

Assignment #2

The goal of this assignment is to recreate the functionality of the tool found at the following webpage:

<http://web.expasy.org/translate/>

This webpage allows you to translate DNA sequence into the encoded amino acids. Note all six possible frames are translated – the three possible forward frames along with the three possible frames of the reverse complement.

This tool also allows you to display the sequences in three possible modes: Verbose, Compact, Include nucleotide sequence. Your code should also do this by accepting a single argument that determines which mode that will be used.

Your code should then prompt the user for DNA sequence, print out the output, and continue until they want to quit.

Your program does not need to accept different genetic codes or perform any color/highlighting to the text in the output. Standard text output is fine.

To help you get started, I have created a template script that you must use for your code. Additional functionality or ambiguity is clarified in the comments of that sample code. I have also shown some sample output from the code that I have written for this assignment.

Example output:

```
$ python3 Assignment2_Solution.py
Invalid number of options
Usage: python3 Assignment2_solution.py <mode>
Mode can be one of the following options:
    COMPACT
    VERBOSE
    DNA

$ python3 Assignment2_Solution.py Compact Verbose
Invalid number of options
Usage: python3 Assignment2_solution.py <mode>
Mode can be one of the following options:
    COMPACT
    VERBOSE
    DNA

$ python3 Assignment2_Solution.py Compac
COMPAC not a valid option
Usage: python3 Assignment2_solution.py <mode>
Mode can be one of the following options:
    COMPACT
    VERBOSE
    DNA

$ python3 Assignment2_Solution.py Compact
Enter DNA sequence (or Exit to quit the program): ;sdja;sdf;lkajsdf
Invalid DNA sequence. Characters must be one of A, a, C, c, G, g, T, or t
Enter DNA sequence (or Exit to quit the program): ASDJFLS:FJKEWL:LKJFKL:
Invalid DNA sequence. Characters must be one of A, a, C, c, G, g, T, or t
Enter DNA sequence (or Exit to quit the program): exit
```

```
$ python3 Assignment2_Solution.py Compact
Enter DNA sequence (or Exit to quit the program):
ATGACGGAGTACAAGCTTGTGGTAGTTGGAGATGGAGGAGTTGGTAAATCAGCACTCACCATTCAACTCATCCAGAATCACTTTGTCTGA
AGAATACGACCCGACCATAGAGGACAGCTACAGAAAGCAGGTTGTGATAGACGGTGAGACATGCCTCCTCGACATATTGGATACCGCCG
GACAAGAAGAATATTCGGCGATGCGTGATCAGTACATGAGGACAGGCGAAGGATTTCTGTTGGTTTTCGCCGTCAACGAGGCTAAATCT
TTCGAGAATGTCGCTAACTACCGCGAGCAGATTTCGGAGGGTAAAGGATTTCAGATGATGTTCTTATGGTCTTGGTAGGGAATAAATGTGA
TTTGTTCATCTCGATCAGTCGACTTCCGAACAGTCAGTGAGACAGCAAAGGGTTACGGTATTCCGAATGTCGACACATCTGCCAAAACGC
GTATGGGAGTTGATGAAGCATTTTACACACTTGTTAGAGAAATTCGCAAGCATCGTGAGCGTCACGACAATAATAAGCCACAAAAGAAG
AAGAAGTGTCAAATAATGTGA
5' to 3' Frame: 0
MTEYKLVVVG DGGV GKSALTIQLIQNH FVEEYDPTIEDSYRKQVVIDGETCLLDILD TAGQEEYSAMRDQYMR TGEGFLLVFAVNEAKS
FENVANYREQIRRVKSDDDVPMVLVGNKCDLSSRSVDFRTVSETAKGYGIPNVDTS AKTRMGVDEAFYTLVREIRK HRE RHDNNKPQKK
KKCQIM-
```

```
5' to 3' Frame: 1
-RSTSLW-LEMEELVNQHSPFNSSRITLSKNTTRP-RTATESRL--TVRHASSTYWIPD KKNIRRCVIST-
GQAKDFCWFSPTRLNLSRMSLT TASRFGG-RIQMMFLWSW-
GINVICHLDQSTSEQSVRQQRVT VFRMSTHLPKR V WELMKHFTHLLEKFASIVSVTTIISHKR RRSVK-C
```

```
5' to 3' Frame: 2
DGVQACGSWRWRWSW-ISTHHSTHPESLCRRIRPDHRGQLQKAGCDRR-DMPPRHIGYRRTRRIFGDA-SVHEDRRRISVGFR RQRG-
IFRECR-LPRADSEGKGFR-CSYGLGRE-M-FVISISRLPNSQ-DSKGLRYSECRHICQNAYGS--SILHTC-RNSQAS-ASRQ--
ATKEEEVSNNV
```

```
3' to 5' Frame: 0
SHYLTLLLLLWLIIVVTLTMLANFSNKC VKCFINSHTRFGRCDIRNTVTLCCLTDCSEVD-SR-QITFIPYQDHRNII-
ILYPNLLAVVSDILERFSLVDGENQQKS FACPHVLITHRRIFFLSGGIQYVEEACLT VYHNLLSVAVLYGRVVFDDKVLDELNGEC-
FTNSSISNYHKLVLRH
```

```
3' to 5' Frame: 1
HII-HFFFFCGLLLLS-RSRLRISLTSV-NASSTPIRVLADVSTFGIP-
PFAVSLTVRKSTDRDDKSHLFPTKTIGTSSESFTLRICSR-LATFSKDLASLTAKTNRNPSPVLMY-
SRIAEYSSCPAVSNMSRRHVSPSITTCFL-LSSMVGYSSTK-FWMS-MVSADLPTPPSP TTTSLYSV
```

```
3' to 5' Frame: 2
TLFDTSSSFVAYYCRDAHDACEFL-QVCKMLHQLPYAFWQMCRHSEYRNPLLSH-LFGSR LIEMTNHIYSLPRP-
EHHLNPLPSESARGS-RHSRKI-PR-RRKPTEILRLSSCTDHASPNILLVRRYPICRGMSHRLSQPAFCSCPLWSGRILRQSDSG-
VEW-VLIYQLLHLQLPQACTPS
```

