Lesson E3, Exercises

Write 4 Python programs to solve the following questions. Please name your program files as *Q1.py*, *Q2.py*, and so on, *i.e.*, according to the serial number of questions. All data files you need can be obtained from the e3 system. As for Question 5, prepare your answer in Word (saved as .doc or .docx) or pure text (.txt) format.

1. Print out the AT content of the DNA sequence recorded in the file *dna.txt*.

2. The file *protein.txt* contains a protein sequence. Compute and print out the number of 20 standard types of amino acids that the protein comprises. Your output may look like this:

   ```
   A: 12
   C: 8
   D: 1
   ... etc ...
   ```

   (Hint: A, C, D, E, F, G, H, I, K, L, M, N, P, Q, R, S, T, V, W, Y)

3. There are several blood pressure values saved in the file *blood_pressures.txt*. Write a program to read the values and print to the screen their average. (Hint: to make a string variable become a numeric variable with an integer value, you can use the function int())

4. FASTA file format is a commonly-used DNA and protein sequence file format. Sequence(s) in FASTA format look like this:

   ```
   >sequence_one
   ATCGATCGATCGATCGAT
   >sequence_two
   ACTAGCTAGCTAGCATCG
   >sequence_three
   ACTGCATCGATCGTACCT
   ```

   where the text after ">" is a header that describes the sequence (the ">" symbol indicates the start of the header line). Write a program that will create a FASTA file containing the following sequences.

Sequence header	DNA sequence
ABC123	ATCGACGATCGATCGATCGCAGACGTATCG
DEF456	ACTGATGACGATGATCGACACGACT
HIJ789	ACTGACACTGTATGTACATGTG

5. Why is "100 * x / y" better than "x / y * 100" in computer programming?