The Assignment

Please hand in the solutions to the following problems on Wednesday, July 28, 2010. Hand in a hard copy and a CD or USB key containing your solutions.

Problem 1

A) The nucleotide sequence of one DNA strand of a double helix is given. Write the complementary sequence found on the other strand. Notice that the new sequence you will write is on the lower strand. Do not forget to label the ends of your sequence.

              5' ––– CACTGTCATGGCTTTTGATCAAAAAAA ––– 3'

B) Search the Web for an on-line tool that will find the complement of a DNA sequence. Write down the URL.

C) Suppose that the DNA molecule from part a) is transcribed and the lower strand (from 3' to 5') is used as the template strand. What is the RNA sequence obtained from the transcription? Label the 5' and 3' ends of the molecule.

D) What is the difference between the RNA molecule you obtained and the given sequence of part a)?

Problem 2

Complete the following table. Assume that
- the columns represent transcriptional and translational alignments;
- the top DNA strand is in the 5' to 3' direction and is the coding strand, while the bottom DNA strand is the template strand and goes from 3' to 5'.

G											DNA double helix
					T	C	A				
	C	U									mRNA transcribed
							C	G	A		Appropriate tRNA anticodon
		Met (M)									Amino acids incorporated into protein

Problem 3

A) The following is a sequence of bases within a segment of a RNA molecule.

           5'––– CACUGUCACGGCUUUAGAUCAAAAAAA ––– 3'

Write the amino acid sequence that would exist in the corresponding segment of the encoded polypeptide molecule. Assume that translation has been initiated and that this sequence is in the proper reading frame (first reading frame).

B) Find an on-line RNA translation tool. Write down its URL.

C) Write all six possible reading frames from 5' to 3' of the following sequence:

5' --- GCACTAGTCATGGCTTTTGAC --- 3'

Problem 4

NCBI has a sample GenBank record at:

http://www.ncbi.nlm.nih.gov/Sitemap/samplerecord.html

Please go to that site, read the example and answer the following questions:

a) Is it better to search by using the actual accession number or the locus name? Why?

b) What are the different types of molecules contained in GenBank? Just write them out – no need to give any explanations.

c) What does GI stand for? How does it differ from accession?

d) There are three occurrences of "CDS" under "Features". Consider the third occurrence of "CDS". Explain why "/translation" ends with the specific sequence of amino acids: GSLF.

Problem 5

Go to the Swiss Institute of Bioinformatics (SIB) at http://www.isb-sib.ch/. Answer the following questions based on the information you find in their website.

a) What are the 2 missions of SIB?
b) How old is SIB?
c) When and where will "SIB/NBIC Summerschool 2010" take place?
d) A few fields of research are mentioned under "Research". Choose two, define them and give applications beyond what is written beyond the webpage.

Problem 6

Go to NCBI and retrieve nucleotide sequence with accession number Z48051.

a) What is the name of the gene and what is its locus (location on chromosome)?

b) How many exons and introns does the gene have?

c) Under CDS, explain why "/translation" ends with the following specific sequence of amino acids: LRNPF?

d) Under the second misc_feature, we have: "polymorphic (TAAA)n". Explain in your own words the meaning of this entry in the annotation of Z48051.

e) Exon 5 is from base pair 11860 to base pair 11880. What are the amino acids produced by exon 5?

f) Exon 7 is from base pair 14658 to base pair 14678. What are the amino acids produced by exon 7?

Problem 7
Consider the following sequence:

```
GCTGGATCCACTGGAGCAGGCAAGACTTCACTTCTAATGATGATTATGGGAGAACTGGAG
CCTTCAGAGGGTAAAATTAAGCACAGTGGAAGAATTTCATTCTGTTCTCAGTTTTCCTGG
ATTATGCCTGGCACCATTAAAGAAAATATCATCTTTGGTGTTTCCTATGATGAATATAGA
TACAGAAGCGTCATCAAAGCATGCCAACTAGAAGAGGACATCTCCAAGTTTGCAGAGAAA
GACAATATAGTTCTTGGAGAAGGTGGAATCACACTGAGTGGAGGTCAACGAGCAAGAATT
```

Go to the NCBI website at http://www.ncbi.nlm.nih.gov
Perform a BLASTN search of the given sequence against the NCBI nonredundant database; and answer the following questions:

a) To what organism does the selected sequence most probably belong?
b) What type of sequence is it?

From the BLASTN output, choose a human, full-length mRNA which significantly aligns with the given sequence.

c) What is the mRNA's accession number?
d) What is the mRNA's GI number?
e) On what chromosome is the mRNA sequence located?
f) What part of the sequence actually encodes for a protein?
g) What is the protein's function?

One of the hits you obtain is with the organism Sus scrofa with GI: 157279742 (NM_ 001104950).

h) When was this data entry last updated?
i) What organism is that?
j) Explain why translation ends with the specific amino acids: "VQETRL".