

Biochemistry and Molecular Biology 400
Final Examination
Fall 1998
December 14, 1998

Instructor: Hardison

This examination has 31 questions for a total of 158 points. **All** are multiple choice. The value of each question ranges from 4 to 10 points. 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 B, and has answers.

Useful information and equations are at the end, on pages 17 and 18.

PLEASE TURN IN ONLY THE ANSWER SHEET; you may keep the exam. Answers will be posted promptly in the hall of the first floor of S. Frear. Good luck!

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.

A14=B1=C6. (4 pts) 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. 2, 3 c. 2, 4 d. 1, 4 e. 1, 2, 3

d is correct, 2 pts for a

A15=B2=C7. (4 pts) 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 α subunit of RNA polymerase.
- [3] RNA polymerase binds more strongly to the *lac* promoter in the presence of cAMP-CAP.
- [4] The α -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, 2 b. 1, 3 c. 1, 2, 3 d. 1, 4 e. 1, 2, 3, 4

c is correct, 2 pts for a or b

A16=B3=C8. (4 pts) Two mutant strains of bacteriophage λ , cI^- and cII^- , 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^- and cII^-), which of the following will result?

- a. No lysogens will be formed.
- b. Lysogens will be formed, with cI^- as the prophage.
- c. Lysogens will be formed, with cII^- as the prophage.
- d. Lysogens will be formed, some with cI^- as the prophage and some with cII^- as the prophage.

C is correct.

A17=B4=C9. (4 pts) Which of these statements about infection of *E. coli* with bacteriophage are correct?

- [1] The *cI* gene is transcribed immediately after infection, from a strong promoter.
- [2] The *N* gene is transcribed immediately after infection, from a strong promoter.

[3] Action of the N protein allows expression of genes *cII*, *cIII*, and *Q*.

[4] The cII protein stimulates transcription of the *cI* gene from an otherwise weak promoter.

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

b is correct, 2 pts for c, 1 pts for d

A18=B5=C10. (4 pts) Which of these statements about antitermination during infection of *E. coli* with bacteriophage are correct?

[1] N protein acts to prevent termination at rho-dependent sites.

[2] Action of N protein requires sequences on the DNA called *nut* sites.

[3] The NusA protein prevents the action of the N protein.

[4] The N protein binds to DNA at rho-independent sites to prevent termination.

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

a is correct, 2 pts for e

A19=B6=C11. (4 pts) If the ATG corresponding to the AUG start codon for the leader peptide in the *E. coli trp* operon is mutated so that translation cannot initiate in the *trp* leader RNA, what will the effects be on the regulation of the operon by attenuation?

- a. The operon will be OFF in the presence of either low or high concentrations of tryptophan.
- b. The operon will be ON when the concentration of tryptophan is low, and it will be OFF when the concentration of tryptophan is high.
- c. The operon will be ON in the presence of either low or high concentrations of tryptophan.
- d. The operon will be OFF when the concentration of tryptophan is low, and it will be ON when the concentration of tryptophan is high.

a is correct.

A20=C7=C12. (4 pts) In wild-type *E. coli*, which of the following events occur when it is starved for the amino acid tryptophan?

[1] Ribosomes translating the leader peptide stall at UGG codons.

[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 terminates at the attenuator, thus blocking expression of *trp EDCBA*..

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

d is correct, 2 pts for e

For the next five questions, consider the following hypothetical data on regulation of an operon in a fungus. The genes encoding enzymes required for repair of UV damage, such as photolyase and an excinuclease, are inducible. Thus the enzymes are produced in abundance only after exposure to radiation such as UV radiation. Four genes or loci, *radA*, *radB*, *radC*, and *radD*, 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 *rad* gene or locus and mutant alleles are indicated by a - under the letter. The activities of the two enzymes, photolyase and excinuclease, were assayed after growth in the absence or presence of UV. The units of enzyme activity are 100 = induced activity of the wild-type enzyme (in the presence of UV), 1 = uninduced activity of the wild-type enzyme (in the absence of UV), and 0 = no measurable activity. In the diploid analysis, one copy of each operon is present in each cell.

