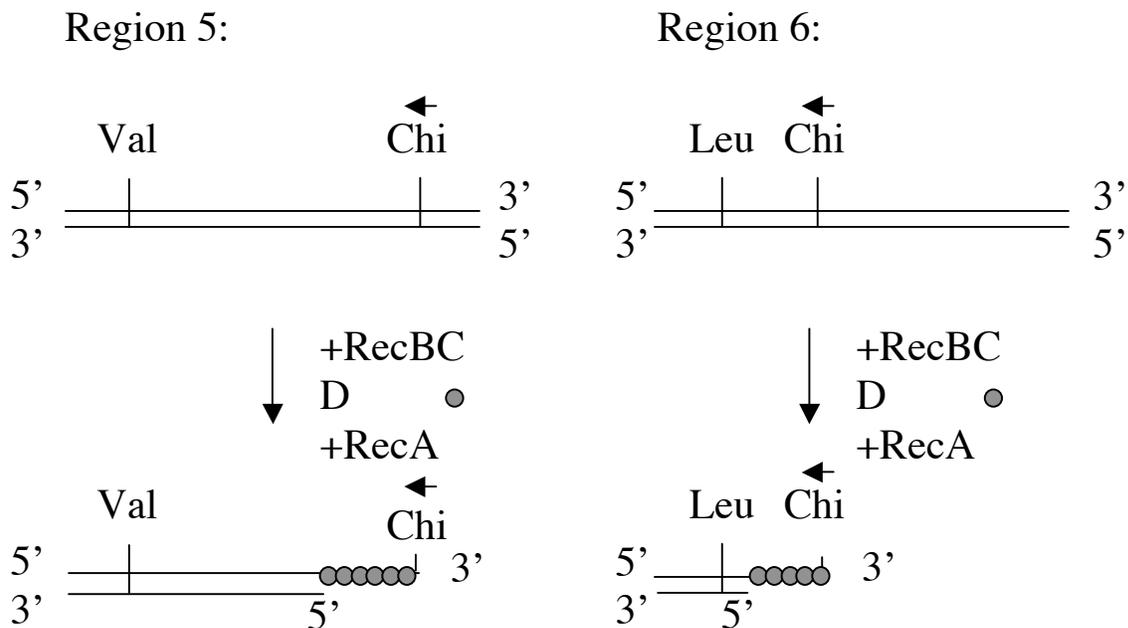
**Region 5 has a low number of Chi sites relative to the other regions, or is further from a Chi site than the other regions are from their closest Chi site. The experimental evidence for this is Table 2. In a *recD* mutant (which no longer responds to Chi sites), the recombination frequency elevates to that of the other regions. (Remember that *recD* mutants can be hyper-recombinogenic; this explains why recombination in Regions 1 and 3 are also elevated.)**

**2C.** Suggest a hypothesis for the mechanistic basis of the aberrant recombination frequency observed in table 1 for region 6. Explain the experimental support for your model.

**Region 6 has a high number of Chi sites relative to other regions, or is closer to a Chi site than the other regions are to their closest Chi site. The experimental evidence for this is Table 2. In a *recD* mutant, which no longer responds to Chi sites, the recombination frequency remains the same.**

**2D.** To explore the mechanism further you isolate a 15 kb fragment of DNA from region 5 and a 15 kb fragment from region 6 and incubate each of these DNA fragments *in vitro* with purified RecBCD enzyme. Diagram the **DNA substrates** (marking the position of the gene and any other important features) and the **DNA products** for each of these two fragments. **Your diagram should be approximately to scale; there is no need to draw**

a gel---just a line drawing of the DNA. You should include the position of bound RecA in your diagram of your final DNA products.



**2E.** Having completed your analysis of regions 5 and 6, you turn back to region 3. At a loss for an explanation for the poor recombination seen in region 3, you talk to your labmates to see if they have any good ideas. One smart person suggests that you examine the *E. coli* genomic sequence immediately surrounding the *lac*<sup>-</sup> mutation on the recipient chromosome. Given that the sequence is readily available, you decide this is an excellent idea. What sequence motif does your friend suggest you look for? What distribution of this motif could provide an explanation for your recombination frequency data?

**You would search for the RuvC consensus sequence because these sites will prevent branch migration from moving through the *lac* coding region and inhibit recombination of that region. There should be RuvC consensus sites immediately surrounding the *lac* coding region.**

**Note: The answer was not a Chi site consensus sequence because the recombination frequency of this region observed in a *recD* mutant was still low relative to other regions.**

**Question 3.** As part of your UROP project, you have identified two spontaneous mutants of your favorite yeast strain, *Saccharomyces spikus*, which have the following types of colony morphology in the haploid mutants.

1. One mutant has a bald colony phenotype instead of the normal hairy colony phenotype. (The “hairy” appearance is caused by hyphae projecting from the surface of the colony.)
2. The second mutant fails to produce the blue pigment observed in wild type colonies, so the colonies are white.

You are interested in further investigating the mutants you have identified. You have investigated the inheritance pattern of each single mutant through several different genetic crosses and have found that each mutant phenotype appears to segregate as a single genetic locus. You call these two genetic loci *HAIRY* and *BLUE*. You have decided to determine the chromosomal position of the two mutations relative to each other. Therefore, you have mated your bald mutant (of mating type 1) with your white mutant (of mating type 2.)

In *S. spikus*, like *S. cerevisiae* and other eukaryotes, DSBs are generated in the chromosomal DNA during meiosis. The MRX complex of *S. spikus* processes these DSBs to generate a substrate suitable for binding by the *S. spikus* RecA homologs, Rad51 and Dmc1. Unlike *E. coli* RecBCD, the *S. spikus* MRX complex only degrades DNA in the 5' to 3' direction. These DSBs are then repaired by recombination between non-sister homologous chromatids. Homologous recombination is required for proper segregation of chromosomes during meiosis.

In addition, *S. spikus*, like *S. cerevisiae* and some other fungi, forms spore sacs that contain all four spores generated from a single meiosis. These spore tetrads can be separated and the colonies formed by each spore can be analyzed. After your yeast mating, you separated the tetrads that were produced and analyzed morphologies of the colonies that grew from the spores. You obtained the following results:

<b>Tetrad Class</b>	<b>Types of tetrad classes observed:</b>	<b># of tetrad class observed:</b>
<b>A</b>	2 hairy, white colonies 2 bald, blue colonies	540
<b>B</b>	1 hairy, white colony 1 bald, blue colony 1 hairy, blue colony 1 bald, white colony	9
<b>C</b>	1 hairy, white colony 2 bald, blue colonies 1 bald, white colony	6
<b>D</b>	2 hairy, white colonies 1 bald, blue colony 1 hairy, blue colony	5

**3A.** Based on this data, do you think the genes encoding the proteins required for hyphae development and blue pigment production are found on the same chromosome? Why or why not?

The *HYPHAE* and *BLUE* genes are found on the same chromosome. If the two genes were on separate chromosomes, you would expect to observe independent assortment of the alleles of the *HYPHAE* and *BLUE* genes. This means that 50% of the time you would observe tetrads of the parental class (2 hairy, white colonies and 2 bald, blue colonies) and the other 50% of the time you would observe tetrads of the non-parental class (2 hairy, blue colonies and 2 bald, white colonies.) However, you observed that 96% of the tetrads you analyzed were of the parental class. This indicates that the two genes are on the same chromosome.

**3B.** What process or processes are responsible for generating tetrad groups B?

The tetrads in group B were generated by recombination between two of the four non-sister homologous chromosomes. This recombination occurred between the *HYPHAE* and *BLUE* genes and was resolved to generate crossover products. This is why you observe one hairy, blue colony and one bald, white colony as well as one hairy, white and one bald, blue colony.

**3C.** What process or processes are responsible for generating tetrad groups C and D?

The tetrads in groups C and D were generated by gene conversion of the *HAIRY* gene alleles. In group C, one of the hairy alleles was converted to a bald allele. In group D, one of the bald alleles was converted to a hairy allele.

Gene conversion can occur during DSB repair. If a DSB occurs near the *HAIRY* gene on one of the non-sister homologous chromatids, genetic information specifying the *HAIRY* allele can be lost during 5' resection. This lost information can be repaired by DNA synthesis using the undamaged non-sister homologous chromosome as a template during DSB repair. This non-sister homologous chromosome carries a different allele of the *HAIRY* gene, so synthesis results in the incorporation of sequence that encodes a different allele of the *HAIRY* gene. (See 3E for more information.)

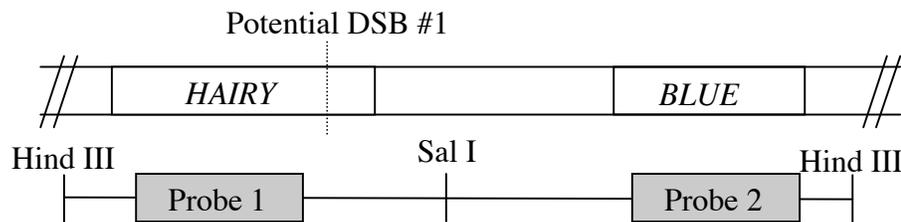
**3D.** Based on this data, can you predict whether DSBs are more likely to occur close to the *BLUE* or *HAIRY* gene? What is the evidence that supports this prediction?

A DSB is more likely to occur close to the *HAIRY* gene, as you observed gene conversion of the alleles of the *HAIRY* gene but did not observe conversion of the alleles of the *BLUE* gene. This indicates that a DSB did not occur in proximity to the *BLUE* gene at a high enough frequency for you to observe tetrads that resulted from gene conversion.

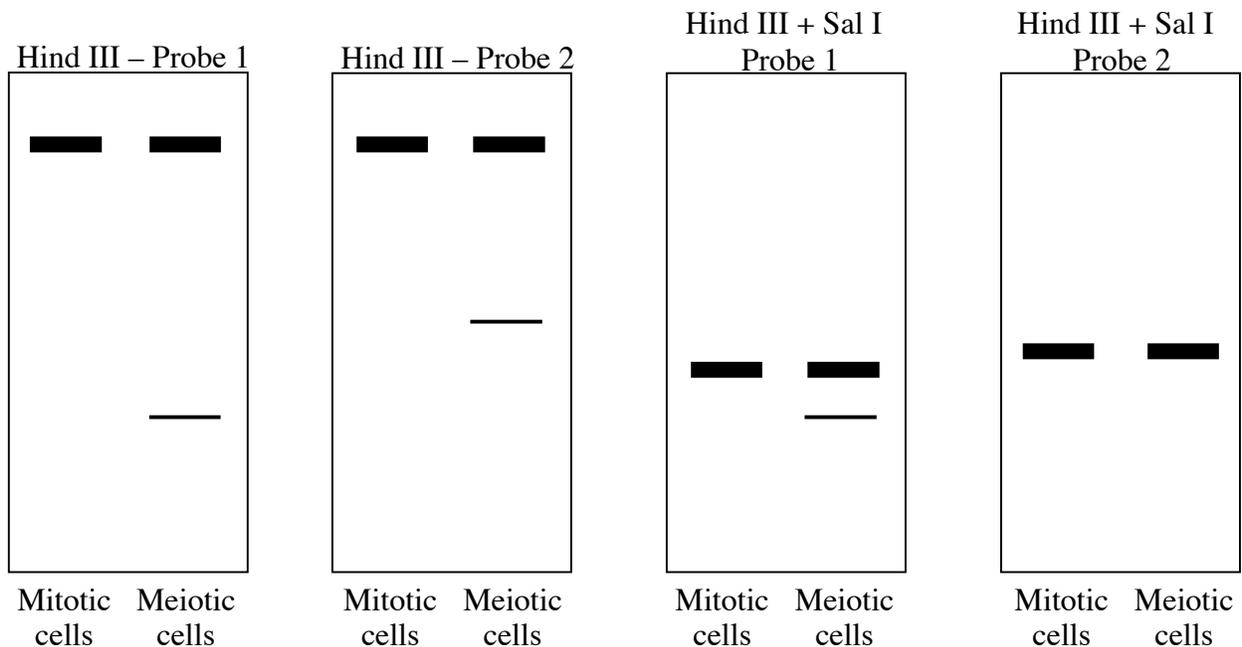
**3E.** How could you test your prediction? Describe an assay you could use to identify sites of DSBs.

You could look for the enrichment of DSBs using a genomic DNA Southern. Here is one potential way this assay could be done:

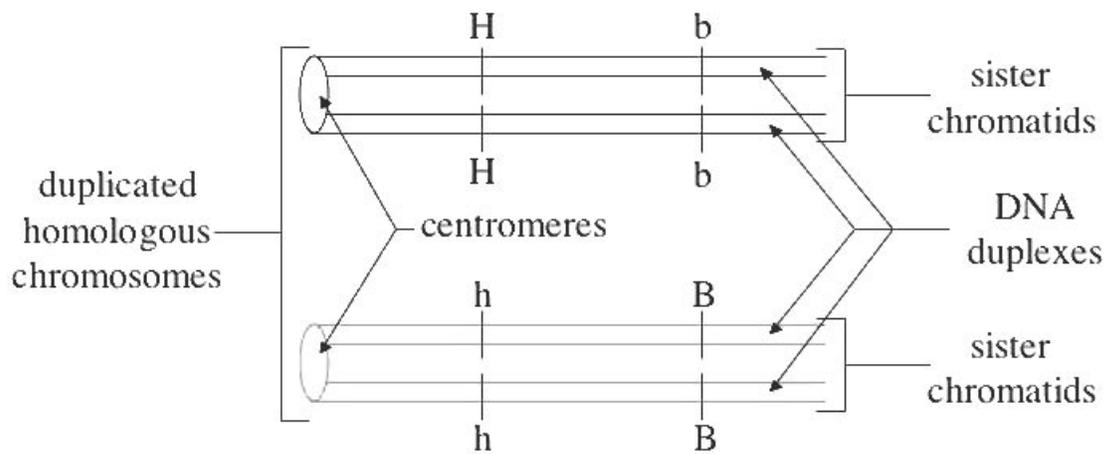
Hypothetical map of *HAIRY/BLUE* region of chromosome:



1. Isolate DNA from cells undergoing meiosis. Isolate DNA from mitotic cells to use as a control.
2. Perform different restriction enzyme digestions on this DNA (Hind III alone; HindIII + SalI based on the map above.)
3. Run the digested DNA on a non-denaturing agarose gel.
4. Transfer this DNA to a filter, denature, and probe with different labeled probes generated from the *HAIRY/BLUE* chromosomal region. (For example, Probe 1 and Probe 2 mapped above.)
5. If a DSB is preferentially formed near the *HAIRY* gene you would observe the following results on your blots.



**3F.** Diagram how class C tetrads were generated. (Remember that the *S. spikus* MRX complex only degrades DNA in the 5' to 3' direction to generate 3' tails for Dmc1/Rad51 binding.)



“H” represents the sequence of the wild type allele of the *HAIRY* gene

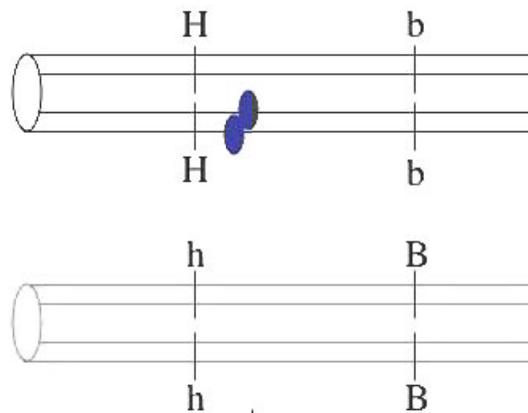
“h” represents the sequence of the mutant allele of the *HAIRY* gene

“B” represents the sequence of the wild type allele of the *BLUE* gene

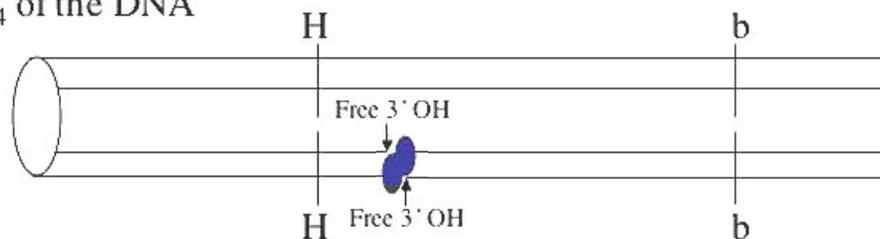
“b” represents the sequence of the mutant allele of the *BLUE* gene

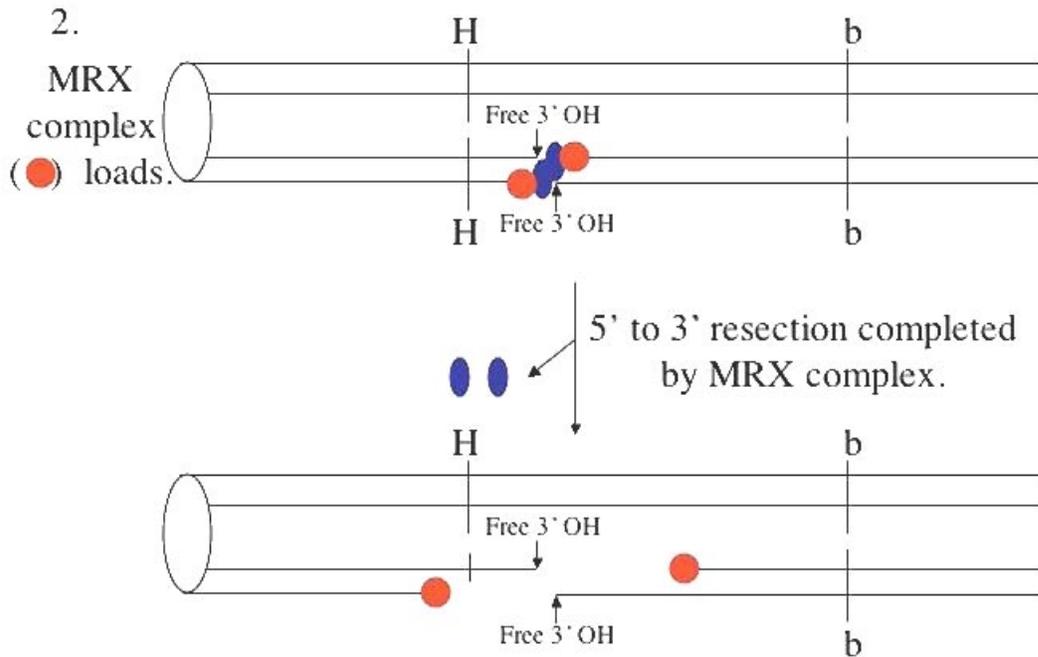
1.

