Anirudh Rao
BE21B004

BT3040 – Bioinformatics

Practical 8

1

The Python code to find the hydrophobicity profiles for the sequences in the file Q1.fasta and identify the α-helices and β-strands is given below:

```python
import matplotlib.pyplot as plt
import numpy as np

values = {
    '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}

f = open("q1.fasta", mode="r")
sequences = [seq.strip() for seq in f.readlines()[1::2]]
f.close()

fig, ax = plt.subplots(nrows=3, dpi=200, figsize=(10, 10))
count = 0
for sequence in sequences:

    hydrophobicity_profile = [values[residue] for residue in sequence]
    average = np.mean(hydrophobicity_profile)
    deviation = [0 if hydrophobicity < average else 1 for hydrophobicity in
hydrophobicity_profile]
    helices = set()
    sheets = set()
    i = 0

    for i in range(len(sequence) - 4 + 1):

        if deviation[i] == deviation[i + 1] and deviation[i] != deviation[i + 2] and
deviation[i + 2] == deviation[i + 3]:
            helices.update({i, i + 1, i + 2, i + 3})
        elif deviation[i] != deviation[i + 1] and deviation[i] == deviation[i + 2] and
deviation[i] != deviation[i + 3]:
            sheets.update({i, i + 1, i + 2, i + 3})

    ax[count].plot(range(len(sequence)), hydrophobicity_profile, c="black", marker="o",
markersize=3)
    ax[count].set_title(f"\nHydrophobicity profile of sequence {count + 1}")
    ax[count].set_xlabel("Residue number")
```

```
    ax[count].set_ylabel("Hydrophobicity value")
    ax[count].axhline(average, c="black", ls="--", lw=0.75)

    for position in helices:
        ax[count].axvspan(position - 0.5, position + 0.5, alpha=0.5, color='blue')
    for position in sheets:
        ax[count].axvspan(position - 0.5, position + 0.5, alpha=0.5, color='red')

    count += 1

fig.tight_layout()
plt.show()
```
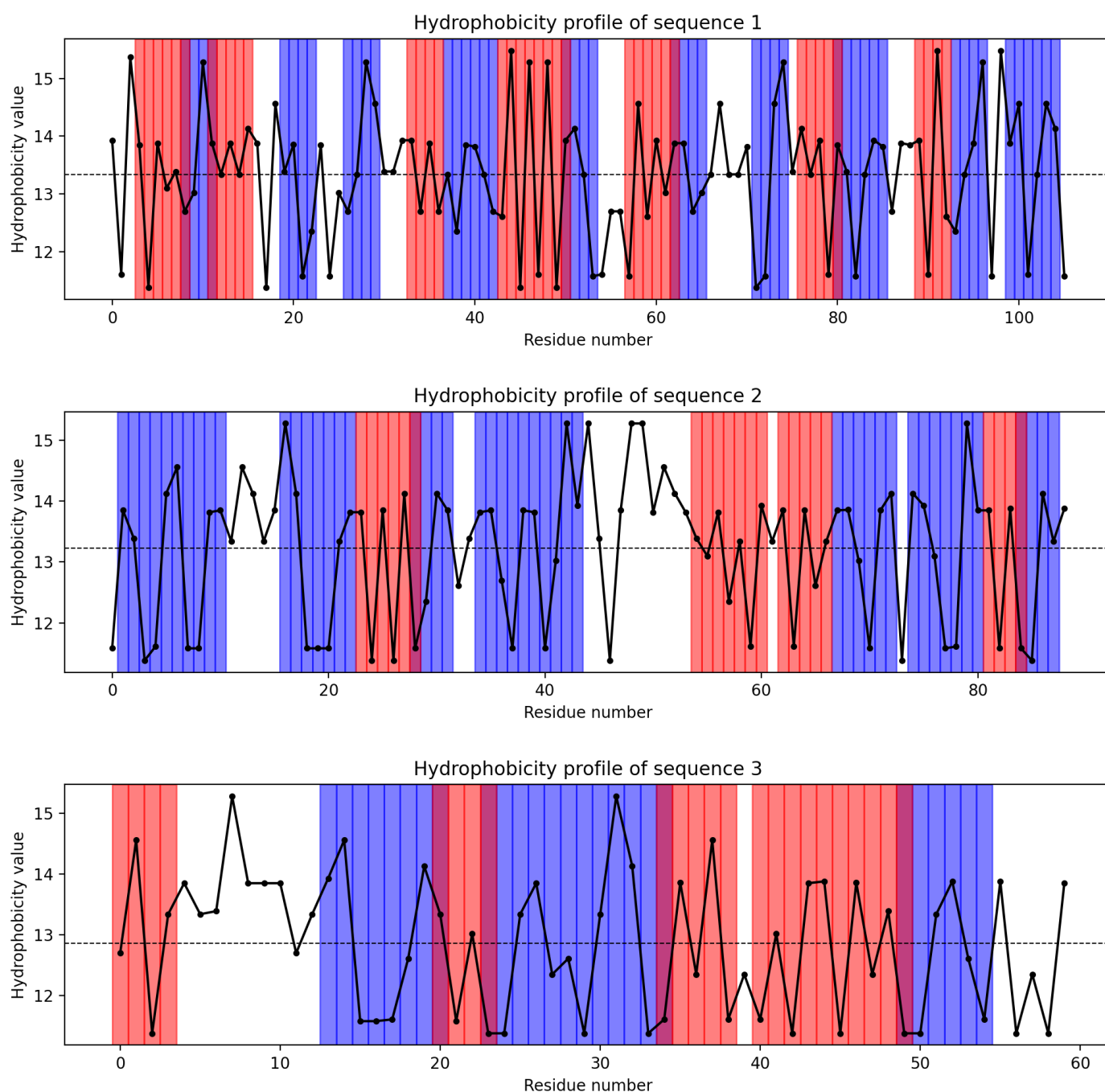
The output of this, with α-helices marked in blue and β-strands in red, is shown below:

The Python code to calculate the amphipathic indices for the helices (with length 8) and the strands (with length 6) found in Q1 is given below:

```python
import numpy as np

values = {
    '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}

f = open("Q1.fasta", mode="r")
sequences = [seq.strip() for seq in f.readlines()[1::2]]
f.close()

count = 1

for sequence in sequences:

    hydrophobicity_profile = [values[residue] for residue in sequence]
    average = np.mean(hydrophobicity_profile)
    deviation = [0 if hydrophobicity < average else 1 for hydrophobicity in
hydrophobicity_profile]
    alphas = [i for i in range(len(sequence) - 8 + 1) if
            ((deviation[i:i + 8] == [0, 0, 1, 1, 0, 0, 1, 1]) or (deviation[i:i + 8]
== [1, 1, 0, 0, 1, 1, 0, 0]))]
    betas = [i for i in range(len(sequence) - 6 + 1) if
            ((deviation[i:i + 6] == [0, 1, 0, 1, 0, 1]) or (deviation[i:i + 6] == [1,
0, 1, 0, 1, 0]))]

    alpha_amphipathicities = []

    for alpha_start in alphas:
        a1 = 0
        a2 = 0
        a3 = 0
        a4 = 0
        for j in range(0, 8, 4):
            a1 += values[sequence[alpha_start + j]]
            a2 += values[sequence[alpha_start + 1 + j]]
            a3 += values[sequence[alpha_start + 2 + j]]
            a4 += values[sequence[alpha_start + 3 + j]]
        a1 = a1 / 2
        a2 = a2 / 2
        a3 = a3 / 2
        a4 = a4 / 2
        if deviation[alpha_start] == deviation[alpha_start + 1]:
```

```python
            alpha_amphipathicities.append(abs((a1 + a2) - (a3 + a4)))
        else:
            alpha_amphipathicities.append(abs((a1 + a4) - (a2 + a3)))

    beta_amphipathicities = []

    for beta_start in betas:
        b1 = 0
        b2 = 0
        for j in range(0, 6, 2):
            b1 += values[sequence[beta_start + j]]
            b2 += values[sequence[beta_start + 1 + j]]
        b1 = b1 / 3
        b2 = b2 / 3
        if deviation[beta_start] == 0:
            beta_amphipathicities.append(b2 - b1)
        else:
            beta_amphipathicities.append(b1 - b2)

    print(f"Sequence {count}")
    if len(alphas) != 0:
        for i in range(len(alphas)):
            print(
                f"Helix of length 8 found at position {alphas[i]} with amphipathicity
{round(alpha_amphipathicities[i], 3)}")
    else:
        print("No helices of length 8 were found")

    if len(betas) != 0:
        for i in range(len(betas)):
            print(
                f"Sheet of length 6 found at position {betas[i]} with amphipathicity
{round(beta_amphipathicities[i], 3)}")
    else:
        print("No sheets of length 6 were found")
    print("")

    count += 1
```

The output of this is given below:

```
Sequence 1
No helices of length 8 were found
Sheet of length 6 found at position 3 with amphipathicity 1.313
Sheet of length 6 found at position 43 with amphipathicity 3.48
Sheet of length 6 found at position 44 with amphipathicity 3.89
Sheet of length 6 found at position 45 with amphipathicity 3.373
Sheet of length 6 found at position 57 with amphipathicity 1.72

Sequence 2
Helix of length 8 found at position 1 with amphipathicity 4.89
Helix of length 8 found at position 3 with amphipathicity 5.105
Helix of length 8 found at position 34 with amphipathicity 3.23
Helix of length 8 found at position 36 with amphipathicity 4.0
Sheet of length 6 found at position 23 with amphipathicity 2.487
Sheet of length 6 found at position 54 with amphipathicity 1.163
Sheet of length 6 found at position 55 with amphipathicity 1.343

Sequence 3
No helices of length 8 were found
Sheet of length 6 found at position 44 with amphipathicity 2.007
```

The Python code to plot the hydrophobicity profile for the sequence in the file Q2.fasta with window lengths 9 and 19 and to identify the transmembrane segments is given below:

```python
import matplotlib.pyplot as plt
import numpy as np

values = {
    '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}

f = open("Q2.fasta", mode="r")
sequences = [seq.strip() for seq in f.readlines()[1::2]]
f.close()

window_size1 = 9
half_width1 = window_size1 // 2
window_size2 = 19
half_width2 = window_size2 // 2

for sequence in sequences:

    hydrophobicity_profile = [np.mean([values[residue] for residue in sequence[i -
half_width1:i + half_width1 + 1]])
                              for i in range(half_width1, len(sequence) - half_width1 -
1)]
    average = np.mean(hydrophobicity_profile)
    deviation = [0 if hydrophobicity < average else 1 for hydrophobicity in
hydrophobicity_profile]
    transmembrane = set()
    i = 0

    for i in range(len(hydrophobicity_profile) - 4 + 1):

        if deviation[i] == 1 and deviation[i] == deviation[i + 1] and deviation[i + 1]
== deviation[i + 2] and \
                deviation[i + 2] == deviation[i + 3]:
            transmembrane.update({i, i + 1, i + 2, i + 3})

    plt.figure(dpi=200, figsize=(10, 5))
    plt.plot(range(len(hydrophobicity_profile)), hydrophobicity_profile, c="black",
marker="o", markersize=3)
    plt.title(f"Hydrophobicity profile of sequence with window length {window_size1}")
    plt.xlabel("Residue number")
    plt.ylabel("Hydrophobicity value")
    plt.axhline(average, c="black", ls="--", lw=0.75)
```

```
    for position in transmembrane:
        plt.axvspan(position - 0.5, position + 0.5, alpha=0.5, color='orange')

    plt.show()

    hydrophobicity_profile = [np.mean([values[residue] for residue in sequence[i -
half_width2:i + half_width2 + 1]])
                              for i in range(half_width2, len(sequence) - half_width2 -
1)]
    average = np.mean(hydrophobicity_profile)
    deviation = [0 if hydrophobicity < average else 1 for hydrophobicity in
hydrophobicity_profile]
    transmembrane = set()
    i = 0

    for i in range(len(hydrophobicity_profile) - 4 + 1):

        if deviation[i] == 1 and deviation[i] == deviation[i + 1] and deviation[i + 1]
== deviation[i + 2] and \
                deviation[i + 2] == deviation[i + 3]:
            transmembrane.update({i, i + 1, i + 2, i + 3})

    plt.figure(dpi=200, figsize=(10, 5))
    plt.plot(range(len(hydrophobicity_profile)), hydrophobicity_profile, c="black",
marker="o", markersize=3)
    plt.title(f"Hydrophobicity profile of sequence with window length {window_size2}")
    plt.xlabel("Residue number")
    plt.ylabel("Hydrophobicity value")
    plt.axhline(average, c="black", ls="--", lw=0.75)

    for position in transmembrane:
        plt.axvspan(position - 0.5, position + 0.5, alpha=0.5, color='orange')

    plt.show()
```
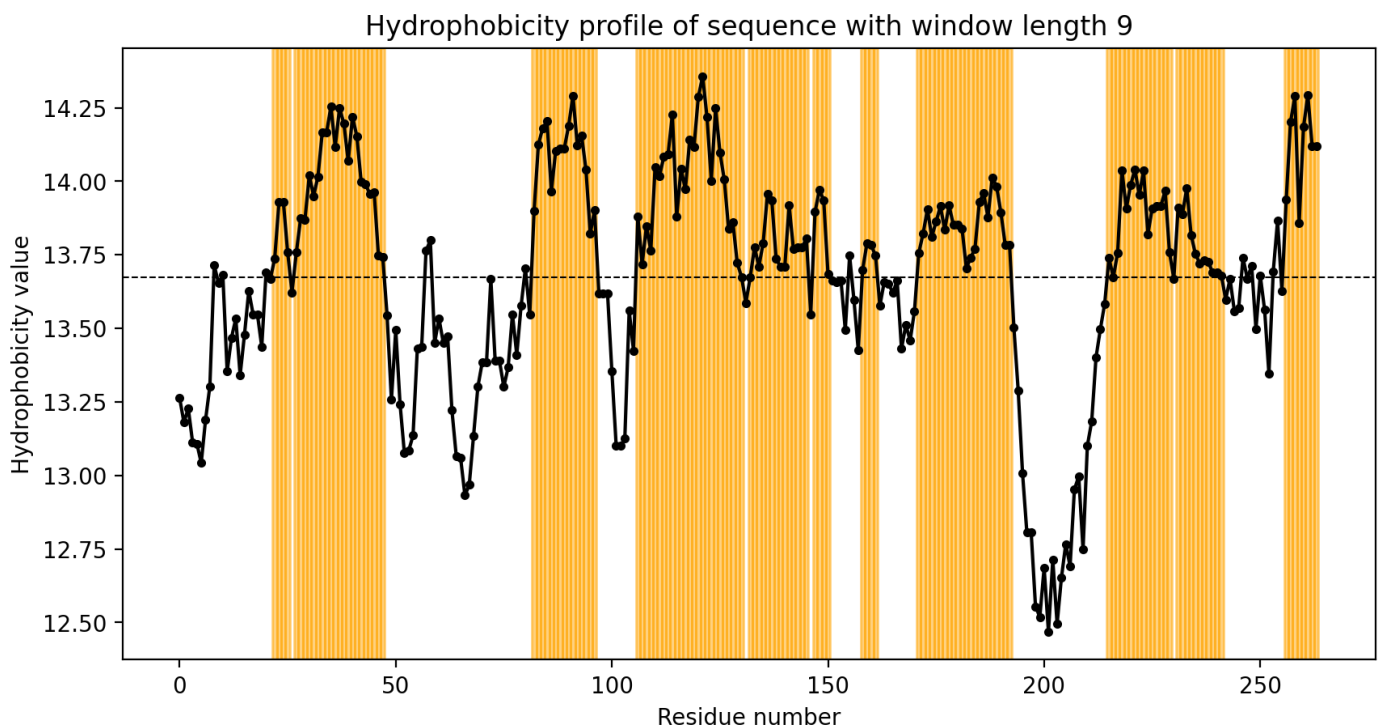
The output of this, with transmembrane segments marked in orange, is shown below:

Hydrophobicity profile of sequence with window length 9

Hydrophobicity profile of sequence with window length 19
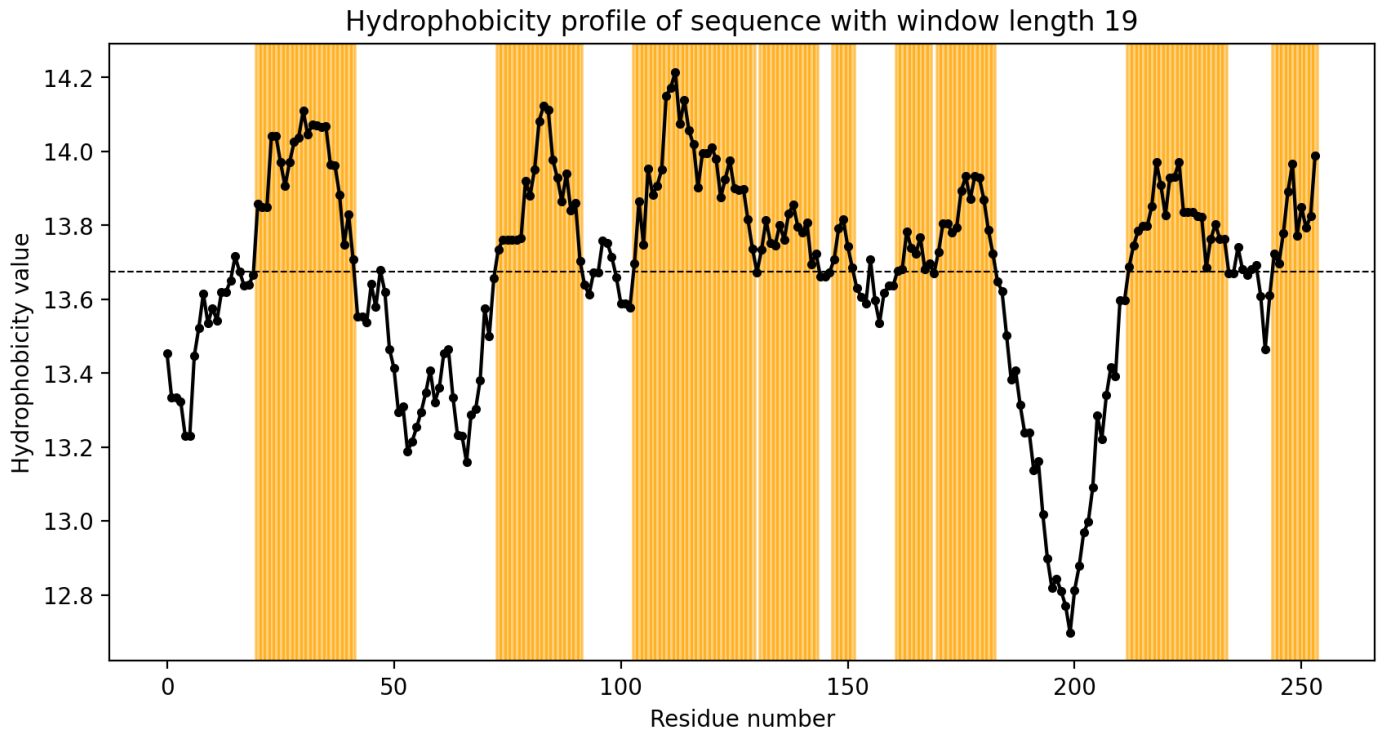
The ScanProsite tool was used to search for the given patterns against UniProtKB.

For the pattern [SV]-T-[VT]-[DERK](2)-{IL}, 11007 hits were found in 10584 sequences.

For the pattern [FILV]Qxxx{RK}Gxxx[RK]xx[FILVWY], 3837 hits were found in 3774 sequences.

The Python code to identify the patterns from Q4 in the file Q4.fasta is given below:

```python
f = open("Q4.fasta", mode="r")
content = f.readlines()
sequences = {content[i][1::].strip(): content[i + 1].strip() for i in range(0,
len(content) - 1, 2)}
f.close()

for header, sequence in sequences.items():

    pattern1_length = 6
    pattern2_length = 14

    for i in range(len(sequence) - pattern1_length + 1):
        sub_sequence = sequence[i:i + pattern1_length]
        if (sub_sequence[0] == "S" or sub_sequence[0] == "V") and (sub_sequence[1] ==
"T") and (
                sub_sequence[2] == "V" or sub_sequence[2] == "T") and (
```

```python
            sub_sequence[3] == "D" or sub_sequence[3] == "E" or sub_sequence[3] ==
"R" or sub_sequence[
            3] == "K") and (sub_sequence[4] == "D" or sub_sequence[4] == "E" or
sub_sequence[4] == "R" or sub_sequence[
            4] == "K") and (sub_sequence[5] != "I" and sub_sequence[5] != "L"):
            print(f"Match for Pattern 1 found in {header} at position {i + 1}")

    for i in range(len(sequence) - pattern2_length + 1):
        sub_sequence = sequence[i:i + pattern2_length]
        if (sub_sequence[0] in ["F", "I", "L", "V"]) and (sub_sequence[1] == "Q") and (
                sub_sequence[5] not in ["R", "K"]) and (sub_sequence[6] == "G") and (
                sub_sequence[10] in ["R", "K"]) and (sub_sequence[13] in ["F", "I",
"L", "V", "W", "Y"]):
            print(f"Match for Pattern 2 found in {header} at position {i + 1}")
```

The output of this is shown below:

```
Match for Pattern 1 found in 4A0C_2|Chains C,E|CULLIN-4B|HOMO SAPIENS (9606) at positio
n 665
Match for Pattern 1 found in 4A0K_1|Chain A|CULLIN-4A|HOMO SAPIENS (9606) at position 6
66
Match for Pattern 2 found in 4FXG_2|Chains B,E|Complement C4-A alpha chain|Homo sapiens
(9606) at position 252
Match for Pattern 2 found in 4FXK_2|Chain B|Complement C4-A Alpha chain|Homo sapiens (9
606) at position 252
Match for Pattern 2 found in 4XAM_2|Chains C,E|Complement C4-A|Homo sapiens (9606) at p
osition 175
Match for Pattern 1 found in 5F0J_3|Chain C|Sorting nexin-3|Homo sapiens (9606) at posi
tion 70
Match for Pattern 1 found in 5F0L_3|Chain C|Sorting nexin-3|Homo sapiens (9606) at posi
tion 70
Match for Pattern 1 found in 5F0M_3|Chain C|Sorting nexin-3|Homo sapiens (9606) at posi
tion 70
Match for Pattern 1 found in 5F0P_3|Chain C|Sorting nexin-3|Homo sapiens (9606) at posi
tion 70
Match for Pattern 2 found in 5JPM_2|Chains B,E|Complement C4-A|Homo sapiens (9606) at p
osition 252
Match for Pattern 2 found in 5JPN_2|Chain B|Complement C4-A|Homo sapiens (9606) at posi
tion 252
Match for Pattern 2 found in 5JTW_2|Chains B,E|Complement C4-A|Homo sapiens (9606) at p
osition 175
Match for Pattern 1 found in 5N69_1|Chains A,B|Myosin-7|Bos taurus (9913) at position 6
9
Match for Pattern 1 found in 5N6A_1|Chain A|Myosin-7|Bos taurus (9913) at position 69
Match for Pattern 1 found in 5TBY_1|Chains A,B|Myosin-7|Homo sapiens (9606) at position
69
Match for Pattern 1 found in 6FSA_1|Chains A,B|Myosin-7|Bos taurus (9913) at position 6
9
Match for Pattern 1 found in 6X5Z_3|Chains D,G,J|Myosin-7|Bos taurus (9913) at position
69
Match for Pattern 2 found in 6YSQ_1|Chains A,B|Complement C4-B,Complement C4-B|Homo sap
iens (9606) at position 854
Match for Pattern 1 found in 7JH7_2|Chains F,G,H|Myosin-7|Sus scrofa (9823) at position
69
```

Match for Pattern 1 found in 1A8J_1|Chains H,L|IMMUNOGLOBULIN LAMBDA LIGHT CHAIN DIMER (MCG)|Homo sapiens (9606) at position 204

Match for Pattern 1 found in 1ADQ_2|Chain L|IGM-LAMBDA RF-AN FAB (LIGHT CHAIN)|Homo sapiens (9606) at position 201

Match for Pattern 1 found in 1AIV_1|Chain A|OVOTRANSFERRIN|Gallus gallus (9031) at position 543

Match for Pattern 1 found in 1AQK_1|Chain L|FAB B7-15A2|Homo sapiens (9606) at position 204

Match for Pattern 1 found in 1BJM_1|Chains A,B|LOC - LAMBDA 1 TYPE LIGHT-CHAIN DIMER|Homo sapiens (9606) at position 204

Match for Pattern 1 found in 1DCL_1|Chains A,B|MCG|Homo sapiens (9606) at position 204

Match for Pattern 1 found in 1DPU_1|Chain A|REPLICATION PROTEIN A (RPA32) C-TERMINAL DOMAIN|Homo sapiens (9606) at position 86

Match for Pattern 1 found in 1E8I_1|Chains A,B|EARLY ACTIVATION ANTIGEN CD69|HOMO SAPIENS (9606) at position 19

Match for Pattern 1 found in 1FM2_2|Chain B|GLUTARYL 7-AMINOCEPHALOSPORANIC ACID ACYLASE|Brevundimonas diminuta (293) at position 112

Match for Pattern 1 found in 1FM5_1|Chain A|EARLY ACTIVATION ANTIGEN CD69|Homo sapiens (9606) at position 40

Match for Pattern 1 found in 1FNT_15|Chains c,d,e,f,g,h,i,j,k,l,m,n,o,p|PROTEASOME ACTIVATOR PROTEIN PA26|Trypanosoma brucei (5691) at position 153

Match for Pattern 1 found in 1FO9_1|Chain A|ALPHA-1,3-MANNOSYL-GLYCOPROTEIN BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE|Oryctolagus cuniculus (9986) at position 19

Match for Pattern 1 found in 1FOA_1|Chain A|ALPHA-1,3-MANNOSYL-GLYCOPROTEIN BETA-1,2-N-ACETYLGLUCOSAMINYLTRANSFERASE|Oryctolagus cuniculus (9986) at position 19

Match for Pattern 1 found in 1GPJ_1|Chain A|Glutamyl-tRNA reductase|Methanopyrus kandleri (2320) at position 316

Match for Pattern 1 found in 1IQ7_1|Chain A|Ovotransferrin|Gallus gallus (9031) at position 202

Match for Pattern 1 found in 1JT1_1|Chain A|FEZ-1, class B3 metallo-beta-lactamase|Fluoribacter gormanii (464) at position 122

Match for Pattern 1 found in 1JVK_1|Chains A,B|IMMUNOGLOBULIN LAMBDA LIGHT CHAIN|Homo sapiens (9606) at position 205

Match for Pattern 1 found in 1JVZ_2|Chain B|cephalosporin acylase beta chain|Brevundimonas diminuta (293) at position 112

Match for Pattern 1 found in 1K07_1|Chains A,B|FEZ-1 beta-lactamase|Fluoribacter gormanii (464) at position 122

Match for Pattern 1 found in 1KEH_1|Chain A|precursor of cephalosporin acylase|Brevundimonas diminuta (293) at position 281

Match for Pattern 1 found in 1KVD_2|Chains B,D|SMK TOXIN|Pichia farinosa (4920) at position 46

Match for Pattern 1 found in 1KVE_2|Chains B,D|SMK TOXIN|Pichia farinosa (4920) at position 46

Match for Pattern 1 found in 1L9Y_1|Chains A,B|FEZ-1 b-lactamase|Fluoribacter gormanii (464) at position 122

Match for Pattern 1 found in 1LGV_1|Chains A,B|IMMUNOGLOBULIN LAMBDA LIGHT CHAIN|Homo sapiens (9606) at position 205

Match for Pattern 1 found in 1LHZ_1|Chains A,B|IMMUNOGLOBULIN LAMBDA LIGHT CHAIN|Homo sapiens (9606) at position 205

Match for Pattern 1 found in 1LIL_1|Chains A,B|LAMBDA III BENCE JONES PROTEIN CLE|Homo sapiens (9606) at position 200

Match for Pattern 1 found in 1MCB_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCC_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCE_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCF_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCH_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCJ_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCL_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCO_1|Chain L|IGG1 MCG INTACT ANTIBODY (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCR_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCS_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1NL0_1|Chain L|anti-factor IX antibody, 10C12, chain L|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1OVT_1|Chain A|OVOTRANSFERRIN|Gallus gallus (9031) at position 543
Match for Pattern 1 found in 1Q1J_1|Chains L,M|Fab 447-52D, light chain|Homo sapiens (9606) at position 205
Match for Pattern 1 found in 1R5M_1|Chain A|SIR4-interacting protein SIF2|Saccharomyces cerevisiae (4932) at position 12
Match for Pattern 1 found in 1RP1_1|Chain A|PANCREATIC LIPASE RELATED PROTEIN 1|Canis lupus familiaris (9615) at position 437
Match for Pattern 1 found in 1RYX_1|Chain A|Ovotransferrin|Gallus gallus (9031) at position 543
Match for Pattern 1 found in 1RZF_1|Chain L|Fab E51 light chain|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1S5J_1|Chain A|DNA polymerase I|Sulfolobus solfataricus (2287) at position 604
Match for Pattern 1 found in 1UYP_1|Chains A,B,C,D,E,F|BETA-FRUCTOSIDASE|THERMOTOGA MARITIMA (243274) at position 371
Match for Pattern 1 found in 1VQT_1|Chain A|Orotidine 5'-phosphate decarboxylase|Thermotoga maritima (2336) at position 72
Match for Pattern 1 found in 1W72_5|Chains L,M|HYB3 LIGHT CHAIN|HOMO SAPIENS (9606) at position 202
Match for Pattern 1 found in 1WF5_1|Chain A|sidekick 2 protein|Homo sapiens (9606) at position 29
Match for Pattern 1 found in 1Z1D_1|Chain A|Replication protein A 32 kDa subunit|Homo sapiens (9606) at position 90
Match for Pattern 1 found in 1Z7Q_15|Chains c,d,e,f,g,h,i,j,k,l,m,n,o,p|proteasome activator protein PA26|Trypanosoma brucei (5691) at position 153
Match for Pattern 1 found in 1ZTM_1|Chains A,B,C|Fusion glycoprotein|Human parainfluenza virus 3 (11216) at position 228
Match for Pattern 1 found in 1ZVO_1|Chains A,B|myeloma immunoglobulin D lambda|Homo sapiens (9606) at position 202

Match for Pattern 1 found in 2APC_1|Chain A|Alpha-1,3-mannosyl-glycoprotein 2-beta-N-acetylglucosaminyltransferase|Oryctolagus cuniculus (9986) at position 13

Match for Pattern 1 found in 2B0S_1|Chain L|Fab 2219, light chain|Homo sapiens (9606) at position 206

Match for Pattern 1 found in 2B1A_1|Chain L|Fab 2219, light chain|Homo sapiens (9606) at position 206

Match for Pattern 1 found in 2B1H_1|Chain L|Fab 2219, light chain|Homo sapiens (9606) at position 206

Match for Pattern 1 found in 2BB0_1|Chains A,B|Imidazolonepropionase|Bacillus subtilis (1423) at position 114

Match for Pattern 1 found in 2BB5_1|Chains A,B|Transcobalamin II|Homo sapiens (9606) at position 335