Strain number	<i>rad</i>				Photolyase .		Excinuclease .		
	A	B	C	D	- UV	+ UV	-UV	+ UV	
1	+	+	+	+	1	100	1	100	
2	-	+	+	+	1	100	0	0	
3	+	-	+	+	0	0	1	100	
4	+	+	-	+	100	100	100	100	
5	+	+	+	-	100	100	100	100	
Diploid	A	B	C	D/A	B	C	D		
6	-	+	+	+/+	-	+	+	1	100
7	-	+	+	+/+	+	+	-	1	100
8	+	+	-	+/-	-	+	+	100	100
9	+	-	+	-/-	+	-	+	1	100

A21=B8=C1. (4 pts) Which option best describes the phenotype with respect to the two enzymes for strains 4 and 5?

- | | |
|-----------------------------|--------------------------|
| <u>Photolyase</u> | <u>Excinuclease</u> |
| a. no activity | inducible |
| b. noninducible | noninducible |
| c. constitutive, high level | constitutive, high level |
| d. inducible | constitutive |

c is correct

A22=B9=C2. (4 pts) Which genes encode the enzymes? Choose the option with the correct gene under each enzyme.

- | | |
|-------------------|---------------------|
| <u>Photolyase</u> | <u>Excinuclease</u> |
| a. <i>radA</i> | <i>radB</i> |
| b. <i>radB</i> | <i>radA</i> |
| c. <i>radA</i> | <i>radC</i> |
| d. <i>radD</i> | <i>radC</i> |
| e. <i>radC</i> | <i>radD</i> |

b is correct

A23=B10=C3. (4 pts) Which gene or locus shows *cis* dominance, i.e. the particular allele that is in *cis* to the reporter gene is dominant?

- a. none of the genes
- b. *radA*
- c. *radB*
- d. *radC*
- e. *radD*

d is correct

A24=B11=C4. (4 pts) Which gene or locus encodes a diffusible factor (e.g. a protein) that regulates the *rad* operon in *trans* ?

- a. none of the genes
- b. *radA*
- c. *radB*
- d. *radC*
- e. *radD*

e is correct

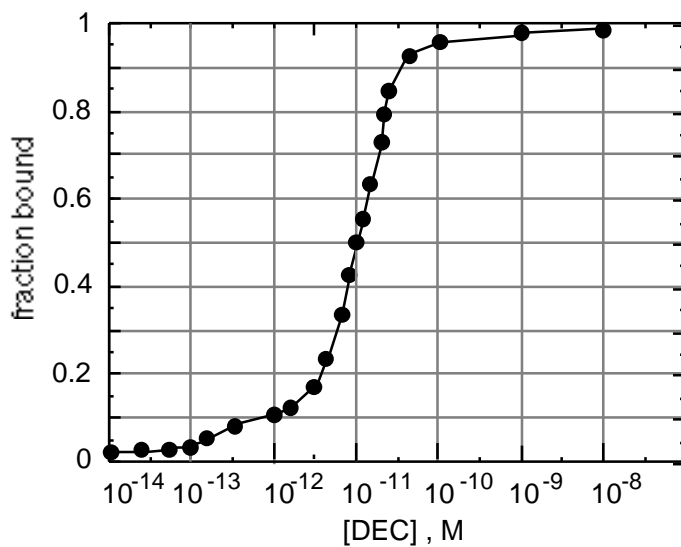
A25=B12=C5. (4 pts) Which statements about the regulation of the *rad* operon in this fungus are supported by the data presented?

