

Biochemistry and Molecular Biology 400  
**Final Examination**  
Fall 2000  
December 11, 2000

Instructor: Hardison

This examination has 40 questions for a total of 200 points. **All** are multiple choice and are worth 5 points each. Please answer the questions on the enclosed answer sheet. **BE SURE TO ENCODE YOUR STUDENT NUMBER AND TEST FORM ON THE ANSWER SHEET!**

***THIS IS FORM A, and has answers.***

Useful information and equations are at the end. You may wish to tear these off and use them throughout the exam.

PLEASE TURN IN ONLY THE ANSWER SHEET; you may keep the exam.

Some questions request you to choose all the correct statements from a list. Partial credit is given for choosing **some** of the correct answers, **maximal** credit is given for choosing **all** the correct statements, but **no** credit is given for a choice that includes **any incorrect** statements. **Pick only one option per question.**

1. Which of the following statements about regulation of the *lac* operon in *E. coli* are correct?

- [1] A repressor negatively regulates the operon when the inducer is absent.
- [2] The *lac* repressor is not bound to any DNA sequence when inducer is bound to repressor.
- [3] The *lac* repressor prevents binding of RNA polymerase to the promoter.
- [4] The *lac* repressor bound to its operator decreases the rate of the closed to open transition of the RNA polymerase-promoter complex.

Correct choices are:

- a. 1                      b. 1, 4                      c. 1, 2, 3                      d. 2, 3                      e. 1, 3, 4

**b is correct, 2 pts for a**

2. Which statements about positive regulation of the *lac* operon in *E. coli* are correct?

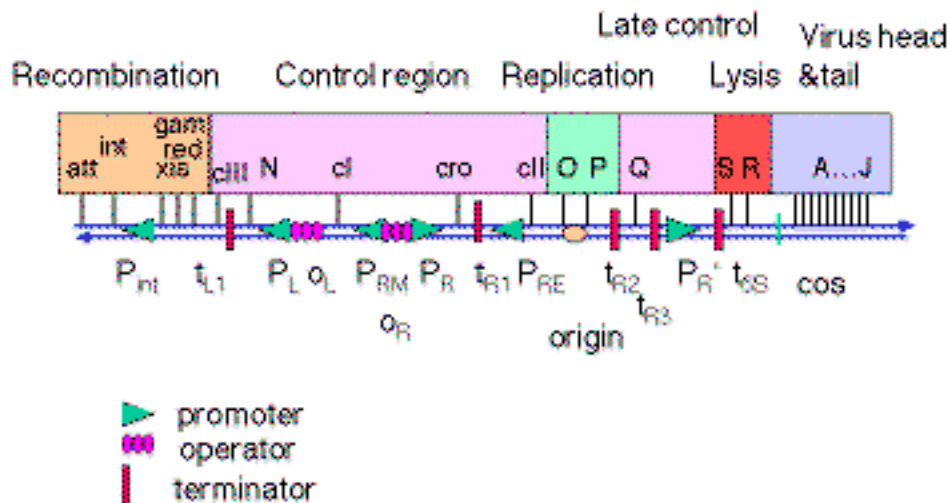
- [1] Binding of the complex cAMP-CAP stimulates transcription of the operon.
- [2] cAMP-CAP interacts with the  $\alpha$  subunit of RNA polymerase.
- [3] RNA polymerase binds more strongly to the *lac* promoter in the presence of cAMP-CAP.
- [4] The  $\alpha$ -subunit of RNA polymerase binds to a specific DNA sequence upstream of the  $-35$  box of the *lac* operon to stimulate transcription.

Correct choices are:

- a. 1                      b. 2                      c. 1, 4                      d. 1, 2, 3                      e. 1, 2, 3, 4

**d is correct, 2 pts for a or b**

The next 3 questions concern bacteriophage  $\lambda$ . A condensed map of the virus is shown below.



3. Which of the following would decrease the chances of lysogeny occurring?

- [1] Mutations that inactivate  $P_{RE}$ .
- [2] Mutations that inactivate  $P_{int}$ .
- [3] Mutations that eliminate the  $t_{6S}$  termination site.

[4] Mutations that prevent the cI repressor from binding to O<sub>RI</sub>.

- a. 3                      b. 1, 2, 4              c. 2, 3                  d. 1, 4                  e. 1

**b is correct, 3 for d, 2 for e**

4. Two mutant strains of bacteriophage, *cI*<sup>-</sup> and *cII*<sup>-</sup>, produce no lysogens, and thus they make clear plaques individually when they infect *E. coli*. If they are coinfecting into *E. coli* (i.e. cells are infected with both *cI*<sup>-</sup> and *cII*<sup>-</sup>), which of the following will result?

- a. No lysogens will be formed.  
b. Lysogens will be formed, with *cI*<sup>-</sup> as the prophage.  
c. Lysogens will be formed, with *cII*<sup>-</sup> as the prophage.  
d. Lysogens will be formed, some with *cI*<sup>-</sup> as the prophage and some with *cII*<sup>-</sup> as the prophage.

**c is correct.**

5. Which of the following is **NOT** necessary for antitermination during lytic infection by bacteriophage ?

- a. N protein  
b. *E. coli* NusA protein  
c. *E. coli* NusG protein  
d. cIII protein  
e. a *nut* site

**d is correct.**

6. Which of the following events occur when *E. coli* is starved for the amino acid tryptophan?

- [1] The ribosome translates the leader peptide completely (to the UGA stop codon).  
[2] A G+C rich stem-loop structure forms in the nascent RNA (regions 3 and 4) at the attenuator site.  
[3] A step-loop structure forms in the nascent RNA (regions 2 and 3) that precludes formation of the G+C rich stem-loop at the attenuator site.  
[4] Transcription reads through the attenuator into *trp EDCBA*.

- a. 1                      b. 1, 2                      c. 3                      d. 3, 4                      e. 1, 2, 4

**d is correct, 2 pts for c.**

7. Which of the following mutations in the leader of the *E. coli trp* operon will result in transcriptional termination at the attenuator, regardless of whether the concentration of tryptophan is high or low?

