Biochemistry 465 **Your Name:** Key
Biological Information Processing **Prof. Jason Kahn. Univ. Maryland**
Exam II (100 points total) **November 20, 2008**

N = 69

You have 80 minutes for this exam.

Exams written in pencil or erasable ink will not be re-graded under any circumstances.

Explanations should be concise and clear. I have given you more space than you should need.

You do not need a calculator for this exam, and no other study aids or materials are permitted.

Generous partial credit will be given, *i.e.*, if you don't know, guess.

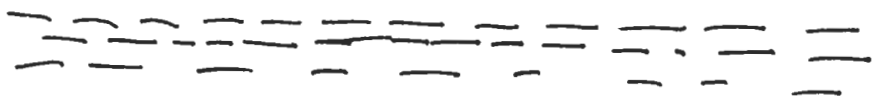
Honor Pledge: At the end of the examination time , please write out the following sentence and sign it, or talk to me about it:

"I pledge on my honor that I have not given or received any unauthorized assistance on this examination."

1. (16 pts) DNA Sequencing

(a; 10 pts) Briefly describe how a whole microbial (i.e. relatively small, ~ 2 Mb) genome is determined, given that individual sequence reads are much shorter (~ 1 kb). No need to describe any details of the sequencing chemistry.

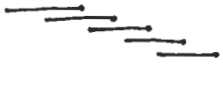
(+4)
- The genome is fragmented randomly



etc.

- Individual fragments are cloned & sequenced

(+3)
- Contigs are constructed by looking for overlapping sequences



(+3)
- Gaps are filled using mapping information from plage or other libraries with bg DNA, or by PCR across the gaps. Or - next-gen sequencing to fill gaps. (+2)

Score for page:_____

(b; 6 pts) Briefly define "bioinformatics" and describe one application.

- Bioinformatics: the use of computers to extract meaning from large volumes of sequence or other data. (+3)

(+3 for any)
- Generating phylogenetic relationships from sequences
- Predicting structure/function of new proteins
- ~~Aligning~~ Searching for SNP's that predispose to disease

2. (10 pts) Protein-DNA interaction

(a; 5 pts) Explain what sequence-specific direct readout is, and why it generally occurs in the DNA major groove rather than the minor groove.

(+2) - Recognition via ~~interaction between~~ an Hydrogen bonding between amino acids and the edges of base pairs

(+3) - The pattern of H-bond donors + acceptors is different for each base pair in the major groove, in the minor groove AT = TA and GC = CG to a first approximation.

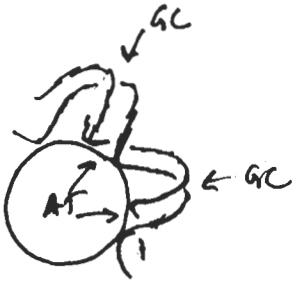
(b; 5 pts) Give an example of protein-DNA recognition by deformability (a.k.a. induced fit) : name the protein and describe in what way deformability plays a role in specific recognition.

(+2)

- TATA box binding protein - TBP - binds sequences that are most easily deformed into unwound and kinked shape (+3)

or

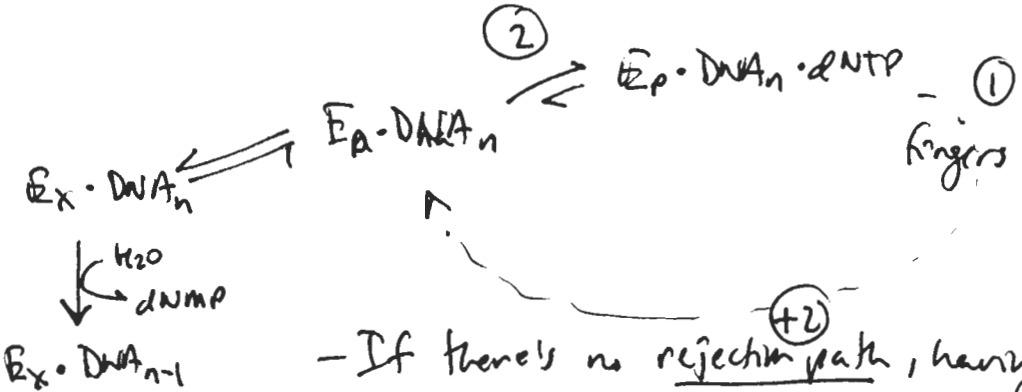
- Nucleosome positioning - Histone octamer (+2) selects sequences of alternating (AT)n/(GC)n (+3) sequences b/c AT is more easily bent to compress the minor groove (or groove bending preference)