During meiosis, the Spo11 homolog (●) binds to DNA and initiates a DSB

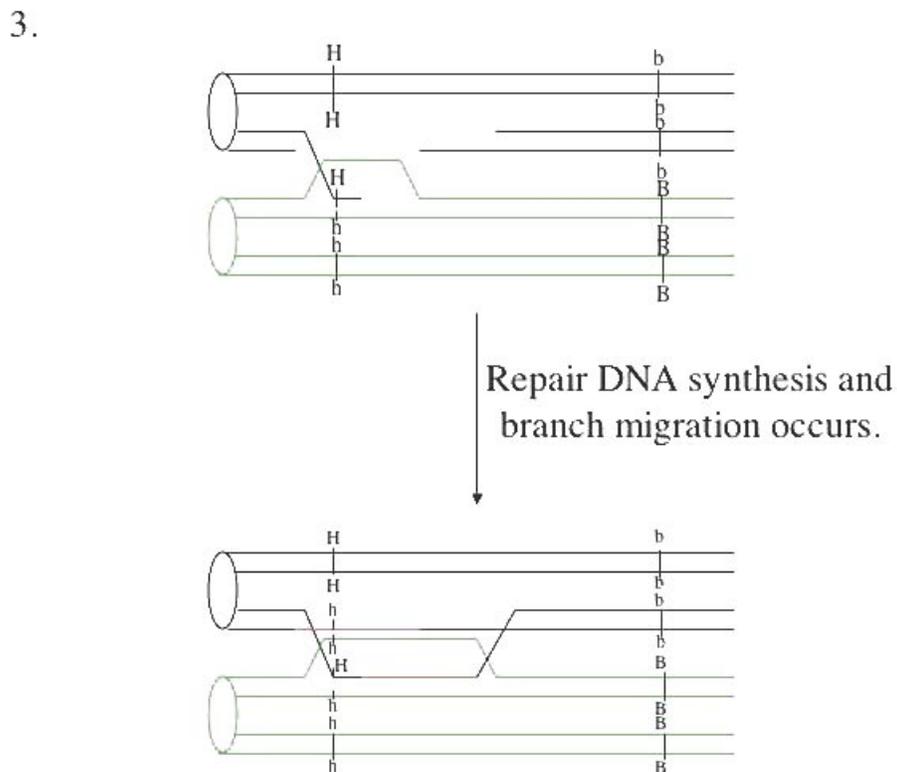


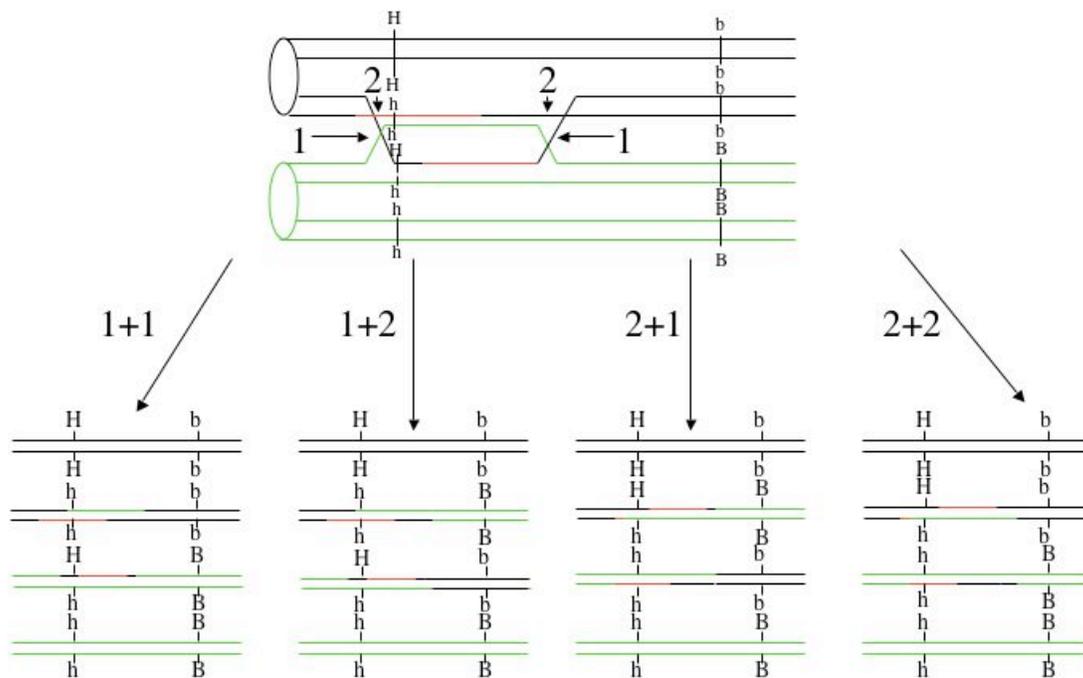
Spo11 is covalently attached to the 5' PO<sub>4</sub> of the DNA





- The sequence of the H allele is lost on one of the DNA strands.
- The Rad51/Dmc1 homologs coat the ssDNA and mediate strand invasion into the chromosome of a non-sister homologous chromatid.





- The double Holliday junction is resolved 1 of 4 ways. Chromosomes are segregated into spores.
- If the heteroduplex sequence is repaired to the sequence of the newly synthesized strand, gene conversion will occur. This generates Class 3 tetrads.

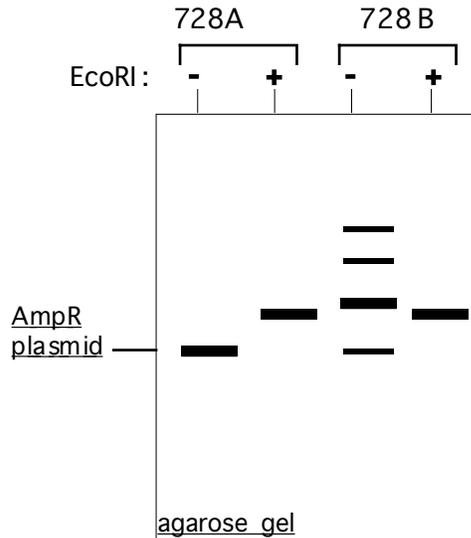
**3G.** How would your results be different if both the bald and blue mutants that you used in your mating were also mutated in the *S. spikus* MutS homologs.

As outlined in Figure 4F, gene conversion can occur as a result of mismatch repair. If your matings were done in *S. spikus* cells mutant in the MutS homologs, you would not observe mismatch repair. As a result, the heteroduplex DNA would not be repaired until it was replicated. As the cells began to replicate and divide, half of the progeny cells would receive the new allele found in the repaired DNA strand, and the other half would receive the old allele. This would generate a variegated colony that exhibits a half hairy/half bald phenotype.

**Question 4.** You are studying mechanisms by which bacteria acquire and maintain genes encoding antibiotic resistance. One strain you are working with (named 728A) is highly resistant to ampicillin; 100% of the cells plated on Amp-containing media form colonies.

Knowing that both plasmids and transposons often carry genes imparting drug resistance, you analyze strain 728A and find that it carries a plasmid. **To explore the mechanisms used by the cells to maintain this plasmid, you mutagenize the strain, and isolate a variant (called 728B) which loses the ampicillin resistance phenotype at high frequency.** When a liquid culture of 728B is plated on Amp-containing media, typically only 10-15% of the cells give rise to colonies.

To investigate why 728B shows reduced inheritance of the ampicillin resistance plasmid, you purify the plasmid DNA from cultures of 728A and 728B cells and analyze the DNA on a native agarose gel. Gel analysis is done both before and after digesting the DNA with EcoRI, which cleaves the plasmid DNA once.



**4A)** Based on this analysis, what is a reasonable explanation for the reduced efficiency of plating seen with the 728B cells on Amp-containing media?

**If plasmids undergo homologous recombination after replicating, they form multimers which must be separated by Site Specific Recombination. If they are not separated, then only one of the daughter cells will receive the plasmids containing the AmpR genes when the cells divide. The plasmids isolated from the 728B cells run as monomers, dimers, trimers and tetramers. When cut with EcorI, which cuts once per plasmid, they all resolve to a band of the same size. Therefore, this is a mutation in the recombinase or the sites it recognizes.**

**Note that although a replicative transposon would form a cointegrant that would be bigger than the original plasmid, it would not cut back to the same size, as the replicated transposon would add extra length.**

4B) Further analysis of the 728B cells reveals that the mutation causing poor plasmid inheritance is **on the plasmid DNA**. Furthermore, the gene encodes a **trans-acting factor**. Briefly describe one or two experiments that would support these conclusions.

**Although there are many possible experiments here, you must test plasmid inheritance, NOT look at transcription from a test promoter.**

**For example, transform 728B cells with a wild-type plasmid isolated from 728A with a different marker, such as KanR. If the mutation is on the plasmid (cis), this new plasmid will be maintained, so the cells will be KanR. If it encodes a trans-acting factor, the cells will also be AmpR.**

**OR:**

**To tell that the mutation is on the plasmid, you could transform the mutant plasmid isolated from 728B into a wild-type strain, and it should then be poorly maintained. You would then have to come up with another experiment to prove that it's a trans-acting factor.**

4C) Given the results in parts a and b, you come up with a hypothesis for the type of protein that may be encoded by the gene which is mutated in the 728B cells. You decide to try to isolate the wild-type protein using a biochemical approach.

Describe:

- (1) The assay you would use to follow the purification.
- (2) What cells you would use as starting material; and
- (3) If there are any small molecule cofactors you would be likely to need for your assay to be functional.

