Übung 6

Benjamin Sae-Chew

2)

Proteinsequenz *„Human Hemoglobin subunit alpha“*:

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLS

HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFK

LLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR

Proteinsequenz *„Human Hemoglobin subunit beta“*:

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS

TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD

PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

3) Unterschied globales zu lokalem Alignement

Lokales Alignement

```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
          |||| |||||| ||||||||||||||||
       5' TACTCACGGATGAGGTACTTTAGAGGC 3'
```

Beim lokalen Alignement vergleicht man einen Substring der Vergleichssequenz mit der Referenz

Globales Alignement

```
5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'
   ||||||||||||   |||||||  |||||||||||||||  |||||||
5' ACTACTAGATT----ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'
```

Beim globalen Alignement wird end-to-end Alignement durchgeführt. Dies bedeutet, dass die Enden der Sequenzen als Start- und Endpunkte agieren. Es können daher Gaps entstehen, wenn die zu vergleichenden Sequenzen nicht die gleiche Länge besitzen. Wenn Bereiche vertauscht sind werden diese sich negativ auf das Alignement auswirken.

4) globales Alignement mit default Parametern

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    90/149 (60.4%)
# Gaps:           9/149 ( 6.0%)
# Score: 292.5
#
#
#=======================================

EMBOSS_001         1 MV-
LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D     48
                     || |:|.:|:.|.|.||||
:..|.|.|||.|:.:.:|.|:.:|..| |
EMBOSS_001         1 MVHLTPEEKSAVTALWGKV--
NVDEVGGEALGRLLVVYPWTQRFFESFGD     48

EMBOSS_001        49 LS-----
HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR     93
                     ||
.|:.:||.|||||..|.:::.:||:|::....:.||:||..||.
EMBOSS_001        49
LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH     98

EMBOSS_001        94
VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR    142

|||.||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..||.
EMBOSS_001        99
VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH    147
```

Globales Alignement mit EPAM10 Substiutionsmatrix

```
# Length: 203
# Identity:      61/203 (30.0%)
# Similarity:    61/203 (30.0%)
# Gaps:         117/203 (57.6%)
# Score: 136.0
#
#
#=======================================

EMBOSS_001         1 MV-LSPADKTNVKAAWGKV------GAHAGEYGAEALERM-------
--F     34
                     || |.|..|..|.|.||||      |        |||.|.
|
```

```
EMBOSS_001         1 MVHLTPEEKSAVTALWGKVNVDEVGG--------
EALGRLLVVYPWTQRF     42

EMBOSS_001        35 LSFPTTKTYFPHF----DLSHGSAQ----------VKGHGKKV--A-
--DA     66
                                |    |||               ||.|||||  |
|.
EMBOSS_001        43 ------------FESFGDLS-----
TPDAVMGNPKVKAHGKKVLGAFSDG     75

EMBOSS_001        67 LTNAVAHVDDMPN-----ALSALSDLHAHKLRVDPVNFKLLSH---
CLLV    108
                      |    ||.|   |     |  .||.||..||.|||.||.||..
|.|
EMBOSS_001        76 L----AHLD---NLKGTFA--
TLSELHCDKLHVDPENFRLLGNVLVCVL-    115

EMBOSS_001       109 TLAAHLPA----EFTPAVHASLDKFLASVSTVLTSKYR---------
---    142
                        ||       ||||.|.|      |         |.
EMBOSS_001       116 ---AH---HFGKEFTPPVQA------A---------
YQKVVAGVANALAH    144

EMBOSS_001       143 ---    142

EMBOSS_001       145 KYH    147
```

globales Alignement mit GAP OPEN penalty 20

```
EMBOSS_001         1 -
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF--     47
                       :.|:|.:|:.|.|.||||
:..|.|.|||.|:.:.:|.|:.:|..|
EMBOSS_001         1 MVHLTPEEKSAVTALWGKV--
NVDEVGGEALGRLLVVYPWTQRFFESFGD     48

