Eindopdracht

Niek, Sander en Melvin

Inleiding

-   Taakverdeling

-   Opdracht in het kort;

    helix & sheet naar output

    helix, sheet en volledige sequence → graph

Alle functies

read()   - Inlezen en openen bestand

pro_sequence()   - Aanmaken proteïne sequentie

helix()   - Aanmaken Helix sequentie

sheet()   - Aanmaken Sheet sequentie

printer()   - Header aanmaken + Wegschrijven naar ander bestand

graph()   - Maken van histogram m.b.v. sequentie

graph_helix()   - Maken van histogram m.b.v. sequentie

graph_sheet()   - Maken van histogram m.b.v. sequentie

helix()

```python
def helix():
    global helix_sequence
    helix_list = []
    helix_start = 0
    helix_stop = 0
    helix_count = 0
    helix_sequence = ''
    file_obj = open(file)
    for line in file_obj:
        helix_count = 0
        if line.startswith('HELIX'):
            helix_list = line.split()
            helix_start = int(helix_list[5])
            helix_stop = int(helix_list[8])
            for char in protein_letters:
                if helix_count >= helix_start and helix_count <= helix_stop:
                    helix_sequence += char
                helix_count += 1
```

sheet()

```python
def sheet():
    global sheet_sequence
    sheet_count = 0
    sheet_sequence = ''
    file_obj = open(file)
    for line in file_obj:
        sheet_count = 0
        if line.startswith('SHEET'):
            letter = False
            sheet_list = []
            sheet_start = ''
            sheet_stop = ''
            sheet_list = line.split()
            for char in sheet_list[6]:
                if char.isalpha():
                    letter = True
                    break
                else:
                    sheet_start += char
                    sheet_stop = sheet_list[9]

            if letter == True:
                for char in sheet_list[5]:
                    if not char.isalpha():
                        sheet_start += char
                for char in sheet_list[7]:
                    if not char.isalpha():
                        sheet_stop += char

            for char in protein_letters:
                if sheet_count >= int(sheet_start) and sheet_count <= int(sheet_stop):
                    sheet_sequence += char
                sheet_count += 1
    file_obj.close()
```

```
SHEET    1 AA1 2 VAL A 393  THR A 398  0
SHEET    2 AA1 2 GLU A 479  LYS A 483 -1  O
SHEET    1 AA2 5 ILE A 488  CYS A 491  0
SHEET    2 AA2 5 LEU A 568  ASP A 572  1  O
SHEET    3 AA2 5 ARG A 600  VAL A 603  1  O
SHEET    4 AA2 5 LEU A 453  GLY A 458  1  N
SHEET    5 AA2 5 LYS A 615  LEU A 619  1  O
SHEET    1 AA3 3 ALA A1225  ILE A1234  0
SHEET    2 AA3 3 MET A1210  LYS A1218 -1  N
SHEET    3 AA3 3 ASN A1262  GLN A1268 -1  N
SHEET    1 AA4 6 PHE A1286  ILE A1289  0
SHEET    2 AA4 6 ILE A1366  ASP A1370  1  O
SHEET    3 AA4 6 THR A1396  CYS A1400  1  O
SHEET    4 AA4 6 ARG A1239  LEU A1243  1  N
SHEET    5 AA4 6 GLN A1412  GLU A1417  1  O
SHEET    6 AA4 6 LYS A1420  TYR A1424 -1  O
```

graph()

```python
def graph():
    h = '#'
    total_p = 0
    count_dict = {'C': 0, 'D': 0, 'S': 0,
                  'Q': 0, 'K': 0, 'I': 0,
                  'P': 0, 'T': 0, 'F': 0,
                  'N': 0, 'G': 0, 'H': 0,
                  'L': 0, 'R': 0, 'W': 0,
                  'A': 0, 'V': 0, 'E': 0,
                  'Y': 0, 'M': 0, '-': 0}
    for p in protein_letters:
        total_p += 1
        count_dict[p] += 1
    print('\nGrafiek voor de protein sequence:')
    for key in count_dict.keys():
        pro_p = count_dict[key] / total_p * 100
        print(key, count_dict[key], ' ', h * int(round(pro_p)), (round(pro_p)), "%")
```

```
Grafiek voor de protein sequence:
C 18    # 1 %
D 58    #### 4 %
S 125   ######## 8 %
Q 68    ##### 5 %
K 92    ###### 6 %
I 119   ######## 8 %
P 45    ### 3 %
T 83    ###### 6 %
F 86    ###### 6 %
N 55    #### 4 %
G 84    ###### 6 %
H 25    ## 2 %
L 185   ############ 12 %
R 78    ##### 5 %
W 23    ## 2 %
A 83    ###### 6 %
V 91    ###### 6 %
E 94    ###### 6 %
Y 40    ### 3 %
M 37    ## 2 %
- 19    # 1 %
```

