

HW 1: Write a Perl program that implements DNA translation to amino acid sequence

Your program should read the sequence from a data file and write the results into another data file

Submission deadline: **24 March 2019**

PS: HWs submitted after the deadline will not be accepted

Email your HW to: nizamettinaydin@gmail.com

You should place **I2B-HW1** in the "Subject" field.

You must name your file by using your matriculation number (for example, [07011068-hw1.pl](#))

Reminder:

Example 1.

Transcribe the following DNA to RNA, then use the genetic code to translate it to a sequence of amino acids.

TCATAATACGTTTTGTATTCGCCAGCGCTTCGGTGT

Answer 1.

To transcribe the DNA, first substitute each DNA for its counterpart (i.e., G for C, C for G, T for A and A for T):

TCATAATACGTTTTGTATTCGCCAGCGCTTCGGTGT
AGTATTATGCAAAACATAAGCGGTCGCGAAGCCACA

Next, remember that the Thymine (T) bases become a Uracil (U). Hence our sequence becomes:

AGUAUUAUGCAAAACAUAAAGCGGUCGCGAAGCCACA

Using the genetic code is also easy – just split the RNA sequence into triplets: :

AGU AUU AUG CAA AAC AUA AGC GGU CGC GAA GCC ACA

then look each triplet (codon) up in the genetic code table. So AGU becomes Serine, which we can write as Ser, or just S. AUU becomes Isoleucine (Ile), which we write as I. Carrying on in this way, we get:

SIMQNISGREAT

		Second letter				
		U	C	A	G	
First letter	U	UUU } Phe UUC } UUA } Leu UUG }	UCU } UCC } Ser UCA } UCG }	UAU } Tyr UAC } UAA Stop UAG Stop	UGU } Cys UGC } UGA Stop UGG Trp	U C A G
	C	CUU } CUC } Leu CUA } CUG }	CCU } CCC } Pro CCA } CCG }	CAU } His CAC } CAA } Gln CAG }	CGU } CGC } Arg CGA } CGG }	U C A G
	A	AUU } AUC } Ile AUA } AUG Met	ACU } ACC } Thr ACA } ACG }	AAU } Asn AAC } AAA } Lys AAG }	AGU } Ser AGC } AGA } Arg AGG }	U C A G
	G	GUU } GUC } Val GUA } GUG }	GCU } GCC } Ala GCA } GCG }	GAU } Asp GAC } GAA } Glu GAG }	GGU } GGC } Gly GGA } GGG }	U C A G

A=Ala=Alanine
C=Cys=Cysteine
D=Asp=Aspartic acid
E=Glu=Glutamic acid
F=Phe=Phenylalanine
G=Gly=Glycine
H=His=Histidine
I=Ile=Isoleucine
K=Lys=Lysine
L=Leu=Leucine
M=Met=Methionine
N=Asn=Asparagine
P=Pro=Proline
Q=Gln=Glutamine
R=Arg=Arginine
S=Ser=Serine
T=Thr=Threonine
V=Val=Valine
W=Trp=Tryptophan
Y=Tyr=Tyrosine