Enter DNA sequence (or Exit to quit the program): exit

```
$ python3 Assignment2_Solution.py Verbose
Enter DNA sequence (or Exit to quit the program):
ATGACGGAGTACAAGCTTGTGGTAGTTGGAGATGGAGGAGTTGGTAAATCAGCACTCACCATTCAACTCATCCAGAATCACTTTGTCTGA
AGAATACGACCCGACCATAGAGGACAGCTACAGAAAGCAGGTTGTGATAGACGGTGAGACATGCCTCCTCGACATATTGGATACCGCCG
GACAAGAAGAATATTCGGCGATGCGTGATCAGTACATGAGGACAGGCGAAGGATTTCTGTTGGTTTTCGCCGTCAACGAGGCTAAATCT
TTCGAGAATGTCGCTAACTACCGCGAGCAGATTTCGGAGGGTAAAGGATTTCAGATGATGTTCTTATGGTCTTGGTAGGGAATAAATGTGA
TTTGTTCATCTCGATCAGTCGACTTCCGAACAGTCAGTGAGACAGCAAAGGGTTACGGTATTCCGAATGTCGACACATCTGCCAAAACGC
GTATGGGAGTTGATGAAGCATTTTACACACTTGTTAGAGAAATTCGCAAGCATCGTGAGCGTCACGACAATAATAAGCCACAAAAGAAG
AAGAAGTGTCAAATAATGTGA
5' to 3' Frame: 0
Met T E Y K L V V V G D G G V G K S A L T I Q L I Q N H F V E E Y D P T I E D S Y R K Q V
V I D G E T C L L D I L D T A G Q E E Y S A Met R D Q Y Met R T G E G F L L V F A V N E A
K S F E N V A N Y R E Q I R R V K D S D D V P Met V L V G N K C D L S S R S V D F R T V S
E T A K G Y G I P N V D T S A K T R Met G V D E A F Y T L V R E I R K H R E R H D N N K P
Q K K K K C Q I Met Stop
```

```
5' to 3' Frame: 1
Stop R S T S L W Stop L E Met E E L V N Q H S P F N S S R I T L S K N T T R P Stop R T A
T E S R L Stop Stop T V R H A S S T Y W I P P D K K N I R R C V I S T Stop G Q A K D F C
W F S P S T R L N L S R Met S L T T A S R F G G Stop R I Q Met Met F L W S W Stop G I N V
I C H L D Q S T S E Q S V R Q Q R V T V F R Met S T H L P K R V W E L Met K H F T H L L E
K F A S I V S V T T I I S H K R R R S V K Stop C
```