Match for Pattern 1 found in 2DD8_2|Chain L|IGG Light Chain|Homo sapiens (9606) at position 201

Match for Pattern 1 found in 2DVV_1|Chain A|Bromodomain-containing protein 2|Homo sapiens (9606) at position 54

Match for Pattern 1 found in 2E3K_1|Chains A,B,C,D|Bromodomain-containing protein 2|Homo sapiens (9606) at position 54

Match for Pattern 1 found in 2E7N_1|Chain A|Bromodomain-containing protein 3|Homo sapiens (9606) at position 61

Match for Pattern 1 found in 2ES7_1|Chains A,B,C,D|putative thiol-disulfide isomerase and thioredoxin|Salmonella typhimurium (99287) at position 25

Match for Pattern 1 found in 2FB4_1|Chain L|IGG1-LAMBDA KOL FAB (LIGHT CHAIN)|Homo sapiens (9606) at position 204

Match for Pattern 1 found in 2FH6_1|Chain A|pullulanase|Klebsiella aerogenes (548) at position 522

Match for Pattern 1 found in 2FH8_1|Chain A|pullulanase|Klebsiella aerogenes (548) at position 522

Match for Pattern 1 found in 2FHB_1|Chain A|pullulanase|Klebsiella aerogenes (548) at position 522

Match for Pattern 1 found in 2FHC_1|Chain A|pullulanase|Klebsiella aerogenes (548) at position 522

Match for Pattern 1 found in 2FHF_1|Chain A|pullulanase|Klebsiella aerogenes (548) at position 522

Match for Pattern 1 found in 2FL5_1|Chains A,C,E,L|Immunoglobulin Igg1 Lambda Light Chain|Homo sapiens (9606) at position 200

Match for Pattern 1 found in 2G3F_1|Chains A,B|Imidazolonepropionase|Bacillus subtilis (1423) at position 114

Match for Pattern 1 found in 2G4A_1|Chain A|Bromodomain-containing protein 2|Homo sapiens (9606) at position 50

Match for Pattern 1 found in 2G75_2|Chains B,D|IGG Light Chain|Homo sapiens (9606) at position 201

Match for Pattern 1 found in 2GAN_1|Chains A,B|182aa long hypothetical protein|Pyrococcus horikoshii (70601) at position 11

Match for Pattern 1 found in 2H32_2|Chain B|Immunoglobulin omega chain|Homo sapiens (9606) at position 109

Match for Pattern 1 found in 2H3N_2|Chains B,D|Ig lambda-5|Homo sapiens (9606) at position 108

Match for Pattern 1 found in 2IDR_1|Chains A,B|Eukaryotic translation initiation factor 4E-1|Triticum aestivum (4565) at position 35

Match for Pattern 1 found in 2IDV_1|Chain A|Eukaryotic translation initiation factor 4E-1|Triticum aestivum (4565) at position 35

Match for Pattern 1 found in 2IG2_1|Chain L|IGG1-LAMBDA KOL FAB (LIGHT CHAIN)|Homo sapiens (9606) at position 204

Match for Pattern 1 found in 2J28_23|Chain O|50S RIBOSOMAL PROTEIN L18|ESCHERICHIA COLI (562) at position 52

Match for Pattern 1 found in 2J42_1|Chain A|C2 TOXIN COMPONENT-II|CLOSTRIDIUM BOTULINUM (1491) at position 338

Match for Pattern 1 found in 2J6E_3|Chains L,M|IGM|HOMO SAPIENS (9606) at position 223

Match for Pattern 1 found in 2JB5_2|Chain L|FAB FRAGMENT MOR03268 LIGHT CHAIN|HOMO SAPIENS (9606) at position 205

Match for Pattern 1 found in 2JB6_1|Chains A,L|FAB FRAGMENT MOR03268 LIGHT CHAIN|HOMO SAPIENS (9606) at position 205

Match for Pattern 1 found in 2JE8_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOMICRON (226186) at position 349

Match for Pattern 1 found in 2JE8_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOMICRON (226186) at position 475

Match for Pattern 1 found in 2JQ3_1|Chain A|Apolipoprotein C-III|Homo sapiens (9606) at position 55

Match for Pattern 1 found in 2KC8_1|Chain A|Toxin relE|Escherichia coli (83333) at position 23

Match for Pattern 1 found in 2KC9_1|Chain A|Toxin relE|Escherichia coli (83333) at position 23

Match for Pattern 1 found in 2LDX_1|Chains A,B,C,D|APO-LACTATE DEHYDROGENASE|Mus musculus (10090) at position 1

Match for Pattern 1 found in 2MCG_1|Chains 1,2|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT CHAIN)|Homo sapiens (9606) at position 204

Match for Pattern 1 found in 2MXC_1|Chain A|Sorting nexin-3|Homo sapiens (9606) at position 75

Match for Pattern 1 found in 2OAJ_1|Chain A|Protein SNI1|Saccharomyces cerevisiae (4932) at position 753

Match for Pattern 1 found in 2OAJ_1|Chain A|Protein SNI1|Saccharomyces cerevisiae (4932) at position 825

Match for Pattern 1 found in 2OLD_1|Chains A,B|Bence Jones KWR Protein - Immunoglobulin Light Chain|Homo sapiens (9606) at position 205

