20.180:Assignment1

Parts

- Promoter – taatacgactcactatagggaga
- RBS – attaaagaggagaaa
- ORF -
  atggcttcctccgaagacgttatcaaagagttcatgcgtttcaaagttcgtatggaaggttccgttaacggtcacgagtt
  cgaaatcgaaggtgaaggtgaaggtcgtccgtacgaaggtacccagaccgctaaactgaaagttaccaaaggtggt
  ccgctgccgttcgcttgggacatcctgtccccgcagttccagtacggttccaaagcttacgttaaacacccggctgac
  atcccggactacctgaaactgtccttcccggaaggtttcaaatgggaacgtgttatgaacttcgaagacggtggtgttg
  ttaccgttacccaggactcctccctgcaagacggtgagttcatctacaaagttaaactgcgtggtaccaacttcccgtc
  cgacggtccggttatgcagaaaaaaaccatgggttgggaagcttccaccgaacgtatgtacccggaagacggtgct
  ctgaaaggtgaaatcaaaatgcgtctgaaactgaaagacggtggtcactacgacgctgaagttaaaaccacctacat
  ggctaaaaaaccggttcagctgccgggtgcttacaaaaccgacatcaaactggacatcacctcccacaacgaagac
  tacaccatcgttgaacagtacgaacgtgctgaaggtcgtcactccaccggtgcttaataa
- Terminator -
  ccaggcatcaaataaaacgaaaggctcagtcgaaagactgggcctttcgttttatctgttgtttgtcggtgaacgctctc
  tactagagtcacactggctcaccttcgggtgggcctttctgcgtttata
- Barcode - CGCTGATAGTGCTAGTGTAGATCGC
- Use Parts.txt as your input.

```
Write your code so that it could take in any input file which has the
following structure:
key1
value1
key2
value2
key3
value3...
```

- Please plan to submit one .py file containing the code for both question 1 and question 2, named as yourathenaname_assignmentnumber.py. For example, for the first assignment, my file would be called spencers_1.py.
- Your code should create two output files, one for question 1, called output1.txt, and one for question 2, called output2.txt.
- NEW! output1.txt should contain only the DNA sequence as a single string.
- NEW! output2.txt should contain one ORF per line, and nothing else.

Submission instructions

- On a paper copy of the pset pdf, please hand write your answers to question 0 as well as the answer to this question: What will this composite part do when placed inside a living bacterium?
- Late psets will NOT be accepted.

Questions and Clarifications

- Note that the stop codon TAA must be in frame, i.e. a multiple of 3 basepairs away from the ATG. For example, ATGxxxxxxTAA would be in frame, but ATGxxxxxTAA would not be. (x is any basepair)

- Is it significant that the barcode is CAPS and the other parts are lower case?
    - *NO/no.*

- Can an ORF be any length over 50, or should its length be a multiple of some small integer?
    - *An ORF should be a length that is a multiple of three, the number of base pairs that comprise a codon*

- Does the ORF include the start ATG and stop TAA? Suppose the DNA string is "ATG...TAA": is the ORF "..." or "ATG..." or "ATG...TAA" or "...TAA"?
    - *The ORF includes the "start" ATG and "stop" TAA.*

- Can ORFs overlap? Suppose the DNA string is "ATG...TAAxxxTAA". The first ORF is obviously (modulo previous question) "ATG...TAA". Is "ATG...TAAxxxTAA" also an ORF? It meets the specification of "a string starting with ATG and ending with TAA". One could imagine a similar situation with overlapping starting tags: "ATG...ATGxxxTAA" might have both "ATG...ATGxxxTAA" and "ATGxxxTAA".
    - *Yes, ORFs can overlap.*
    - **Although "ATG...TAAxxxTAA" has a small chance of occurring in biology, for the purposes of this programming assignment, please end ORFs at the first in-frame TAA.**

- For Q2, ATG...TAA...TAA isn't an ORF, but what if ATG...TAA is less than 50 bp and ATG...TAA...TAA is >50bp?
    - *Still not an ORF (assuming the TAA's are in frame). The >50bp is something humans have used as a qualifier to weed out things that are not ORFs, since we've observed that ORFs are usually >50bp. The biology of translation will still see TAA as a stop codon and stop translation at the first TAA, making the sequence less than 50bp.*