

# Assignment 4

---

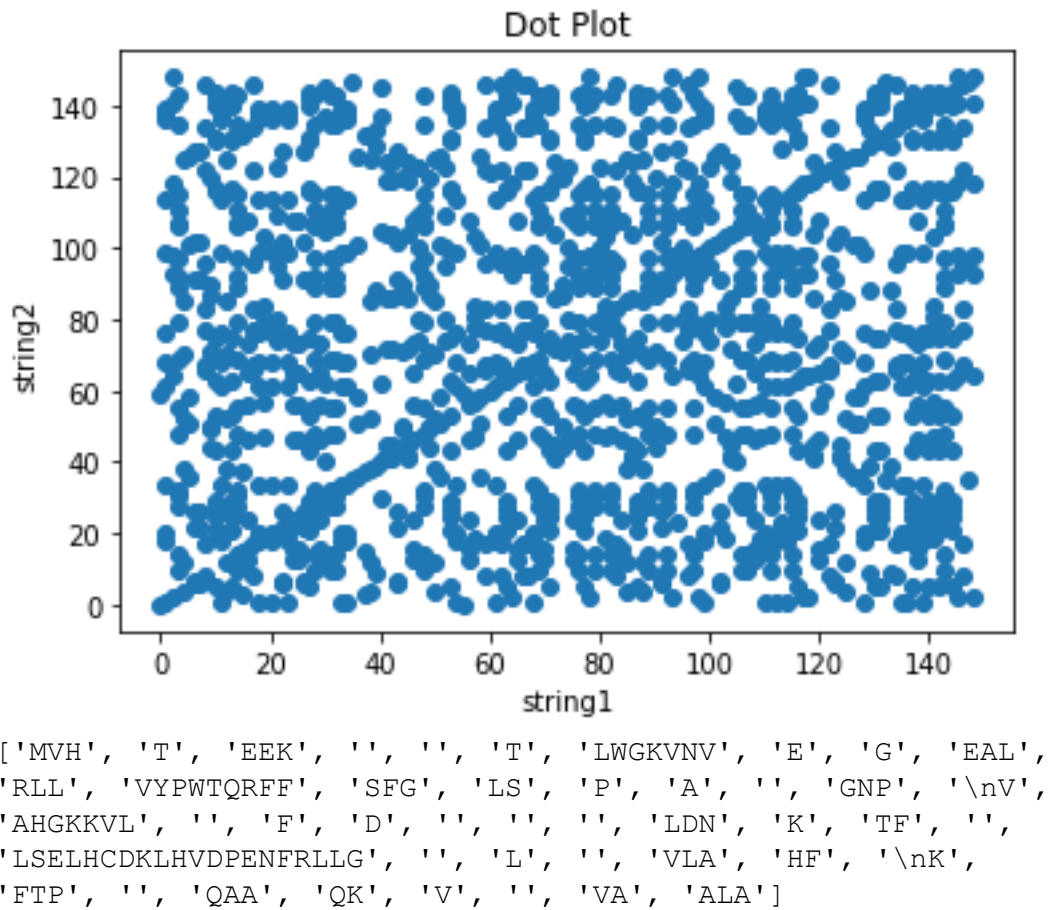
1. a. Write a program to construct a dot plot for the alignment of human and chicken hemoglobin  $\beta$  chain. Identify the segments, which are same in both sequences.

```
import matplotlib.pyplot as plt
def matchstr(s1, s2):
    seq=[]
    seqrev=[]
    tempstr=''
    for i in range(len(s1)):
        if s1[i]==s2[i]:
            tempstr+=s1[i]
        else:
            seq.append(tempstr)
            tempstr=''
    return seq
def plot(s1, s2):
    x=[]
    y=[]
    for i in range(len(s1)):
        for j in range(len(s2)):
            if s1[i]==s2[j]:
                x.append(i)