Score for page: _____

3. (25 pts) DNA replication

(a; 9 pts) DNA polymerase fidelity relies on two independent checking steps. The first checking step is the fingers closing around the bound triphosphate, which is much slower for the incorrect triphosphate. The second checking step is very slow extension of a mismatch. Explain why the nearly nonspecific 3′→ 5′ exonuclease is essential to the actual ability to use the second checking step. Name but do not draw the type of mechanism we described for all transesterifications. State why it is important that the exonuclease reaction be irreversible.



– If there's no rejection path, having a second checking step is useless – it simply delays the inevitable sealing in of the mismatch (+2)

(+2) Two-metal-ion mechanism

(+3) If the exo were reversible, we would feed in errors by reversing the pathway and decrease overall fidelity

(b; 6 pts) Why is a polymerase error rate of $1/10^5$ adequate for a virus but not for *E. coli*? Why can an error rate of $1/10^5$ or more be tolerated in transcription and translation?

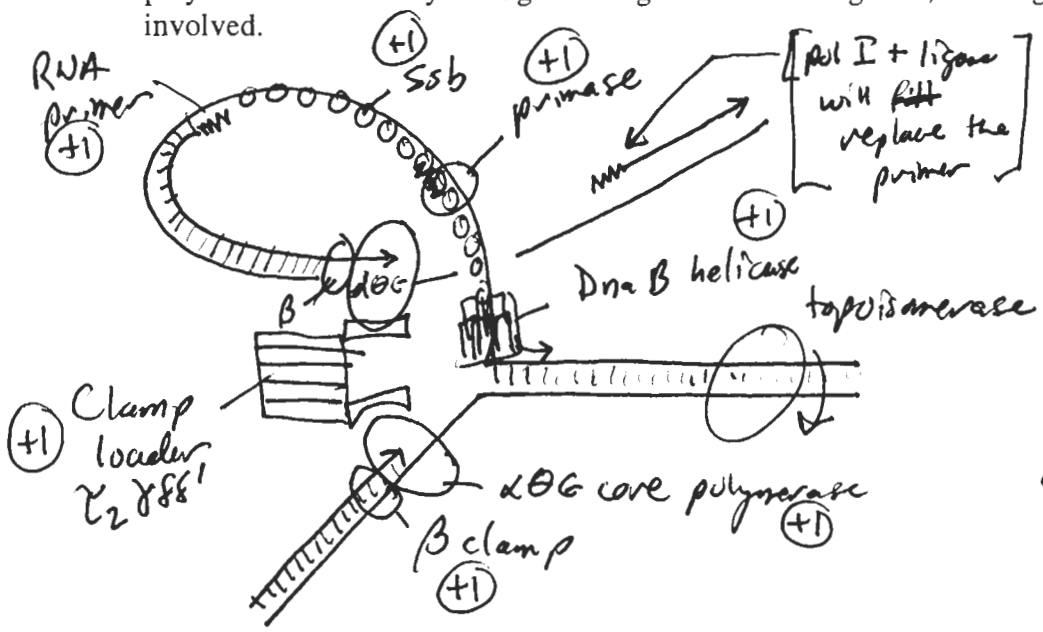
→ should have specified HIV

(+3) – A virus has a relatively small genome, and many progeny are produced at once – even if 10% of progeny are defective it's no big deal. [Also allows for rapid evolution]

(+3) – We can tolerate much larger error rates in transcription + translation because the products are transient – if they are defective they can simply be destroyed. Many mutants will still have some function anyway

Score for page:_____

(c; 10 pts) Sketch the trombone model for DNA replication when the lagging strand polymerase is half way through making an Okazaki fragment, labeling all the proteins involved.