**Assay: Incubate plasmid DNA from 728B with fractionated cell extracts and see if multimers(bigger bands) go away. Note that biochemical complementation is not in itself an assay. You still need to indicate what assay that mutant cell extract can't do that the wild-type protein will then complement.**

**Starting material: Fractionate 728A cell extracts to get the wild-type protein.**

**Cofactors: If you were looking for a recombinase, no cofactors were needed.**

**4D)** While continuing with your protein purification, you also return to using genetics to understand the mechanism of plasmid maintenance. Therefore, you decide to see if you can isolate a mutation that suppresses the phenotype of the 728B strain. To do this, you start with a culture of 728B cells, mutagenize these cells, and screen for a new mutant that restores the plating efficiency on Amp-containing media to nearly 100%. You succeed in finding a strain with this phenotype, and name this strain 728C. This time the new mutation responsible for suppression maps to the cellular chromosome rather than the plasmid.

Suggest one chromosomal gene that might be mutated in the 728C cells and would cause this suppression. Explain your answer.

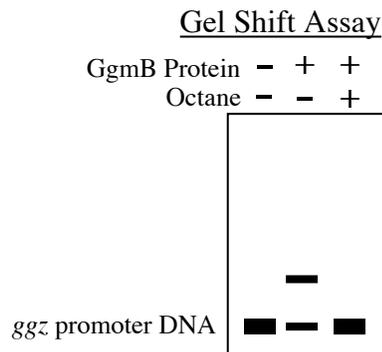
**A gene involved in homologous recombination, such as RecA/B/C/D. The plasmid multimerization was caused by homologous recombination between the plasmids.**

What is another phenotype you would expect the 728C cells to have if your hypothesis is correct?

**728C cells would also be sensitive to agents that cause double strand breaks, such as x-rays.**

**Question 5.** You are studying the regulation of the *gasguzzler* (*ggz*) gene in a novel bacteria that is able to use Octane as a carbon source. *ggz* is expressed at basal levels in the absence of octane and is induced by addition of octane. You have isolated a mutant, which you call *ggmB*, which results in constitutive expression of the *ggz* operon. Based on genetic evidence, you think that the *ggmB* encodes a transcriptional repressor.

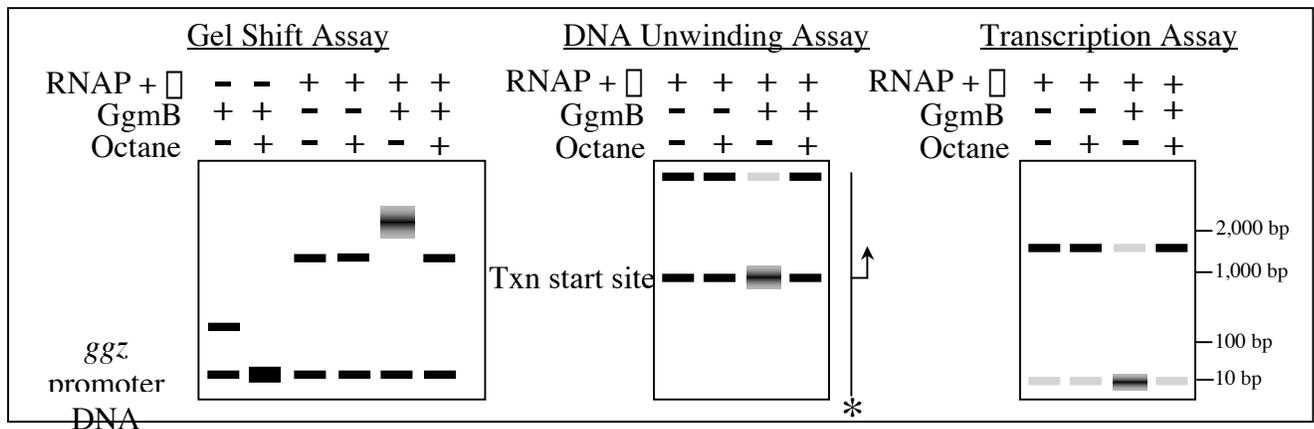
To further study how the GgmB protein represses transcription in the absence of octane, you have purified the GgmB protein. You use this purified protein to measure binding to the *ggz* promoter region *in vitro* using gel mobility shift assays. You use radiolabeled dsDNA that contains ~200 basepairs upstream of the *ggz* gene as a substrate for DNA binding. The results of your assay are shown below:



5A. Based on these results, how do you think octane regulates GgmB activity?

**Octane causes GgmB to dissociate from the promoter DNA. Since addition of octane activates gene expression, dissociation of GgmB must relieve repression.**

You want to determine how GgmB represses transcription of *gasguzzler*. To this end you perform assays to look at RNAP binding, promoter unwinding, and transcription product length in an *in vitro* assay that shows octane-dependent expression of the *ggz* promoter. Your results are shown below.



5B . Based on this data, describe a model for the mechanism that GgmB uses to regulate *ggz* promoter function. Include in the presence of GgmB and the absence of octane, a larger, more intense band is observed than with RNA pol. holoenzyme alone or RNA pol. holoenzyme + GgmB + octane. The larger band indicates that both GgmB and RNA pol. holoenzyme are bound to the promoter. The more intense band indicates that more of the promoter DNA molecules are occupied by RNA pol. holoenzyme + GgmB. This shows that GgmB stimulates closed complex formation.

-The DNA unwinding assay shows that in the absence of octane, GgmB also stimulates open complex formation by RNA pol. holoenzyme so that you observe unwinding on more DNA molecules. This is evidenced by the more intense band around the transcription start site and the decreased full length band at the top of the gel. (This gel is a lower resolution gel than the gels that were in the paper you read, this is why you don't see distinct bands at the site of promoter unwinding.)

-The transcription assay shows that in the absence of octane, GgmB also stimulates transcription of the *ggz* promoter. This is evidenced by the more intense band around the transcription start site and the decreased full length band at the top of the gel. (This gel is a lower resolution gel than the gels that were in the paper you read, this is why you don't see distinct bands at the site of promoter unwinding.)

This is also why you see the full length primer extension product at the top of the gel.)

-The transcription assay shows that in the absence of octane, GgmB reduces the amount of full length transcript formed, although the amount of abortive transcript is increased. This indicates that the RNA pol. holoenzyme is defective in promoter clearance. This result would be observed if GgmB stabilized the unstable ternary complex.

-Together, all of these results fit the following model. GgmB (in the absence of octane) strongly interacts with RNA pol. holoenzyme. This interaction stimulates closed and open complex formation by RNA pol. holoenzyme. This interaction also stabilizes the unstable ternary complex. Therefore, in the absence of octane the RNA pol holoenzyme is bound to the promoter and unable to transcribe the *ggz* gene in the absence of octane, but is primed for *ggz* transcription when octane is added to the cells. This mechanism is similar to the mechanism observed by N-HS repression of rRNA transcription in *E. coli*.

**Question 6** You are a graduate student in a lab working on prokaryotic transcription. You decide that you are interested in creating a very strong promoter that gives very high transcription rates. As you think about how to do this, you start thinking about the conundrum that RNA polymerase holoenzyme must deal with. It must bind DNA in a site-specific manner at the promoter but then must move away from the promoter and transcribe DNA into RNA in a non-site-specific manner.