Output command line

```
Grafiek voor de protein sequence:    Grafiek voor de helix sequence:      Grafiek voor de sheet sequence:
C 18     # 1 %                        C 11     # 1 %                        C 2      ## 2 %
D 58     #### 4 %                     D 25     ### 3 %                      D 4      #### 4 %
S 125    ######## 8 %                 S 56     ####### 7 %                  S 7      ######## 8 %
Q 68     ##### 5 %                    Q 36     #### 4 %                     Q 2      ## 2 %
K 92     ###### 6 %                   K 49     ###### 6 %                   K 4      #### 4 %
I 119    ######## 8 %                 I 72     ######### 9 %                I 12     ############# 13 %
P 45     ### 3 %                      P 18     ## 2 %                       P 1      # 1 %
T 83     ###### 6 %                   T 37     #### 4 %                     T 5      ##### 5 %
F 86     ###### 6 %                   F 58     ####### 7 %                  F 3      ### 3 %
N 55     #### 4 %                     N 20     ## 2 %                       N 2      ## 2 %
G 84     ###### 6 %                   G 37     #### 4 %                     G 6      ####### 7 %
H 25     ## 2 %                       H 15     ## 2 %                       H 2      ## 2 %
L 185    ############ 12 %            L 124    ############### 15 %         L 15     ################ 16 %
R 78     ##### 5 %                    R 48     ###### 6 %                   R 1      # 1 %
W 23     ## 2 %                       W 17     ## 2 %                       W 0      0 %
A 83     ###### 6 %                   A 60     ####### 7 %                  A 4      #### 4 %
V 91     ###### 6 %                   V 59     ####### 7 %                  V 9      ########## 10 %
E 94     ###### 6 %                   E 34     #### 4 %                     E 8      ######### 9 %
Y 40     ### 3 %                      Y 23     ### 3 %                      Y 3      ### 3 %
M 37     ## 2 %                       M 29     #### 4 %                     M 1      # 1 %
- 19     # 1 %                        - 0      0 %                          - 0      0 %
```

Uitkomst printer()

```
>5UAK MEMBRANE PROTEIN, HYDROLASE HELIX
VVSKLFFSTRPILRKGYRQRADNLSEKLEREWDRELASKKLINALRRCFFWRFMFYGIFLYLGEVTKAVQ
LLLGRIIASYDRSIAIYLGIGLCLLFIVRTLLLHPAIFGLHHIGMQMRIAMFSLIYKKTLKLSRVLDKIG
LVSLLSNNLNKFDEGLALAHFVWIAPLQVALLMGLIWSAFCGLGFLIVLALFQAGLGRMMMKYRDQRAGK
SERLVITSEMIENIIQSVKAYCWAMEKMIENLRQTELKLTRKAAYVRYFNSSAFFFSGFFVVFLSVLPYA
IKGRKIFTTISFCIVLRMAVTRQQFPWAVQTWYDSLGAINKIQDFLQKKTSLLMVIMGEIKENIIFGEYR
RSVIKACQLEEDISKFAGGQRARISLARAVYKDVLTEKEIFESCVCKLMAMEHLKKADFSELQNLQPPDF
SKLMGWNTYLRYITVSLIFVLIWCLVIFLAEVAASLVVLWLLYIYVGVAADTLLAPLVHTLITVSKILHH
MLHSVLQMSTLNTLKAGGILNRFSKDIAILDDLLLPLTIFDFIQLLLIVIGAIAVVAVLQPIFVATVPVI
AFIMLRAYFLQTSQQLKQLESEGRSPIFTHLVTSLKGLLWTLRAFGRQPYFETLFHKALNLHTANWFLYL
TLRWFQMRIEMIFVIFFIAVTFISILTTGRVGIILTLAMNIMMSTLQWAVNSSIDVDSLMRSVSRVFKFI
MPKSTLLSAFLRLLQQWRKFRKNLDPDQEIWKVADEVGRSVIEQFVDGGCHGHKQLMCLARSVLSKPSAH
DPVTYQIIRRTLKQAFAEAMLEIQKLLNERSLFLEKASVVSKLFFSW

>5UAK MEMBRANE PROTEIN, HYDROLASE SHEET
MENVTAGKIKHSFCSYLLDSILVTLAVAGSILILHILENISFSISTVKDLTAKYTEGEIQIGVIPLLLDE
ILCEVGLLGFLVIEEVRQYD

```