**7.013 Central Dogma Section - Replication/Transcription/Translation**

**Part 1**

Shown below is a 240 base pair segment of a modified version of an *E. coli* gene. It includes the promoter and the first codons of the gene. The sequences of both strands of the DNA duplex are shown in Figure 1. The top strand reads 5' to 3' left to right (1 to 240); the bottom, complimentary, strand reads 5' to 3' right to left (240 to 1).

```
5' - ATGAGTTAGCTCACTTAGGGCACCCAGGCTTTACACTTTATGTTCCGGCTCGTA
1    ---------+-------------+--------------------------+  60
3' - TACACTCAATCGAGTGAATTCCGGTGGGTCGAAATGTGAAATACGAAGGGCCGAGCAT
```

```
61    --a--------+------------b+---------c-----------+  120
ACAAACACACTCTACGCTATTTGTTACAGGTGCTCTTTGTGATTTCTGATACAAA
```

```
121    +------------g-+--------------------------+  180
TGCGGTTCGAGCCTTAATTGGGAGTTTCTGACCTCGAGGTGGCAGGAGCCAC
```

```
181    +-------------x+--------------------------+  240
CGCCGGCGCATCTAGAATTAGTGGATCCCCCGGGCTGCAGCATTCGATATCA
```

**a)** RNA polymerase binds to the sequence (underlined above) and shown below.

```
5' -...CTTTACACTTT...14bp space....TATGTTG...3'
      ||||||||||        ||||||||
3' -...GAAATGTGAAA...14bp space....ATACAAC...5'
```

Once bound, RNA polymerase starts making mRNA at the 6th nucleotide after the end of the sequence (at position a, also underlined above). Synthesis of the mRNA proceeds 5' to 3' left to right on the sequence above. Write the sequence of the first 10 nucleotides of the resulting mRNA.

**b)** What are the first five amino acids of the resulting protein?

**c)** Does translation terminate at the underlined TAA at position 108 (c, bold)? Why or why not?
d) How would your answer to b) change if the C/G base pair at position 95 (d, bold) was deleted?

e) How would your answer to b) change if an A/T base pair were added between 98 & 99 (e, bold)?

f) How would your answer to b) change if the A/T base pair at position 103 (f, bold) were changed to G/C?

g) Give a single base change (substitution, deletion, or addition of a single base and it's partner on the other strand) that would cause termination of the polypeptide chain at TAA codon 147 (g, underlined).

h) Give a nonsense mutation (codon --> stop codon).

i) Give a missense mutation (codon --> codon for another amino acid).

j) Give a silent mutation (codon ---> codon for the same amino acid).

**The Genetic Code**

<table>
<thead>
<tr>
<th></th>
<th>U</th>
<th>C</th>
<th>A</th>
<th>G</th>
</tr>
</thead>
<tbody>
<tr>
<td>U</td>
<td>UUU phe</td>
<td>UCU ser</td>
<td>UAU tyr</td>
<td>UGU cys</td>
</tr>
<tr>
<td></td>
<td>UUC phe</td>
<td>UCC ser</td>
<td>UAC tyr</td>
<td>UGC cys</td>
</tr>
<tr>
<td></td>
<td>UUA leu</td>
<td>UCA ser</td>
<td>UAA STOP</td>
<td>UGA STOP</td>
</tr>
<tr>
<td></td>
<td>UUG leu</td>
<td>UCG ser</td>
<td>UAG STOP</td>
<td>UGG trp</td>
</tr>
<tr>
<td>C</td>
<td>CUU leu</td>
<td>CCU pro</td>
<td>CAU his</td>
<td>CGU arg</td>
</tr>
<tr>
<td></td>
<td>CUC leu</td>
<td>CCC pro</td>
<td>CAC his</td>
<td>CGC arg</td>
</tr>
<tr>
<td></td>
<td>CUA leu</td>
<td>CCA pro</td>
<td>CAA gln</td>
<td>CGA arg</td>
</tr>
<tr>
<td></td>
<td>CUG leu</td>
<td>CCG pro</td>
<td>CAG gln</td>
<td>CGG arg</td>
</tr>
<tr>
<td>A</td>
<td>AUU ile</td>
<td>ACU thr</td>
<td>AAU asn</td>
<td>AGU ser</td>
</tr>
<tr>
<td></td>
<td>AUC ile</td>
<td>ACC thr</td>
<td>AAC asn</td>
<td>AGC ser</td>
</tr>
<tr>
<td></td>
<td>AUA ile</td>
<td>ACA thr</td>
<td>AAA lys</td>
<td>AGA arg</td>
</tr>
<tr>
<td></td>
<td>AUG met</td>
<td>ACG thr</td>
<td>AAA lys</td>
<td>AGG arg</td>
</tr>
<tr>
<td>G</td>
<td>GUU val</td>
<td>GCU ala</td>
<td>GAU asp</td>
<td>GGU gly</td>
</tr>
<tr>
<td></td>
<td>GUC val</td>
<td>GCC ala</td>
<td>GAC asp</td>
<td>GCC gly</td>
</tr>
<tr>
<td></td>
<td>GUA val</td>
<td>GCA ala</td>
<td>GAA glu</td>
<td>GGA gly</td>
</tr>
<tr>
<td></td>
<td>GUG val</td>
<td>GCG ala</td>
<td>GAG glu</td>
<td>GGG gly</td>
</tr>
</tbody>
</table>
Part 2

