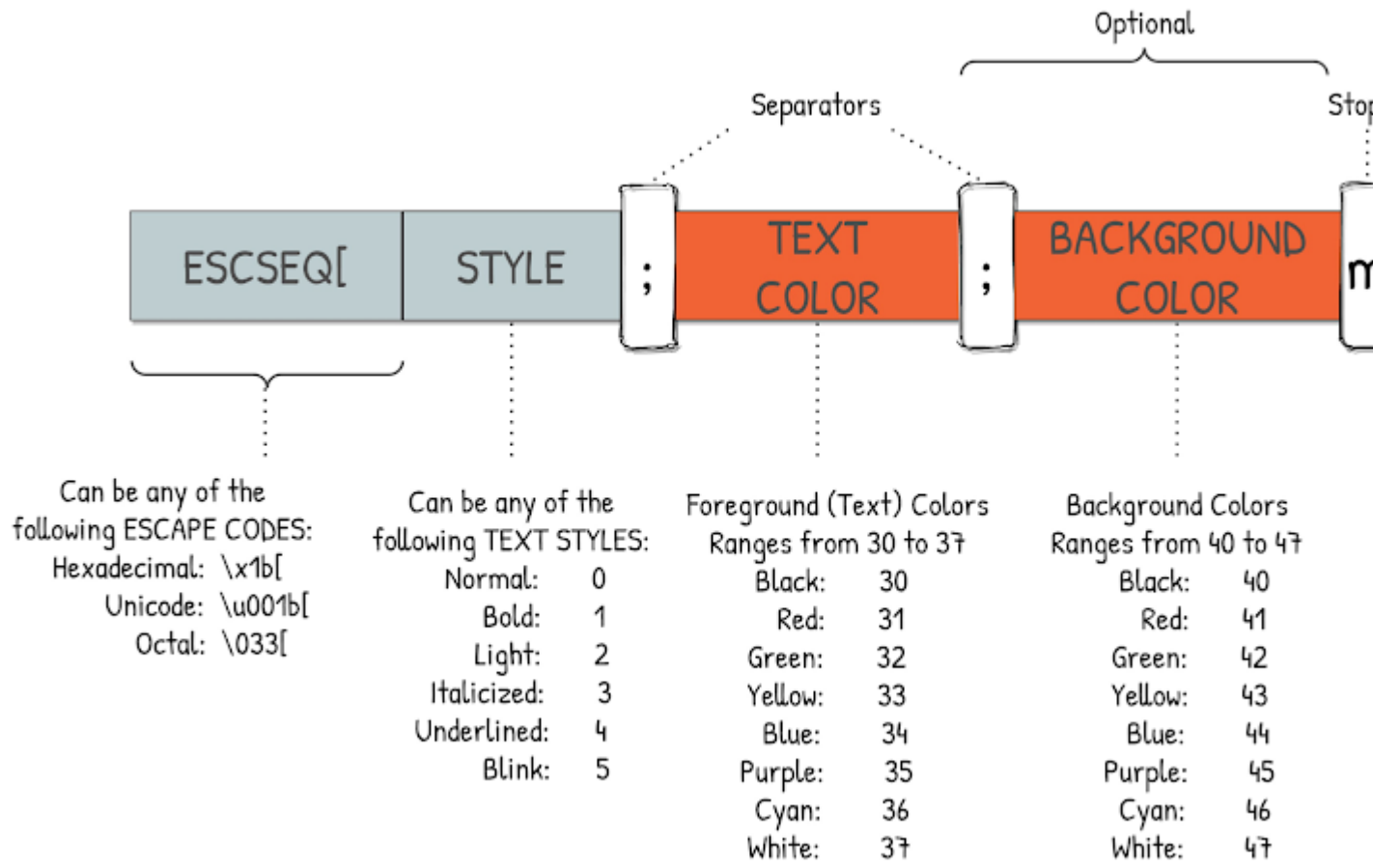


Assignment 4

DNA -> Protein (Translation) There are 64 codons which code for 20 different amino acids. Amino Acids join to form a protein sequence. There are 6 reading frames (3 forward and 3 reverse) in DNA. All these codes for different protein sequences.