- a. The gene *radD* encodes a repressor that binds to the operator sequence *radC* in the absence of UV but is inactive in the presence of UV.
- b. The gene *radC* encodes a repressor that binds to the operator sequence *radB* in the presence of UV but is inactive in the absence of UV.
- c. The gene *radD* encodes an activator that binds to the DNA sequence at *radC* in the presence of UV but not in the absence of UV.
- d. The gene *radB* encodes an activator that binds to the DNA sequence at *radA* in the presence of UV but not in the absence of UV.

a is correct

For the next two questions, let's imagine that you are studying a protein called DEC that binds to a specific site on DNA called DRE. On Dec. 31, 1999, you do an experiment to measure the binding constant (K_B) for binding of DEC to DRE DNA. You incubate labeled DRE DNA with increasing concentrations of DEC and measure the amount of DNA bound to DEC as the fraction of labeled DNA in complex with DEC (i.e. $[\text{DEC}\cdot\text{DRE complex}]/[\text{DRE}]_{\text{total}}$). After obtaining the numbers, you go to a New Year's Eve party.

A26=B13C17. (10 pts) On Jan. 01, 2000, you go in to the lab to calculate the K_B . Unfortunately, none of the computers to which you have access were ready for the year 2000, and they will not work. You promised to have a value determined by Jan. 02, so you decide to



calculate a value for K_B from the following graph.

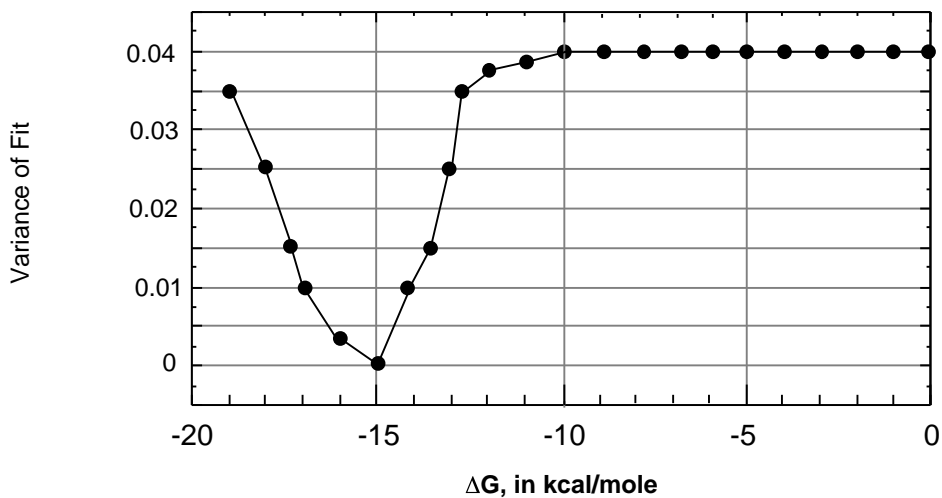
What is the best value for K_B that can be determined from this graph?

- a. $K_B = 10^{10} \text{ M}^{-1}$
- b. $K_B = 10^{11} \text{ M}^{-1}$
- c. $K_B = 10^{12} \text{ M}^{-1}$
- d. $K_B = 10^{-11} \text{ M}$

e. $K_B = 10^{-10} \text{ M}$

b is correct

A27=B14=C17. (10 pts) Later in that week, the computers are finally cured of the year 2000 “bug” and you can run a non-linear least squares analysis of your data on DEC binding to DRE. You obtain the following graph for variance of fit at a range of possible values for G .



What is the most accurate value for K_B based on this analysis?

a. $K_B = 10^{-15} \text{ M}$

b. $K_B = 10^{-10} \text{ M}$

c. $K_B = 2 \times 10^7 \text{ M}^{-1}$

d. $K_B = 10^{12} \text{ M}^{-1}$

e. $K_B = 10^{11} \text{ M}^{-1}$

e is correct

A28=B15=C13. (5 pts) The lag time for abortive initiation of transcription was measured as various concentrations of RNA polymerase both in the presence and the absence of a repressor. When this lag time was plotted against the reciprocal of the concentration of RNA polymerase (i.e. $1/[RNAP]$), the y-intercept was found to be 0.1 sec in the absence of the repressor but it was 10 sec in the presence of the repressor. What can you conclude about the mechanism by which this repressor affects transcription by RNA polymerase?

