Translating RNA into protein

Orr Shomroni

27 September, 2022

Problem

The 20 commonly occurring amino acids are abbreviated by using 20 letters from the English alphabet (all letters except for B, J, O, U, X, and Z). **Protein strings** are constructed from these 20 symbols. Henceforth, the term **genetic string** will incorporate protein strings along with DNA strings and RNA strings.

The **RNA codon table** dictates the details regarding the encoding of specific codons into the amino acid alphabet.

Given: An RNA string ***s*** corresponding to a strand of mRNA (of length at most 10 kbp).

Return: The protein string encoded by ***s***.

Sample Dataset

```
AUGGCCAUGGCGCCCAGAACUGAGAUCAAUAGUACCCGUAUUAACGGGUGA
```

Sample Output

```
MAMAPRTEINSTRING
```

```
f=open("rosalind_prot.txt",'r')
text=f.readlines()[0].replace("\n","")
prot=''
RNA_codon_table = {
# U
'UUU': 'Phe', 'UCU': 'Ser', 'UAU': 'Tyr', 'UGU': 'Cys', # UxU
'UUC': 'Phe', 'UCC': 'Ser', 'UAC': 'Tyr', 'UGC': 'Cys', # UxC
'UUA': 'Leu', 'UCA': 'Ser', 'UAA': '---', 'UGA': '---', # UxA
'UUG': 'Leu', 'UCG': 'Ser', 'UAG': '---', 'UGG': 'Trp', # UxG

# C
'CUU': 'Leu', 'CCU': 'Pro', 'CAU': 'His', 'CGU': 'Arg', # CxU
'CUC': 'Leu', 'CCC': 'Pro', 'CAC': 'His', 'CGC': 'Arg', # CxC
'CUA': 'Leu', 'CCA': 'Pro', 'CAA': 'Gln', 'CGA': 'Arg', # CxA
'CUG': 'Leu', 'CCG': 'Pro', 'CAG': 'Gln', 'CGG': 'Arg', # CxG

# A
'AUU': 'Ile', 'ACU': 'Thr', 'AAU': 'Asn', 'AGU': 'Ser', # AxU
'AUC': 'Ile', 'ACC': 'Thr', 'AAC': 'Asn', 'AGC': 'Ser', # AxC
'AUA': 'Ile', 'ACA': 'Thr', 'AAA': 'Lys', 'AGA': 'Arg', # AxA
```

```
'AUG': 'Met', 'ACG': 'Thr', 'AAG': 'Lys', 'AGG': 'Arg', # AxG

# G
'GUU': 'Val', 'GCU': 'Ala', 'GAU': 'Asp', 'GGU': 'Gly', # GxU
'GUC': 'Val', 'GCC': 'Ala', 'GAC': 'Asp', 'GGC': 'Gly', # GxC
'GUA': 'Val', 'GCA': 'Ala', 'GAA': 'Glu', 'GGA': 'Gly', # GxA
'GUG': 'Val', 'GCG': 'Ala', 'GAG': 'Glu', 'GGG': 'Gly'  # GxG
}
aa_letter = {'Cys': 'C', 'Asp': 'D', 'Ser': 'S', 'Gln': 'Q', 'Lys': 'K',
'Trp': 'W', 'Asn': 'N', 'Pro': 'P', 'Thr': 'T', 'Phe': 'F', 'Ala': 'A',
'Gly': 'G', 'Ile': 'I', 'Leu': 'L', 'His': 'H', 'Arg': 'R', 'Met': 'M',
'Val': 'V', 'Glu': 'E', 'Tyr': 'Y', '---': '*'}

subtext=text
while len(subtext)>0:
  tmpText=subtext[0:3]
  prot=prot+aa_letter[RNA_codon_table[tmpText]]
  subtext=subtext[3:len(subtext)]

print("RNA string: "+str(text))
```

```
## RNA string: AUGUGUAGAAAGGAAGUGUCUACCGCGUGCAUGCUCACGCCGACACCUCCUUAUGAAGGGGACCUGGAGAUUCGCACCGUAUCUAGGAU
```

```
print("RNA string: "+str(prot))
```

```
## RNA string: MCRKEVSTACMLTPTPPYEGDLEIRTVSRIVMLDIELDLVIEAIEAYVPPIGAMLSAANARWVLPGVRAQFLEGLLVLNSSIDQAPLA
```