**6A.** How does the RNA polymerase holoenzyme accomplish these tasks?

**The holoenzyme contains the sigma factor that help the RNAP bind the promoter site-specifically. Once that occurs, the holoenzyme goes through conformational changes that cause it to bind sigma less tightly (or not at all) so that the RNAP can release the promoter and continue elongation.**

You know that strong prokaryotic promoters often have -10 and -35 sites that have sequences that are close to the consensus sequence and that they also often have an UP element. You decide to take a strong prokaryotic promoter, whose sigma factor is called  $\sigma^{7.28}$ , with all of these elements and make it stronger by putting in the exact consensus sites for the -10 and -35 sites. You decide to call this promoter your *gnu* promoter.

**6B.** How would you test your *gnu* promoter *in vivo*? Don't forget your controls!

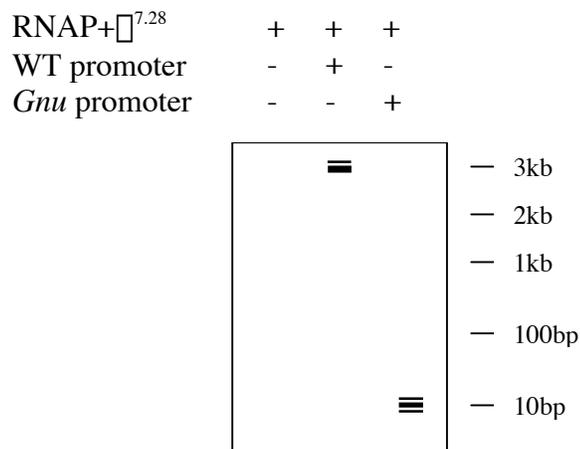
**Take a prokaryotic plasmid and create constructs with both your wildtype and *gnu* promoter fused to the LacZ gene (or any other reporter gene). Transform the plasmids into the cell and test for transcription of the reporter gene.**

You find that you cannot detect any transcription *in vivo* and so decide to try some *in vitro* experiments. First, you want to measure closed complex formation using Core RNAP and  $\sigma^{7.28}$  (the RNAP holoenzyme).

6C. What experiment would you do to test binding and how would you test to see if this holoenzyme binds more or less tightly to your *gnu* promoter versus a wildtype promoter?

**Do a gel shift assay to test for RNAP holoenzyme binding to the promoter. To test strength of binding, add cold competitor promoter to test how tightly DNA is bound to the protein. Compare wildtype to *gnu* promoter DNA.**

Your experiment tells you that the RNAP holoenzyme binds your *gnu* promoter even better than your wildtype promoter. You decide to do an incorporation assay to see if you can get any transcription *in vitro*. Your results are shown below.



6D. What step in the transcription initiation process is the *gnu* promoter defective for?

**There is a lot of abortive transcript made, suggesting that the holoenzyme is getting stuck at the promoter (defective in promoter clearance). The *gnu* promoter seems to maintain the unstable ternary complex stage of initiation.**

6E. You decide that you might still be able to get super duper transcription levels from this promoter but will need to make a mutation in the sigma factor. Based on your answer in part D and your knowledge of sigma factor function, what part of  $\square^{7.28}$  would you target for mutation so that it would work well with your new promoter?

**Region 3.2 of the sigma factor must be ejected from the RNA exit channel before the RNAP can move beyond the abortive transcript phase. Deleting this region of the sigma factor may make it easier for RNAP to be released from the *gnu* promoter. Also, making mutations in region 2 or 4 of the sigma factor that causes it to bind the promoter less tightly would also alleviate the problem.**

**Question 7.** You are studying the regulation of a eukaryotic gene involved in memory called Absent Minded. Animals with mutants in this gene have severe short-term memory loss.

You want to understand what controls the levels of Absent Minded gene expression to see if you can increase its expression in normal animals and therefore increase short-term memory. Using mutational mapping of the promoter, you identify two 10 bp regions of the promoter that when mutated strongly reduce expression of the Absent Minded promoter *in vivo*. You purify the proteins that bind these promoter elements and call the proteins ABR1 and ABR2.

To be sure that these proteins are involved in activation of the Absent Minded promoter, you fuse the Absent Minded promoter to the LacZ gene and test the levels of expression in strains that lack the ABR1 and ABR2 genes. You get the following results:

<u>Activator</u>	<u><i>in vivo</i></u> <u>Transcription Units</u>
Wild Type	2000 U
□ABR1	<b>50 U</b>
□ABR2	<b>200 U</b>
□ABR1+ □ABR2	50 U

**7A** Based on these data, what can you conclude about the function of ABR1 and ABR2 at the Absent Minded promoter.

**From this data, you can conclude that both ABR1 and ABR2 are needed to activate transcription of the Absent Minded promoter to wild type levels *in vivo*. You can also conclude that ABR1 activity is required for ABR2 to activate transcription, because the same level of transcription from the Absent Minded promoter occurs in both a □ABR1 mutant and a □ABR1 □ABR2 mutant. Finally, you can conclude that ABR1 can activate transcription without ABR2, because a 4-fold increase in transcription of the Absent Minded promoter is seen in an □ABR2 mutant relative to a □ABR1 □ABR2 mutant.**

Your next step is to purify the ABR1 and ABR2 proteins. You first test their ability to activate transcription *in vitro* by adding the purified activators to a plasmid containing the Absent Minded promoter. You then add purified RNA Pol II and auxiliary factors (TFIIA, TFIIB, TFIID, TFIIE, TFIIIF, and TFIIH). You get the following results:

<u>Activator</u>	<u>Absent Minded mRNA (pmoles/minute)</u>
None	500
ABR1	500
ABR2	3,000
ABR1+ ABR2	3,000

**7B** Based on the differences between the reconstituted *in vitro* transcription experiments and the *in vivo* promoter fusion experiments, propose a function for the ABR1 regulator. Explain your reasoning.

**This experiment demonstrates that ABR1 is required for activation *in vivo* but is not required in your *in vitro* assay for activation. Therefore, at least one condition that is important for ABR1 to activate transcription must be different in your *in vitro* assays. One difference in the *in vitro* transcription reactions is the absence of nucleosomes on the DNA. Therefore, one possible model for ABR1's activity is that it activates transcription of the Absent Minded promoter *in vivo* by recruiting nucleosome remodeling proteins to reposition nucleosomes around the promoter region and make the promoter more accessible to the replication machinery. Nucleosome repositioning could both stimulate transcription by the general transcription machinery, which is why you see a 4-fold increase in transcription from the Absent Minded promoter *in vivo* in the presence of ABR1, and allow ABR2 to bind and activate transcription.**

**Other models for ABR1 function were also accepted if they provided a reasonable explanation of the difference between *in vivo* and *in vitro* assays.**

You decide to map the transcriptional activation domain of ABR1. Since a lab mate had recently mapped the DNA binding domain of ABR2, you decide to use this DNA binding domain to map the transcriptional ABR1 transcriptional activation domain.

**7C** Describe how you would test for the ABR1 activation domain using the ABR2 DNA binding domain.

**To test for the ABR1 activation domain you would:**

- 1. Fuse the ABR1 protein to the ABR2 DNA binding domain. (Note: This fusion is created by joining together the coding sequences of the two proteins and allowing the cells to transcribe and translate the fusion protein, not by physically fusing polypeptide chains.)**
- 2. Test the ability of this fusion protein to activate transcription from a hybrid promoter containing an ABR2 binding site and promoter upstream of a reporter gene. (Ideally, this hybrid promoter would lack ABR1 binding sites, as your full length ABR1-ABR2 DNA binding domain protein could possibly bind to either DNA binding site if both were present.)**
- 3. If the protein containing the full length ABR1 fused to the ABR2 binding domain activates transcription *in vivo*, make fusion proteins that contain**

smaller regions of ABR1 fused to the ABR2 DNA binding domain and test the ability of these fusion proteins to activate transcription from the hybrid promoter *in vivo*.