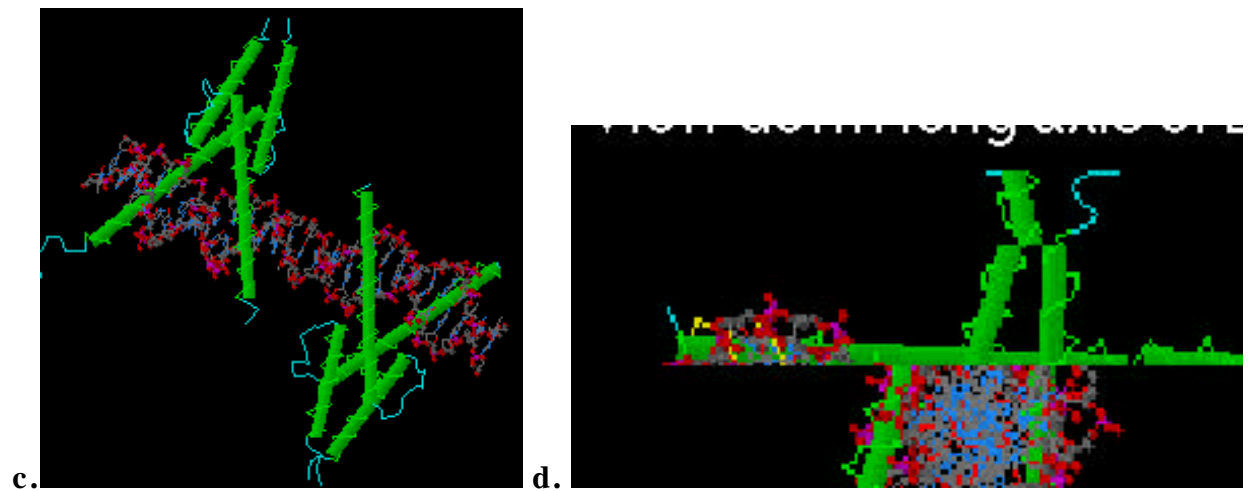
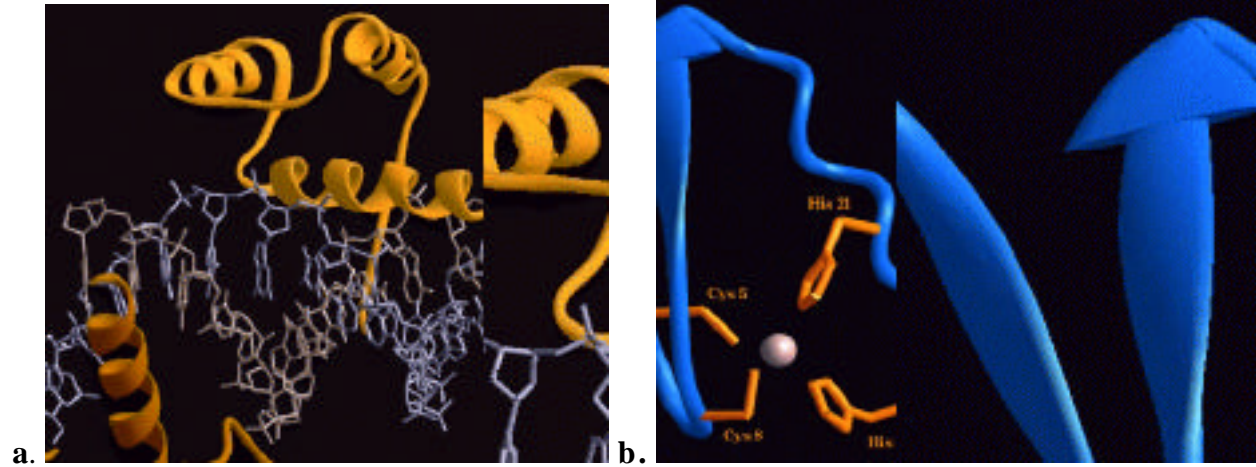
- [1] Mutations in region 1 that prevent it from binding to region 2.  
[2] Mutations in region 2 that prevent it from binding to region 3 but not region 1.  
[3] Mutations in region 3 that prevent it from binding to region 4, but not region 2.  
[4] Mutations in region 3 that prevent it from binding to either region 2 or region 4.

Correct choice(s) are:

- a. 2                      b. 1                      c. 3                      d. 3, 4                      e. 2, 3

**a is correct.**

Use the following pictures to answer the next two questions.



8. Which is the structure of a basic helix-loop-helix DNA binding domain?

**c is correct**

9. Which is the structure of a  $C_2H_2$  Zn finger?

**b is correct**

For the next four questions, consider the following hypothetical data on regulation of an operon in a bacterium. The genes encoding enzymes required for metabolism of galactose, such as galactokinase and galactose phosphate epimerase, are inducible. Thus the enzymes are produced in abundance in the presence of this sugar, but in the absence of galactose, they are produced at very low levels. Four genes or loci, *galA*, *galB*, *galC*, and *galD*, affecting the activity or regulation of these enzymes were studied in a series of haploid and diploid strains. In the following table, wild-type alleles of the genes or loci are indicated by a + under the letter of the *gal* gene or locus and mutant alleles are indicated by a - under the letter. The activities of the two enzymes, galactokinase and galactose phosphate epimerase, were assayed after growth in the absence or presence of galactose. The units of enzyme activity are 100 = induced activity of the wild-type enzyme (in the presence of galactose), 1 = uninduced activity of the wild-type enzyme (in the absence of galactose), and 0 = no measurable activity. In the diploid analysis, one copy of each operon is present in each cell.

Strain number	<i>gal</i>	Galactokinase .		Galactose phosphate epimerase .	
		- galactose	+ galactose	-galactose	+ galactose
Haploid	A B C D				
1	+ + + +	1	100	1	100
2	- + + +	0	0	1	100
3	+ - + +	1	1	1	1
4	+ + - +	1	1	1	1
5	+ + + -	1	100	0	0
Diploid	A B C D/A B C D				
6	- + + +/+ + + -	1	100	1	100
7	+ - + +/- + + +	1	1	2	101
8	+ + - +/- + + +	1	100	2	200
9	- - + +/+ + - -	1	100	1	1

10. Which option best describes the phenotype with respect to the two enzymes for strains 3 and 4?

- | <u>Galactokinase</u>        | <u>Galactose phosphate epimerase</u> |
|-----------------------------|--------------------------------------|
| a. activated                | activated                            |
| b. no activity              | inducible                            |
| c. noninducible             | noninducible                         |
| d. constitutive, high level | constitutive, high level             |
| e. inducible                | no activity                          |

**c is correct**

11. Which genes encode the enzymes? Choose the option with the correct gene under each enzyme.

- | <u>Galactokinase</u> | <u>Galactose phosphate epimerase</u> |
|----------------------|--------------------------------------|
| a. <i>galA</i>       | <i>galB</i>                          |
| b. <i>galC</i>       | <i>galD</i>                          |
| c. <i>galA</i>       | <i>galC</i>                          |
| d. <i>galB</i>       | <i>galC</i>                          |
| e. <i>galA</i>       | <i>galD</i>                          |

**e is correct**

12. Which gene or locus shows *cis* dominance, i.e. the particular allele that is in *cis* to the reporter gene is dominant?