- a. The repressor **decreases** the forward rate constant for the closed to open transition by a factor of 100.
- b. The repressor **increases** the forward rate constant for the closed to open transition by a factor of 100.
- c. The repressor **decreases** the forward rate constant for the closed to open transition by a factor of 100 **and decreases** the binding constant for polymerase to promoter by a factor of 10.
- d. The repressor **decreases** the binding constant for polymerase to promoter by a factor of 100.
- e. The repressor **increases** the binding constant for polymerase to promoter by a factor of 10.

A is correct.

A29=B16=C14. (5 pts) Which statement(s) about chromatin remodeling 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 protein PCAF is a histone acetyl transferase involved in regulation of many genes in mammalian cells.
- [4] The ATPase SWI/SNF alters nucleosomal structure such that transcription factors can bind more readily.

a. 1 b. 1, 2 c. 1, 2, 3 d. 1, 2, 3, 4 e. 2, 4
d is correct, 2 pt for a, 3 pts for b or e, 4 pts for c.

A30=B17=C15. (10 pts) For a nuclear protein called BOS1 that binds to a specific site on DNA called BRE, the binding of BOS1 to a specific site and to nonspecific sites is described by the following equations.

Let $P = \text{BOS1}$

$D_s = \text{a specific binding site in DNA}$

$D_{ns} = \text{a nonspecific binding site in the genomic DNA}$



$$K_s = \frac{[PD_s]}{[P][D_s]} = 10^{10} \text{ M}^{-1} \quad (\text{eqn 2})$$

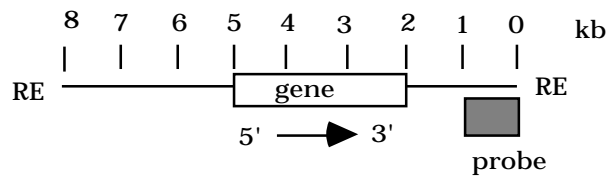
$$K_{ns} = \frac{[PD_{ns}]}{[P][D_{ns}]} = 10^5 \text{ M}^{-1} \quad (\text{eqn 3})$$

In a nucleus with a volume of 5×10^{-13} L, and a total concentration of BOS1 (i.e. $[P_{\text{total}}]$) of 1×10^{-7} M, what fraction of BRE sites are occupied by BOS1? Consider the total [BRE] (i.e. $[D_s_{\text{total}}]$) to be 5×10^{-12} M and the total concentration of nonspecific binding sites (i.e. $[D_{ns}]$) to be 5×10^{-3} M. The fraction of BRE sites occupied by BOS1 (i.e. $[PD_s]/[D_s_{\text{total}}]$) is

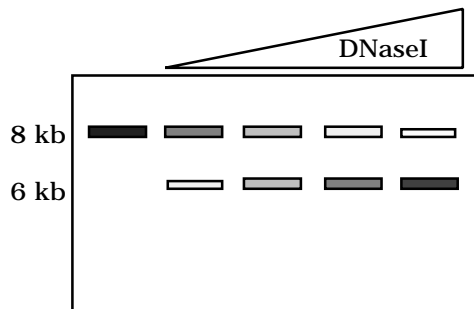
- a. 0.01
- b. 0.1
- c. 0.67
- d. 1.0
- e. 2.0

c is correct, 5 pts for e

A31=B18=C18. (5 pts) DNase hypersensitive sites around a gene were mapped by treating nuclei from cells that express that gene with increasing amounts of DNaseI. The partially digested DNA was isolated, cut to completion with a restriction enzyme (RE), and analyzed by Southern blot-hybridization using a radioactive probe that is located 3' to the gene. Cleavage of genomic DNA with the restriction enzyme generates an 8 kb fragment that contains the gene, and the probe for the blot hybridization is located at the right end of the fragment (left to right defined as the direction of transcription of the gene). The results of this indirect end-labeling assay shows a gradual fade-out of the 8 kb fragment with increasing [DNaseI], and the appearance of a new band at 6 kb with DNaseI treatment.



Result of the indirect end-label assay:



What can you infer from this information?