                y.append(j)
    plt.scatter(x, y)
    plt.xlabel("string1")
    plt.ylabel("string2")
    plt.title("Dot Plot")
    plt.show()

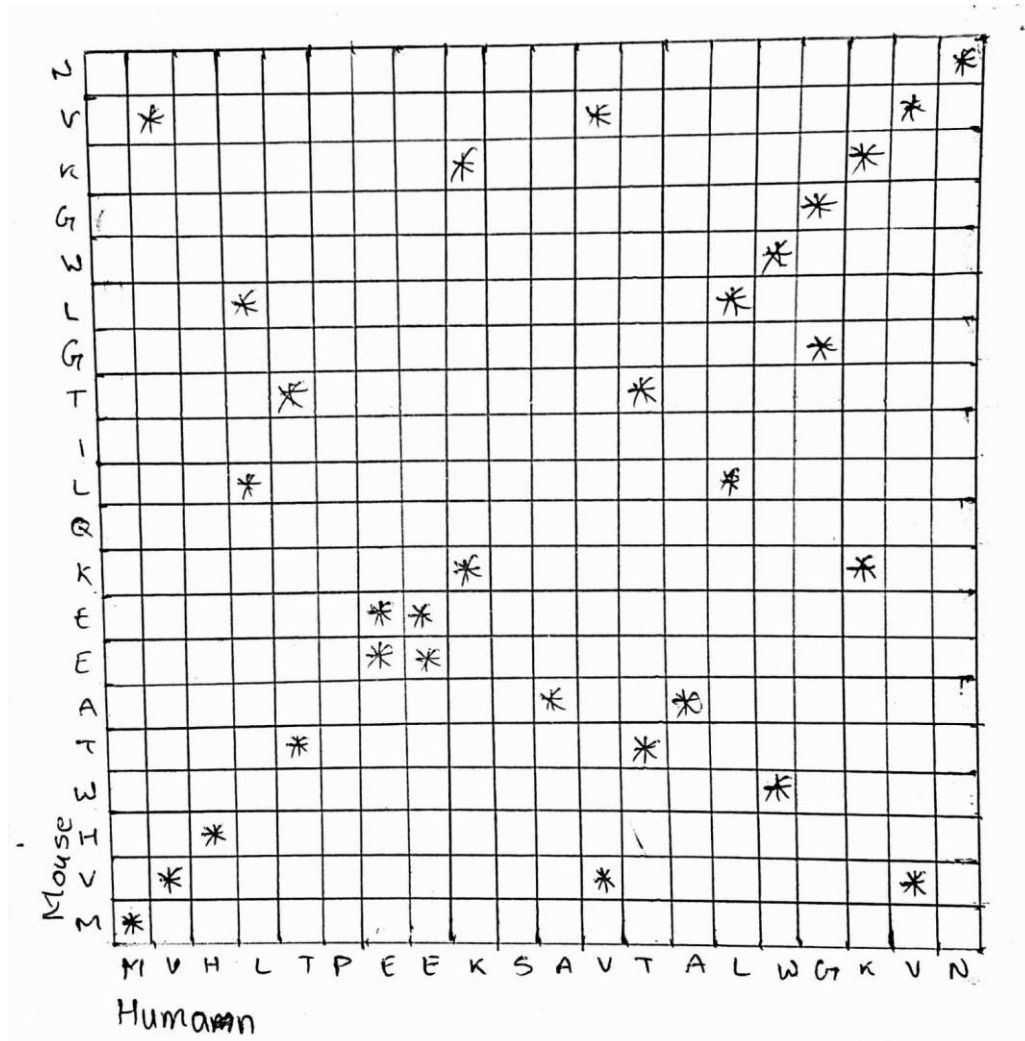
if __name__=="__main__":

    string1='' 'MVHLTPEEKSAVTALWGKVNVDVEVGGEALGRLLVVYPWTQRFFESFGDLST
PDAVMGNPK
VKAHGKKVLGAFSDGLAHL DNLKGTFTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTTPVQAA YQKV VAGVANALAHKYH' ''

    string2='' 'MVHWTAE EKQLITGLWGKVNVAECGAEALARLLIVYPWTQRFFASFGNLSS
PTAILGNPM
VRAHGKKVLTSFGDAVK NLDNIKNTFSQLSELHCDKLHVDPENFRLLGDILIIVLA AHFS
KDFTPECQAAWQKLVRVVAHALARKYH' ''
    plot(string1, string2)
    match=matchstr(string1, string2)
    print(match)
Output:
```