- a. *galA*
- b. *galB*
- c. *galC*
- d. *galD*
- e. none of the genes

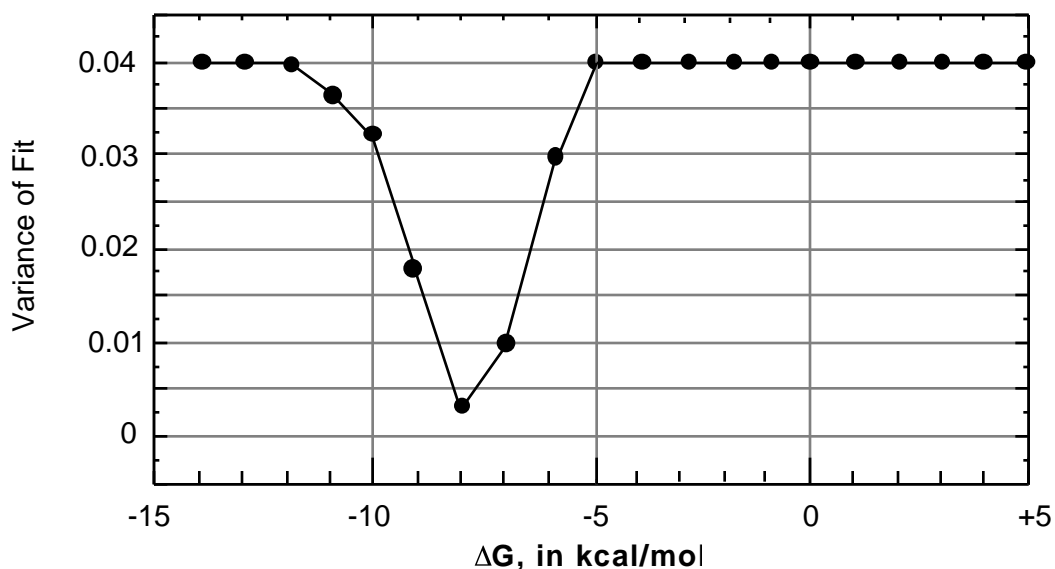
**b is correct**

13. Which statements about the regulation of the *gal* operon in this bacterium are supported by the data presented?

- a. The gene *galC* encodes an activator that binds to the DNA sequence at *galB* in the presence of galactose but not in the absence of galactose.
- b. The gene *galB* encodes an activator that binds to the DNA sequence at *galD* in the presence of galactose but not in the absence of galactose.
- c. The gene *galD* encodes a repressor that binds to the operator sequence *galA* in the absence of galactose but dissociates in the presence of galactose.
- d. The gene *galC* encodes a repressor that binds to the operator sequence *galB* in the absence of galactose but dissociates in the presence of galactose.

**a is correct, 2 points for d**

For the next two questions, let's imagine that you mixed increasing amounts of the DNA binding protein called AP7 with a constant amount of a labeled duplex oligonucleotide containing the binding site. After measuring the fraction of DNA bound by AP7 (i.e. the fractional occupancy) as a function of [AP7], the data were analyzed by nonlinear, least squares regression analysis at a wide range of possible values for  $\Delta G$ . The error associated with the fit of each of those values to experimental data is shown below; the higher the variance of fit, the larger the error.



14. What is the most accurate value of  $\Delta G$  for binding of AP7 to this duplex oligonucleotide?

- a. -14 kcal/mole
- b. -11 kcal/mole
- c. -8 kcal/mole
- d. -5 kcal/mole
- e. +5 kcal/mole

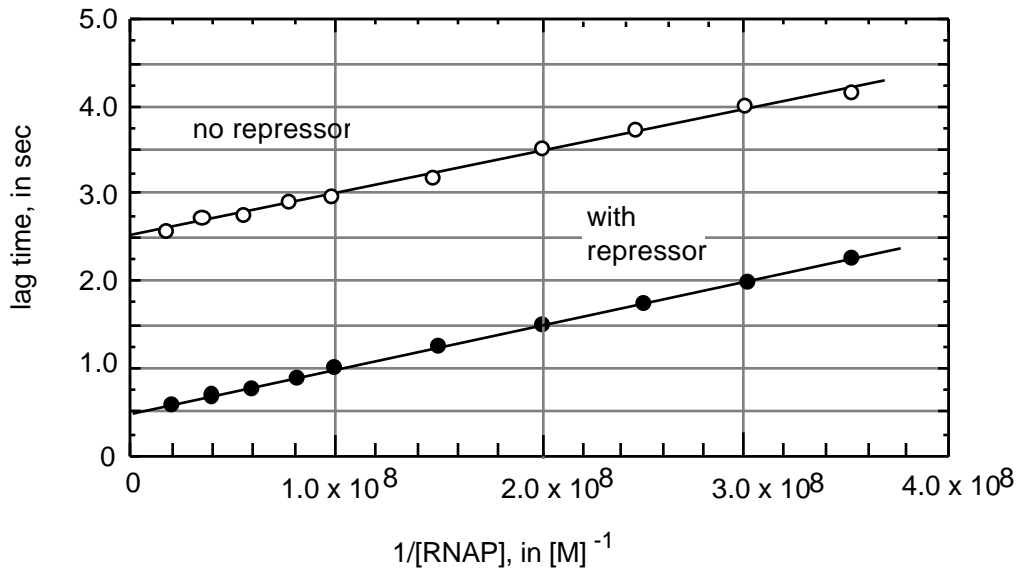
**c. is correct**, the minimum in error is the maximum in accuracy.

15. What is the most accurate measure of the equilibrium constant,  $K_s$ , for binding of AP7 to this duplex oligonucleotide?

- a.  $2.09 \times 10^{-4} \text{ M}^{-1}$
- b.  $4.79 \times 10^3 \text{ M}^{-1}$
- c.  $7.74 \times 10^5 \text{ M}^{-1}$
- d.  $1.25 \times 10^8 \text{ M}^{-1}$
- e.  $2.02 \times 10^{10} \text{ M}^{-1}$

**c. is correct** ( $7.74 \times 10^5 \text{ M}^{-1}$ ;  $G = -RT \ln K_s$ )

For the next two problems, consider a hypothetical eubacterial operon in which the operator extends from  $-5$  to  $+15$ , where  $+1$  is the start site for transcription. Measurement of the lag time before production of abortive transcripts as a function of the inverse of the RNA polymerase concentration ( $1/[\text{RNAP}]$ ) gave the results shown below. The **open** circles are the results of the assay in the **absence** of repressor, and the **closed** circles are the results in the **presence** of repressor bound to the operator.



