

Homework 1

Elliott Pryor

7 September 2021

Problem 1

For the following mRNA sequence, can you extract its 5' UTR, 3' UTR and the protein sequence? ACTTGTCATGGTAACTCCGTCGTACCAGTAGGTCATG

So the above is actually DNA, since it has T's. So we assume that the T's are actually U's.

So in RNA we have: ACUUGUCAUGGUAACUCCGUCGUACCAGUAGGUCAUG

We then look for the start codon (Met) which is AUG. So we get:

ACUUGUC AUG GUA ACU CCG UCG UAC CAG UAG GUC AUG

The 5' UTR is everything left of the first AUG which is: ACUUGUC. The stop codons are either UAA, UAG, or UGA. We see UAG, so the 3' UTR is everything right of this: GUCAUG

Thus our coding region is:

AUG GUA ACU CCG UCG UAC CAG UAG → MET VAL THR PRO SER TYR GLN Stop

Problem 2 Implement the Z-algorithm in the language of your choice. Also, code up the Exact Pattern Matching algorithm that makes use of the Z algorithm. Demonstrate both algorithms on some test strings (use screen shots to show runs).

Below are some screenshots showing functionality. They were run in a python shell and imported the python file that I made for concision and clarity.

```
>>> import HW1.z_algorithm as z
>>> z.z_alg('aabbabaaa')
[0, 1, 0, 0, 1, 0, 2, 2, 1]
>>> z.z_alg('aaaaaaaa')
[0, 7, 6, 5, 4, 3, 2, 1]
>>> z.z_alg('ababbba')
[0, 0, 2, 0, 0, 0, 1]
>>> z.z_alg('aaaacaaaaaa')
[0, 4, 3, 2, 1, 0, 5, 5, 4, 3, 2, 1]
```

Figure 1: Examples using base z algorithm

```
>>> import HW1.z_algorithm as z
>>> z.z_pattern_match("aaa", "abaaabaabaaa")
[2, 9]
>>> z.z_pattern_match('aaa', 'aaaaaaaaaaaa')
[0, 1, 2, 3, 4, 5, 6, 7, 8, 9]
>>> z.z_pattern_match('ab', 'hello world')
[]
```

Figure 2: Examples using pattern matching

I also did the bonus, which runs via the command line. Here is an example below showing functionality.