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).

   - The genome is fragmented randomly
   - The fragments are cloned and sequenced
   - Contigs are assembled by looking for overlapping sequences
   - Gaps are filled using mapping information from phage or other large-insert libraries with big DNA, or by PCR across the gaps. Or - next-gen sequencing to fill gaps.

Score for page:__________
Briefly define “bioinformatics” and describe one application.

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

- Generating phylogenetic relationships from sequences
- Predicting structural function of novel proteins
- Searching for sites 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.

- Recognition via interactions between base pairs and the edges of base pairs
- The pattern of H-bond donor 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.

- TATA box binding protein: TBP - binds sequences that are most easily deformed into unwound and linked shape

- Nucleosome positioning: Histone octamer selects sequence of alternating AT-rich, GC-rich segments. AT is more easily bent to compress the minor groove (or great 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 fengers closing around the bound triphosphate, which is much slower for the incorrect triphosphate. The second checking step is a 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 translesionifications. State why it is important that the exonuclease reaction be irreversible.

(b; 6 pts) Why is a polymerase error rate of 1/10^9 adequate for a virus but not for E. coli?
Why can an error rate of 1/10^9 or more be tolerated in transcription and translation?

- 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]

- We can tolerate much larger error rates in translation because the products are transient— if they are defective they can simply be destroyed. Many methods 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 halfway through making an Okazaki fragment, labeling all the proteins involved.

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

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

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

- Abortive initiation - use free energy from NTP → react to translocate active site forward - stored energy used to break promoter contacts

- CTD phosphorylation - uses ATP energy to break contacts of polymer at free promoter

- The transcription bubble is unstable about 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.

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

- **It is the rapid repair of lesions in the coding strand of a transcribed gene.**
- Recognition of arrested RNA polymerase by transcription termination and recruitment of NER machinery.
- DNA damage is only a problem when the information actually needs to be used (DSBs are bad because they are recombinogenic.)

<table>
<thead>
<tr>
<th>Page</th>
<th>Score</th>
</tr>
</thead>
<tbody>
<tr>
<td>1</td>
<td></td>
</tr>
<tr>
<td>2</td>
<td></td>
</tr>
<tr>
<td>3</td>
<td></td>
</tr>
<tr>
<td>4</td>
<td></td>
</tr>
<tr>
<td>5</td>
<td></td>
</tr>
<tr>
<td>6</td>
<td></td>
</tr>
<tr>
<td>Total</td>
<td></td>
</tr>
</tbody>
</table>

Score for page: