BT 3040: BIOINFORMATICS

Assignment 7

Atharva Mandar Phatak | BE21B009
Department of Biotechnology

Indian Institute of Technology
Madras

Q1) Compute the amino acid composition of the following sequences. Provide the output as a table of amino acid percentage values for each sequence and comment on the results.

Calculated using EMBOSS

a) Sequence 1:
RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLA
AL KQKSRHAA

```
#
# Output from 'compseq'
#
# The Expected frequencies are calculated on the (false) assumption that every
# word has equal frequency.
#
# The input sequences are:
#


Word size 1
Total count         67

#
# Word    Obs Count  Obs Frequency       Exp Frequency       Obs/Exp Frequency
#
A         12                  0.1791045 0.0476190 3.7611940
C         1                   0.0149254 0.0476190 0.3134328
D         2                   0.0298507 0.0476190 0.6268657
E         2                   0.0298507 0.0476190 0.6268657
F         1                   0.0149254 0.0476190 0.3134328
G         4                   0.0597015 0.0476190 1.2537313
H         3                   0.0447761 0.0476190 0.9402985
I         1                   0.0149254 0.0476190 0.3134328
K         5                   0.0746269 0.0476190 1.5671642
L         6                   0.0895522 0.0476190 1.8805970
M         0                   0.0000000 0.0476190 0.0000000
N         1                   0.0149254 0.0476190 0.3134328
P         5                   0.0746269 0.0476190 1.5671642
Q         1                   0.0149254 0.0476190 0.3134328SS
R         4                   0.0597015 0.0476190 1.2537313
S         4                   0.0597015 0.0476190 1.2537313
T         7                   0.1044776 0.0476190 2.1940299
U         0                   0.0000000 0.0476190 0.0000000
V         1                   0.0149254 0.0476190 0.3134328
W         4                   0.0597015 0.0476190 1.2537313
Y         3                   0.0447761 0.0476190 0.9402985

Other     0                   0.0000000 0.0000000 100000000000.0000000
```

b) Sequence 2:
AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGL
GL YVMFAVA

```
#
# Output from 'compseq'
#
# The Expected frequencies are calculated on the (false) assumption that every
# word has equal frequency.
#
# The input sequences are:
#


Word size 1
Total count          68

#
# Word     Obs Count  Obs Frequency        Exp Frequency        Obs/Exp Frequency
#
A          13                   0.1911765  0.0476190  4.0147059
C          0                    0.0000000  0.0476190  0.0000000
D          2                    0.0294118  0.0476190  0.6176471
E          1                    0.0147059  0.0476190  0.3088235
F          4                    0.0588235  0.0476190  1.2352941
G          10                   0.1470588  0.0476190  3.0882353
H          0                    0.0000000  0.0476190  0.0000000
I          8                    0.1176471  0.0476190  2.4705882
K          1                    0.0147059  0.0476190  0.3088235
L          9                    0.1323529  0.0476190  2.7794118
M          5                    0.0735294  0.0476190  1.5441176
N          0                    0.0000000  0.0476190  0.0000000
P          3                    0.0441176  0.0476190  0.9264706
Q          2                    0.0294118  0.0476190  0.6176471
R          2                    0.0294118  0.0476190  0.6176471
S          0                    0.0000000  0.0476190  0.0000000
T          1                    0.0147059  0.0476190  0.3088235
U          0                    0.0000000  0.0476190  0.0000000
V          6                    0.0882353  0.0476190  1.8529412
W          0                    0.0000000  0.0476190  0.0000000
Y          1                    0.0147059  0.0476190  0.3088235

Other      0                    0.0000000  0.0000000  10000000000.0000000
```

c) Sequence 3:
AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGD
EG
AGKHSLSFAPVFVYEFAGDSIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLK
FAN GQSVGVRAIHYSNAGLKQPNDGIESYSLFYKIPI

```
#
# Output from 'compseq'
#
# The Expected frequencies are calculated on the (false) assumption that every
# word has equal frequency.
#
# The input sequences are:
#


Word size 1
Total count          151

#
# Word    Obs Count  Obs Frequency        Exp Frequency        Obs/Exp Frequency
#
A         16                   0.1059603  0.0476190  2.2251656
C         0                    0.0000000  0.0476190  0.0000000
D         9                    0.0596026  0.0476190  1.2516556
E         6                    0.0397351  0.0476190  0.8344371
F         8                    0.0529801  0.0476190  1.1125828
G         23                   0.1523179  0.0476190  3.1986755
H         2                    0.0132450  0.0476190  0.2781457
I         8                    0.0529801  0.0476190  1.1125828
K         6                    0.0397351  0.0476190  0.8344371
L         9                    0.0596026  0.0476190  1.2516556
M         1                    0.0066225  0.0476190  0.1390728
N         5                    0.0331126  0.0476190  0.6953642
P         4                    0.0264901  0.0476190  0.5562914
Q         5                    0.0331126  0.0476190  0.6953642
R         5                    0.0331126  0.0476190  0.6953642
S         15                   0.0993377  0.0476190  2.0860927
T         7                    0.0463576  0.0476190  0.9735099
U         0                    0.0000000  0.0476190  0.0000000
V         8                    0.0529801  0.0476190  1.1125828
W         6                    0.0397351  0.0476190  0.8344371
Y         8                    0.0529801  0.0476190  1.1125828

Other     0                    0.0000000  0.0000000  10000000000.0000000
```