**16.** What is the value of the forward rate constant ( $k_f$ ) for closed to open complex formation under the two different conditions?

- | $k_f$ , no repressor | $k_f$ with repressor |
|----------------------|----------------------|
| a. 2.0 per sec       | 0.4 per sec          |
| b. 2.5 per sec       | 0.5 per sec          |
| c. 2.0 per sec       | 2.0 per sec          |
| d. 0.5 per sec       | 0.4 per sec          |
| e. 0.4 per sec       | 2.0 per sec          |

**e is correct**, 0.4 per sec without repressor, 2 per sec with repressor; the y-intercept is  $1/k_f$

**17.** From these data, how does the equilibrium constant ( $K_B$ ) for binding of the RNA polymerase to the promoter change in the presence or absence of repressor?

- $K_B$  is decreased 50-fold in the presence of repressor.
- $K_B$  is decreased 5-fold in the presence of repressor.
- $K_B$  is decreased 2-fold in the presence of repressor.
- $K_B$  is the same in the absence and presence of repressor.
- $K_B$  is increased 10-fold in the presence of repressor.

**b is correct.**

For the next three questions, consider a repressor protein, called R, that binds to the operator, called O, of a eubacterial operon with the following binding constants. The subscript S refers to binding to the specific DNA sequence at the operator, the subscript NS refers to non-sequence specific binding. In the presence of an inducer, the binding constant to the specific site is reduced by a factor of 100, as indicated by the equation for  $K_{S,ind}$ , whereas the value for  $K_{NS,r}$  does not change in the presence of inducer.

This bacterial genome has about  $4.6 \times 10^6$  bp, and with only one genome per cell and a cell volume of  $1 \times 10^{-15}$  L, this gives a concentration of nonspecific binding sites of  $7.6 \times 10^{-3}$  M, and a concentration of specific binding sites of  $1.7 \times 10^{-9}$  M (this is  $[D_S]_{total}$  which is also  $[O]_{total}$ ). There are 450 molecules of the repressor per cell, giving a total repressor concentration of  $7.5 \times 10^{-7}$  M.

The following equations apply:



$$K_{S,r} = \frac{[RO]}{[R][O]} = 10^{10} \text{ M}^{-1} \quad (\text{eqn 2})$$

$$K_{NS,r} = \frac{[RD_{NS}]}{[R][D_{NS}]} = 10^5 \text{ M}^{-1} \quad (\text{eqn 3})$$

$$K_{S,ind} = \frac{[RO]}{[R][O]} = 10^8 \text{ M}^{-1} \quad (\text{eqn 4})$$

**18.** What fraction of the operator sites (or  $D_S$ ) are bound by repressor in the **absence** of inducer?

a.  $1 \times 10^{-10}$

b. 0.001

c. 0.91

d. 0.50

e.  $1 \times 10^{+10}$

**c is correct, 0.91**

**19.** What fraction of the operator sites are bound by repressor in the **presence** of inducer?

a.  $1 \times 10^{-8}$

b. 0.090

c. 0.0091

d. 0.005

e.  $1 \times 10^8$

**b is correct, 0.090**

**20.** Assuming that the operon is expressed when an operator is no longer bound by the repressor, the 100 fold decrease in  $K_S$  in the presence of the inducer leads to how much induction of the operon? The ratio of expression in the presence of inducer to that in the absence of inducer is:

a. 100

b. 10

c. 1

d. 0.1

e. 0.01

**b is correct**, the operon is expressed at a 10-fold higher level ( $0.91/0.09 = 10.1$ ; this is the ratio of free operators in the presence of inducer, which is  $1.00 - 0.09 = 0.91$ , to free operators in the absence of inducer, which is  $1.00 - 0.91 = 0.09$ )

**21.** Which statement(s) about acetylation of histone N-terminal tails are true?

[1] Acetylation of the histone N-terminal tails is associated with active transcription.

[2] The protein Gcn5p is a histone acetyl transferase involved in regulation of many genes in yeast.

[3] The SAGA complex contains several types of proteins, including Gcn5p, ADA proteins, Spt proteins and some TAFs.

[4] Acetylation of histones is a stable modification, i.e. once the acetyl group is added, it is rare for it to be removed.

a. 1

b. 1, 2

c. 1, 3, 4

d. 1, 2, 3, 4

e. 1, 2, 3

**e is correct, 3 pts for b, 2 pts for a.**

**22.** Which statements about the SWI/SNF complex are true?

[1] It is only found in yeast.

[2] It remodels nucleosomes by phosphorylating them.

[3] It uses ATP hydrolysis to remodel nucleosomes.

[4] The remodeled nucleosomes have lost H2A and H2B.

[5] The products of the reaction are remodeled dimers of nucleosomes.

a. 3, 5

b. 1, 2

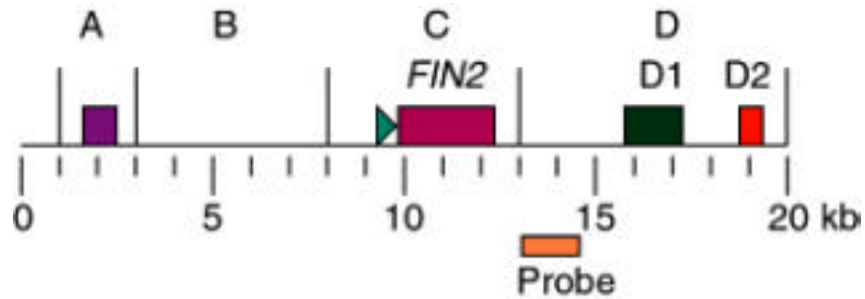
c. 3, 4

d. 4, 5

e. 3

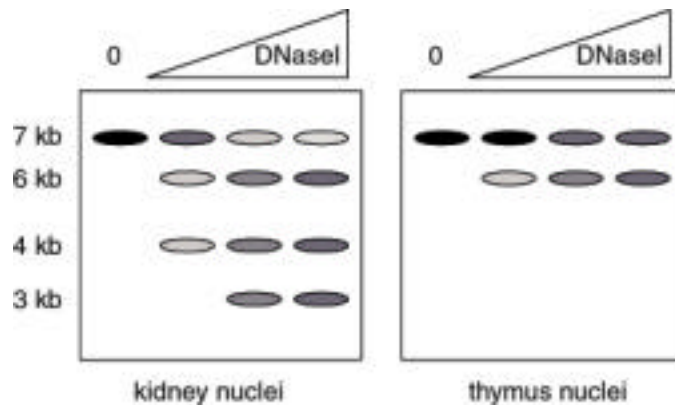
**a is correct, 2 pts for e**

For the next 9 questions, consider this map of the region containing a hypothetical “human” gene called *FIN2* and some regulatory elements.



