Please put your name and student ID # on the first page.

1. (20 pts.) Find the one letter below that best matches the following statements. You may use each letter only once.

   A. major groove  
   B. DNA glycosylase  
   C. dnaA  
   D. DNA ligase  
   E. DNA pol III  
   F. 5'-->3'  
   G. DNA ligase  
   H. adenine  
   I. plasmid  
   J. ribonuclease H  
   K. stimulates primase  
   L. minor groove  
   M. dnaC  
   N. replicative helicase  
   O. Dpn1  
   P. uracil  
   Q. 3'-->5'  
   R. primer  
   S. junk DNA  
   T. LINE  

   i)  ___C___  recognizes bacterial origin of DNA replication  
   ii) ___R___  location of the mutation in site-directed mutagenesis  
   iii) ___E___  binds sliding clamp  
   iv) ___K___  SSB  
   v)  ___J___  hydrolyzes RNA in a RNA:DNA heteroduplex  
   vi) ___D/G___  performs last step in base-excision, nucleotide-excision and mismatch repair pathways  
   vii) ___T___  the most common repeated motif in the human genome  
   viii) ___P___  the RNA base without an exocyclic amino group  
   ix)  ___F___  direction of DNA synthesis in a DNA sequencing reaction  
   x)  ___A___  displays H-bonding groups specific for each base pair  

2. (10 pts.) Your graduate advisor has given you three tubes containing solutions of either ssDNA, dsDNA or ssRNA of the same length. The labels have washed off the tubes. Briefly describe two experiments you would do to determine which solution is in which tube. How would these two experiments distinguish the samples?

   1. Distinguish ss from ds by thermal denaturation (ds melts most)  
   2. Distinguish RNA from DNA by base hydrolysis (RNA gets hydrolyzed)  

Other combinations will work
3. A BLAST search of the bacterial genome database using the sequence of the 76-amino-acid, human ubiquitin (Ub) gave the following results:

<table>
<thead>
<tr>
<th>E Value</th>
<th>Ref</th>
<th>Description</th>
</tr>
</thead>
<tbody>
<tr>
<td>5e-28</td>
<td>ref</td>
<td>YP_973019.1</td>
</tr>
<tr>
<td>3e-22</td>
<td>ref</td>
<td>YP_213464.1</td>
</tr>
<tr>
<td>5.3</td>
<td>ref</td>
<td>ZP_01741221.1</td>
</tr>
<tr>
<td>7.0</td>
<td>ref</td>
<td>ZP_01514519.1</td>
</tr>
</tbody>
</table>

3A. (8 pts.) Which of these bacterial proteins is most closely related to human ubiquitin? Explain your answer.

YP_973019. The E-value, which denotes the number of random hits of this quality in a database of this size, is the lowest for this sequence.

3B. (8 pts.) The top hit produced the following alignment:

<table>
<thead>
<tr>
<th>Length=287</th>
<th>Identities = 60/73 (82%), Positives = 71/73 (97%), Gaps = 0/73 (0%)</th>
</tr>
</thead>
<tbody>
<tr>
<td>Query 1</td>
<td>MQIFVKTLTGKITLEVEPSDTIENVKAKIQDKEGIPPDQRLIFAGKQLEDGRTLSDYN 60</td>
</tr>
<tr>
<td></td>
<td>MQ+FVK L+G+T+TL+VEPSD+IENVKAKIQD++ IPP++QRLIFAGKQLEDGRTLSDYN</td>
</tr>
<tr>
<td>Sbjct 40</td>
<td>MQVFVKMLSGETLTLDVEPSDSIENVKAKIQDQKDIPPERQRLIFAGKQLEDGRTLSDYN 99</td>
</tr>
<tr>
<td>Query 61</td>
<td>IQKESTLHLVLRL 73</td>
</tr>
<tr>
<td></td>
<td>IQK+STLHLVLRL</td>
</tr>
<tr>
<td>Sbjct 100</td>
<td>IQKDSTLHLVRL 112</td>
</tr>
</tbody>
</table>

Because Ub is thought to occur only in eukaryotes and a Ub homolog was found in only 2/867 bacterial genomes, Dr. Alber suggested that this sequence might be an error in the database due to contaminating DNA from another species. Based on the BLAST results, do you think this “bacterial” sequence could come from contaminating human DNA? Why or why not?

It’s not human DNA. There are too many changes between the bacterial and human sequences.

3C. (8 pts.) Suggest one computational experiment and one laboratory experiment you would do to test the idea that this ubiquitin gene does not occur in the sequence of the bacterial genome, but rather comes from contaminating eukaryotic DNA. Briefly describe what results you expect from these specific experiments if Dr. Alber was wrong (i.e. if the bacterial DNA really does contain this gene and the Ub sequence is not a contaminant).

Computational experiment:

BLAST the nr database with the bacterial sequence and see if there is any eukaryotic ubiquitin with the identical sequence. All the eukaryotic sequences would be different if Alber was wrong.

