RNA to protein — X95757_en

Recall that the primary structure of a protein can be represented as a sequence over the alphabet of amino acids A (alanine, Ala), R (arginine, Arg), N (asparagine, Asn), D (aspartate, Asp), C (cysteine, Cys), E (glutamate, Glu), Q (glutamine, Gln), G (glycine, Gly), H (histidine, His), I (isoleucine, Ile), L (leucine, Leu), K (lysine, Lys), M (methionine, Met), F (phenylalanine, Phe), P (proline, Pro), S (serine, Ser), T (threonine, Thr), W (tryptophan, Trp), Y (tyrosine, Tyr), and V (valine, Val).

A codon of three nucleotides is translated into a single amino acid within a protein, with translation beginning with a start codon (AUG) and ending with a stop codon (UAA, UAG, or UGA). The $4^3 = 64$ different nucleotide triplets code for 20 amino acids, one translation start signal (methionine, one of these amino acids) and three translation stop signals, with some redundancies. The genetic code defines a mapping between codons and amino acids, and despite variations in the genetic code across species, there is a standard genetic code common to most species.

AAA	K	AAC	N	AAG	K	AAU	N	ACA	T	ACC	T	ACG	T	ACU	T
AGA	R	AGC	S	AGG	R	AGU	S	AUA	I	AUC	I	AUG	M	AUU	I
CAA	Q	CAC	H	CAG	Q	CAU	H	CCA	P	CCC	P	CCG	P	CCU	P
CGA	R	CGC	R	CGG	R	CGU	R	CUA	L	CUC	L	CUG	L	CUU	L
GAA	E	GAC	D	GAG	E	GAU	D	GCA	A	GCC	A	GCG	A	GCU	A
GGA	G	GGC	G	GGG	G	GGU	G	GUA	V	GUC	V	GUG	V	GUU	V
UAA	-	UAC	Y	UAG	-	UAU	Y	UCA	S	UCC	S	UCG	S	UCU	S
UGA	-	UGC	C	UGG	W	UGU	C	UUA	L	UUC	F	UUG	L	UUU	F

Write pseudocode, Python code, and C++ code for the protein translation problem. Make two submissions, including the pseudocode as a comment to both the Python and the C++ code.

Input

The input is a string s (genomic sequence) over the alphabet $\{A, C, G, U\}$.

Output

The output is the translation of a minimal substring of s from a start codon to a stop codon to a string (proteomic sequence) over the alphabet $\{A, R, N, D, C, E, Q, G, H, I, L, K, M, F, P, S, T, W, Y, V\}$.

Sample input

```
GUCGCCAUGAUGGUGGUUAUUAUACCGUCAAGGACUGUGUGACUA
```

Sample output

```
MVVIIPSRTV
```

Problem information

Author : Gabriel Valiente
Generation : 2021-10-05 11:41:12