The region is divided into 4 fragments (A, B, C, and D) by the restriction endonuclease *KpnI*; vertical lines above the horizontal line mark cleavage sites. The entire *FIN2* gene (rectangle) is in fragment C, along with its promoter (triangle pointing in the direction of transcription). D1 and D2 are smaller restriction fragments that are part of fragment D.

23. The *FIN2* gene is expressed in kidney cells but not in thymus. Nuclei isolated from kidney cells and from thymus cells were digested with increasing amounts of DNase I, then DNA was isolated, digested with *KpnI*, run on a gel, blotted and hybridized with the fragment labeled “Probe” on the map. The results diagrammed below show the intact 7 kb fragment D plus additional fragments. The “0” lane has DNA from nuclei not treated with DNase I but cut with *KpnI* after isolation. The intensity of gray is proportional to the amount of DNA in the band. Which of the following is a correct interpretation of these results?

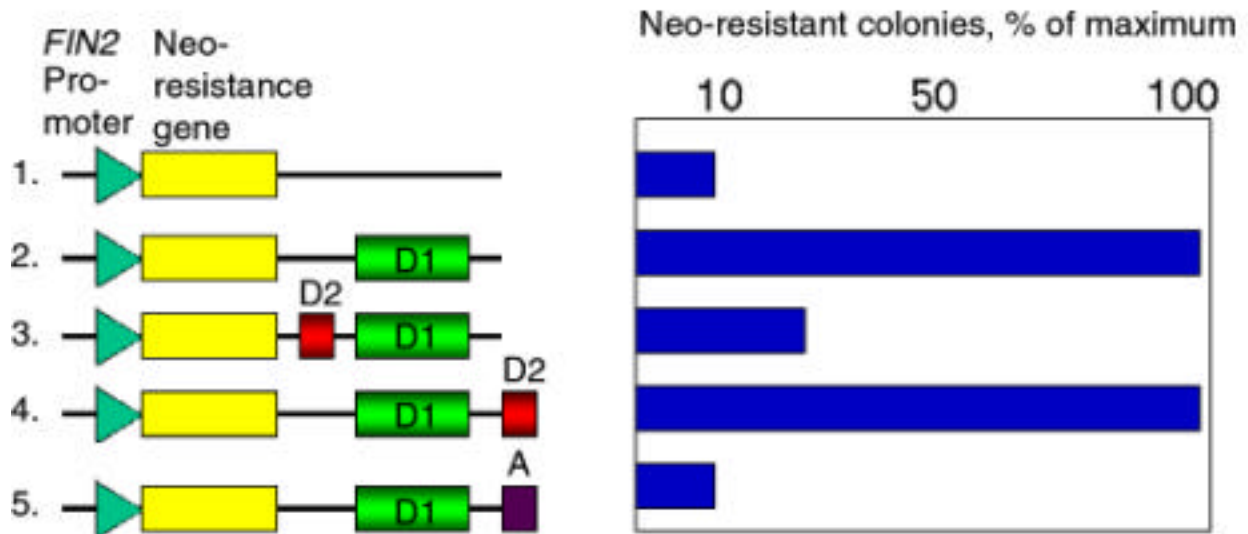


- [1] No DNase hypersensitive sites are present in thymus nuclei.
- [2] A DNase hypersensitive site located at 19 kb on the map is present in both kidney and thymus nuclei.
- [3] DNase hypersensitive sites located at 16 and 17 kb on the map are present only in kidney nuclei.
- [4] A DNase hypersensitive site located at 14 kb on the map is present in both kidney and thymus nuclei.

- a. 2, 3
- b. 1
- c. 3, 4
- d. 2, 4
- e. 3

**a is correct, 2 points for e**

Consider the following experiment and the map above for the next 3 questions. Restriction fragments A, D1 and D2 were placed into a plasmid with a neomycin resistance gene expressed from the *FIN2* promoter, in the positions shown below (left). Plasmids 1-4 were transfected into a kidney cell line, and plasmid 5 was transfected into a thymus cell line. For all 5 experiments, clones of cells expressing the neomycin resistance gene were selected in media containing the neomycin analog G418. The number of neomycin-resistant clones (colonies) is plotted on the right, below.



24. Which DNA fragment(s) is/are (an) enhancer(s)?

- a. A                      b. A and D2                      c. D2                      d. D1                      e. D1 and D2

**d is correct.**

25. Which DNA fragment(s) is/are (a) silencer(s)?

- a. A                      b. A and D2                      c. D2                      d. D1                      e. D1 and D2

**a is correct.**

26. Which DNA fragment(s) is/are (an) insulator(s)?

- a. A                      b. A and D2                      c. D2                      d. D1                      e. D1 and D2

**c is correct.**

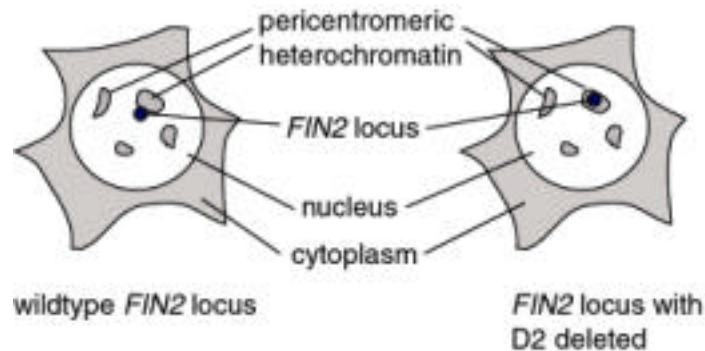
27. A human DNA segment spanning fragments B, C and D was integrated into two different locations in the genome of transgenic mice. Line 1 has the *FIN2* segment integrated into a locus that is in the euchromatic portion of kidney nuclei, whereas Line 2 has the *FIN2* segment integrated into centromeric repeats, so it is in heterochromatin. The human *FIN2* gene is expressed at the same level as the endogenous mouse gene in kidney cells of transgenic Line 1, whereas it is expressed at only 25% that of the endogenous mouse gene in transgenic Line 2. Examination of human *FIN2* in single cells by fluorescent in situ hybridization to *FIN2* RNA showed that 100% of the kidney cells in transgenic Line 1 hybridized, but only 25% of the kidney cells in transgenic Line 2 hybridized. What do you conclude about the human *FIN2* gene from these data?

