

Question 1

I've used the emboss cli on MacOS

Question 2

Sequence = CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG

Complement : GAGCCTAAACATTTCTAGTACTAGAGTATGTATCATGGATCGGTAAC

Reverse Complement : CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG

```
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % echo ">Seq1\nCTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG" > file1
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % more file1
>Seq1
CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % revseq
Reverse and complement a nucleotide sequence
Input (gapped) nucleotide sequence(s): file1
Error: Failed to open filename 'file1'
Error: Unable to read sequence 'file1'
Input (gapped) nucleotide sequence(s): file
output sequence(s) [seq1.rev]: ans
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % more ans
>Seq1 Reversed:
CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG
ans (END)
```

Question 3

In [8]:

```
seq = "CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG"
d = {a : b for a,b in zip('ATGC', 'TACG')}
complement = ''.join([d[base] for base in seq])
print(f"Complement : {complement}")
print(f'Reverse Complement : {complement[::-1]}')
```

Complement : GAGCCTAAACATTTCTAGTACTAGAGTATGTATCATGGATCGGTAAC

Reverse Complement : CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG

Question 4

Sequence =

GACATTGTGAACAGTAAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAGGAAGCAGGGCAAGC
TCCCATGGACTACTCTCCTCTGCCCATCGACAAGCATGAGCCTGAATTTGGTCCATGCAGAAGAA

Part (i)

Protein Sequence =

DIVNSKKVHAMRKEQKRKQGKQRSMGSPMDYSPLPIDKHEPEFGPCRRKLDG

```
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % echo ">seq1\nGACATTGTGAACAGTAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAG
GAAGCAGGGCAAGCAGCGCTCCATGGGCTCTCCCATGGACTACTCTCCTCTGCCCATGACAAGCATGAGCCTGAATTTGGTCCATGCAGAAGAAAACCTGGATGGG" > file
844567688557
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % transeq ZTznZz09
Translate nucleic acid sequences
Input nucleotide sequence(s): file
protein output sequence(s) [seq1.pep]: ans
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % more ans
>seq1_1
DIVNSKKVHAMRKEQKRKQKQKQSMGSPMDYSPLPIDKHEPEFGPCRRKLDG
ans (END)
```

Part (ii)

ORF = [156 - 1] (REVERSE SENSE)

```
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % getorf file
Find and extract open reading frames (ORFs)
protein output sequence(s) [seq1.orf]: file
(base) Ajeet@MacBook-touch BT3040 Bioinformatics % more file
>seq1_1 [18 - 116]
KSPCNAQGAEEAGQAALHGLSHGLLSSAHRQA
>seq1_2 [11 - 154]
TVKKSQCARSRRGSRASSAPWALPWTLLCPSTMSLNLVHAEENWH
>seq1_3 [126 - 155]
IWSMQKKTGW
>seq1_4 [1 - 156]
DIVNSKKVHAMRKEQKRKQKQKQSMGSPMDYSPLPIDKHEPEFGPCRRKLDG
>seq1_5 [155 - 93] (REVERSE SENSE)
PSSFLLHGPNSGCLSMGRGE
>seq1_6 [89 - 3] (REVERSE SENSE)
SMGEPMERCLPCFLFCSLRIAWTFLLFTM
>seq1_7 [154 - 2] (REVERSE SENSE)
HPVFFCMDQIQAHACRWAEESPWESPWSAACPASSAPCALHGLFYCSQC
>seq1_8 [156 - 1] (REVERSE SENSE)
PIQFSSAWTKFRLMLVDGQRRVVHGRAHGALLALLPLLLAHCMDFFTVHNV
file (END)
```

Question 5

```
In [1]: # Saved the codon information in a text file
import re
seq = "GACATTGTGAACAGTAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAGGAAGCAGGGCAAGCAGC
GACAAGCATGAGCCTGAATTTGGTCCATGCAGAAGAAAACCTGGATGGG"
with open('codon.txt') as file:
    string = file.read()
    codon = re.findall("[A-z]{3}\s[A-z]{1,4}", string)
    codon = dict([tuple(x.split(' ')) for x in codon])
seq = seq.replace('T', 'U')
protein = ''.join([codon[seq[i:i+3]] for i in range(0, len(seq), 3)])
print(f'Translated sequence = {protein}')
```