5' to 3' Frame: 2

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

3' to 5' Frame: 0

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

3' to 5' Frame: 1

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

3' to 5' Frame: 2

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

Enter DNA sequence (or Exit to quit the program): exit

\$ python3 Assignment2_Solution.py DNA

Enter DNA sequence (or Exit to quit the program):

ATGACGGAGTACAAGCTTGTGGTAGTTGGAGATGGAGGAGTTGGTAAATCAGCACTCACCATTCAACTCATCCAGAATCACTTTGTGCGA
AGAATACGACCCGACCATAGAGGACAGCTACAGAAAGCAGGTTGTGATAGACGGTGAGACATGCCTCCTCGACATATTGGATACCGCCG
GACAAGAAGAATATTCGGCGATGCGTGATCAGTACATGAGGACAGGCGAAGGATTTCTGTTGGTTTTCGCCGTCAACGAGGCTAAATCT
TTCGAGAATGTCGCTAACTACCGCGAGCAGATTTCGGAGGGTAAAGGATTTCAGATGATGTTCTATGGTCTTGGTAGGGAATAAATGTGA
TTTGTTCATCTCGATCAGTCGACTTCCGAACAGTCAGTGAGACAGCAAAGGGTTACGGTATTCCGAATGTCGACACATCTGCCAAAACGC
GTATGGGAGTTGATGAAGCATTTTACACACTTGTTAGAGAAATTCGCAAGCATCGTGAGCGTCACGACAATAATAAGCCACAAAAGAAG
AAGAAGTGTCAAATAATGTGA

5' to 3' Frame: 0

ATGACGGAGTACAAGCTTGTGGTAGTTGGAGATGGAGGAGTTGGTAAATCAGCACTCACC
M T E Y K L V V V G D G G V G K S A L T
ATTCAACTCATCCAGAATCACTTTGTGCGAAGAATACGACCCGACCATAGAGGACAGCTAC
I Q L I Q N H F V E E Y D P T I E D S Y
AGAAAGCAGGTTGTGATAGACGGTGAGACATGCCTCCTCGACATATTGGATACCGCCGGA
R K Q V V I D G E T C L L D I L D T A G
CAAGAAGAATATTCGGCGATGCGTGATCAGTACATGAGGACAGGCGAAGGATTTCTGTTG
Q E E Y S A M R D Q Y M R T G E G F L L
GTTTTCGCCGTCAACGAGGCTAAATCTTTCGAGAATGTCGCTAACTACCGCGAGCAGATT
V F A V N E A K S F E N V A N Y R E Q I
CGGAGGGTAAAGGATTTCAGATGATGTTCTATGGTCTTGGTAGGGAATAAATGTGATTTG
R R V K D S D D V P M V L V G N K C D L
TCATCTCGATCAGTCGACTTCCGAACAGTCAGTGAGACAGCAAAGGGTTACGGTATTCCG
S S R S V D F R T V S E T A K G Y G I P
AATGTCGACACATCTGCCAAAACGCGTATGGGAGTTGATGAAGCATTTTACACACTTGTT
N V D T S A K T R M G V D E A F Y T L V
AGAGAAATTCGCAAGCATCGTGAGCGTCACGACAATAATAAGCCACAAAAGAAGAAG
R E I R K H R E R H D N N K P Q K K K K
TGTCAAATAATGTGA
C Q I M -

