Practical 7

Date: 13-03-2022
Day: Sunday

Name: Aswin. V
Roll Number: BS19B003

Q1 –

```
def Q1(A):
    AA = {'A':0, 'C':0, 'D':0, 'E':0, 'F':0, 'G':0, 'H':0, 'I':0, 'K':0,
'L':0,
          'M':0, 'N':0, 'P':0, 'Q':0, 'R':0, 'S':0, 'T':0, 'V':0, 'W':0,
'Y':0}
    for i in A:
        if i in AA:
            AA[i] += 1*(100/len(A))
    return AA

if __name__ == "__main__":
    s1 =
'RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLAALKQKSRHAA'
    s2 =
'AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYVMFAVA'
    s3 =
'AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGKHSLSFAPVFVYEFAGD
SIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVRAIHYSNAGLKQPNDGIESYSLFYKI
PI'

    seq1, seq2, seq3 = Q1(s1), Q1(s2), Q1(s3)
    st = "{:<15} {:<25} {:<25} {:<25}"
    print(st.format('Amino acids', 'Sequence 1', 'Sequence 2', 'Sequence
3'))
    d = ['A', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'K', 'L',
         'M', 'N', 'P', 'Q', 'R', 'S', 'T', 'V', 'W', 'Y']
    for j in range(len(d)):
        print(st.format(d[j], seq1[d[j]], seq2[d[j]], seq3[d[j]]))
```

Output:

https://drive.google.com/file/d/1nHyCkHsGme1IIo96K98OVKAEumSYm21S/view?usp=sharing

*above file has the output

Observations:

- ✓ Alanine, Threonine, Leucine, Lysine, and Proline are more abundant in Sequence 1. This indicates that the sequence is both hydrophobic and hydrophilic in nature (due to presence of Threonine). As a result, this polypeptide sequence must be found in the inner core of a protein in a typical hydrophilic environment, such as the exterior space, where water is abundant.

- ✓ Because it has a higher percentage of Alanine, Glycine, Isoleucine, and Leucine, Sequence 2 is particularly hydrophobic. As a result, this polypeptide sequence must be found in the inner core of a protein in a typical hydrophilic environment, such as the exterior space, where water is abundant
- ✓ Glycine, Serine, and Alanine are more abundant in Sequence 3. Serine renders this protein slightly hydrophilic, despite the presence of Gly and Ala, which would ordinarily make it hydrophobic.

Q2 –

```
def Q2(A):
    AAMW = {'A': 85, 'C': 115, 'D': 130, 'E': 145, 'F': 160,
            'G': 70, 'W': 200, 'H': 150, 'I': 125, 'K': 145,
            'L': 125, 'M': 143, 'N': 130, 'Y': 175, 'P': 110,
            'Q': 140, 'R': 170, 'S': 100, 'T': 115, 'V': 110}
    w = 0
    for i in A:
        w += AAMW[i]
    w -= (len(A) - 1) * 18
    return w

if __name__ == "__main__":
    s1 =
'RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLAALKQKSRHAA'
    s2 =
'AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYVMFAVA'
    s3 =
'AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGKHSLSFAPVFVYEFAGD
SIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVRAIHYSNAGLKQPNDGIESYSLFYKI
PI'

    for k in [s1, s2, s3]:
        print(f"The molecular weight for\n{k}\n is {Q2(k)} units.")
```

Output:

The molecular weight for

RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLAALKQKSRHAA

is 7127 units.

The molecular weight for

AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYVMFAVA

is 6529 units.

The molecular weight for

AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGKHSLSFAPVFVYEF
AGDSIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVRAIHYSNAGLKQPNDGIESYSLFYKI
PI

is 15453 units.

Q3 –

https://drive.google.com/file/d/1dW0n8q7HulhWy3tp2UU8mEvnAzW0uit7/view?usp=sharing

*above file has the code for this question

Output:

Deviation score from Group A: 55.84402985074628

Deviation score from group B: 58.523731343283586

The given sequence is more close to Group A.

Deviation score from Group A: 74.51470588235291

Deviation score from group B: 76.83823529411764

The given sequence is more close to Group A.

