

## Assignment One

Please hand in the solutions to the following problems on Thursday, September 11, 2008. Hand in a hard copy (required) and a disk or CD (optional) 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' --- CACTGTCATGGCCCTTGATCAAAAAAA --- 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 sequence from part a) is transcribed and the lower strand 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

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

5'--- CACUGUCAUGGCCCUUGAUCAAAAAAA --- 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 3

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

- b) What does GI stand for? How does it differ from accession?
- c) What does CDS stand for? Explain.
- d) There are three occurrences of “CDS” under “Features”. Consider the first occurrence of “CDS”. One of its subfields is “/translation”.

Explain why “/translation” starts with the specific sequence of amino acids: SSIYN.

### **Problem 4**

CLUSTALW is available for PCs and also on a Web site at EMBL-EBI in the United Kingdom: <http://www.ebi.ac.uk/clustalw/>

The file “globin\_sequences.txt” contains seven amino acid sequences in FASTA format. Copy and paste in the sequence file into the CLUSTALW data window. Use the default conditions provided by the program, except for “Color Alignment”, choose “yes”, and run it to align the four sequences.

Note that the seven sequences are:

- HBB\_HUMAN = human hemoglobin beta chain
- HBB\_HORSE = horse hemoglobin beta chain
- HBA\_HUMAN = human hemoglobin alpha chain
- HBA\_HORSE = horse hemoglobin alpha chain
- GLB5\_PETMA = sea lamprey hemoglobin
  - Globin V is the major component of the six globins found in the sea lamprey.
- MYG\_PHYCA = sperm whale myoglobin
  - FUNCTION: serves as a reserve supply of oxygen and facilitates the movements of oxygen within muscles.
- LGB2\_LUPLU = leghaemoglobin from lupin root
  - FUNCTION: Provides oxygen to the bacteroids. This role is essential for symbiotic nitrogen fixation.
  - SUBUNIT: Monomer.
  - TISSUE SPECIFICITY: Root nodules.
  - SIMILARITY: belongs to the plant globin family.

To hand in: Copy the alignment obtained (found under: “CLUSTAL W (1.82) multiple sequence alignment”), paste it and submit it as the solution to this problem.

Note that at the top of the page that contains the alignments, there is a link to “JaIView”, a JAVA alignment viewer and editor. Go there and increase the font to view a different way of showing the alignment of the seven sequences.

## **Problem 5**

Huntington disease is a neurodegenerative disorder caused by a triplet repeat expansion of the nucleotides CAG in the huntingtin gene. This causes an expansion of a string of glutamine residues (also called polyglutamine repeats) in the huntingtin gene. We would like to learn more about the Huntington disease. A good starting place is the Online Mendelian Inheritance in Man (OMIM) that can be accessed through the National Center for Biotechnology Information (NCBI).

- Go to the NCBI web site: [www.ncbi.nlm.nih.gov](http://www.ncbi.nlm.nih.gov).
- Click on the OMIM logo (at the top) to get to the OMIM home page
- Choose 'Statistics' from under the heading 'OMIM Facts' from the list on the left hand side.
- Examine the top table. The various links in the table will call up a list of the genes fulfilling a particular mode of inheritance.

1) What is meant by "Autosomal", "X-linked", and "Y-linked"?

- Click on the green OMIM logo at the top of the page to get back to the OMIM homepage.
- Type 'Huntington' into the OMIM search window
- Press 'Go' to see the display for the OMIM record for this gene.
- Go to "Links" on the right of the first entry with accession number \*143100.
- Click on the 'Protein' link to take us to the Entrez protein record for Huntington.
- Find the entry for accession number NP\_002102 either by scrolling through the results (close to number 38), or by performing another search in this database.
- Click on accession number 'NP\_002102' to see the record.
- Examine the record and answer the following questions:

2) When was this database entry last updated in GenBank?

3) How many glutamine repeats does the protein contain?

4) Where are the repeats located? Give the precise positions in the protein.

5) How many repetitions would constitute a pathological case (and would eventually lead to Huntington disease)?