***Practicals – 7***

*-BS19B032*

*-R. Vasantha Kumar*

1) I created a python code to compute the amino acid composition, the molecular weight, and find to which group it belongs. I attached the code in submission.

***Amino acid composition:***

Sequence 1:

|  |  |
| --- | --- |
| Alanine | 17.91044776119403 |
| Cysteine | 1.4925373134328357 |
| Aspartic acid | 2.9850746268656714 |
| Glutamic acid | 2.9850746268656714 |
| Phenylalanine | 1.4925373134328357 |
| Glycine | 5.970149253731343 |
| Histidine | 4.477611940298507 |
| Isoleucine | 1.4925373134328357 |
| Lysine | 7.462686567164178 |
| Leucine | 8.955223880597014 |
| Methionine | 0.0 |
| Asparagine | 1.4925373134328357 |
| Proline | 7.462686567164178 |
| Glutamine | 1.4925373134328357 |
| Arginine | 5.970149253731343 |
| Serine | 5.970149253731343 |
| Threonine | 10.44776119402985 |
| Valine | 1.4925373134328357 |
| Tryptophan | 5.970149253731343 |
| Tyrosine | 4.477611940298507 |

Sequence 2:

|  |  |
| --- | --- |
| Alanine | 19.11764705882353 |
| Cysteine | 0.0 |
| Aspartic acid | 2.941176470588235 |
| Glutamic acid | 1.4705882352941175 |
| Phenylalanine | 5.88235294117647 |
| Glycine | 14.705882352941178 |
| Histidine | 0.0 |
| Isoleucine | 11.76470588235294 |
| Lysine | 1.4705882352941175 |
| Leucine | 13.23529411764706 |
| Methionine | 7.352941176470589 |
| Asparagine | 0.0 |
| Proline | 4.411764705882353 |
| Glutamine | 2.941176470588235 |
| Arginine | 2.941176470588235 |
| Serine | 0.0 |
| Threonine | 1.4705882352941175 |
| Valine | 8.823529411764707 |
| Tryptophan | 0.0 |
| Tyrosine | 1.4705882352941175 |

Sequence 3:

|  |  |
| --- | --- |
| Alanine | 10.596026490066226 |
| Cysteine | 0.0 |
| Aspartic acid | 5.960264900662252 |
| Glutamic acid | 3.9735099337748347 |
| Phenylalanine | 5.298013245033113 |
| Glycine | 15.2317880794702 |
| Histidine | 1.3245033112582782 |
| Isoleucine | 5.298013245033113 |
| Lysine | 3.9735099337748347 |
| Leucine | 5.960264900662252 |
| Methionine | 0.6622516556291391 |
| Asparagine | 3.3112582781456954 |
| Proline | 2.6490066225165565 |
| Glutamine | 3.3112582781456954 |
| Arginine | 3.3112582781456954 |
| Serine | 9.933774834437086 |
| Threonine | 4.635761589403973 |
| Valine | 5.298013245033113 |
| Tryptophan | 3.9735099337748347 |
| Tyrosine | 5.298013245033113 |

From the results, it is clear that, cysteine is a rare amino acid, with very minimum compositions. Methionine composition is also very low in these sequences.

Also, the amino acids leucine and glycine has high compositions. Alanine also has high composition in these sequences.

2) Using the given values of molecular weights of amino acid residue and the below formula, I calculated molecular weight:

Mol. Weight = ∑f(i)\*w(i) – w(w)\*(n-1)

Where,

f(i) = frequency of residue

w(i) = weight of residue

w(w) = weight of water = 18

n = length of sequence

***Molecular weight:***

Sequence 1:

The molecular weight is: 7127

Sequence 2:

The molecular weight is: 6529

Sequence 3:

The molecular weight is: 15453

3) Using the given values and the below formula for deviation, I calculated the deviations of query sequence with group A and group B.

Then, the protein belongs to group with low deviation.

std = ∑|fA(i) – f(i)|

where,

fA(i) = amino acid composition of residue in group A/B

f(i) = amino acid composition of residue in query sequence

***Group A/B:***

Sequence 1:

std with Group A: 55.84402985074627

std with Group B: 58.523731343283586

The protein sequence belongs to: GROUP A

Sequence 2:

std with Group A: 74.51470588235294

std with Group B: 76.83823529411765

The protein sequence belongs to: GROUP A

Sequence 3:

std with Group A: 38.332251655629136

std with Group B: 32.59781456953643

The protein sequence belongs to: GROUP B

4) I created a python code to compute the residue pair preference for the three query sequences, with given three normalisations.

(a) Nij\*100/(Ni+Nj)

(b) Nij\*100/(N-1)

(c) Nij\*/(Ni\*N)

I have submitted a total of 9 excel files, containing 9 tables, 3 for each sequence.

The top 10 preferred residues from each of the three pair-preferences are:

Sequence 1:

|  |  |
| --- | --- |
| 1. | Alanine |
| 2. | Lysine |
| 3. | Glycine |
| 4. | Threonine |
| 5. | Serine |
| 6. | Tryptophan |
| 7. | Proline |
| 8. | Tyrosine |
| 9. | Histidine |
| 10. | Arginine |

Sequence 2:

|  |  |
| --- | --- |
| 1. | Alanine |
| 2. | Leucine |
| 3. | Glycine |
| 4. | Isoleucine |
| 5. | Valine |
| 6. | Phenylalanine |
| 7. | Methionine |
| 8. | Glutamine |
| 9. | Proline |
| 10. | Arginine |

Sequence 3:

|  |  |
| --- | --- |
| 1. | Serine |
| 2. | Glycine |
| 3. | Alanine |
| 4. | Phenylalanine |
| 5. | Isoleucine |
| 6. | Threonine |
| 7. | Valine |
| 8. | Tyrosine |
| 9. | Aspartic acid |
| 10. | Glutamic acid |

5) I created a python code to calculate average hydrophobicity, helical contact area and total non-bond energy, using given values in website. I attached the code in submission.

Avg. hydrophobicity = (∑f(i)\*Hgm(i))/N

Where,

f(i) = frequency of residue

Hgm(i) = hydrophobicity values of residues N = number of residues

Helical contact area = ∑f(i)\*Ca(i)

Where,

f(i) = frequency of residue

Ca(i) = helical contact area values of residues

Total non-bonded energy = ∑f(i)\*Et(i)

Where,

f(i) = frequency of residue

Et(i) = total non-bounded energy values of residues

Sequence 1:

Average hydrophobicity: 13.352537313432833

Helical contact area: 2156.0

Total non-bonded energy: 117.74000000000001

Sequence 2:

Average hydrophobicity: 13.771617647058823

Helical contact area: 2067.0

Total non-bonded energy: 126.65999999999998

Sequence 3:

Average hydrophobicity: 13.418675496688744

Helical contact area: 4616.0

Total non-bonded energy: 267.75

From the results, we could find that as the number of residues in a sequence increases, the non-bound energy and helical surface area of the sequence, also increases.

But, when we consider the average hydrophobicity, it remains fairly similar to all the three sequences.