EMBOSS_001        48 ----
DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR     93

|...|:.:||.|||||..|.:::.:||:|::....:.||:||..||.
EMBOSS_001        49
LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH     98

EMBOSS_001        94
VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR    142

|||.||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..||.
EMBOSS_001        99
VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH    147
```

Lokales Alignement LALIGN mit default Parametern

```
>>EMBOSS_001                                                (147 aa)
 Waterman-Eggert score: 375;  106.6 bits; E(1) <  1.7e-28
43.4% identity (74.5% similar) in 145 aa overlap (3-141:4-146)

             10        20        30        40         50
EMBOSS LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-----HGSAQV
       :.: .:. : : ::::  .. : :.::: :... .: :. .:  : :::      :. .:
EMBOSS LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPKV
            10          20        30        40        50        60

         60        70        80        90       100       110
EMBOSS KGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPA
       :.::::  :.....::.:.. .....::.::  ::.::: ::.::.. :. .:: :.
EMBOSS KAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGK
              70        80        90       100       110       120

        120       130       140
EMBOSS EFTPAVHASLDKFLASVSTVLTSKY
       :::: :.:. .: .:.:...:. ::
EMBOSS EFTPPVQAAYQKVVAGVANALAHKY
             130       140

>--
 Waterman-Eggert score: 42;  15.9 bits; E(1) <  0.29
33.3% identity (76.2% similar) in 21 aa overlap (53-73:124-144)

             60        70
EMBOSS SAQVKGHGKKVADALTNAVAH
       .  :..  .::. ...::.::
EMBOSS TPPVQAAYQKVVAGVANALAH
           130       140

>--
 Waterman-Eggert score: 37;  14.5 bits; E(1) <  0.58
35.5% identity (67.7% similar) in 31 aa overlap (78-108:6-35)

        80        90       100
EMBOSS PNALSALSDLHAHKLRVDPVNFKLLSHCLLV
       :.  ::.. : . :. :: :. . :.. :.:
EMBOSS PEEKSAVTALWG-KVNVDEVGGEALGRLLVV
          10        20        30

>--
 Waterman-Eggert score: 32;  13.2 bits; E(1) <  0.89
50.0% identity (83.3% similar) in 12 aa overlap (79-90:67-78)

       80        90
EMBOSS NALSALSDLHAH
       ..:.:.::  ::
EMBOSS KVLGAFSDGLAH
         70

>--
 Waterman-Eggert score: 29;  12.4 bits; E(1) <  0.98
29.4% identity (64.7% similar) in 17 aa overlap (78-94:52-68)
```

```
        80        90
EMBOSS PNALSALSDLHAHKLRV
       :.:. .   ..::  .:
EMBOSS PDAVMGNPKVKAHGKKV
             60
```

```
>--
 Waterman-Eggert score: 29;  12.4 bits; E(1) <  0.98
60.0% identity (80.0% similar) in 10 aa overlap (11-20:127-136)
```

```
               20
EMBOSS VKAAWGKVGA
       :.::. :: :
EMBOSS VQAAYQKVVA
        130
```

```
>--
 Waterman-Eggert score: 28;  12.1 bits; E(1) <  0.99
41.7% identity (66.7% similar) in 12 aa overlap (31-42:40-51)
```

```
               40
EMBOSS ERMFLSFPTTKT
       .:.: ::   .:
EMBOSS QRFFESFGDLST
      40        50
```

```
>--
 Waterman-Eggert score: 28;  12.1 bits; E(1) <  0.99
37.5% identity (62.5% similar) in 16 aa overlap (119-134:58-73)
```

```
      120       130
EMBOSS TPAVHASLDKFLASVS
       .: :.:   : :.. :
EMBOSS NPKVKAHGKKVLGAFS
        60        70
```

```
>--
 Waterman-Eggert score: 27;  11.8 bits; E(1) <  1
46.2% identity (53.8% similar) in 13 aa overlap (112-124:2-14)
```

```
             120
EMBOSS AHLPAEFTPAVHA
       .::  :   :: :
EMBOSS VHLTPEEKSAVTA
              10
```

```
>--
 Waterman-Eggert score: 26;  11.6 bits; E(1) <  1
28.6% identity (66.7% similar) in 21 aa overlap (5-25:126-146)
```