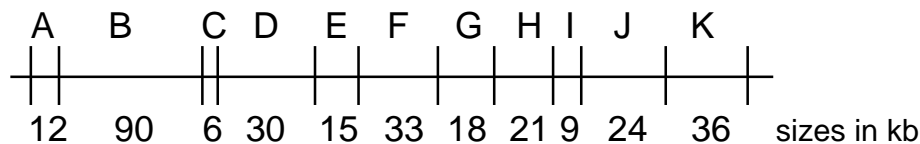
- The chromatin is very accessible to nucleases at a discrete site within the gene, located 2 kb 3' to the start site.

- b. The chromatin is very accessible to nucleases at a position 1 kb 5' to the gene, at a possible regulatory site.
- c. The chromatin is very accessible to nucleases at a discrete site at the 3' end of the gene, at a possible termination site.
- d. The chromatin is highly resistant to nucleases at a position 1 kb 5' to the gene, at a possible regulatory site.
- e. The chromatin is highly resistant to nucleases at a discrete site at the 3' end of the gene, at a possible termination site.

b is correct.

Mutations in the gene encoding the cystic fibrosis transmembrane conductance regulator (CFTR, gene name is *CFTR*) can cause cystic fibrosis, an autosomal recessive condition that is the most common genetic disease in Caucasian populations. CFTR functions as a chloride channel and controls the regulation of other transport pathways in nasal, tracheal, and bronchial epithelial tissues. For the next 8 questions, consider these simplified data that illustrate some of the current information and some potential future experiments on *CFTR*.

A1=B19=C24. (5 pts) Let's imagine that you have cloned all of the *CFTR* gene on a yeast artificial chromosome. You map the *EcoRI* sites as follows, and discover that fragments B, C, E, G and J hybridize to labeled cDNA made from mRNA isolated from normal human bronchial epithelium.



Furthermore, the non-repetitive DNA in fragments B, C, E, G and J all hybridize to the same 6500 nucleotide long mRNA in bronchial epithelium. Assuming that the 6500 nucleotide long mRNA is from the *CFTR* gene, what can you conclude about the structure of the *CFTR* gene?

- a. The *CFTR* gene is at least 132 kb long and has at least 3 introns.
- b. The *CFTR* gene is no more than 132 kb long and has at least 5 introns.
- c. The *CFTR* gene is at least 294 kb long and has at least 6 introns.
- d. The *CFTR* gene is no more than 294 kb long and has NO introns.

A is correct.

A2=B20=C25. (5 pts) Fragment C labeled at the 5' end at the B/C junction is not protected from digestion by S1 nuclease when it is hybridized to *CFTR* mRNA, but fragment B labeled at the 5' end at the B/C junction is protected from digestion by S1 nuclease when it is hybridized to *CFTR* mRNA. What is the direction of transcription with respect to the map above?

- a. Cannot be determined.
- b. 5' to 3' transcription is right to left.
- c. 5' to 3' transcription is left to right.
- d. Transcription is bi-directional.

C is correct.

A3=B21=C26. (5 pts) Analysis of replication of the *CFTR* locus using 2-dimensional gels of replicating molecules in two different tissues showed the following:

- In DNA from replication cells in bronchial epithelium, a bubble arc was detected by a probe from fragment I, and all other fragments in the YAC detected Y arcs.
- In DNA from replicating bone marrow cells, probes from all the fragments in the YAC detected Y arcs.

What can you conclude about replication of the *CFTR* locus?

- a. Multiple origins of replication are used in both are used in both bronchial epithelium and bone marrow cells.
- b. Fragment I contains an origin of replication used in bronchial epithelium but not in the bone marrow cells.
- c. Fragment I contains an terminus of replication used in bronchial epithelium but not in the bone marrow cells.

- d. Fragment I contains an origin of replication used in both bronchial epithelium and bone marrow cells.
- e. All fragments in the YAC contain origins of replication used in both bronchial epithelium and bone marrow cells.

B is correct.

A4=B22=C27. (5 pts) The most common mutation in the *CFTR* gene is similar to the following, where part of the protein-coding region is shown, and a hyphen (-) means a deletion.

