Name – Ankeet Jena

Roll No. –

200106006

# Lab 2

1. Can you figure out the name and organism of the proteins (*Hint: use the program BLAST*)?

Ans. Protein 1 is

erythropoietin precursor [Homo sapiens]                          Homo sapiens

Protein 2 is

erythropoietin isoform 1 precursor [Mus musculus]               Mus musculus

2. Use the program Emboss-Needle to carry out **global alignment.**
   (a) Note down the sequence identity and similarity, gaps and scores. Why is the sequence similarity higher than identity? *Note: default set of parameters.*

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   167/193 (86.5%)
# Gaps:           1/193 ( 0.5%)
# Score: 765.0
#
#
#=======================================

EMBOSS_001         1 MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE     50
                     |||.|.|. |.||||||.:||||||.||||||||||||||:|||||||
EMBOSS_001         1 MGVPERPT-LLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAKEAE     49

EMBOSS_001        51 NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA    100
                     |:|.||||...|:||||||||||||||||||||||.:||:|||||:|||||
EMBOSS_001        50 NVTMGCAEGPRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEA     99

EMBOSS_001       101 VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD    150
                     :|:.||||.|||||.|.||||:|||:||||||||:|||.||||||.:||||
EMBOSS_001       100 ILQAQALLANSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPD    149

EMBOSS_001       151 AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR    193
                     ....|||||:|.|||.||||||:|||||||||||||||.||.|||
EMBOSS_001       150 TTPPAPLRTLTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR    192
```

Similarity in sequence alignment is the resemblance between two sequences when compared while Identity in sequence alignment is the number of characters that match exactly between two different sequences.

In the above sequence 167 characters are similar whereas 152 characters are identical.

(b) Now, change the scoring matrices (PAM10, PAM250, BLOSUM30 and BLOSUM90). Note down the differences in the results, if any.

## PAM10 results

```
"
# Length: 201
# Identity:     152/201 (75.6%)
# Similarity:   152/201 (75.6%)
# Gaps:          17/201 ( 8.5%)
# Score: 989.0
#
#
#=====================================

EMBOSS_001         1 MGVHECPAWLWLLLSLLSL---PLGLPVLGAPPRLICDSRVLERYLLEAK      47
                     |||.|.|.   ||| |||||    |||||||.||||||||||||||||.||||
EMBOSS_001         1 MGVPERPT---LLL-LLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAK      46

EMBOSS_001        48 EAENITTGCAEHCS---LNENITVPDTKVNFYAWKRMEVGQQAVEVWQGL      94
                     |||||.|.|||||       |.|||||||||||||||||||||||..||.||||||
EMBOSS_001        47 EAENVTMGCAE---GPRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGL      93

EMBOSS_001        95 ALLSEAVLRGQALLVNSSQPWEP--LQLHVDKAVSGLRSLTTLLRALGAQ     142
                     .||||||.|..|||||.|||||| |   |||||.||||.||||||||||.|||.||||
EMBOSS_001        94 SLLSEAILQAQALLANSSQP--PETLQLHIDKAISGLRSLTSLLRVLGAQ     141

EMBOSS_001       143 KEAISPPDAASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGD     192
                     ||..|||||....||||||.|.|||.||||||.||||||||||||||.||.||
EMBOSS_001       142 KELMSPPDTTPPAPLRTLTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGD     191

EMBOSS_001       193 R     193
                     |
EMBOSS_001       192 R     192
```

The score is increases to 989 whereas the identity and similarity is same now (75.6%) and the gaps are increased to 17.

## PAM250 results

```
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   173/193 (89.6%)
# Gaps:           1/193 ( 0.5%)
# Score: 775.0
#
#
#=======================================

EMBOSS_001         1 MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE     50
                     |||.|.|: |.||||||.:||||||.||||||||||||||||:|||||||
EMBOSS_001         1 MGVPERPT-LLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAKEAE     49

EMBOSS_001        51 NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA    100
                     |:|.||||...|:|||||||||||||||||||||.:||:|||||||:|||||
EMBOSS_001        50 NVTMGCAEGPRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEA     99