d) Combined Output

Sequence 1		
Residue	Frequency	Percentage Occurance
A	12	17.91045
C	1	1.49254
D	2	2.98507
E	2	2.98507
F	1	1.49254
G	4	5.97015
H	3	4.47761
I	1	1.49254
K	5	7.46269
L	6	8.95522
M	0	0
N	1	1.49254
P	5	7.46269
Q	1	1.49254
R	4	5.97015
S	4	5.97015
T	7	10.44776
U	0	0
V	1	1.49254
W	4	5.97015
Y	3	4.47761
Total		**100.00001**

Sequence 2		
Residue	Frequency	Percentage Occurance
A	13	19.11765
C	0	0
D	2	2.94118
E	1	1.47059
F	4	5.88235
G	10	14.70588
H	0	0
I	8	11.76471
K	1	1.47059
L	9	13.23529
M	5	7.35294
N	0	0
P	3	4.41176
Q	2	2.94118
R	2	2.94118
S	0	0
T	1	1.47059
U	0	0
V	6	8.82353
W	0	0
Y	1	1.47059
Total		**100.00001**

Sequence 3		
Residue	Frequency	Percentage Occurance
A	16	10.59603
C	0	0
D	9	5.96026
E	6	3.97351
F	8	5.29801
G	23	15.23179
H	2	1.3245
I	8	5.29801
K	6	3.97351
L	9	5.96026
M	1	0.66225
N	5	3.31126
P	4	2.64901
Q	5	3.31126
R	5	3.31126
S	15	9.93377
T	7	4.63576
U	0	0
V	8	5.29801
W	6	3.97351
Y	8	5.29801
Total		**99.99998**

Q2) Assume the molecular weights of the 20 amino acid residues as given below. Compute the molecular weight of the three sequences given in question 1

Code:

```
# BT3040 Assignment 7| Atharva Mandar Phatak | BE21B009
#Q2

def aa_molwt_calc(AA_input):
    Mol_Wt_dict={'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}
    mol_wt_aa=0
    for i in AA_input:
        mol_wt_aa=mol_wt_aa+Mol_Wt_dict[i]
    return mol_wt_aa



print(f"Mol Wt for Sequence 1 is {aa_molwt_calc(seq1)} units")
print(f"Mol Wt for Sequence 2 is {aa_molwt_calc(seq2)} units")
print(f"Mol Wt for Sequence 3 is {aa_molwt_calc(seq3)} units")
```

Output:

```
Mol Wt for Sequence 1 is 8315 units
Mol Wt for Sequence 2 is 7735 units
Mol Wt for Sequence 3 is 18153 units
```