Translated sequence = DIVNSKKVHAMRKEQKRKQKQKQSMGSPMDYSPLPIDKHEPEFGPCRRKLDG

Question 6

```
# Note that the position is indexed from 0
import re
seq = "GACATTGTGAACAGTAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAGGAAGCAGGGCAAGCAGCAGACAAGCATGAGCCTGAATTTGGTCCATGCAGAAGAAAAC TGGATGGG"

# Input a string from the user
find = input("Input a string")
positions = [(i.start(), i.end()) for i in re.finditer(find, seq)]
print(f"Searching for : '{find}'\n")
print(f"Number of matches = {len(positions)}")
print(f"Positions of matches = {positions}")
```

```
Positions of matches = [(0, 3), (87, 90), (108, 111)]
```

Pepstats - Calculate properties of protein sequences such as molecular weight

```
seq = "CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG"
stacking = {"AA": -4, "AT": -7, "AC": -5, "AG": -11, "TA": -7, "TT": -2, "TC": -5,
            "CA": -9, "CT": -5, "CC": -6, "CG": -7, "GA": -9, "GT": -6, 'GC': -5}
energy = sum([stacking[seq[i:i+2]] for i in range(len(seq)-1)]) / (len(seq)-1)
print(f'Average Stacking Energy = {energy}')
```

Question 9

ATATATATAT

| Properties | Scale/unit | Average value |
|--|--------------------------------------|---------------|
| Stacking energy | kcal/mol | 1.8 |
| Enthalpy | kcal/mol | 6.04444 |
| Entropy | cal/mol/K | 16.6222 |
| Flexibility_shift | kJ mol ⁻¹ Å ⁻² | 2.53 |
| Flexibility_slide | kJ mol ⁻¹ Å ⁻² | 9.66333 |
| Free energy | kcal/mol | 0.655556 |
| Melting Temperature | degree | 48.0022 |
| Mobility to bend towards major groove | mu | 1.09778 |
| Mobility to bend towards minor groove | mu | 1.03333 |
| Probability contacting nucleosome core | % | 6.75556 |
| Rise stiffness | kcal/mol angstrom | 7.80778 |
| Roll stiffness | kcal/mol degree | 19.3333 |
| Shift stiffness | kcal/mol angstrom | 0.892222 |
| Slide stiffness | kcal/mol angstrom | 2.66111 |
| Tilt stiffness | kcal/mol degree | 28 |
| Twist stiffness | kcal/mol degree | 25.8889 |

Your input seq is:

GCGCGCGCGC

Physicochemical Properties:

| Properties | Scale/unit | Average value |
|--|--------------------------------------|---------------|
| Stacking energy | kcal/mol | 1.75556 |
| Enthalpy | kcal/mol | 11.0778 |
| Entropy | cal/mol/K | 27.5556 |
| Flexibility_shift | kJ mol ⁻¹ Å ⁻² | 6.49111 |
| Flexibility_slide | kJ mol ⁻¹ Å ⁻² | 4.19778 |
| Free energy | kcal/mol | 1.85889 |
| Melting Temperature | degree | 107.867 |
| Mobility to bend towards major groove | mu | 0.997778 |
| Mobility to bend towards minor groove | mu | 1.20556 |
| Probability contacting nucleosome core | % | 3.37778 |
| Rise stiffness | kcal/mol angstrom | 8.06333 |
| Roll stiffness | kcal/mol degree | 21.5556 |
| Shift stiffness | kcal/mol angstrom | 1.14667 |
| Slide stiffness | kcal/mol angstrom | 2.33889 |
| Tilt stiffness | kcal/mol degree | 31.5556 |
| Twist stiffness | kcal/mol degree | 20.1111 |

part (i) ATATATATAT

melting temperature = 48.0022

part (ii) GCGCGCGCGC

melting temperature = 107.867

Question 10

GC content of AAATGGCCCTAA : 41.666667%

Your input seq is:

AAATGGCCCTAA

Nucleotide Content:

| | Nucleotide content in % |
|-------------------|-------------------------|
| AT_content | 58.333333 |
| Adenine_content | 41.666667 |
| Cytosine_content | 25.000000 |
| GC_content | 41.666667 |
| Guanine_content | 16.666667 |
| Keto_GT_content | 33.333333 |
| Purine_AG_content | 58.333333 |
| Thymine_content | 16.666667 |