

## Python Project: Sequence File

[FASTQ](#) is a common format for nucleotide (DNA/RNA) sequence data. Besides the actual sequence, it carries information regarding the experimental quality (certainty) for every position in the sequence, as well as miscellaneous information such as name.

There are several tools for working with this format but you should avoid using any of them for this project. This will allow you to work with a relevant format without just using a black-box solution that won't help you gain programming experience.

### Instructions:

1. Use only standard Python modules. OK to use “import” but not to download external tools for this project.
2. Download FASTQ data file named “project0.fq” from Studium (Python Course Outline). From your computer, this file can be copied to the course Jupyter server by dragging it into the file area (below the existing folders).
3. Learn the details about FASTQ format from Wikipedia (or other relevant source).
4. Subsequent steps should be solved by a Python script that you write. Only use modules from the standard Python distribution.
5. For every sequence in the file, make a table consisting of the following:
  - a. sequence name
  - b. %GC content (% of the bases which are G or C)
  - c. average quality score
6. Convert the FASTQ file to FASTA format.
7. *Extra:* If time permits, convert the nucleotide sequences to protein fragments. Assume the reading frame starting from the first position and take *stop codons* into consideration (to determine how to split the fragments).