**6 kyu**

**RNA to Protein Sequence Translation**

55491% of 69374 of766[torret](https://www.codewars.com/users/torret)

Python

* [TRAIN AGAIN](https://www.codewars.com/kata/rna-to-protein-sequence-translation/train/python)
* [NEXT KATA](https://www.codewars.com/trainer/python)

Details

[Solutions](https://www.codewars.com/kata/rna-to-protein-sequence-translation/solutions/python)

[Discourse (47)](https://www.codewars.com/kata/rna-to-protein-sequence-translation/discuss/python)

* Add to Collection
* |
* Share this kata:

The central dogma of molecular biology is that DNA is transcribed into RNA, which is then tranlsated into protein. RNA, like DNA, is a long strand of nucleic acids held together by a sugar backbone (ribose in this case). Each segment of three bases is called a *codon*. Molecular machines called ribosomes translate the RNA codons into amino acid chains, called polypeptides which are then folded into a protein.

Protein sequences are easily visualized in much the same way that DNA and RNA are, as large strings of letters. An important thing to note is that the 'Stop' codons do not encode for a specific amino acid. Their only function is to stop translation of the protein, as such they are not incorporated into the polypeptide chain. 'Stop' codons should not be in the final protein sequence. To save a you a lot of unnecessary (and boring) typing the keys and values for your amino acid dictionary are provided.

Given a string of RNA, create a funciton which translates the RNA into its protein sequence. Note: the test cases will always produce a valid string.

protein('UGCGAUGAAUGGGCUCGCUCC') returns 'CDEWARS'

Included as test cases is a real world example! The last example test case encodes for a protein called green fluorescent protein; once spliced into the genome of another organism, proteins like GFP allow biologists to visualize cellular processes!

Amino Acid Dictionary

# Phenylalanine

'UUC':'F', 'UUU':'F',

# Leucine

'UUA':'L', 'UUG':'L', 'CUU':'L', 'CUC':'L','CUA':'L','CUG':'L',

# Isoleucine

'AUU':'I', 'AUC':'I', 'AUA':'I',

# Methionine

'AUG':'M',

# Valine

'GUU':'V', 'GUC':'V', 'GUA':'V', 'GUG':'V',

# Serine

'UCU':'S', 'UCC':'S', 'UCA':'S', 'UCG':'S', 'AGU':'S', 'AGC':'S',

# Proline

'CCU':'P', 'CCC':'P', 'CCA':'P', 'CCG':'P',

# Threonine

'ACU':'T', 'ACC':'T', 'ACA':'T', 'ACG':'T',

# Alanine

'GCU':'A', 'GCC':'A', 'GCA':'A', 'GCG':'A',

# Tyrosine

'UAU':'Y', 'UAC':'Y',

# Histidine

'CAU':'H', 'CAC':'H',

# Glutamine

'CAA':'Q', 'CAG':'Q',

# Asparagine

'AAU':'N', 'AAC':'N',

# Lysine

'AAA':'K', 'AAG':'K',

# Aspartic Acid

'GAU':'D', 'GAC':'D',

# Glutamic Acid

'GAA':'E', 'GAG':'E',

# Cystine

'UGU':'C', 'UGC':'C',

# Tryptophan

'UGG':'W',

# Arginine

'CGU':'R', 'CGC':'R', 'CGA':'R', 'CGG':'R', 'AGA':'R', 'AGG':'R',

# Glycine

'GGU':'G', 'GGC':'G', 'GGA':'G', 'GGG':'G',

# Stop codon

'UAA':'Stop', 'UGA':'Stop', 'UAG':'Stop'

<https://www.codewars.com/kata/rna-to-protein-sequence-translation/python>

# Hello World program in Python

#print "Hello World!\n"

#s = "0123456789"

#print (s[4:6])

def protein(rna):

# your code here

dic = {

# Phenylalanine

'UUC':'F', 'UUU':'F',

# Leucine

'UUA':'L', 'UUG':'L', 'CUU':'L', 'CUC':'L','CUA':'L','CUG':'L',

# Isoleucine

'AUU':'I', 'AUC':'I', 'AUA':'I',

# Methionine

'AUG':'M',

# Valine

'GUU':'V', 'GUC':'V', 'GUA':'V', 'GUG':'V',

# Serine

'UCU':'S', 'UCC':'S', 'UCA':'S', 'UCG':'S', 'AGU':'S', 'AGC':'S',

# Proline

'CCU':'P', 'CCC':'P', 'CCA':'P', 'CCG':'P',

# Threonine

'ACU':'T', 'ACC':'T', 'ACA':'T', 'ACG':'T',

# Alanine

'GCU':'A', 'GCC':'A', 'GCA':'A', 'GCG':'A',

# Tyrosine

'UAU':'Y', 'UAC':'Y',

# Histidine

'CAU':'H', 'CAC':'H',

# Glutamine

'CAA':'Q', 'CAG':'Q',

# Asparagine

'AAU':'N', 'AAC':'N',

# Lysine

'AAA':'K', 'AAG':'K',

# Aspartic Acid

'GAU':'D', 'GAC':'D',

# Glutamic Acid

'GAA':'E', 'GAG':'E',

# Cystine

'UGU':'C', 'UGC':'C',

# Tryptophan

'UGG':'W',

# Arginine

'CGU':'R', 'CGC':'R', 'CGA':'R', 'CGG':'R', 'AGA':'R', 'AGG':'R',

# Glycine

'GGU':'G', 'GGC':'G', 'GGA':'G', 'GGG':'G',

# Stop codon

'UAA':'Stop', 'UGA':'Stop', 'UAG':'Stop'

}

concat = ""

i = 0

subs = ""

#'UAA':'Stop', 'UGA':'Stop', 'UAG':'Stop'

while i + 3 < len(rna):

subs = rna[i : i + 3]

#'UAA':'Stop', 'UGA':'Stop', 'UAG':'Stop'

if subs == 'UAA' or subs == 'UGA' or subs == 'UAG': break

#print(subs + " " + dic[subs])

if subs in dic:

concat += str(dic[subs])

print(subs + " " + dic[subs])

i += 3

return concat

#print(protein('AUGCAGCUUUCUAUGGAGGGUAGUGUUAACACCACGCCUGA'))

print(protein('AUGCAGCUUUCUAUGGAGGGUAGUGUUAACUACCACGCCUGA'))

#print(protein('AUGUGA'))

#print(protein('UGCGAUGAAUGGGCUCGCUCC'))

#, 'MQLSMEGSVNYHA')