EMBOSS_001       101 VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD    150
                     :|::||||.||||||.|.||||:|||:||||||||:|||.|||||||.:|||
EMBOSS_001       100 ILQAQALLANSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPD    149

EMBOSS_001       151 AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR         193
                     ::::||||:|.|||.||||||:||||||||||||||.||.|||
EMBOSS_001       150 TTPPAPLRTLTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR         192
```

The score is 775 whereas the similarity (89.6%) is greater than identity (78.8%) and there is only 1 gap.

# BLOSUM30

```
#
# Length: 193
# Identity:      152/193 (78.8%)
# Similarity:    174/193 (90.2%)
# Gaps:            1/193 ( 0.5%)
# Score: 1011.0
#
#
#=======================================

EMBOSS_001         1 MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE      50
                     |||:|.|: |.||||||.:|||||||.||||||||||||||||:||||||
EMBOSS_001         1 MGVPERPT-LLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAKEAE      49

EMBOSS_001        51 NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA     100
                     |:|.||||...|.|||||||||||||||||||||.:||:||||||:|||||
EMBOSS_001        50 NVTMGCAEGPRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEA      99

EMBOSS_001       101 VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD     150
                     :|:.|||||:||||||.|.||||:|||:||||||||:|||:|||||||.:||||
EMBOSS_001       100 ILQAQALLANSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPD     149

EMBOSS_001       151 AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR         193
                     ::..|||||:|:|||.|||||||:|||||||||||||:||.|||
EMBOSS_001       150 TTPPAPLRTLTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR         192
```

The score is maximum (1011) in this case whereas the similarity (90.2%) is greater than identity (78.8%) and there is only 1 gap.

## BLOSUM90

```
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   166/193 (86.0%)
# Gaps:           1/193 ( 0.5%)
# Score: 886.0
#
#
#=======================================

EMBOSS_001         1 MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE     50
                     |||.|.|. |.|||||||.:||||||||.||||||||||||||||:|||||||
EMBOSS_001         1 MGVPERPT-LLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAKEAE     49

EMBOSS_001        51 NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA    100
                     |:|.||||...|.|||||||||||||||||||||||.:||:||||||:|||||
EMBOSS_001        50 NVTMGCAEGPRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEA     99

EMBOSS_001       101 VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD    150
                     :|:.||||.|||||.|.||||:|||:|||||||||:|||.|||||||.:||||
EMBOSS_001       100 ILQAQALLANSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPD    149

EMBOSS_001       151 AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR         193
                     ....||||||:|.|||.|||||||:|||||||||||||.||.|||
EMBOSS_001       150 TTPPAPLRTLTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR         192
```

The score is 886 whereas the similarity (86%) is greater than identity (78.8%) and there is only 1 gap.

(c) Show that by increasing Gap Open Penalty, the total alignment score decreases? *Note: PAM160 and BLOSUM62.*

Ans. For PAM160

**When gap penalty is 10**

```
..
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   173/193 (89.6%)
# Gaps:           1/193 ( 0.5%)
# Score: 690.0
..
```

**When gap penalty is 20**

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 20.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   173/193 (89.6%)
# Gaps:           1/193 ( 0.5%)
# Score: 680.0
#
#
```

Thus, we can see that as the gap penalty increases the alignment score decreases in PAM160 matrix.

For BLOSUM62

When gap penalty is 10

When gap penalty is 20

```
..
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   167/193 (86.5%)
# Gaps:           1/193 ( 0.5%)
# Score: 765.0
#
#
..
```

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 20.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   167/193 (86.5%)
# Gaps:           1/193 ( 0.5%)
# Score: 755.0
#
#
```

Thus, we can see that as the gap penalty increases the alignment score decreases in Blosum62 matrix.

(d) Are the alignments same after changing Gap Open Penalty? *Note: PAM160 and BLOSUM62.*

Ans. No, the alignment changes by changing the Gap Open Penalty as reflected by the alignment score.

(e) Show that by changing End Gap Open Penalty, the alignments as well as the total score remain almost similar. Why do you think it is reasonably correct? *Note: PAM160 and BLOSUM62.*

For PAM160

When end gap penalty is 10

When end gap penalty is 20