- b. Construct the dot plot manually for the residues 1-20 and verify with the plot obtained using program



2. Calculate the score for the following alignments using code:

AATCTATA

AAG--ATA

Assume that the match score is 1, mismatch score is 0, origination penalty is -2, and length penalty is -1.

```
def score(s1, s2):
    score=0
    for i in range(len(s1)):
        if (s1[i]==s2[i]):
            score+=1
        elif (s1[i]=='-' or s2[i]=='-'):
            if (s1[i-1]=='-' or s2[i-1]=='-'):
                score-=1
            else:
                score-=3
    return score
string1='AATCTATA'
string2='AAG--ATA'
print(score(string1, string2))
```

Output:

1

### 3. Verify the Q2 manually

String1=AATCTATA

String2=AAG--ATA

Starting from first character of two strings:

- |                     |                 |
|---------------------|-----------------|
| 1. A - A = Match    | score=+1 (0+1)  |
| 2. A - A = Match    | score=+2 (1+1)  |
| 3. T - G = mismatch | score=+2 (0+2)  |
| 4. C - - = OP + LP  | score=-1 (-3+2) |
| 5. T - - = LP       | score=-2 (-1-1) |
| 6. A - A = Match    | score=-1 (1-2)  |
| 7. T - T = Match    | score= 0 (1-1)  |
| 8. A - A = Match    | score=+1 (1+1)  |

### 4. Using the Needleman and Wunsch dynamic programming method, construct the partial alignment score table and align the following two sequences (using code):

ACAGTCGAACG and ACCGTCCG

use the scoring parameters: match score: +2; mismatch score: -1 and gap penalty: -2

```
import numpy as np
import pandas as pd

def NnWa(s1, s2):
    arr=np.zeros((len(s1)+1, len(s2)+1))
    arr[:,0]=np.arange(0,2*(-len(s1)-1), -2)
    arr[0,:]=np.arange(0,2*(-len(s2)-1), -2)
    temp=[0, 0, 0]
    for i in range(len(s1)):
        for j in range(len(s2)):
            if (s1[i]==s2[j]):
                temp[0]=arr[i,j]+2
            else:
                temp[0]=arr[i,j]-1
                temp[1]=arr[i,j+1]-2
                temp[2]=arr[i+1,j]-2
                arr[i+1,j+1]=max(temp)
    str1=" "+s1
    str2=" "+s2
    panda_df=pd.DataFrame(data=arr.astype(int), index=[i for i in str1], columns=[i for i in str2])
    print(panda_df)
    x=[]
    y=[]
    i=len(s1)
    j=len(s2)
    x.append(s1[i-1])
    y.append(s2[j-1])
    i-=1
    j-=1
    while i>0 or j>0:
```

```

        if max(arr[i][j],
arr[i][j+1],arr[i+1][j])==arr[i][j]:
            x.append(s1[i-1])
            y.append(s2[j-1])
            i-=1
            j-=1
        elif max(arr[i][j],
arr[i][j+1],arr[i+1][j])==arr[i][j+1]:
            x.append(s1[i-1])
            y.append('-')
            i-=1
        elif max(arr[i][j],
arr[i][j+1],arr[i+1][j])==arr[i+1][j]:
            x.append('-')
            y.append(s1[j-1])
            j-=1
    x=''.join(x[::-1])
    y=''.join(y[::-1])
    print(x)
    print(y)

#-----MAIN PROGRAM-----
string1='ACAGTCGAACG'
string2='ACCGTCCG'
NnWa(string1, string2)

```

Output:

		A	C	C	G	T	C	C	G
	0	-2	-4	-6	-8	-10	-12	-14	-16
A	-2	2	0	-2	-4	-6	-8	-10	-12
C	-4	0	4	2	0	-2	-4	-6	-8
A	-6	-2	2	3	1	-1	-3	-5	-7
G	-8	-4	0	1	5	3	1	-1	-3
T	-10	-6	-2	-1	3	7	5	3	1
C	-12	-8	-4	0	1	5	9	7	5
G	-14	-10	-6	-2	2	3	7	8	9
A	-16	-12	-8	-4	0	1	5	6	7
A	-18	-14	-10	-6	-2	-1	3	4	5
C	-20	-16	-12	-8	-4	-3	1	5	3
G	-22	-18	-14	-10	-6	-5	-1	3	7