CF mutant: ...ATTGCT---TGTACG...
Normal: ...ATTGCTTTCTGTACG...

If the first A in the sequence above is the beginning of a codon, what can you conclude about this mutation?

- a. A frameshift mutation changes the entire amino acid sequence encoded 3' to the deletion.
- b. The deletion brings a termination codon into frame, so the polypeptide is not translated completely.
- c. Two codons, one for leucine and one for serine, are altered by the deletion.
- d. A codon for phenylalanine has been deleted.

d is correct.

A5=B23=C28. (5 pts) Let's suppose that you have obtained the sequence of the nontemplate strand (i.e. synonymous with the mRNA) of a protein-coding exon internal to the *CFTR* gene (i.e. far downstream of the initiation codon). 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.

type fragment E is non-repetitive and hybridizes to total genomic DNA from normal individuals with a $C_{ot_{1/2}}$ of 10^4 . However, the 21kb fragment E from the mutant allele hybridizes to total genomic DNA from normal individuals with a $C_{ot_{1/2}}$ of 0.01. What do you conclude about the generation of this mutation?

- a. A 6 kb segment of highly repeated DNA inserted into fragment E.
- b. A 6 kb segment of DNA duplicated in fragment E.
- c. A 6 kb segment of DNA was deleted from fragment E.
- d. A unique segment of DNA within fragment E was amplified 10^6 times in the mutant.

A is correct.

A8=B26=C31. (5 pts) From the measured mass of polyA⁺ RNA in a bronchial epithelium cell, you determine that each cell contains about 300,000 mRNA molecules. Hybridization of labeled polyA⁺ RNA to an excess of *CFTR* cDNA drives 0.02% of the polyA⁺ RNA into duplex. What is the abundance of *CFTR* mRNA in this cell?

- a. 30,000 molecules per cell
- b. 6000 molecules per cell
- c. 300 molecules per cell
- d. 60 molecules per cell
- e. 6 molecules per cell

d is correct

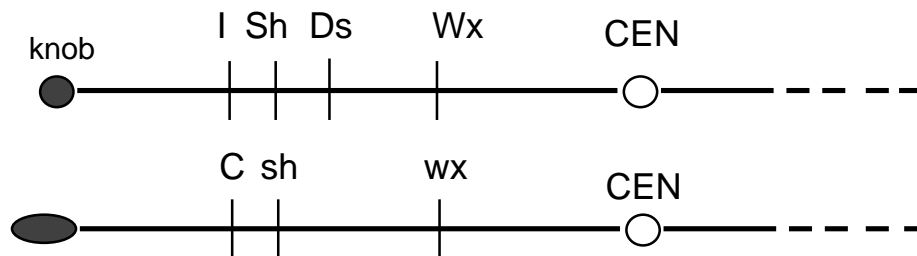
A9=B27=C19. (5 pts) Which statements about the nucleosome core (i.e. the basic structural unit of chromatin with 146 bp of DNA) are true?

- (1) Histone H1 is found in the nucleosomal core.
- (2) Histones H3 and H4 dimerize via interactions of the histone fold.
- (3) Heterodimers of H2A and H2B are in the nucleosomal core.
- (4) Histones bind to specific DNA sequences by interactions of the histone fold with the bases in the major groove of the DNA.

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

c is correct, 2 pts for d

A10=B28=C20. (5 pts) Several loci on the short arm of chromosome 9 of maize have been intensively studied. At the *C* locus, allele *C* generates a colored corn kernel, whereas the allele *I* (inhibitor) blocks *C* activity in a dominant fashion and generates colorless kernels. At the *shrunken* locus, the nonshrunken allele *Sh* is dominant to the shrunken allele *sh*. At the *waxy* locus, the dominant allele *Wx* generates starch in the endosperm, whereas the recessive allele *wx* has no detectable starch. *Ds* is a transposable element under control of *Ac*.



In a plant with the two chromosomes diagramed above and an active *Ac* element in all kernels, what do you expect to see?