5' to 3' Frame: 1

TGACGGAGTACAAGCTTGTGGTAGTTGGAGATGGAGGAGTTGGTAAATCAGCACTCACCA
- R S T S L W - L E M E E L V N Q H S P
TTCAACTCATCCAGAATCACTTTGTCTGAAGAATACGACCCGACCATAGAGGACAGCTACA
F N S S R I T L S K N T T R P - R T A T
GAAAGCAGGTTGTGATAGACGGTGAGACATGCCTCCTCGACATATTGGATACCGCCGGAC
E S R L - - T V R H A S S T Y W I P P D
AAGAAGAATATTCGGCGATGCGTGATCAGTACATGAGGACAGGCGAAGGATTTCTGTTGG
K K N I R R C V I S T - G Q A K D F C W
TTTTTCGCCGTCAACGAGGCTAAATCTTTTCGAGAATGTCGCTAACTACCGCGAGCAGATT
F S P S T R L N L S R M S L T T A S R F
GGAGGGTAAAGGATTCAGATGATGTTTCTATGGTCTTGGTAGGGAATAAATGTGATTTGT
G G - R I Q M M F L W S W - G I N V I C
CATCTCGATCAGTCGACTTCCGAACAGTCAGTGAGACAGCAAAGGGTTACGGTATTCCGA
H L D Q S T S E Q S V R Q Q R V T V F R
ATGTCGACACATCTGCCAAAACGCGTATGGGAGTTGATGAAGCATTTTACACACTTGTTA
M S T H L P K R V W E L M K H F T H L L
GAGAAATTCGCAAGCATCGTGAGCGTCACGACAATAATAAGCCACAAAAGAAGAAGAAGT
E K F A S I V S V T T I I S H K R R R S
GTCAAATAATGT
V K - C

5' to 3' Frame: 2

GACGGAGTACAAGCTTGTGGTAGTTGGAGATGGAGGAGTTGGTAAATCAGCACTCACCAT
D G V Q A C G S W R W R S W - I S T H H
TCAACTCATCCAGAATCACTTTGTCTGAAGAATACGACCCGACCATAGAGGACAGCTACAG
S T H P E S L C R R I R P D H R G Q L Q
AAAGCAGGTTGTGATAGACGGTGAGACATGCCTCCTCGACATATTGGATACCGCCGGACA
K A G C D R R - D M P P R H I G Y R R T
AGAAGAATATTCGGCGATGCGTGATCAGTACATGAGGACAGGCGAAGGATTTCTGTTGGT
R R I F G G D A - S V H E D R R R I S V G
TTTCGCCGTCAACGAGGCTAAATCTTTTCGAGAATGTCGCTAACTACCGCGAGCAGATTCTG
F R R Q R G - I F R E C R - L P R A D S
GAGGGTAAAGGATTCAGATGATGTTTCTATGGTCTTGGTAGGGAATAAATGTGATTTGTC
E G K G F R - C S Y G L G R E - M - F V
ATCTCGATCAGTCGACTTCCGAACAGTCAGTGAGACAGCAAAGGGTTACGGTATTCCGAA
I S I S R L P N S Q - D S K G L R Y S E
TGTCGACACATCTGCCAAAACGCGTATGGGAGTTGATGAAGCATTTTACACACTTGTTAG
C R H I C Q N A Y G S - - S I L H T C -
AGAAATTCGCAAGCATCGTGAGCGTCACGACAATAATAAGCCACAAAAGAAGAAGAAGTG
R N S Q A S - A S R Q - - A T K E E E V
TCAAATAATGTG
S N N V

3' to 5' Frame: 0

TCACATTATTTGACACTTCTTCTTTCTTTTGTGGCTTATTATTGTCGTGACGCTCACGATG
S H Y L T L L L L W L I I V V T L T M
CTTGCGAATTTCTCTAACAAGTGTGTAATAATGCTTCATCAACTCCCATACGCGTTTTGGC
L A N F S N K C V K C F I N S H T R F G
AGATGTGTCTGACATTTCGGAATACCGTAACCCCTTTGCTGTCTCACTGACTGTTTCGGAAGTC
R C V D I R N T V T L C C L T D C S E V
GACTGATCGAGATGACAAATCACATTTATTCCCTACCAAGACCATAGGAACATCATCTGA
D - S R - Q I T F I P Y Q D H R N I I -
ATCCTTTTACCCTCCGAATCTGCTCGCGGTAGTTAGCGACATTCTCGAAAGATTTAGCCTC
I L Y P P N L L A V V S D I L E R F S L
GTTGACGGCGAAAACCAACAGAAATCCTTCGCCTGTCCTCATGTACTGATCACGCATCGC
V D G E N Q Q K S F A C P H V L I T H R
CGAATATTCTTCTTGTCCGGCGGTATCCAATATGTCGAGGAGGCATGTCTCACCGTCTAT
R I F F L S G G I Q Y V E E A C L T V Y
CACAACCTGCTTTCTGTAGCTGTCCTCTATGGTCGGGTCGTATTCTTCGACAAAGTGATT
H N L L S V A V L Y G R V V F F D K V I
CTGGATGAGTTGAATGGTGAGTGCTGATTTACCAACTCCTCCATCTCCAACCTACCACAAG
L D E L N G E C - F T N S S I S N Y H K
CTTGACTCCGTCAT