- [1] It is subject to position effect variegation in Line 1.
- [2] It is subject to position effect variegation in Line 2.
- [3] It is expressed in only 25% of the cells in Line 2.
- [4] In the *expressing cells* in Line 2, it is expressed at the same level as the endogenous mouse gene.
- [5] In the *expressing cells* in Line 2, it is expressed at 25% of the level as the endogenous mouse gene.

- a. 2, 3, 5      b. 2, 3      c. 1, 2, 4      d. 3      e. 2, 3, 4

**e is correct, 3 pts for b, 2 pts for d**

**28.** The position of the *FIN2* gene in interphase nuclei was determined by hybridization to a *FIN2* fluorescent probe (dark dot below) and hybridization to the gamma satellite DNA at the centromere (gray patches in the nuclei). After examination by confocal microscopy, the *FIN2* gene was found to be at the edge of the pericentromeric heterochromatin in all kidney cells examined in transgenic Line 2. The D2 region then was deleted from the human *FIN2* locus in a derivative of transgenic Line 2. Analysis of this derivative showed that the *FIN2* gene was embedded in pericentromeric heterochromatin in the interphase nuclei of all kidney cells examined.



What do you conclude from these data?

- a. The D2 enhancer keeps the *FIN2* gene far away from pericentromeric heterochromatin.
- b. The D1 silencer acts by embedding the *FIN2* gene into pericentromeric heterochromatin.
- c. The *FIN2* locus is randomly engulfed in pericentromeric heterochromatin, regardless of the presence or absence of D2.
- d. The D2 region is needed to prevent *FIN2* from being completely embedded in heterochromatin.
- e. The D1 region is sufficient to prevent *FIN2* from being embedded in heterochromatin.

**d is correct.**

**29.** Combining the information from the previous 6 problems, what can you conclude about regulation of the *FIN2* gene?

- [1] The A region is a silencer that keeps *FIN2* off in nonexpressing cells.
- [2] The A region is an enhancer of *FIN2* expression in kidney cells.
- [3] The D1 region, marked by tissue-specific hypersensitive sites, is an enhancer of expression.
- [4] The combination of A, D1 and D2 can prevent position effect variegation even when *FIN2* is located in centromeric repeats.
- [5] The D2 region, marked by a hypersensitive site present in both expressing and nonexpressing cells, is an insulator that can prevent a *FIN2* gene from being embedded in pericentromeric heterochromatin.

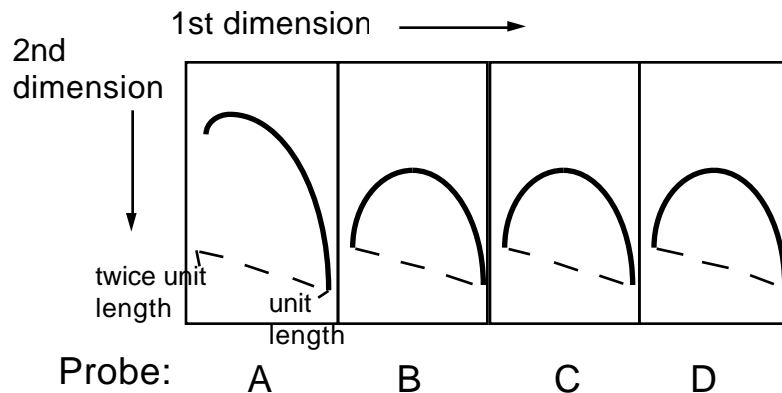
- a. 2, 3                      b. 2, 3, 4                      c. 1, 3                      d. 1, 4, 5                      e. 1, 3, 5  
**e is correct, 3 points for c.** Statements 1, 3, and 5 are correct.

**30.** An *EcoRI* site at position 11 kb on the map (site not shown) cuts within an exon of *FIN2*. The DNA is cut at the *EcoRI* site, labeled at the 3' ends, cut with *KpnI*, and 3 kb and 2 kb *KpnI* to *EcoRI* fragments (3' end-labeled only at the *EcoRI* sites) were isolated. These fragments were hybridized to *FIN2* RNA and then digested with the single-strand-specific nuclease S1. Which labeled fragment(s) generated (a) product(s) protected from the nuclease?

- a. The 3 kb *KpnI* to *EcoRI* fragment  
 b. The 2 kb *KpnI* to *EcoRI* fragment  
 c. Both the 3 kb and 2 kb *KpnI* to *EcoRI* fragments  
 d. Neither the 3 kb nor the 2 kb *KpnI* to *EcoRI* fragments

**b is correct.**

**31.** The replication origin(s) around the *FIN2* gene was (were) mapped using two dimensional gel analysis. DNA was isolated from replicating kidney cells, digested with *KpnI*, separated on Brewer and Fangman two-dimensional gels, blotted onto a filter and hybridized with probes from fragments A, B, C and D (see map above). The results are shown below.



What do these data tell you?

- a. Fragment A contains a replication origin, which is located 5' to the gene.  
 b. Fragment A contains a replication origin, which is located 3' to the gene.  
 c. Fragments B, C and D contain replication origins.  
 d. Replication is unidirectional, proceeding in the same direction as transcription of the *FIN2* gene.  
 e. Replication is bidirectional.

**a is correct.**

**32.** Suppose that you have obtained the sequence of the nontemplate strand (i.e. synonymous with the mRNA) of a protein-coding region of a gene. You also learn that a pathogenic mutation of G changing to T occurs at position 24 in the sequence below. You want to know the effects of that nucleotide substitution on the amino acid sequence. Assuming that the longest open reading frame is the one used for translation, what is the sequence of the wild-type and the mutant proteins encoded by this segment of DNA? An open reading frame is a string of nucleotides without a translation stop codon.

mutant

T

10                      20                      30

wt     5'     TAAGCTGGTGGTGGTGGAGCGCCGGCGGTGTCA

In the choices below, the amino acid encoded by the altered codon is underlined.

- |   |   |
|---|---|
| <p>Wild-type:</p> <p>a. AlaGlyGlyGlyGluArg<u>Arg</u>ArgCys</p> <p>b. SerTrpTrpTrpStop</p> <p>c. ThrPro<u>Pro</u>AlaLeuThrThrThrSerLeu</p> <p>d. AspThr<u>Ala</u>GlyGlyHisHisHisGlnLeu</p> <p>e. LysLeuValValValSerAla<u>Gly</u>GlyVal</p> | <p>Mutant:</p> <p>AlaGlyGlyGlyGluArg<u>Arg</u>ArgCys</p> <p>SerTrpTrpTrpStop</p> <p>ThrPro<u>Thr</u>AlaLeuThrThrThrSerLeu</p> <p>AspThr<u>Asp</u>GlyGlyHisHisHisGlnLeu</p> <p>LysLeuValValValSerAla<u>Val</u>GlyVal</p> |
|---|---|

**e is correct; 2 pts for c and d.** This sequence and mutation are actually from the (proto)oncogene *HRAS*, and the designated mutation is tumorigenic.

