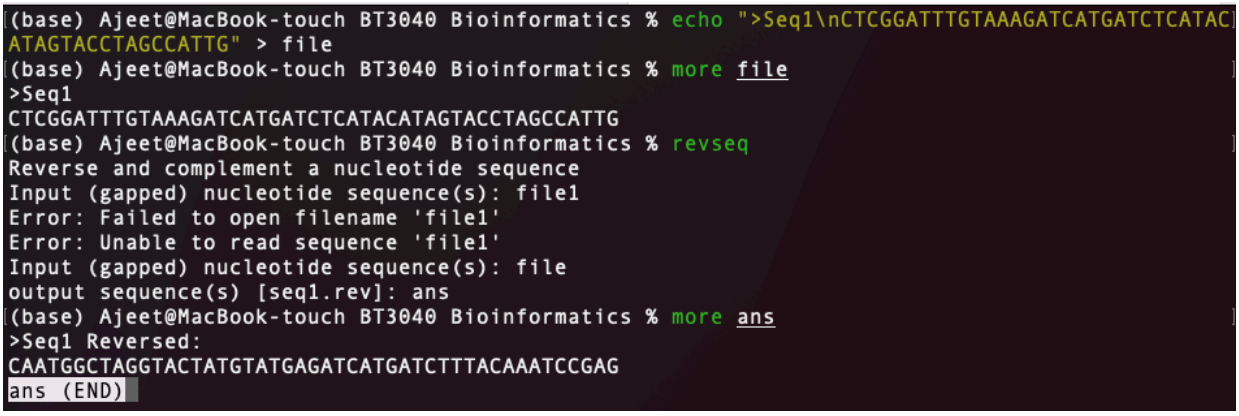
Question 1

I've used the emboss cli on MacOs

Question 2

Sequence = CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG

Complement : GAGCCTAAACATTTCTAGTACTAGAGTATGTATCATGGATCGGTAAC
Reverse Complement : CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG



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]}')
```

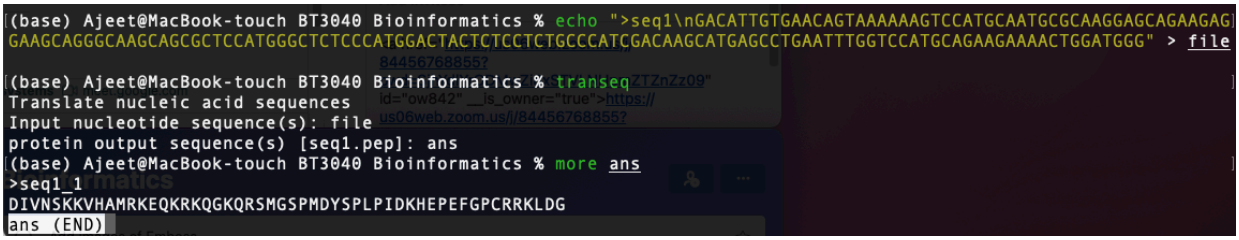
```
Complement : GAGCCTAAACATTTCTAGTACTAGAGTATGTATCATGGATCGGTAAC
Reverse Complement : CAATGGCTAGGTACTATGTATGAGATCATGATCTTTACAAATCCGAG
```

Question 4

Sequence =
GACATTGTGAACAGTAAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAGGAAGCAGGGCAAGC
TCCCATGGACTACTCTCCTCTGCCCATCGACAAGCATGAGCCTGAATTTGGTCCATGCAGAAGAA

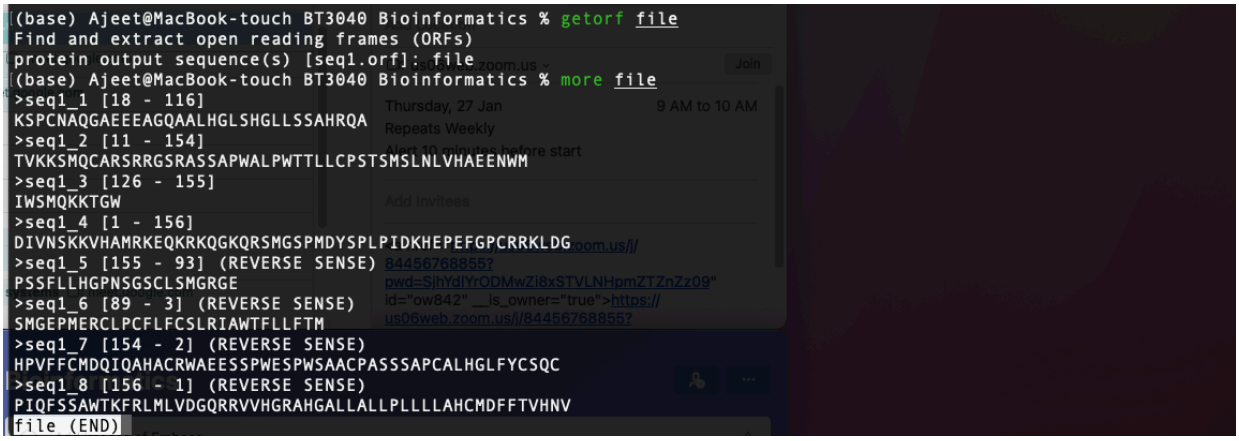
Part (i)

Protein Sequence =
DIVNSKKVHAMRKEQKRKQGKQRSMGSPMDYSPLPIDKHEPEFGPCRRKLDG



Part (ii)

ORF = [156 - 1] (REVERSE SENSE)



Question 5

In [1]:

```
# Saved the codon information in a text file
import re
seq = "GACATTGTGAACAGTAAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAGGAAGCAGGGCAAGCAGCG
GACAAGCATGAGCCTGAATTTGGTCCATGCAGAAGAAAACTGGATGGG"
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 = DIVNSKKVHAMRKEQKRKQGKQRSMGSPMDYSPLPIDKHEPEFGPCRRKLDG
```

Question 6

In [3]:
```
# Note that the positon is indexed from 0
import re
seq = "GACATTGTGAACAGTAAAAAAGTCCATGCAATGCGCAAGGAGCAGAAGAGGAAGCAGGGCAAGCAGCC
GACAAGCATGAGCCTGAATTTGGTCCATGCAGAAGAAAACTGGATGGG"

# 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}")
```

```
Searching for : 'GAC'

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

Question7

dan - Calculate nucleic acid melting temperature

banana - Plot bending and curvature data for B-DNA

Pepstats - Calculate properties of protein sequences such as molecular weight

Question 8

In [57]:
```
seq = "CTCGGATTTGTAAAGATCATGATCTCATACATAGTACCTAGCCATTG"
stacking = {"AA": -4, "AT": -7, "AC": -5, "AG": -11, "TA": -7, "TT": -2, "
            "CA": -9, "CT": -5, "CC": -6, "CG": -7, "GA": -9, "GT": -6, 'GC
energy = sum([stacking[seq[i:i+2]] for i in range(len(seq)-1)]) / (len(seq
print(f'Average Stacking Energy = {energy}')
```

```
Average Stacking Energy = -5.804347826086956
```

Question 9

Your input seq is:

ATATATATAT

Physicochemical Properties:

Properties	Scaleunit	Average value
Stacking energy	kcal/mol	1.8
Enthalpy	kcal/mol	6.04444
Entropy	cal/mol/K	16.6222
Flexibility_shift	kJ mol^-1 A^-2	2.53
Flexibility_slide	kJ mol^-1 A^-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 angstroem	7.80778
Roll stiffness	kcal/mol degree	19.3333
Shift stiffness	kcal/mol angstroem	0.892222
Slide stiffness	kcal/mol angstroem	2.66111
Tilt stiffness	kcal/mol degree	28
Twist stiffness	kcal/mol degree	25.8889

Your input seq is:

GCGCGCGCGC

Physicochemical Properties:

Properties	Scaleunit	Average value
Stacking energy	kcal/mol	1.75556
Enthalpy	kcal/mol	11.0778
Entropy	cal/mol/K	27.5556
Flexibility_shift	kJ mol^-1 A^-2	6.49111
Flexibility_slide	kJ mol^-1 A^-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 angstroem	8.06333
Roll stiffness	kcal/mol degree	21.5556
Shift stiffness	kcal/mol angstroem	1.14667
Slide stiffness	kcal/mol angstroem	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