- The kernels are colorless, nonshrunken and have starch.
- Approximately equal numbers of three different phenotypes are seen: (i) colored, nonshrunken, with starch, (ii) colorless, shrunken, with starch, and (iii) noncolored, nonshrunken, but no detectable starch.
- The kernels show a variegated phenotype with sectors that are colored and shrunken and have no detectable starch in the endosperm.
- The kernels show a variegated phenotype with sectors that are colored and shrunken but still have starch in the endosperm.

d. is correct.

A11=B29=C21. (5 pts) Which of the following statements about RecA is/are correct?

- [1] In the presence of ATP, RecA forms a filament on single-stranded DNA.
- [2] The RecA-coated single strand finds a region of homology in duplex DNA.
- [3] In a step utilizing ATP hydrolysis, RecA mediates assimilation of a single strand of DNA into a duplex, i.e. strand exchange.
- [4] RecA binds to Holliday junctions.

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

b is correct, 3 pts for c, 2 pts for d

A12=B30=C22. (5 pts) Which statement(s) about TBP is/are correct?

- [1] TBP binds to a TATA sequence located in the promoter of many genes transcribed by RNA polymerase II.
- [2] TBP binds in the narrow groove of DNA and bends the DNA.
- [3] TFIIA can bind to the TBP-DNA complex.

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

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

A13=B31=C23. (5 pts) Which statements about splicing of eukaryotic pre-mRNA are true?

- [1] The mechanism of splicing involves cutting of the RNA and an ATP-dependent ligation of the exons.
- [2] Introns almost invariably begin with 5' GU and end with AG 3'.
- [3] A particular snRNP binds to the branch point.
- [4] A different snRNP binds to the 5' splice site.

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

b. is correct, 2 pts for d

The Genetic Code

Position in Codon									
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.

List of Restriction Endonucleases and their Cleavage Sites.

A ' means that the nuclease cuts between these 2 nucleotides to generate a 3' hydroxyl and a 5' phosphate.

<u>Enzyme</u>	<u>Site</u>	<u>Enzyme</u>	<u>Site</u>
<i>AluI</i>	AG'CT	<i>NotI</i>	GC'GGCCGC
<i>BamHI</i>	G'GATCC	<i>PstI</i>	CTGCA'G
<i>BglII</i>	A'GATCT	<i>PvuII</i>	CAG'CTG
<i>EcoRI</i>	G'AATTC	<i>SalI</i>	G'TCGAC
<i>HaeIII</i>	GG'CC	<i>Sau3AI</i>	'GATC
<i>HhaI</i>	GCG'C	<i>SmaI</i>	CCC'GGG
<i>HincII</i>	GTY'RAC	<i>SpeI</i>	A'CTAGT
<i>HindIII</i>	A'AGCTT	<i>TaqI</i>	T'CGA

<i>Hinf</i> I	G'ANTC	<i>Xba</i> I	T'CTAGA
<i>Hpa</i> II	C'CGG	<i>Xho</i> I	C'TCGAG
<i>Kpn</i> I	GGTAC'C	<i>Xma</i> I	C'CCGGG
<i>Mbo</i> I	'GATC		

N = A,G,C or T

R = A or G

Y = C or T

Equations for complexity (N) and repetition frequency (R):

$$N_n = C_0 t_{1/2}^{\text{pure},n} \times \frac{N^{\text{std}}}{C_0 t_{1/2}^{\text{std}}}$$

$$R_n = \frac{f_n G}{N_n} = \frac{C_0 t_{1/2}^{\text{mix single copy}}}{C_0 t_{1/2}^{\text{mix } n}}$$

Equation for specificity of binding of a protein (P) to DNA (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}):

$$\text{specificity} = \frac{K_s}{K_{ns}} = \frac{[PD_s]}{[D_s]} \times \frac{[D_{ns}]}{[PD_{ns}]} = \frac{[PD_s]}{[D_s]} \times \frac{[D_{ns}]}{[P \text{ total}] - [D_s \text{ total}]}$$

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

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

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

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

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):

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

Map of the “simplified” *CFTR* locus:

