Hand-in 4, due Tuesday, September 30

Background

You have been provided with a file in Absalon, which contains amino acid sequences for all known proteins in the Ecoli organism. It is in the so-called Fasta format, which means that for every protein, we first have a line starting with > containing the name:

```
>YBGC_ECOLI
```

and then some lines containing the sequence of amino acids:

```
MNTTLFRWPVRVYYEDTDAGGVVYHASYVAFYERARTEMLRHHHFSQQALMAERVAFVVRKMTVEYYAPA
RLDDMLEIQTEITSMRGTSLVFTQRIVNAENTLLNEAEVLVVCVDPLKMKPRALPKSIVAEFKQ
```

Exercises

1. Create a module called `handin4.py`, and inside that module create a function called `read_fasta` that reads a fasta file and creates a dictionary, where the keys are the names of the proteins and the sequences are the values.

   Create a new file, called `handin4_test.py` where you import the handin4 module and write some code that tests whether the function works on the Ecoli data file mentioned above.

2. In the same module, create another function called `find_prot` that takes a dictionary as first argument, and as second argument takes a protein name. The function should return the sequence corresponding to the provided name and provide an error message if the name is not present.

   Again, write some test code in `handin4_test.py` that repeatedly (e.g. 3 times) calls this function on the dictionary from before - include one case where the name is not present.

3. Finally, in the same module, create a function called `find_prot2` that takes a dictionary and regular expression (as a string), and returns all of the keys in the dictionary that the pattern matches.

   In the test program, use this function to count the number of the protein names in Ecoli that only consist of three letters before `_ECOLI` (e.g. `VSR_ECOLI`).

Please include both the Python code and the output produced from running the code. The code should be in a .py file, while the output should be a simple text file. You can easily save your output by copy&pasting from the PyCharm console window.

If your code produces external output files, please include those as well.

Please remember to comment your code, use doc-strings in function definitions and to use meaningful variable and function names!