The next two questions concern a sequenced cosmid clone containing a genomic DNA fragment from human chromosome 22. It was obtained from GenBank, accession number AC000068.

**33.** All the known repetitive elements in it were analyzed using a program called RepeatMasker, which compares the sequence to a database of repetitive elements. It found the results in the following table. Interspersed repetitive elements are organized into four major classes in this output, followed by a tabulation of small RNAs, simple and low complexity repeats.

```

sequences:                      1
total length:            43934 bp
GC level:                    47.03 %
=====

```

	number of elements	length occupied	percentage of sequence
-----			
SINES:	35	8975 bp	20.43 %
ALUs	33	8792 bp	20.01 %
MIRs	2	183 bp	0.42 %
LINEs:	17	5172 bp	11.77 %
LINE1	9	3440 bp	7.83 %
LINE2	7	1242 bp	2.83 %
LTR elements:	5	1590 bp	3.62 %
MaLRs	4	1183 bp	2.69 %
Retroviral	1	407 bp	0.93 %
MER4_group	0	0 bp	0.00 %
DNA elements:	4	818 bp	1.86 %
MER1_type	3	482 bp	1.10 %
MER2_type	1	336 bp	0.76 %
Mariners	0	0 bp	0.00 %
Unclassified:	0	0 bp	0.00 %
Total interspersed repeats:		16555 bp	37.68 %

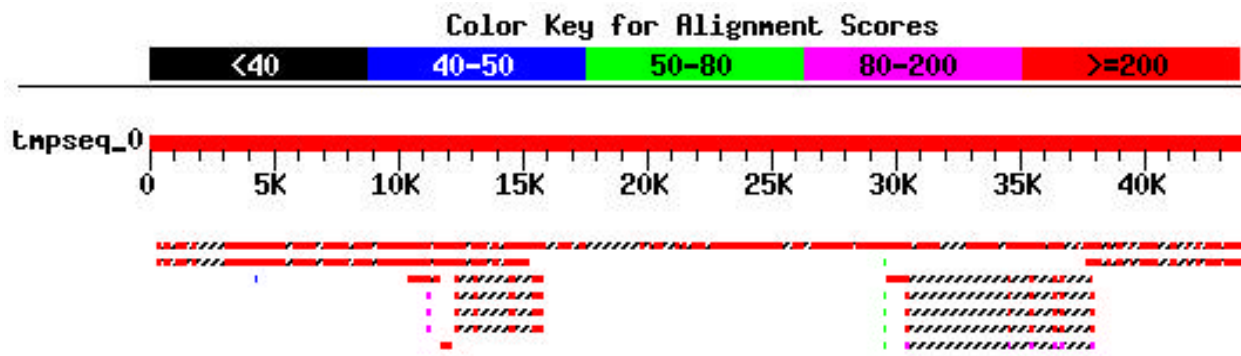
Small RNA:	1	40 bp	0.09 %
Satellites:	0	0 bp	0.00 %
Simple repeats:	7	429 bp	0.98 %
Low complexity:	2	65 bp	0.15 %

Which conclusion is accurate about repeats in AC000068?

- The most frequently occurring repeat is a LINE.
- The average SINE is shorter than an average MIR.
- LINE1 repeats occupy more of this sequence than any other repeat.
- The average LINE1 repeat is longer than the average ALU.
- Interspersed repeats occupy 20% of AC000068.

**d is correct.**

34. After masking the repeats, known genes were identified in the sequence AC000068 using the BLAST server at NCBI. The following table summarizes the high-scoring matches to all “finished” human DNA sequences, which includes other genomic DNAs and mRNAs (cDNAs). Matching segments are filled, whereas regions separating matches are cross-hatched.



The mRNA for a gene called “nuclear localization signal deleted in velocardiofacial syndrome” (*NLVCF*) is identical to the genomic DNA in 4 segments extending from positions 12306 to 15837. The mRNA for a gene called “ubiquitin fusion degradation 1-like” (*UFDIL*) is identical to the genomic DNA in 6 segments extending from positions 30387 to 37904. These are shown as matches to multiple entries of the mRNAs on lines 3-6 on the summary figure.

What do you conclude about AC000068 from these data?

- It has 37 exons.
- It has 4 exons of the *NLVCF* gene and 6 exons of the *UFDIL* gene.
- It has 3 exons of the *NLVCF* gene and 5 exons of the *UFDIL* gene.
- It does not have any genes.
- It has 3 genes in addition to *NLVCF* and *UFDIL*.

- 4
- 2
- 3
- 2, 5
- 1, 3, 5

**b is correct**

35. Which of the following statements about protein synthesis in *E. coli* is true?

- The peptidyl transferase step requires hydrolysis of high energy phosphate bonds in GTP.

- b. Binding of f-Met-tRNA to the mRNA on the small ribosomal subunit utilizes ATP and an elongation factor.
- c. Translocation of the peptidyl-tRNA from the A site to the P site of the ribosome is catalyzed exclusively by an RNA component of the ribosome.
- d. Base pairing between a short segment of the 16S rRNA (in the small subunit) and its complement in the 5' untranslated region of the mRNA is used to recognize the initiator AUG.

**d. is correct**

**36.** Which of the following statements about introns and splicing of the precursors to mRNA are correct?

- [1] The initiating nucleophile for splicing of nuclear pre-mRNA is the 2' hydroxyl of an internal adenine nucleotide.
- [2] Introns almost always begin with GU and end with AG.
- [3] Particular snRNPs bind at the branch point and the 5' splice site.
- [4] Assembly of a spliceosome requires the cleavage of high-energy bonds from ATP.

- a. 2, 3
- b. 1, 4
- c. 2, 3, 4
- d. 1, 2, 3
- e. 1, 2, 3, 4

**e. is correct**, 3 pts for c or d, 2 pts for a or b.

**37.** How does DNA polymerase III achieve high processivity? Choose the correct statement.

- a. The core ( ) is inherently highly processive.
- b. The  $\beta$  subunit of the DNA polymerase III holoenzyme serves as a "sliding clamp" holding the polymerase onto the primer-template.
- c. The  $\beta$  complex is a "clamp" that holds the polymerase core onto the template.
- d. The  $\beta$  "sliding clamp" is unloaded by the  $\beta'$  subunits.

**b. is correct.**

**38.** Which of the following proteins catalyze the polymerization of nucleotides? Do not choose an option with any incorrect responses.

- a. DnaA and PriA
- b. DnaG and telomerase
- c. DnaG and RecA
- d. UvrD and DnaB

**b. is correct.**

**39.** Consider a circular, B form DNA molecule with a superhelical density of -0.05. Suppose that during initiation of replication, 50 base pairs unwind (untwist) without breaking the phosphodiester backbone. What will be the effect on the linking number (  $L$  ) and the supercoiling (  $W$  )?

- a.  $L = 0$  and  $W = +5$
- b.  $L = 0$  and  $W = -5$

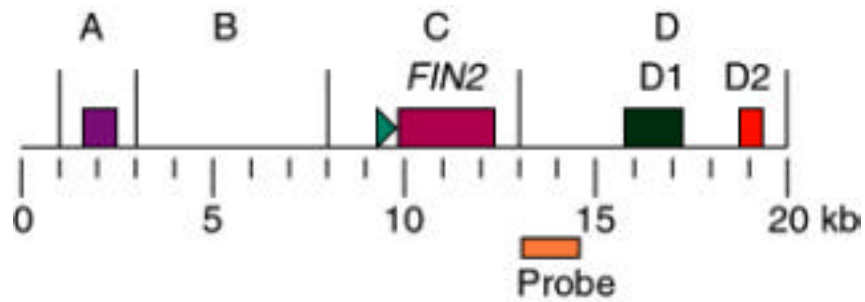
- c.  $L = -5$  and  $W = +5$
- d. The effect cannot be determined without knowing the length of the DNA molecule.

**a. is correct.  $\Delta T = -5$  (50 bp/10 bp per twist are removed),  $\Delta L = 0$  since no covalent bonds in the DNA are broken and reformed, so  $\Delta W = -\Delta T = +5$ .**

**40.** Which of the following enzymes catalyzes homologous pairing and invasion of single stranded DNA into a duplex during recombination?

- a. RecBCD
  - b. RuvC
  - c. RuvB
  - d. The 5' to 3' exonuclease activity of DNA polymerase I
  - e. RecA
- e is correct.**

## Map of the *FIN2* locus



## The Genetic Code

1st	2nd								3rd
	U		C		A		G		
U	UUU	Phe	UCU	Ser	UAU	Tyr	UGU	Cys	U
	UUC	Phe	UCC	Ser	UAC	Tyr	UGC	Cys	C
	UUA	Leu	UCA	Ser	UAA	Term	UGA	Term	A
	UUG	Leu	UCG	Ser	UAG	Term	UGG	Trp	G
C	CUU	Leu	CCU	Pro	CAU	His	CGU	Arg	U
	CUC	Leu	CCC	Pro	CAC	His	CGC	Arg	C
	CUA	Leu	CCA	Pro	CAA	Gln	CGA	Arg	A
	CUG	Leu	CCG	Pro	CAG	Gln	CGG	Arg	G
A	AUU	Ile	ACU	Thr	AAU	Asn	AGU	Ser	U
	AUC	Ile	ACC	Thr	AAC	Asn	AGC	Ser	C
	AUA	Ile	ACA	Thr	AAA	Lys	AGA	Arg	A
	AUG*	Met	ACG	Thr	AAG	Lys	AGG	Arg	G
G	GUU	Val	GCU	Ala	GAU	Asp	GGU	Gly	U
	GUC	Val	GCC	Ala	GAC	Asp	GGC	Gly	C
	GUA	Val	GCA	Ala	GAA	Glu	GGA	Gly	A
	GUG*	Val	GCG	Ala	GAG	Glu	GGG	Gly	G

\* Sometimes used as initiator codons.

## Equations for BMB400 Fall 2000

### Binding equilibria

The equilibrium constant for the binding of a protein (P) to a DNA fragment (D) is  $K_B$ .

$$K_B = \frac{[DP]}{[D][P]}$$

The fraction of DNA in a complex with the protein is a function of  $K_B$  and  $[P]$ :

$$\frac{[DP]}{[D]_{total}} = \frac{K_B[P]}{1 + K_B[P]}$$

The specificity of binding of a protein (P) to DNA is described by the following equation.  $K_S$  is the equilibrium binding constant to specific DNA sites, or  $D_S$ .  $K_{NS}$  is the equilibrium binding constant to nonspecific DNA sites, or  $D_{NS}$ . Concentrations are moles/L, i.e. molar.

$$\text{specificity} = \frac{K_S}{K_{NS}} = \frac{[PD_S]}{[D_S]} \times \frac{[D_{NS}]}{[P]_{total} - [D_S]_{total}}$$

Relationship between  $G$  and  $K_{eq}$ , where  $K_{eq}$  is an equilibrium constant:

$$\Delta G = -RT \ln K_{eq}$$

$$R = 1.98 \times 10^{-3} \text{ kcal deg}^{-1} \text{ mol}^{-1}$$

$$T = 298^\circ \text{ K}$$

$$RT = 0.59 \text{ kcal/mol}$$

In an abortive transcription assay, the lag time between the mixing of reagents and the optimal rate of abortive transcript production is related to the concentration of RNA polymerase (or [RNAP]) by the following equation.  $K_B$  is the equilibrium constant for binding of RNAP to the promoter, and  $k_f$  is the forward rate constant for the closed to open transition.

$$\text{lag time} = \frac{1}{K_B k_f} \times \frac{1}{[RNAP]} + \frac{1}{k_f}$$

### **$C_0t$ analysis**

$$\frac{N_{unknown}}{N_{standard}} = \frac{C_0 t_{\gamma_2, unknown}}{C_0 t_{\gamma_2, standard}}$$

### **Supercoiling:**

**$T$  = Twisting Number**

For B form DNA, it is + (# bp/10 bp per twist)

For A form DNA, it is + (# bp/11 bp per twist)

For Z form DNA, it is - (# bp/12 bp per twist)

$$\Delta L = \Delta W + \Delta T$$

Superhelical density =  $\sigma = W/T = -0.05$  for natural bacterial DNA

### **Recombinant DNA:**

$$N = \frac{\ln(1-P)}{\ln(1-f)}$$

$N$  is the number of recombinants required to obtain a given probability ( $P$ ) of having a particular sequence in the library.  $f$  = the fraction of the genome in a single recombinant molecule, i.e.  $f$  = insert size / genome size.