Given the sequences on these next two pages, your goal is to draw a schematic of the con-6 gene. Determine the transcription start and stop sites, start and stop codons, untranslated regions, introns and exons.

5' -
CGTGAATAATACGTATGCAGGTGCTGTGCAGCATCTCATCGATAGGGAGCGAACAACCAAACCTAACATCGGATTGCA
  1 +------------------------------------------------------------------+-
  -
3' -
CCACTTTATTTATGCAGTACTGCCACGAGTCGTAGTAGCTATCCATCTCGCTGTGTTTTTGGAGTTTAGGCCTAACGT

GGACCGCCGGCGAGATTGGCTCCGGGTATCGAGACCTGGGAGTTGGAGTGGGATACCTGCCAGTTGGAGGGGATAGAAGG
  81 +------------------------------------------------------------------+-
  -
CCTGGCCGCCCTCTAAACAGGCGCACAAAATGACTGAACCTCCTTCATTCCACCTACTAATCCCATTCTTCCACTAC

GGTGCCAACTTGGCGAGAAAAGTATAAAGGCTCTTTGCTCCATCACACTCACAACCAACCAACCTCCTTCTCAATCAACCA
  161 +------------------------------------------------------------------+-
  -
CCCACGTTGAAACCGCTCCTTTCACTAATTCCAGAGAAAGGGATTGCTGTTGAGTTGAGTTTCTTGGAGAGTTTCTTCC

ATCAGAAACAAACAAACACTAACATTCCACAAAACAAACCAAACAAACAAACAAACAAACAAACAAACAAACAAACAAACAA
  241 +------------------------------------------------------------------+-
  -
TAGCTGTGGTTTGGTGGTGAAGTTAGGTTCTGATGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGTGGT

AAAATCGAAATATGTGCTGAGATGACCTGCTTGGGAGACCACCGACTCCTTGGGAGAGGAAAGGGAGTTGCTGGGAAGTTG
  321 +------------------------------------------------------------------+-
  -
TTTAGTCTGGTTATACAGGCTGAAGCCTCTTGTGTTTGTTGGGGTACGAGAAACCGCTCCTGTTCCGGTGGGAAGTTGTTG

CTAGTAGTGTATCTTTGAGACCTCAGCCTACCTCTCGACATTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCTCT
  401 +------------------------------------------------------------------+-
  -
GATCATACTAGGAGGAGTCGGAGGCAGGGAGCAGGCTTAAGAAGAAGAAAAAGATATATTGATTAGGTGTTG

TCCCAAACTCATGACTAAACCAGAATATACAGATGTCCGAGGAAACCGACTCACAAGGCGACAAGGCGACACCCTTAACC
  481 +------------------------------------------------------------------+-
  -
AGGGGTTAGTACTTGATTTGGTCTTTATAGTCTACAAAGGCTTCTCGGTCTCCTCAGGAGTTTTCTCACCAGAACCTT

CGGCCGCAGGCTTCACGATGAGCTCTTTCTCGGGCGAGGACCCGGAGCGAGCCAGCAAGAACCCGGGAAACGGTGCGG
  561 +------------------------------------------------------------------+-
  -
CGGCCGCTCCGAGTACTCAGAAGAAGCCTCGTTGTGGTGCTGCTCCGCTGTTTGGGCTTTGCAACGCC
Figure 1: Genomic DNA sequence of con-6 gene from Neurospora crassa. The sequence of both strands (5' to 3' on top, 3' to 5' on bottom) is shown above with nucleotides numbered 1 to 901. The dashed lines are interrupted every tenth nucleotide with a "+".
Figure 2: Sequence alignment of con-6 genomic DNA and mRNA sequences. The top line of each pair of sequences is the sequence of con-6 genomic DNA. The genomic DNA nucleotides are numbered as in figure 1. The bottom line is the sequence of a con-6 mRNA isolated from Neurospora crassa. The nucleotide numbers of the mRNA begin at the 5' end with #1, and end with #539 at the 3' end. Vertical dashes indicate nucleotides identical in both sequences. Dots indicate nucleotides in the genomic sequence that are not found in the mRNA sequence. (@ represents 5' G-cap)
Given the previous figures draw a schematic of the con-6 gene below. Include the transcription start and stop sites, start and stop codons, untranslated regions, introns and exons.