Deviation score from Group A: 38.33225165562914

Deviation score from group B: 32.597814569536425

The given sequence is more close to Group B.

Q4 –

https://drive.google.com/file/d/1vQhxamyNKw3VH1JJ0ZINNY_kcoTZgFrB/view?usp=sharing

*above file has the code for this question

https://drive.google.com/file/d/10ca8-BVUM6YI0U4CoIv2N2eG_2tBvnU0/view?usp=sharing

*above file has the output

Q5 –

- The average hydrophobicity is computed by adding the "Hgm" values from the database for each amino acid and dividing by the total number of amino acids.

- The helical contact area is computed by adding the "Ca" values from the database for each amino acid.
- Total non-bonded energy is calculated by adding the "Et" values from the database for each amino acid.

```
def Q5(A):
    Hgm = {'A': 13.85, 'D': 11.61, 'C': 15.37, 'E': 11.38, 'F': 13.93,
           'G': 13.34, 'H': 13.82, 'I': 15.28, 'K': 11.58, 'L': 14.13,
           'M': 13.86, 'N': 13.02, 'P': 12.35, 'Q': 12.61, 'R': 13.10,
           'S': 13.39, 'T': 12.70, 'V': 14.56, 'W': 15.48, 'Y': 13.88}
    Ca = {'A': 20.00, 'D': 26.00, 'C': 25.00, 'E': 33.00, 'F': 46.00,
          'G': 13.00, 'H': 37.00, 'I': 39.00, 'K': 46.00, 'L': 35.00,
          'M': 43.00, 'N': 28.00, 'P': 22.00, 'Q': 36.00, 'R': 55.00,
          'S': 20.00, 'T': 28.00, 'V': 33.00, 'W': 61.00, 'Y': 46.00}
    Et = {'A': 1.90, 'D': 1.52, 'C': 2.04, 'E': 1.54, 'F': 1.86,
          'G': 1.90, 'H': 1.76, 'I': 1.95, 'K': 1.37, 'L': 1.97,
          'M': 1.96, 'N': 1.56, 'P': 1.70, 'Q': 1.52, 'R': 1.48,
          'S': 1.75, 'T': 1.77, 'V': 1.98, 'W': 1.87, 'Y': 1.69}

    avg_Hgm, ca, et = 0, 0, 0

    for i in range(len(A)):
        avg_Hgm += (Hgm[A[i]] / len(A))
        ca += Ca[A[i]]
        et += Et[A[i]]

    print(f"Average hydrophobicity of the sequence: {avg_Hgm}"
          f"\nHelical contact area of the sequence: {ca}"
          f"\nTotal non-bonded energy of the sequence: {et}\n")

if __name__ == "__main__":
    s1 =
'RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLAALKQKSRHAA'
    s2 =
'AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYVMFAVA'
    s3 =
'AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGKHSLSFAPVFVYEFAGD
SIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVRAIHYSNAGLKQPNDGIESYSLFYKI
PI'

    for m in [s1, s2, s3]:
        Q5(m)
```

Output:

Average hydrophobicity of the sequence: 13.352537313432833

Helical contact area of the sequence: 2156.0

Total non-bonded energy of the sequence: 117.74000000000005

Average hydrophobicity of the sequence: 13.77161764705882

Helical contact area of the sequence: 2067.0

Total non-bonded energy of the sequence: 126.66000000000003

Average hydrophobicity of the sequence: 13.418675496688742

Helical contact area of the sequence: 4616.0

Total non-bonded energy of the sequence: 267.7500000000001

Observations:

- ✓ All the given sequences have comparable hydrophobicity. This is due to the high composition of Alanine/other aliphatic amino acids in all the polypeptide sequences.
- ✓ Sequence 3 has a much larger helix contact area than sequences 1 and 2. This is due to the fact that sequence 3 could have a right-handed alpha-helical structure, whereas other peptides could have a beta-sheet structure.
- ✓ Sequence 3 has a higher total non-bonded energy than sequences 1 and 2. For sequence 3, large values of "Et" for Gly and Ser result in a high amount of total non-bonded energy.