Laboratory experiment:

1. Isolate genomic DNA from *Acidovorax* and *Bacteroides* and do PCR with Ub primers or a Southern blot with a Ub probe. If Alber was wrong, PCR products would be produced and the expected band would hybridize to the probe in the Southern.
2. A microarray of cDNA made from the bacterial RNA preps could also show if the Ub genes are expressed.
4A. (8 pts.) John Watson says that to use RFLP analysis to distinguish DNA from different people (for example to identify a person from a sample collected at a crime scene), it’s best to use a unique-sequence probe. On the other hand, Frances Crick says that it’s easier to use a minisatellite probe. Who’s right and why?

Frances was right. The minisatellite probe gives more shots on goal (potential differences) with every restriction enzyme used to digest the genomic DNA.

4B. (8 pts.) In searching for the golden mutation in zebra fish, scientists found a RFLP that separated from the light pigmentation trait once in 1100 fish. If 1 centimorgan represents 740,000 base pairs in the zebra fish genome, what was the approximate distance from the RFLP to the mutation?

\[
\frac{1 \text{ recombinant}}{1100 \text{ fish}} \times \frac{1 \text{ centimorgan}}{1 \text{ recombinant}} \times 740,000 \text{ bp} \times \frac{1 \text{ centimorgan}}{1 \text{ recombinant}} = 67,272 \text{ base pairs}
\]

4C. (8 pts.) Given this distance, do you think the RFLP could be in the golden gene? Why or why not?

Yes, this could easily be in the same gene, because some genes are big and the recombination frequency could easily be overestimated with only one recombinant.

Or

No, the RFLP is likely in a different gene, because most genes are smaller than 67,000 base pairs and the recombination frequency could easily be underestimated with only one recombinant.

4D. (8 pts.) Inactivating mutations in the ras oncogene are found in many tumors. Surprisingly, much milder, partially inactivating ras mutations were recently found in a small number of patients with Noonan’s syndrome, an inherited disease characterized by abnormal skull features and cardiac problems. In broad terms, how would you identify candidate genes that might harbor mutations in the Noonan’s patients with normal ras genes?

Any of these answers
1. ras homologs
2. proteins in the ras pathway
3. proteins that regulate ras transcription and localization

5. (8 pts.) Please mark the following statements true or false.

__T__ Telomerase is a RNA-dependent, DNA polymerase without proof reading.

__T__ Replication forks form in pairs and move in opposite directions away from a common point of origin.

__F__ The ter protein is a symmetric dimer that binds to palindromic tus sites.

__F__ Primase uses ATP to switch conformations and determine the timing of primer synthesis.
6. Briefly contrast (explain differences between)

6A. (8 pts.) the roles of the clamp loader and the helicase loader in DNA replication

Clamp loader is a 5-subunit complex that repeatedly loads the sliding clamp in replication. It also is central to the replisome assembly. In contrast, helicase loader is comprised of a single protein in bacteria and loads the hexameric, replicative helicase at the start of replication.

6B. (8 pts.) the roles of DNA polymerase I and DNA polymerase III in bacterial DNA replication

Pol III replicates the genome as part of the replisome complex. Pol III is a large protein (>800 kDa) with many domains in addition to the polymerase active site. Pol III contains no intrinsic exonuclease activity. In contrast, Pol I completes replication of Okazaki fragments and functions in repair. It contains 5’-3’ and 3’-5’ exonuclease activities in addition to the polymerase activity. Pol I is much less processive than pol III.

6C. (8 pts.) the roles of telomerase and DNA polymerase δ in human DNA replication

Telomerase finishes the ends of chromosomes by adding tandem repeats of a hexanucleotide templated by the telomerase RNA. Telomerase is an RNP and functions as a reverse transcriptase (RNA-dependent DNA polymerase). In contrast, pol δ is the main replicative polymerase that copies the genome. The template is DNA, and pol δ requires many other proteins to function efficiently.

6D. (8 pts.) the roles of the centromere and the kinetochore in eukaryotic chromosome segregation

The centromere is the DNA sequence bound by the kinetochore protein complex. This assembly combines to mediate chromosome partitioning.
7. Nucleic-acid mimics called morpholinos are used to form heteroduplexes with RNA to block RNA function. Morpholinos have the following general structure, with the normal bases attached to a six-membered ring and the monomer units joined by a nonionic backbone linkage.

7A. (6 pts.) Do you expect the morpholino to be more or less flexible than single-stranded DNA of the same length and sequence? Briefly justify your answer.

Less flexible. It has 4 rotatable single bonds per subunit in the backbone instead of 5 in DNA. Also, the dimethyl amino group is bulkier than the equivalent O- in the DNA backbone. This bulk will cause more hindrance.

7B. (6 pts.) Do you expect the morpholino to be more or less chemically stable than RNA? Briefly explain your answer.

More chemically stable. There is no 2’-OH in the to trigger base hydrolysis.

7C. (6 pts.) Do you expect the morpholino to be more or less stable than a RNA oligonucleotide of the same sequence inside a living cell? Briefly explain your answer.

More biologically stable. RNA is rapidly degraded by nucleases in vivo. These enzymes won’t hydrolyze the morpholino backbone.

7D. (6 pts.) Do you think a heteroduplex of a morpholino with a complementary RNA strand would form a B-form double helix? Why or why not?

No. The 2’-OH in RNA blocks formation of the B-form structure.