- a. These small regions of the ABR1 protein are generated by making fusion proteins that contain less and less of the N or C-terminal regions of the ABR1 protein. Making these deletions in a systematic way will allow you to identify the minimal region that is necessary to activate.
- b. It was not appropriate to propose that you would use sequence similarity to identify the activation domain of ABR1 and clone this into the fusion protein, as activation domains usually lack conserved sequences.

You set up your assay but find that **none** of the constructs can activate transcription *in vivo*. Frustrated, you alter your assay to use the more traditional Gal4-DNA binding domain instead and observe the expected activation.

**7D** Propose a model to explain the different results for the ABR2 and Gal4 DNA binding domain fusion proteins.

It is clear that even though the ABR2 binding domain binds to its binding site *in vitro* it is not binding appropriately and allowing the ABR1 activation domain to function *in vivo*. However, the Gal4 DNA binding domain, when fused to the ABR1 activation domain, does function *in vivo*. Several possible explanations for this were accepted. A few of these explanations were:

1. The Gal4 DNA binding domain can bind to DNA in the presence of nucleosomes, but the ABR2 binding domain cannot.
2. The wild type ABR2 protein present in your cells is competing with your hybrid protein for occupancy of the binding site and preventing it from activating transcription.
3. Your ABR1-ABR2 fusion protein is not properly folded *in vivo* and so isn't able to activate transcription, while your ABR1-Gal4 fusion protein folds properly.

**7E** Based on all of the observations concerning ABR1 and ABR2, propose a model that explains their synergistic function in the cell.

**One possible model is the following:**

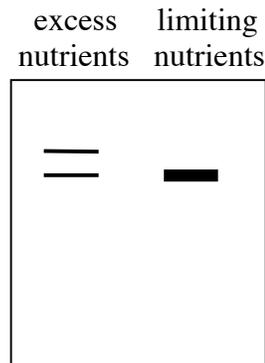
**ABR1 binds to the Absent Minded promoter *in vivo* and recruits chromatin remodeling factors to the promoter. Once the nucleosomes are repositioned, recruitment of the general transcription machinery to the promoter is enhanced. In**



**8A.** What can you conclude about the use of the two TATA boxes? Identify the +1 site for each TATA box.

**Both TATA boxes are used in the WT template, and transcription initiates at two sites. The +1 site for each TATA box is indicated above by the blue boxes. Normally, the most upstream site will be called the +1 site and the second site will be numbered relative to the first (+7 in this case.) Also, as noted in class the 5' most residue of the mRNA is most likely to be A, but we didn't think of that when making the template!**

**8B.** You go on vacation for a few days and forget some yeast you left growing. They must be starving! However, you notice that the yeast on these plates are a darker, more appealing shade of purple. You decide to compare levels of *PUR3* RNA prepared from cells grown in media with excess nutrients to those grown in media with limiting nutrients using northern blots. You observe the following autoradiogram:



How can you explain the different bands seen in the two samples?

**Expression is upregulated from the T2 TATA box when the cells are grown in nutrient-limiting media, and expression is repressed from the T1 TATA box.**

**8C.** You believe that a starvation-inducible gene product could activate *PUR3* transcription. Describe how you could use microarrays to identify starvation-inducible genes.

- isolate RNA from cells grown in nutrient rich media, and separately isolate RNA from cells grown in nutrient limiting media.
- label each sample with a different fluorescent probe (for example use red for rich and green for limiting)
- apply the labeled material to a microarray
- measure the ratio of the two labeled samples at each spot

**-in this case, spots corresponding to genes that are induced by starvation will be green**

**8D.** Using the microarrays, you identify about a dozen starvation-inducible genes that could potentially be activators of *PUR3*. Describe either a simple genetic or biochemical assay that you could use to determine if any of these gene products activates *PUR3*?

**Genetic:** clone the potential activators such that you can induce their expression (by a means other than starvation) *in vivo*, and test for the dark purple phenotype of the yeast or test for use of the T2 TATA box by northern blot or S1 nuclease protection. You could also knock out the potential activator, and look for lack of induction of Pur3 in nutrient limiting media.

**Biochemical:** purify the potential activator, and determine if it binds to part of the promoter region of the Pur3 gene by band shift assay or DNase I footprinting. You could also add the purified potential activator to *in vitro* transcription reactions containing the full length *PUR3* promoter. You can determine whether the activator changes the TATA box selection in this reaction using your S1 nuclease protection assay. (A variety of reasonable answers would be okay for this question)

**8E.** Using this approach you identify a protein (Pta1) that can activate *PUR3* transcription. You want to know the stage at which Pta1 acts to increase *PUR3* transcription. A friend down the hall has developed new methods that allow her to measure the association and dissociation of the holoenzyme complex, the transition to processive elongation, and the rate of elongation (80 bases/sec).

She reports that on average for *PUR3* transcription, it takes the holoenzyme 0.2 seconds to associate and 0.3 seconds to dissociate from the DNA. Once a closed complex is formed, it takes the holoenzyme 5 seconds to transition to processive elongation. She cannot measure the individual steps in this process, but can tell you that once the open complex is formed it takes 9 seconds to generate the 600 nucleotide long transcript. Use this information to predict the steps at which Pta1 could act.

**A transcriptional activator must increase the rate of the rate limiting step, so use the information above to determine the rate limiting step:**

**-formation of a closed complex is fast relative to the other stages in initiation.**

**-the process of getting from a closed complex to an open complex to an unstable ternary complex to a stable ternary complex takes 5 seconds.**

**-It takes 9 seconds to get from open complex formation to the end of a 600 nucleotide long transcript, which will take 7.5 seconds (600 bases/80 bases per second).**

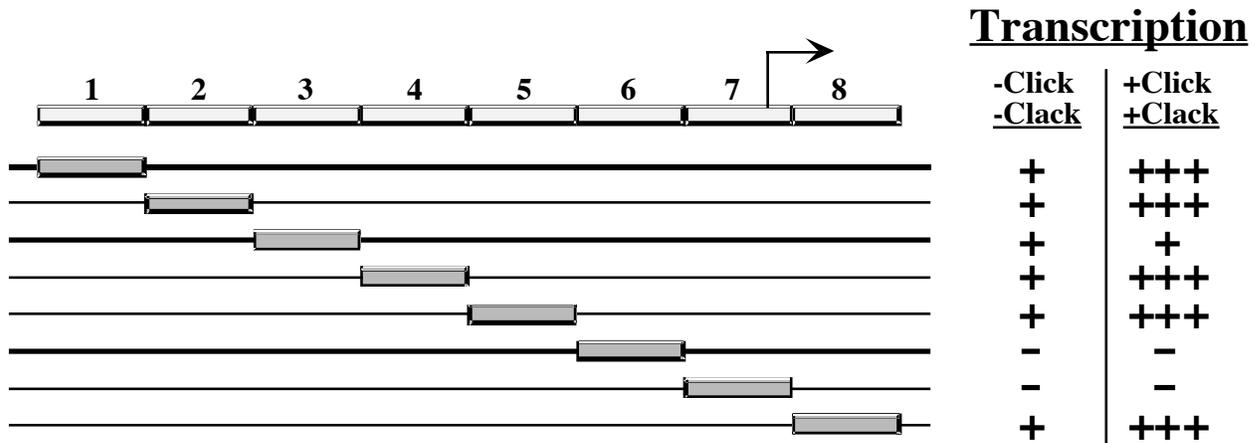
**-Therefore, it must take 1.5 seconds ( $9 - 7.5 = 1.5$ ) to get from the open complex to the stable ternary complex.**

**-Formation of the open complex then must take 3.5 seconds ( $5 - 1.5 = 3.5$ ), making this the rate limiting step.**

**Question 9.** You have developed a transcription assay that recapitulates the *in vivo* regulation of the promoter for the RNA Pol II-transcribed Ratchet gene. By fractionating

this extract, you have purified and cloned two specific transcription factors that, in addition to the general RNA Pol II transcription factors, are required for correct regulation of the promoter.

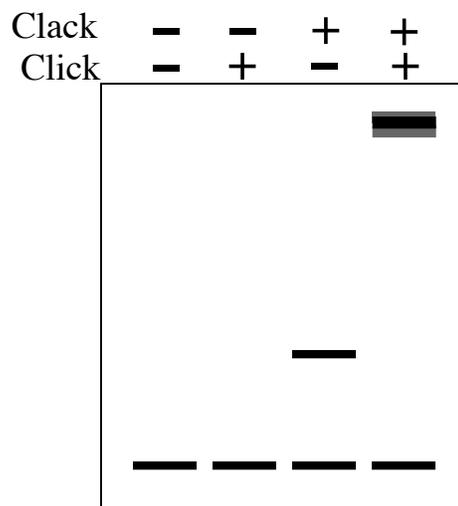
As a first step to understand the function of these factors you make a series of deletion mutants of the Ratchet promoter and test their effect on transcription in the presence and absence of the Click and Clack factors. You find the following results.



9A. Based on this data, what would you predict the function of elements 3, 6, and 7 are

**Region 3 is required for activation by Click and Clack, but not for basal levels of transcription, so it is likely to be an activator binding site. Regions 6 and 7 are required for transcription, and so these regions likely make up the core promoter which includes elements such as the TATA box and INR sequence.**

To further characterize the factors that you have identified, you perform a gel shift assay to look at the ability of the different factors to bind to the Ratchet promoter. You obtain the following results.



**9B.** Based on this data, what can you conclude concerning the function of Click and Clack? How is this finding consistent with the data obtained in part 7a?

**Clack is a DNA binding protein. Click is either a DNA binding protein that binds only in the presence of Clack, or is a factor that increases the amount of Clack binding. Antibody super shift experiments could be used to distinguish between these two models.**

**In 7a, we detected only one specific sequence for activator binding. This shows that the initial binding by Clack is sequence specific and the Click binding or additional Clack binding is sequence non-specific.**

**9C.** How would you determine the region of the Clack factor required for activation of the Ratchet promoter?

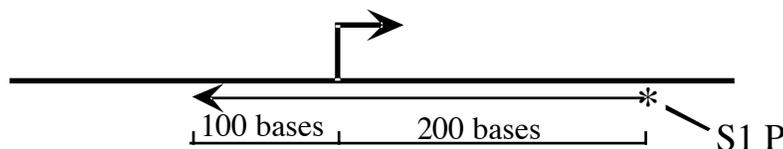
**To map the Clack activation domain:**

**-Fuse Clack to a known DNA binding protein (such as LexA DNA binding domain) so you can measure activation independent of effects on DNA binding.**

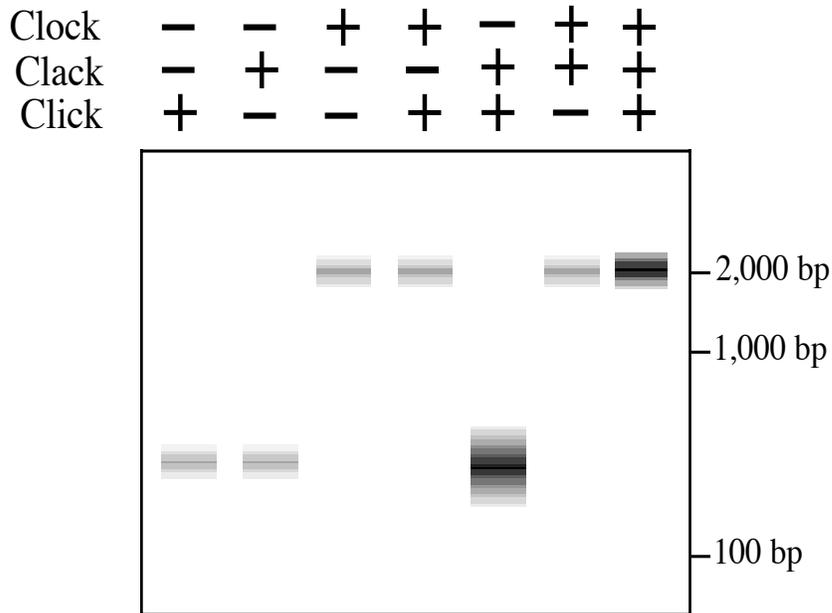
**-Make a series of N and C-terminal deletions of Clack, also fused to the LexA DNA binding domain.**

**-Test the ability of all of these fusions to activate a test promoter. The full length Clack-LexA fusion is a positive control that should show full reporter activity. The test promoter should contain LexA binding sites in the region upstream of the core promoter of a reporter gene (such as LacZ). The minimal region of Clack fused to LexA that activates lacZ expression fully is the activation domain.**

Throughout your analysis of the Ratchet promoter you have been frustrated by the lack of effect of a third transcription factor, **Clack**, which is known to activate this promoter *in vivo*. All of your previous assays for transcription were performed using S1 assays and a ssDNA probe of 300 bases in length that overlaps the ratchet start site by 100 bases.



Thinking you may have missed something, you analyze the products of a set of transcription reactions performed in the presence of radiolabeled UTP on a denaturing agarose gel. To cause termination of RNA Pol II artificially, you cut the template DNA 2 kb downstream of the promoter with a restriction enzyme. You obtain the following results (hint: promoter clearance is complete after 100 bases of RNA synthesis).



**9D.** What aspect of RNA Pol II function is affected by addition of Clock? Why was this effect missed by the S1 assay and detected by the transcription run off assay?

**Clock is a transcription elongation factor. It allows RNA Pol II to transcribe past a pause or termination site to generate full-length transcripts.**

**You didn't see the affects of Clock in the S1 nuclease protection assay because this assay only looks at the beginning of transcripts. You used a 300 base long probe that hybridized to the beginning of the transcript, and would see the same result if the transcript was 200 bases or 2000 bases long.**

**In the run-off assay you measure the incorporation of rNTPs. This assay allows you to determine RNA length information, so you will see different products in the presence and absence of Clock.**