Python exercise 3

Analysing a data file

3.1. Download the file called `1abc.sec`. Use it to answer the following questions by writing and then adapting a simple Python script. You will find it helpful to open the file in a text editor first to understand what it contains.

a) Write a Python script that opens this file (you can "hardcode" this in at this stage by using the name of the file in the open statement) and simply count the number of lines, and print this back to the user.

b) Adapt your script so that the user can specify the name of the file you wish to read in. You can use the "input" function. Then adapt your program further, so that it also asks for the 3-letter code of an amino acid. The program should then only count the lines which contain this 3 letter code. Again, print the answer at the end. Bonus points if you can deal with incorrect file names or amino acid codes elegantly.

c) Next, adapt your program further, so that instead of asking for the name of an amino acid, it counts the number of times each one appears (i.e. for all the common 20 you normally encounter in proteins). You will need to think about using a dictionary for this. If you can capture the 3 letter code info from each line, you can use this as the key to this dictionary, and add one to the value for the key. At the end of the programme (after having read in all the lines), your programme should be able to loop through all the amino acids and print how many times they occur. You will probably need a few extras such as the `string.split()` method, a simple list, and a few things more.