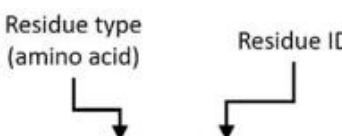


Exercise of Programming Language, Quiz E6

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.

1. Create two target folders, one is **DNA** and another is **Prot**. Copy each **.seq** file to the target folder according to the type of sequence it contains (DNA sequences to the **DNA** folder; protein sequences to the **Prot** folder). After copying, the original **.seq** files should be removed.
2. Protein Data Bank (PDB) is a database for researchers to deposit the protein structures they determined. Structure data of PDB are stored in the PDB file format. In a PDB file, information of an atom is described by a line starting with "ATOM". As shown in the following figure, in an ATOM line, the "residue type" and "residue ID" of the atom is saved at positions "18–20" and "23–27", respectively. Because a residue is composed of several atoms, it usually takes several ATOM lines to describe a residue. For instance, in the following figure, there are three residues, serine 24, histidine 25 and methionine 26.



ATOM	1	N	SER	A	24	-8.951	-8.250	-13.831	1.00	0.00	N
ATOM	2	CA	SER	A	24	-8.966	-8.695	-12.448	1.00	0.00	C
ATOM	3	C	SER	A	24	-8.430	-10.125	-12.352	1.00	0.00	C
ATOM	4	O	SER	A	24	-8.954	-11.032	-12.997	1.00	0.00	O
ATOM	11	N	HIS	A	25	-7.394	-10.281	-11.541	1.00	0.00	N
ATOM	12	CA	HIS	A	25	-6.782	-11.586	-11.353	1.00	0.00	C
ATOM	13	C	HIS	A	25	-7.626	-12.413	-10.381	1.00	0.00	C
ATOM	14	O	HIS	A	25	-7.998	-13.546	-10.685	1.00	0.00	O
ATOM	27	N	MET	A	26	-7.905	-11.815	-9.232	1.00	0.00	N
ATOM	28	CA	MET	A	26	-8.698	-12.481	-8.214	1.00	0.00	C
ATOM	29	C	MET	A	26	-9.591	-11.484	-7.474	1.00	0.00	C
ATOM	30	O	MET	A	26	-9.098	-10.539	-6.860	1.00	0.00	O

There are 9 PDB files in the PDB folder. Write a program to count the number of residues each file has and print the results to the screen. Your output may look like this (no need to sort the results):

```
1mp1A.pdb: 111 residues
1or4A.pdb: 169 residues
... (etc.)
```

3. Use **recursion** to write a function that take two arguments, one is the target folder and another is its output file. What this function do is reading all files and their contents in target folder, taking out all the contents, and then rewriting them into an output file. You may use the **Sequence** folder to test your function which should successfully make an output named **all.fasta**.

(Noting that under the target folder may have several subdirectories, each subdirectory may also have its subdirectory as well. All the file contents should be taking out either in target file or in its subdirectory.)

4. Please use **self-defined function** and **dictionary** to write a program that computes how many times each “word” occurs in the input sequence. For example:

If two witches would watch two watches, which witch would watch which watch?

Please sort the “words” in your final output (separated with TAB), which should look like this:

Word	Frequency
If	1
Two	2
watch	3
watches	1
which	2
witch	1
witches	1
would	2

The program should also pass the following article.

I see trees of green, red roses too.
I see them bloom for me and you.
And I think to myself what a wonderful world.
I see skies of blue and clouds of white.
The bright blessed day, the dark sacred night.
And I think to myself what a wonderful world.
The colors of the rainbow so pretty in the sky.
Are also on the faces of people going by.
I see friends shaking hands saying how do you do.
They're really saying I love you.
I hear babies crying, I watch them grow.
They'll learn much more than I'll never know.
And I think to myself what a wonderful world.
Yes I think to myself what a wonderful world.