Reference1: Color usage syntax in python



Or you can use colorama python lib.

```
>>> print('\033[1;30;43mA\033[0;0m\033[1;30;42mT\033[0;0m')
```

AT

```
>>> print('\033[1;30;43mA\033[0;0m\033[1;30;42mT\033[0;0m')
```

AT

Reference 2: [Codon Table \(genscript.com\)](http://genscript.com)

Write a PYTHON program to generate 192 bases(characters) using random of "ATGC"

1. Read the DNA sequence in all the 6 reading frames and print their corresponding protein sequence.

Forward frame 1

Exon sequence: ATG GAG GAG CCG CAG TCA GAT CCT AGC GTC GAG CCC CCT CTG AGT CAG GAA ACA TTT TCA GAC CTA T

AA sequence: M E E P Q S D P S V E P P L S Q E T F S D L

Forward frame 2

Exon sequence: GAG GAG CCG CAG TCA GAT CCT AGC GTC GAG CCC CCT CTG AGT CAG GAA ACA TTT TCA GAC CTA TGG A

AA sequence: E E P Q S D P S V E P P L S Q E T F S D L W

2. Print the number of start (ATG) and stop codons (TAA, TAG, TGA) and their positions in all the six reading frames
3. Translate the DNA sequence from the start codon

Note:

64 Codons = $64 \times 3 = 192$ Amino acids

ATGCAATTGGCC.....GTGATT (192 bases)

Forward 1...192

ATG.....

Reverse 192..1

TTAGTG.....TA

Amino Acids
Alanine (A)
Arginine (R)
Asparagine (N)
Aspartate (D)
Cysteine (C)
Glutamate (E)
Glutamine (Q)
Glycine (G)
Histidine (H)
Isoleucine (I), Leucine (L)
Lysine (K)
Methionine (M)
Phenylalanine (F)
Proline (P)
Serine (S)
Threonine (T)
Tryptophan (W)
Tyrosine (Y)
Valine (V)

Expected output

Forward:

Dna seq:

ACC AAG GTA ACA TGT CTC TAT CTC CCG TTT GCT TAT TTA TCC CTC CCA ATT GGA GGC TAC TAC AGC GCC TGG TCA CTG
T AAA GCG ACA GCA TGT GGT TCT CTA GAA GCA AGT TTC CTG CAA TAG CAA TGT GGC CTC GCC TTA TAG GCG ATC ACT CA
AGC TAC CCT AAA AGT CCC GGT

protein:

T K V T C L Y L P F A Y L S L P I G G Y Y S A W S L P L P K A T A C G
A S F L Q Q C G L A L A I T H R Y S Y P K S P G

Dna seq:

AAG GTA ACA TGT CTC TAT CTC CCG TTT GCT TAT TTA TCC CTC CCA ATT GGA GGC TAC TAC AGC GCC TGG TCA CTG CCT
A GCG ACA GCA TGT GGT TCT CTA GAA GCA AGT TTC CTG CAA TAG CAA TGT GGC CTC GCC TTA TAG GCG ATC ACT CAC AG
TAC CCT AAA AGT CCC GGT

protein:

K V T C L Y L P F A Y L S L P I G G Y Y S A W S L P L P K A T A C G S
S F L Q Q C G L A L A I T H R Y S Y P K S P G

Reverse:

Dna seq:

TGG CCC TGA AAA TCC CAT CGA CAT GGA CAC TCA CTA GCG GAT ATT CCG CTC CGG TGT AAC GAT AAC GTC CTT TGA ACG
T TGG TGT ACG ACA GCG AAA TCC GTC TCC GTC ACT GGT CCG CGA CAT CAT CGG AGG TTA ACC CTC CCT ATT TAT TCG TT
TAT CTC TGT ACA ATG GAA CCA

protein:

W P K S H R H G H S L A D I P L R C N D N V L T K I S W C T T A K S V
G F R H H R R L T L P I Y S F A L Y L C T M E P

Dna seq:

TGG CCC TGA AAA TCC CAT CGA CAT GGA CAC TCA CTA GCG GAT ATT CCG CTC CGG TGT AAC GAT AAC GTC CTT TGA ACG
T TGG TGT ACG ACA GCG AAA TCC GTC TCC GTC ACT GGT CCG CGA CAT CAT CGG AGG TTA ACC CTC CCT ATT TAT TCG TT
TAT CTC TGT ACA ATG GAA