ACAGTCGAACG  
ACCGTC---CG

## 5. Verify Q4 manually

Verification of points:

$pt(0,0) \Rightarrow A-A$  (match)

$$\begin{array}{c} 0 \quad -2 \\ \swarrow \quad \searrow \\ -2 \quad -2 \end{array} \quad \begin{array}{c} 2, -4, -4 \\ \max = 2 \end{array}$$

$pt(1,0) \Rightarrow A-C$  (mm)

$$\begin{array}{c} -2 \quad -4 \\ \swarrow \quad \searrow \\ 2 \quad -2 \end{array} \quad \begin{array}{c} -3, -6, -6 \\ \max = 0 \end{array}$$

$pt(2,0) \Rightarrow A-C$  (mm)

$$\begin{array}{c} -4 \quad -6 \\ \swarrow \quad \searrow \\ 0 \quad -2 \end{array} \quad \begin{array}{c} -3, -8, -2 \\ \max = -2 \end{array}$$

$pt(3,0) \Rightarrow A-G$  (mm)

$$\begin{array}{c} -6 \quad -8 \\ \swarrow \quad \searrow \\ -2 \quad -2 \end{array} \quad \begin{array}{c} -7, -10, -4 \\ \max = -4 \end{array}$$

$pt(4,0) \Rightarrow A-T$  (mm)

$$\begin{array}{c} -8 \quad -10 \\ \swarrow \quad \searrow \\ -4 \quad -2 \end{array} \quad \begin{array}{c} -9, -12, -6 \\ \max = -6 \end{array}$$

$pt(5,0) \Rightarrow A-C$  (mm)

$$\begin{array}{c} -10 \quad -12 \\ \swarrow \quad \searrow \\ -6 \quad -4 \end{array} \quad \begin{array}{c} -11, -14, -8 \\ \max = -8 \end{array}$$

$pt(6,0) \Rightarrow A-C$  (mm)

$$\begin{array}{c} -12 \quad -14 \\ \swarrow \quad \searrow \\ -8 \quad -6 \end{array} \quad \begin{array}{c} -13, -16, -10 \\ \max = -10 \end{array}$$

$pt(7,0) \Rightarrow A-G$  (mm)

$$\begin{array}{c} -14 \quad -16 \\ \swarrow \quad \searrow \\ -10 \quad -8 \end{array} \quad \begin{array}{c} -15, -18, -12 \\ \max = -12 \end{array}$$

$pt(1,1) \Rightarrow C-A$  (mm)

$$\begin{array}{c} -2 \quad 2 \\ \swarrow \quad \searrow \\ -4 \quad -2 \end{array} \quad \begin{array}{c} -3, 0, -6 \\ \max = 0 \end{array}$$

$pt(1,1) \Rightarrow C-C$  (match)

$$\begin{array}{c} 2 \quad 0 \\ \swarrow \quad \searrow \\ 0 \quad -2 \end{array} \quad \begin{array}{c} 4, -2, -2 \\ \max = 4 \end{array}$$

$pt(1,2) \Rightarrow C-C$  (match)

$$\begin{array}{c} 0 \quad -2 \\ \swarrow \quad \searrow \\ 4 \quad -2 \end{array} \quad \begin{array}{c} 2, -4, 2 \\ \max = 2 \end{array}$$

$pt(1,3) \Rightarrow C-G$  (mm)

$$\begin{array}{c} -2 \quad -4 \\ \swarrow \quad \searrow \\ 2 \quad -2 \end{array} \quad \begin{array}{c} -3, -6, 0 \\ \max = 0 \end{array}$$

In the 1<sup>st</sup> way, we can find the points...

Final alignment:

A C A G T C G A A C G  
A C C G T C - - C G

Scores:

0  $\rightarrow$  2  $\rightarrow$  4  $\rightarrow$  3  $\rightarrow$  5  $\rightarrow$  7  $\rightarrow$  9  $\rightarrow$  8  $\downarrow$  6  $\downarrow$  4  $\downarrow$  5  $\rightarrow$  7