L V L R H

3' to 5' Frame: 1

CACATTATTTGACACTTCTTCTTCTTTTGTGGCTTATTATTGTCGTGACGCTCACGATGC
H I I - H F F F F C G L L L S - R S R C
TTGCGAATTTCTCTAACAAGTGTGTAAAATGCTTCATCAACTCCCATACGCGTTTTGGCA
L R I S L T S V - N A S S T P I R V L A
GATGTGTCGACATTTCGGAATACCGTAACCCTTTGCTGTCTCACTGACTGTTTCGGAAGTCG
D V S T F G I P - P F A V S L T V R K S
ACTGATCGAGATGACAAATCACATTTATTCCCTACCAAGACCATAGGAACATCATCTGAA
T D R D D K S H L F P T K T I G T S S E
TCCTTTACCCTCCGAATCTGCTCGCGGTAGTTAGCGACATTCTCGAAAGATTTAGCCTCG
S F T L R I C S R - L A T F S K D L A S
TTGACGGCGAAAACCAACAGAAATCCTTCGCCTGTCCTCATGTACTGATCACGCATCGCC
L T A K T N R N P S P V L M Y - S R I A
GAATATTCTTCTTGTCCGGCGGTATCCAATATGTCGAGGAGGCATGTCTACCGTCTATC
E Y S S C P A V S N M S R R H V S P S I
ACAACCTGCTTTCTGTAGCTGTCCTCTATGGTCGGGTCGTATTCTTCGACAAAGTGATTCT
T T C F L - L S S M V G S Y S S T K - F
TGGATGAGTTGAATGGTGAGTGCTGATTTACCAACTCCTCCATCTCCAACCTACCACAAGC
W M S - M V S A D L P T P P S P T T T S
TTGTACTCCGTC
L Y S V

3' to 5' Frame: 2

ACATTATTTGACACTTCTTCTTCTTTTGTGGCTTATTATTGTCGTGACGCTCACGATGCT
T L F D T S S S F V A Y Y C R D A H D A
TGCGAATTTCTCTAACAAGTGTGTAAAATGCTTCATCAACTCCCATACGCGTTTTGGCAG
C E F L - Q V C K M L H Q L P Y A F W Q
ATGTGTCGACATTTCGGAATACCGTAACCCTTTGCTGTCTCACTGACTGTTTCGGAAGTCGA
M C R H S E Y R N P L L S H - L F G S R
CTGATCGAGATGACAAATCACATTTATTCCCTACCAAGACCATAGGAACATCATCTGAAT
L I E M T N H I Y S L P R P - E H H L N
CCTTTACCCTCCGAATCTGCTCGCGGTAGTTAGCGACATTCTCGAAAGATTTAGCCTCGT
P L P S E S A R G S - R H S R K I - P R
TGACGGCGAAAACCAACAGAAATCCTTCGCCTGTCCTCATGTACTGATCACGCATCGCCG
- R R K P T E I L R L S S C T D H A S P
AATATTCTTCTTGTCCGGCGGTATCCAATATGTCGAGGAGGCATGTCTACCGTCTATCA
N I L L V R R Y P I C R G G M S H R L S
CAACCTGCTTTCTGTAGCTGTCCTCTATGGTCGGGTCGTATTCTTCGACAAAGTGATTCT
Q P A F C S C P L W S G R I L R Q S D S
GGATGAGTTGAATGGTGAGTGCTGATTTACCAACTCCTCCATCTCCAACCTACCACAAGCT
G - V E W - V L I Y Q L L H L Q L P Q A
TGTACTCCGTCA
C T P S

Enter DNA sequence (or Exit to quit the program): exit