```
"
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   173/193 (89.6%)
# Gaps:           1/193 ( 0.5%)
# Score: 690.0
"
```

```
"
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   173/193 (89.6%)
# Gaps:           1/193 ( 0.5%)
# Score: 690.0
```

Thus, we can see that upon changing the gap penalty in PAM160 matrix the score remains same

For BLOSUM62

When end gap penalty is 10            When end gap penalty is 20

```
#                                        #
# Length: 193                            # Length: 193
# Identity:     152/193 (78.8%)          # Identity:     152/193 (78.8%)
# Similarity:   167/193 (86.5%)          # Similarity:   167/193 (86.5%)
# Gaps:           1/193 ( 0.5%)          # Gaps:           1/193 ( 0.5%)
# Score: 765.0                           # Score: 765.0
#                                        #
```

Thus, we can see that upon changing the gap penalty in PAM160 matrix the score remains same

Reason – The end gap penalty is the score taken away when an end gap is created but in case of a global alignment no end gaps are created hence the score remains same.

3. Use the program Emboss-Water to carry out **local alignment**.
   (a) Note down the sequence identity and similarity, gaps and scores. *Note: default set of parameters.*

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   167/193 (86.5%)
# Gaps:           1/193 ( 0.5%)
# Score: 765.0
#
#
#=======================================

EMBOSS_001         1 MGVHECPAWLWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAE     50
                     |||.|.|. |.||||||.:|||||||.|||||||||||||||:|||||||
EMBOSS_001         1 MGVPERPT-LLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAKEAE     49

EMBOSS_001        51 NITTGCAEHCSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEA    100
                     |:|.||||...|:||||||||||||||||||||||.:||:|||||:|||||
EMBOSS_001        50 NVTMGCAEGPRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEA     99

EMBOSS_001       101 VLRGQALLVNSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPD    150
                     :|:.||||.|||||.|.||||:|||:|||||||||:|||.|||||.:|||
EMBOSS_001       100 ILQAQALLANSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPD    149

EMBOSS_001       151 AASAAPLRTITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR    193
                     ....|||||:|.|||.|||||||:||||||||||||.||.|||
EMBOSS_001       150 TTPPAPLRTLTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR    192


#--------------------------------------
```

(b) Are the alignment statistics of two alignments (**global and local**) similar? If yes, what do you conclude? *Note: default set of parameters.*

Ans. Yes, this means that both the sequences are almost smilar

(c) Show that by increasing Gap Open Penalty, the total alignment score decreases up to a particular value? *Note: PAM160 and BLOSUM62*

Ans. For PAM160

## When gap penalty is 10

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   173/193 (89.6%)
# Gaps:           1/193 ( 0.5%)
# Score: 690.0
..
```

## When gap penalty is 20

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 20.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   173/193 (89.6%)
# Gaps:           1/193 ( 0.5%)
# Score: 680.0
#
#
```

## When gap penalty is 25

```
..
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 25.0
# Extend_penalty: 0.5
#
# Length: 184
# Identity:     147/184 (79.9%)
# Similarity:   167/184 (90.8%)
# Gaps:           0/184 ( 0.0%)
# Score: 679.0
#
```

## When gap penalty is 50

```
#========================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 50.0
# Extend_penalty: 0.5
#
# Length: 184
# Identity:     147/184 (79.9%)
# Similarity:   167/184 (90.8%)
# Gaps:           0/184 ( 0.0%)
# Score: 679.0
#
```

We can see that after increasing the gap penalty beyond 25, the score doesn't change and reaches a minimum value of 679.