Match for Pattern 1 found in 2OMB_1|Chains A,B,C,D|Bence Jones KWR Protein - Immunoglobulin Light Chain|Homo sapiens (9606) at position 205

Match for Pattern 1 found in 2OMN_1|Chains A,B|Bence Jones KWR Protein - Immunoglobulin Light Chain|Homo sapiens (9606) at position 205

Match for Pattern 1 found in 2PI2_1|Chains A,B,C,D|Replication protein A 32 kDa subunit|Homo sapiens (9606) at position 257

Match for Pattern 1 found in 2QA2_1|Chain A|Polyketide oxygenase CabE|Streptomyces (1883) at position 166

Match for Pattern 1 found in 2RCJ_2|Chains C,D,G,H,K,L,O,P,S,T|IgA1 heavy chain|Homo sapiens (9606) at position 202

Match for Pattern 1 found in 2RDO_14|Chain O|50S ribosomal protein L18|Escherichia coli (562) at position 52

Match for Pattern 1 found in 2RDO_34|Chain 8|Ribosome recycling factor|Escherichia coli (562) at position 57

Match for Pattern 1 found in 2VJX_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOMICRON (226186) at position 347

Match for Pattern 1 found in 2VJX_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOMICRON (226186) at position 473

```
Match for Pattern 1 found in 2VL4_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 347
Match for Pattern 1 found in 2VL4_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 473
Match for Pattern 1 found in 2VMF_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 347
Match for Pattern 1 found in 2VMF_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 473
Match for Pattern 1 found in 2VO5_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 347
Match for Pattern 1 found in 2VO5_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 473
Match for Pattern 1 found in 2VOT_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 347
Match for Pattern 1 found in 2VOT_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 473
Match for Pattern 1 found in 2VQT_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (818) at position 347
Match for Pattern 1 found in 2VQT_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (818) at position 473
Match for Pattern 1 found in 2VQU_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (818) at position 347
Match for Pattern 1 found in 2VQU_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (818) at position 473
Match for Pattern 1 found in 2VR4_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 347
Match for Pattern 1 found in 2VR4_1|Chains A,B|BETA-MANNOSIDASE|BACTEROIDES THETAIOTAOM
ICRON (226186) at position 473
Match for Pattern 1 found in 1AA0_1|Chain A|FIBRITIN|Enterobacteria phage T4 (10665) at
position 41
Match for Pattern 1 found in 1FO8_1|Chain A|ALPHA-1,3-MANNOSYL-GLYCOPROTEIN BETA-1,2-N-
ACETYLGLUCOSAMINYLTRANSFERASE|Oryctolagus cuniculus (9986) at position 14
Match for Pattern 1 found in 1JKM_1|Chains A,B|BREFELDIN A ESTERASE|Bacillus subtilis (
1423) at position 345
Match for Pattern 1 found in 1JW0_2|Chain B|cephalosporin acylase beta chain|Brevundimo
nas diminuta (293) at position 112
Match for Pattern 1 found in 1MCD_1|Chains A,B|Immunoglobulin lambda-1 light chain|Homo
sapiens (9606) at position 204
Match for Pattern 1 found in 1MCI_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT C
HAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCK_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT C
HAIN)|Homo sapiens (9606) at position 204
Match for Pattern 1 found in 1MCN_1|Chains A,B|IMMUNOGLOBULIN LAMBDA DIMER MCG (LIGHT C
HAIN)|Homo sapiens (9606) at position 204
```

6

A FASTA file containing all the beta barrel membrane proteins in SwissProt was downloaded first.

First, the Sequence Manipulation Suite is used to identify the given pattern in the FASTA file downloaded. The pattern is modified to [KRHQFE].G[IVLFAC].[IVLFMYW].[IVLFW] to work well with the server. 466 hits were found. The results of this are shown below:

Next, ScanProsite is used to identify the given pattern in the FASTA file downloaded. The pattern is modified to [KRHQFE]-x-G-[IVLFAC]-x-[IVLFMYW]-x-[IVLFW] to work well with the server. 466 hits were found here as well. The results of this are shown below:

ScanProsite results:

Output format: Text

Hits for USERPAT1 "[KRHQFE]-x-G-[IVLFAC]-x-[IVLFMYW]-x-[IVLFW]" on UyxaXAgOi9 custom database sequences:

found: 466 hits in 321 sequences

Graphical View (graphical view with feature detection): https://p-prosite.expasy.org/cgi-bin/prosite/ScanView.cgi?scanfile=443236654235.scan.gz&sig=[KRHQFE]-x-G-[IVLFAC]-x-[IVLFMYW]-x-[IVLFW]
(link will be valid for 12hours)

Hits for:
  >USERPAT1 (user pattern):
  Pattern: [KRHQFE]-x-G-[IVLFAC]-x-[IVLFMYW]-x-[IVLFW]
  Approximate number of expected random matches in ~ 100'000 sequences (50'000'000 residues): 26784

>sp|Q06305|AER3_AERHY (492 aa)
Aerolysin-3 (Hemolysin-3). [Aeromonas hydrophila]
MKKLKITGLSLIISGLLMAQAQAAEPVYPDQLRLFSLGQEVCGDKYRPVNREEAQSVKSNIVGMMGQWQISGLANGWVIM
GPGYNGEIKPGSASSTWCYPTNPATGEIPTLSALDIPDGDEVDVQWRLVHDSANFIKPTSYLAHYLGYAWVGGNHSQYVG