+4 for idea of loop
+1 for approx ½ way through the fragment

extra credit
+2 for Pol I, ligase, topoisomerase if other points are missed

4. (26 pts) DNA Repair

(a; 14 pts) Briefly describe the types of DNA lesions repaired by BER, NER, and MMR. Which one has the most general scope of action, and how is it believed to recognize its target?

+3 BER: Recognizes common types of base damage like 8-oxoG or dU

+3 NER: Recognizes bulky adducts like BPDE

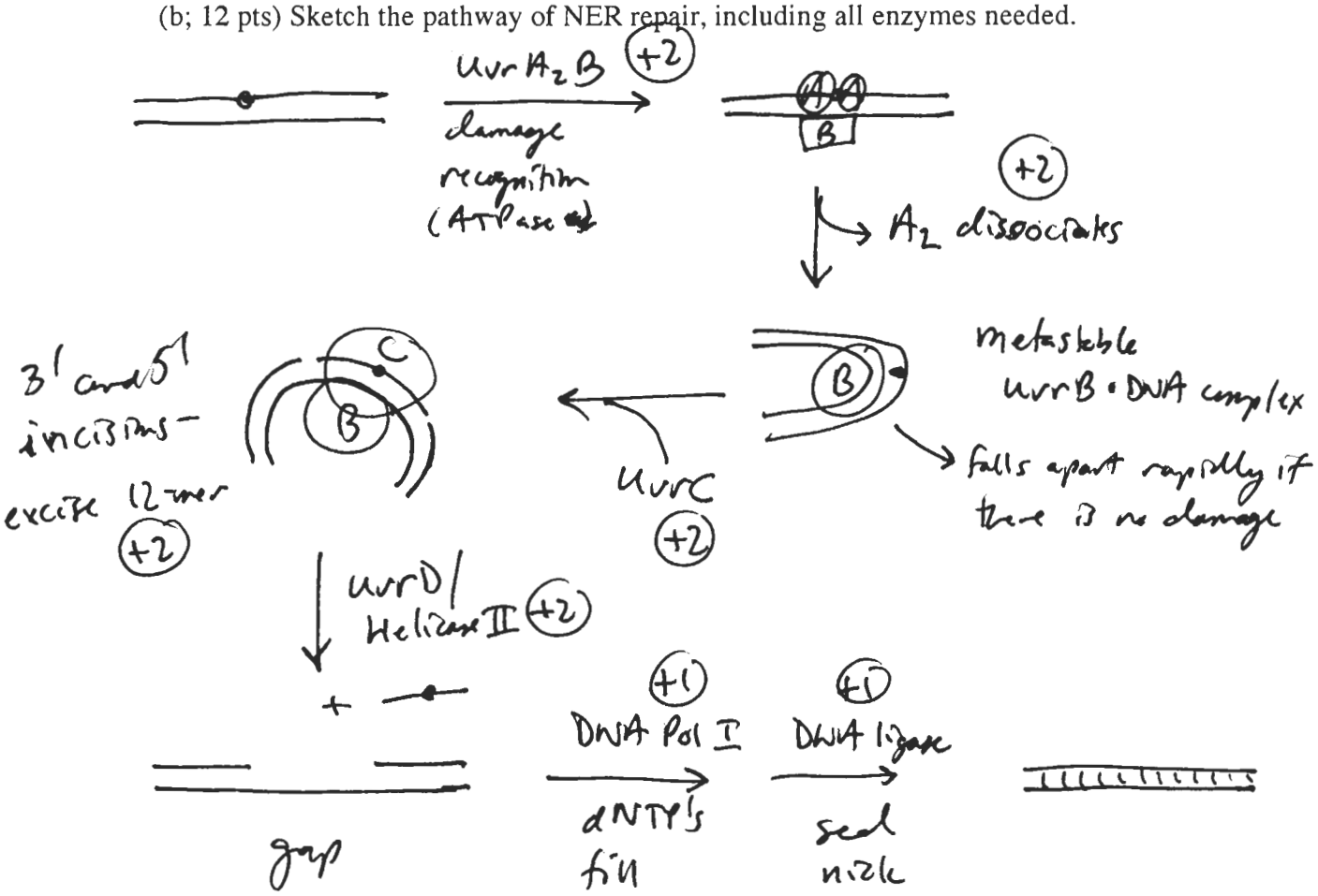
+3 MMR: Mismatches

+2 NER is the most general – it does not recognize any particular type of lesion – must work even on xenobiotics never before seen

+3 Recognition is via deformability / bendability of the DNA being altered by the lesion. [Or if a polymerase was halted at the site]

Score for page:_____

(b; 12 pts) Sketch the pathway of NER repair, including all enzymes needed.



5. (23 pts) Transcription

(a; 9 pts) How do prokaryotic and eukaryotic RNA polymerases manage to escape the promoter and progress into processive elongation (two answers)? Why must any RNA polymerase be processive?

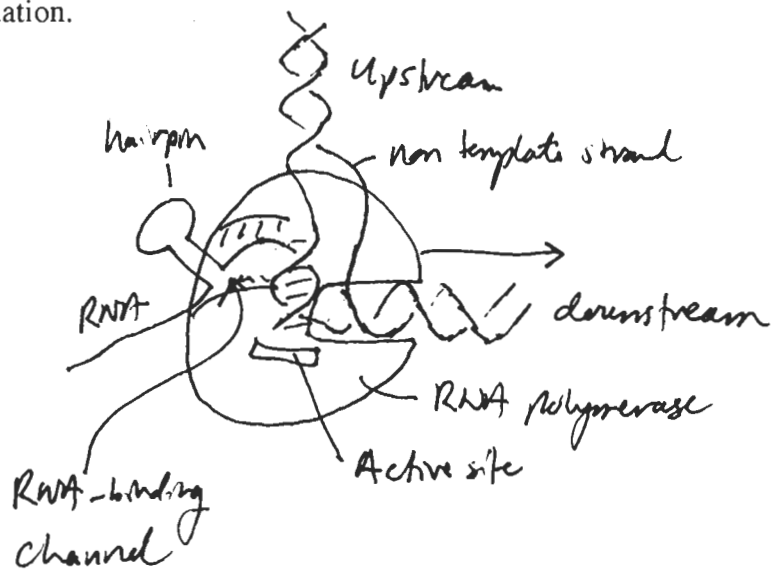
(+3) - Abortive initiation – use free energy from NTP → RNA to translocate active site forward – stored energy used to break promoter-contacts

(+3) - CTD phosphorylation – uses ATP energy to break contacts w/ proteins at the promoter

(+3) - The transcription bubble is unstable w/out bound protein – if it leaves, there's no place to return to!

Score for page:_____

(b; 8 pts) Sketch and label the E. coli ternary elongation complex (TEC) in a paused conformation.



+2 for txn bubble
+1 for RNA-DNA hybrid
+1 for active site
+2 for hairpin
+1 for RNA-binding channel
+1 for direction of motion or clarity of presentation

(c; 6 pts) What is transcription-coupled DNA repair and why does it make biological sense?

(+3) – It is the rapid repair of lesions on the ~~coding~~ template strand of transcribed genes

or

(+3) – Recognition of arrested RNAP by Mfd with subsequent termination and recruitment of NER machinery

(+3) – DNA damage is only a problem when the information actually needs to be used (DSB's are bad because they are recombinogenic.)

Page	Score
1	
2	
3	
4	
5	
6	
Total	

Score for page:_____