## For BLOSUM62

When gap penalty is 10

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   167/193 (86.5%)
# Gaps:           1/193 ( 0.5%)
# Score: 765.0
#
```

When gap penalty is 20

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 20.0
# Extend_penalty: 0.5
#
# Length: 193
# Identity:     152/193 (78.8%)
# Similarity:   167/193 (86.5%)
# Gaps:           1/193 ( 0.5%)
# Score: 755.0
#
```

When gap penalty is 25

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 25.0
# Extend_penalty: 0.5
#
# Length: 184
# Identity:     147/184 (79.9%)
# Similarity:   162/184 (88.0%)
# Gaps:           0/184 ( 0.0%)
# Score: 753.0
#
```

When gap penalty is 50

```
#========================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 50.0
# Extend_penalty: 0.5
#
# Length: 184
# Identity:     147/184 (79.9%)
# Similarity:   162/184 (88.0%)
# Gaps:           0/184 ( 0.0%)
# Score: 753.0
#
#
```

We can see that after increasing the gap penalty beyond 25, the score doesn't change and reaches a minimum value of 753.

(d) Show that keeping the Gap Open Penalty its maximum value; you get the best gap-less local alignment. **Note: PAM160 and BLOSUM62**

Ans. For PAM160

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 100.0
# Extend_penalty: 0.5
#
# Length: 184
# Identity:     147/184 (79.9%)
# Similarity:   167/184 (90.8%)
# Gaps:           0/184 ( 0.0%)
# Score: 679.0
#
#
#=======================================

EMBOSS_001        10 LWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEH     59
                     |.|||||||.:||||||||.||||||||||||||:|||||||:|.||||.
EMBOSS_001         9 LLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAKEAENVTMGCAEG     58

EMBOSS_001        60 CSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQALLV    109
                     ..|:|||||||||||||||||||||.:||:|||||||:|||||:|::||||.
EMBOSS_001        59 PRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEAILQAQALLA    108

EMBOSS_001       110 NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRT    159
                     ||||||.|.||||:|||:|||||||||:|||.|||||||.:|||||::::|||||
EMBOSS_001       109 NSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPDTTPPAPLRT    158

EMBOSS_001       160 ITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR    193
                     :|.||||.|||||||:|||||||||||||||.||.|||
EMBOSS_001       159 LTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR    192
```

We can see that by keeping the Gap Open Penalty to its maximum value, we get the alignment with zero gaps and a near max alignment score

## For BLOSUM62

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 100.0
# Extend_penalty: 0.5
#
# Length: 184
# Identity:     147/184 (79.9%)
# Similarity:   162/184 (88.0%)
# Gaps:           0/184 ( 0.0%)
# Score: 753.0
#
#
#=======================================

EMBOSS_001        10 LWLLLSLLSLPLGLPVLGAPPRLICDSRVLERYLLEAKEAENITTGCAEH     59
                     |.||||||.:|||||||.||||||||||||||||:||||||||:|.||||.
EMBOSS_001         9 LLLLLSLLLIPLGLPVLCAPPRLICDSRVLERYILEAKEAENVTMGCAEG     58

EMBOSS_001        60 CSLNENITVPDTKVNFYAWKRMEVGQQAVEVWQGLALLSEAVLRGQALLV    109
                     ..|:||||||||||||||||||||||.:||:|||||:|||||:|:.||||.
EMBOSS_001        59 PRLSENITVPDTKVNFYAWKRMEVEEQAIEVWQGLSLLSEAILQAQALLA    108

EMBOSS_001       110 NSSQPWEPLQLHVDKAVSGLRSLTTLLRALGAQKEAISPPDAASAAPLRT    159
                     ||||||.|.||||:|||:|||||||:|||.||||||.:||||....|||||
EMBOSS_001       109 NSSQPPETLQLHIDKAISGLRSLTSLLRVLGAQKELMSPPDTTPPAPLRT    158

EMBOSS_001       160 ITADTFRKLFRVYSNFLRGKLKLYTGEACRTGDR         193
                     :|.|||.|||||||:||||||||||||||.||.|||
EMBOSS_001       159 LTVDTFCKLFRVYANFLRGKLKLYTGEVCRRGDR         192
```

We can see that by keeping the Gap Open Penalty to its maximum value, we get the alignment with zero gaps and a near max alignment score