Q3) The amino acid composition of a standard set of Group A (first value) and Group B (second value) proteins are given below. Identify whether the given sequences in Question 1 belong to Group A or Group B and write your answer it.

```
# BT3040 Assignment 7| Atharva Mandar Phatak | BE21B009
#Q3
import numpy as np

group_A = {'A': 8.47, 'D': 5.97, 'C': 1.39, 'E': 6.32, 'T': 5.79, 'F': 3.91, 'G': 7.82, 'H': 2.26, 'I': 5.71, 'V': 7.02,
        'K': 5.76, 'L': 8.48, 'M': 2.21, 'N': 4.54, 'W': 1.44, 'P': 4.63, 'Q': 3.82, 'R': 4.93, 'S': 5.94, 'Y': 3.58}
group_B = {'A': 8.95, 'D': 5.91, 'C': 0.47, 'E': 4.78, 'T': 6.54, 'F': 3.68, 'G': 8.54, 'H': 1.25, 'I': 4.77, 'V': 6.76,
        'K': 4.93, 'L': 8.78, 'M': 1.56, 'N': 5.74, 'W': 1.24, 'P': 3.74, 'Q': 4.75, 'R': 5.24, 'S': 8.05, 'Y': 4.13}

strings =
['"RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLAALKQK
SRHAA"',
        '"AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGL
YVMFAVA"',
        '"AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEG
AGKHSLSFAPVFVYEFAGDSIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSV
GVRAIHYSNAGLKQPNDGIESYSLFYKIPI"']


def remse_calc(seq,seqnum):
    group_A_diff=0
    group_B_diff=0
    for _ in mwdict:
        group_A_diff=group_A_diff+np.abs((pandadict[_][seqnum]-group_A[_]))

        group_B_diff=group_B_diff+np.abs((pandadict[_][seqnum]-group_B[_]))

    if group_A_diff<group_B_diff:
        print(f"Sequence {seqnum +1} is in Group A")

    else:
        print(f"Sequence {seqnum+1} is in Group B")


remse_calc(strings[0],0)
remse_calc(strings[1],1)
remse_calc(strings[2],2)
```

Output:

```
Sequence 1 is in Group A
Sequence 2 is in Group A
Sequence 3 is in Group B
```

Q4) Compute the residue pair preference for the three sequences given in question 1. The required output is a 20x20 table showing the pair preferences (a) [Nij*100/(Ni+Nj)], (b) [Nij*100/(N-1)] and (c) [Nij*100/(Ni*Nj)]. List the top 10 preferred residues from each of the three pair-preferences.

```
import pandas as pd

def ResPairPref(seq):
    AA_all = ['A', 'C', 'D', 'E', 'F', 'G', 'H', 'I', 'K', 'L', 'M', 'N', 'P', 'Q', 'R', 'S', 'T', 'V', 'W', 'Y']

    Pref_1 = [[0 for _ in range(len(AA_all))] for _ in range(len(AA_all))]
    Pref_2 = [[0 for _ in range(len(AA_all))] for _ in range(len(AA_all))]
    Pref_3 = [[0 for _ in range(len(AA_all))] for _ in range(len(AA_all))]
    composition = [0] * len(AA_all)

    for aa in seq:
        ind = AA_all.index(aa)
        composition[ind] += 1

    for i in AA_all:
        for j in AA_all:
            pair = i + j
            for k in range(len(seq) - 1):
                if pair == seq[k:k + 2]:
                    Pref_1[AA_all.index(i)][AA_all.index(j)] += 1
                    Pref_2[AA_all.index(i)][AA_all.index(j)] += 1
                    Pref_3[AA_all.index(i)][AA_all.index(j)] += 1

            if composition[AA_all.index(i)] == 0 or composition[AA_all.index(j)] == 0:
                Pref_1[AA_all.index(i)][AA_all.index(j)] = 0
                Pref_2[AA_all.index(i)][AA_all.index(j)] = 0
                Pref_3[AA_all.index(i)][AA_all.index(j)] = 0
            else:
                Pref_1[AA_all.index(i)][AA_all.index(j)] = (Pref_1[AA_all.index(i)][AA_all.index(j)] * 100) / (
                        composition[AA_all.index(i)] + composition[AA_all.index(j)])
                Pref_2[AA_all.index(i)][AA_all.index(j)] = (Pref_2[AA_all.index(i)][AA_all.index(j)] * 100) / (
                        len(seq) - 1)
                Pref_3[AA_all.index(i)][AA_all.index(j)] = (Pref_3[AA_all.index(i)][AA_all.index(j)] * 100) / (
                        composition[AA_all.index(i)] * composition[AA_all.index(j)])

    print('\nPair-wise preference score for the given sequence:')
    print('\nFormula 1 : ')
    print(pd.DataFrame(Pref_1, columns=AA_all, dtype=float).to_string())

    print('\nPair-wise preference score for the given sequence:')
    print('\nFormula 2 : ')
    print(pd.DataFrame(Pref_2, columns=AA_all, dtype=float).to_string())
```

```
    print('\nPair-wise preference score for the given sequence:')
    print('\nFormula 3 : ')
    print(pd.DataFrame(Pref_3, columns=AA_all, dtype=float).to_string())

strings = [
    '"RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLAALKQ
KSRHAA"',
    '"AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYV
MFAVA"',
    '"AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGK
HSLSFAPVFVYEFAGDSIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVR
AIHYSNAGLKQPNDGIESYSLFYKIPI"'
]

for seq in strings:
    print("For Sequence", strings.index(seq))
    ResPairPref(seq)
```

Output:

(Complete output is given in notepad file: ‘Assignment7_Q4’

```
Pair-wise preference score for the given sequence:

Formula 1 :
           A    C     D         E     F         G         H     I         K         L    M     N         P         Q         R         S
0   12.500000  0.0   0.0  0.000000   0.0  6.250000  0.000000   0.0  0.000000  11.111111  0.0   0.0  0.000000  0.000000  0.000000  6.250000
1    0.000000  0.0   0.0  0.000000  50.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
2    0.000000  0.0   0.0 25.000000   0.0 16.666667  0.000000   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
3    7.142857  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
4    0.000000  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0  50.0  0.000000  0.000000  0.000000  0.000000
5    0.000000 20.0   0.0  0.000000   0.0  0.000000  0.000000  20.0  0.000000  10.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
6   13.333333  0.0   0.0  0.000000   0.0  0.000000 16.666667   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
7    0.000000  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0 16.666667   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
8    5.882353  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0 10.000000 16.666667  0.000000 11.111111
9    5.555556  0.0  12.5  0.000000   0.0  0.000000 11.111111   0.0  9.090909   8.333333  0.0   0.0  0.000000  0.000000  0.000000  0.000000
10   0.000000  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
11   0.000000  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000 20.000000  0.000000
12   0.000000  0.0   0.0 14.285714   0.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000  0.000000 11.111111
13   0.000000  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0 16.666667   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
14   6.250000  0.0   0.0  0.000000   0.0  0.000000 14.285714   0.0  0.000000   0.000000  0.0   0.0 11.111111  0.000000  0.000000  0.000000
15   0.000000  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0 11.111111  10.000000  0.0   0.0  0.000000  0.000000 12.500000  0.000000
16  10.526316  0.0   0.0  0.000000   0.0  9.090909  0.000000   0.0  8.333333   0.000000  0.0   0.0 16.666667  0.000000  9.090909  0.000000
17   0.000000  0.0   0.0  0.000000   0.0 20.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0  0.000000  0.000000  0.000000  0.000000
18   0.000000  0.0   0.0  0.000000   0.0  0.000000  0.000000   0.0  0.000000   0.000000  0.0   0.0 11.111111  0.000000  0.000000 12.500000
19   6.666667  0.0  20.0  0.000000   0.0  0.000000  0.000000   0.0  0.000000  11.111111  0.0   0.0  0.000000  0.000000  0.000000  0.000000

Pair-wise preference score for the given sequence:
```

For sequence 1

1. Top 10 pairs: {'CF': 50.0, 'FN': 50.0, 'DE': 25.0, 'VG': 20.0, 'GC': 20.0, 'NR': 20.0, 'YD': 20.0, 'GI': 20.0, 'EY': 20.0, 'WT': 18.18}
2. Top 10 pairs: {'AT': 4.54545, 'AA': 4.54545, 'TP': 3.0303, 'WT':3.0303, 'AL': 3.0303, 'TA': 3.0303, 'HA':3.0303, 'RA': 1.51515, 'PT': 1.51515, 'TR':1.51515}
3. Top 10 pairs: {'CF': 100.0, 'FN': 100.0, 'VG': 25.0, 'GC': 25.0, 'NR': 25.0, 'DE': 25.0, 'GI': 25.0, 'PV': 20.0, 'IK':20.0, 'KQ': 20.0}

For sequence 2

1. Top 10 pairs: {'RT': 33.3333, 'TQ': 33.3333, 'VM': 27.2727, 'RQ': 25.0, 'GL': 21.0526, 'KF': 20.0, 'QP': 20.0, 'PD': 20.0, 'AA': 19.2308, 'IP':18.1818}
2. Top 10 pairs: {'AA': 7.46269, 'GL': 5.97015, 'AV': 4.47761, 'VM':4.47761, 'AI': 4.47761, 'IG': 4.47761, 'MG':2.98507, 'GA': 2.98507, 'GI': 2.98507, 'LG':2.98507}
3. Top 10 pairs: {'RT': 50.0, 'TQ': 50.0, 'KF': 25.0, 'RQ': 25.0, 'QP': 16.6667, 'PD':16.6667, 'YV': 16.6667, 'QF': 12.5, 'LE': 11.1111, 'LY':11.1111}

For sequence 3

1. Top 10 pairs: {'WD': 20.0, 'AG': 15.38, 'RL': 14.29, 'YW': 14.29, 'IE': 14.29, 'TY': 13.3333, 'LK': 13.3333, 'MT': 12.5, 'KH': 12.5, 'SL': 12.5}
2. Top 10 pairs: {'AG': 4.0, 'AA': 2.0, 'VG': 2.0, 'GA': 2.0, 'TG': 2.0, 'GL': 2.0, 'WD': 2.0, 'GD': 2.0, 'SL': 2.0, 'FA': 2.0}
3. Top 10 pairs: {'MT': 14.29, 'KH': 8.333, 'IH': 6.25, 'HY': 6.25, 'WD': 5.556, 'QP': 5.0, 'PN': 5.0, 'RL': 4.444, 'GM': 4.348, ‘YW': 4.167}

Q5) Compute average hydrophobicity (Hgm), Helical contact area (Ca) and Total non-bonded energy(Et) for the sequences in Q1 and comment on the results. (Refer www.iitm.ac.in/bioinfo/fold_rate/prop_orig.html for the properties)

```
# BT3040 Assignment 7| Atharva Mandar Phatak | BE21B009
#Q5


hgmdict = {'A': 13.85, 'C': 15.37, 'D': 11.61, 'E': 11.38, 'F': 13.93, 'G': 13.34, 'W': 15.48,
          'H': 13.82, 'I': 15.28, 'K': 11.58, 'L': 14.13, 'M': 13.86, 'N': 13.02, 'Y': 13.88,
          'P': 12.35, 'Q': 12.61, 'R': 13.10, 'S': 13.39, 'T': 12.70, 'V': 14.56}

cadict = {'A': 20, 'C': 25, 'D': 26, 'E': 33, 'F': 46, 'G': 13, 'W': 61,
         'H': 37, 'I': 39, 'K': 46, 'L': 35, 'M': 43, 'N': 28, 'Y': 46,
         'P': 22, 'Q': 36, 'R': 55, 'S': 20, 'T': 28, 'V': 33}

etdict = {'A': 1.9, 'C': 2.04, 'D': 1.52, 'E': 1.54, 'F': 1.86, 'G': 1.9, 'W': 1.87,
         'H': 1.76, 'I': 1.95, 'K': 1.37, 'L': 1.97, 'M': 1.96, 'N': 1.56, 'Y': 1.69,
         'P': 1.7, 'Q': 1.52, 'R': 1.48, 'S': 1.75, 'T': 1.77, 'V': 1.98}

strings = [
   '"RATPTRWPVGCFNRPWTKWSYDEALDGIKAAGYAWTGLLTASKPSLHHATATPEYLAALKQ
KSRHAA"',
   '"AAAVMMGLAAIGAAIGIGILGGKFLEGAARQPDLIPLLRTQFFIVMGLVDAIPMIAVGLGLYV
MFAVA"',
   '"AADVSAAVGATGQSGMTYRLGLSWDWDKSWWQTSTGRLTGYWDAGYTYWEGGDEGAGK
HSLSFAPVFVYEFAGDSIKPFIEAGIGVAAFSGTRVGDQNLGSSLNFEDRIGAGLKFANGQSVGVR
AIHYSNAGLKQPNDGIESYSLFYKIPI"'
]

for seq in strings:
   hgm = sum(hgmdict[aa] for aa in seq) / len(seq)
   ca = sum(cadict[aa] for aa in seq)
   et = sum(etdict[aa] for aa in seq)

   print("For Sequence", strings.index(seq)+1)
   print("Hydrophobicity:", hgm)
   print("Helical Contact Area:", ca)
   print("Total Non-Bonded Energy:", et)
```

Output:

```
For Sequence 1
Hydrophobicity: 13.352537313432844
Helical Contact Area: 2156
Total Non-Bonded Energy: 117.74000000000005
For Sequence 2
Hydrophobicity: 13.77161764705882
Helical Contact Area: 2067
Total Non-Bonded Energy: 126.66000000000003
For Sequence 3
Hydrophobicity: 13.418675496688737
Helical Contact Area: 4616
Total Non-Bonded Energy: 267.7500000000001
```

Comment:

- The hydrophobicity levels across all mentioned sequences are comparable, mainly due to the prevalence of Alanine and other aliphatic amino acids within the polypeptide sequences.
- Sequence 3 exhibits a notably larger helix contact area compared to sequences 1 and 2. This difference may indicate that sequence 3 potentially adopts a right-handed alpha-helical structure, while the other peptides may favor a beta-sheet structure.
- Furthermore, the total non-bonded energy of sequence 3 surpasses that of sequences 1 and 2. This increase in energy is primarily attributed to significant values of "Et" for Glycine and Serine in sequence 3, resulting in a higher overall non-bonded energy content.