```
           10        20
EMBOSS PADKTNVKAAWGKVGAHAGEY
       :.. .  :.. : ..: : .:
EMBOSS PVQAAYQKVVAGVANALAHKY
         130       140
```

```
>>>///

142 residues in 1 query   sequences
147 residues in 1 library sequences
 Scomplib [36.3.8g Dec, 2017]
 start: Sun Jul  8 11:16:30 2018 done: Sun Jul  8 11:16:30 2018
 Total Scan time:  0.000 Total Display time:  0.000

Function used was LALIGN [36.3.8g Dec, 2017]
```

Vergleich mit Alignement aus der Vorlesung (Folie 11)

```
HBA_HUMAN   ---------VLSPADKTNVKAAWGKVGA--HAGEYGAEALERMFLSFPTTKTYFPHF
HBB_HUMAN   --------VHLTPEEKSAVTALWGKV----NVDEVGGEALGRLLVVYPWTQRFFESF
HBA_HUMAN   -DLS-----HGSAQVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKL-
HBB_HUMAN   GDLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHL---D--NLKGTFATLSELHCDKL-
HBA_HUMAN   -RVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR------
HBB_HUMAN   -HVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH------
```

In der Vorlesung wurden noch andere Sequenzen zum Alignement hinzugezogen. Diese Sequenzen beeinflussen das Ergebnis, da alle Sequenzen eine gleiche Gewichtung haben. Die Betrachtung zweier Sequenzen, wie in unserem Fall, führt zu einem anderen Ergebnis als der Vergleich mit zusätzlichen Sequenzen. Es kommt zu beispielsweise zu anderen Gaps.

b) (1)

beim ersten Alignement wurde die BLOSUM62 Substitutionsmatrix mit default Einstellungen verwendet. BLOSUM basiert auf der Verwendung lückenloser Blöcke von ähnlicher Aminosäuresequenzen. Die 62 steht für die Homologie. Dies bedeutet, wenn die Homologie der Sequenzen mindestens 62 % entspricht wurden diese zusammengeführt bzw. geclustert. Damit wurde der Einfluss homologer Sequenzen gemindert. Alle Sequenzen unter dem Schwellenwert 62 % wurden für die Erstellung der Matrix verwendet. Die Sequenzen, welche mit Hilfe dieser Matrix aligned werden sollen, sollten optimalerweise unter 62% Ähnlichkeit sein.

Die Gap Penalty bei der default Einstellung war 10. Eine Gap wurde mit -10 Punkten „bestraft“. Desto höher die Gap Penalty desto geringer ist der Einsatz von Gaps, da diese den Score stark negativ beeinflussen.

(2)

Beim zweiten Alignement wurde PAM10(Point accepted Mutation verwendet. PAM ist das ersetzen eine Aminosäure der Primärstruktur eines Proteins unter den Prozessen der natürlichen Selektion. Die Daten für die Erstellung einer PAM-Matrix stammen meist aus den Mutationen in phylogenetischen Stammbäumen von nah verwandten Proteinen. Die Verwandtschaft der Proteine ist wichtig, da man dadurch ausschließen kann, dass die vorhandene Mutation nicht das Ergebnis einer Reihe von Mutation an derselben Stelle. Abhängig von welcher PAM Matrix ist ein Alignement zweier Sequenzen sinnvoll oder sinnlos. Beispielsweise gibt eine PAM1-Matrix die Wahrscheinlichkeit einer Substitution an, bei einer Gesamtmutation von 1% des Proteins. PAM verwendet keine Insertionen oder Deletionen.

Die Gap Penalty betrug aufgrund der default Einstellungen ebenfalls 10.

(3)

Beim dritten Alignement wurde erneut die BLOSUM62 Matrix verwendet, jedoch wurde die Gap Penalty auf 20 gesetzt. Es wurden konsequenterweise weniger Gaps eingefügt.

(4)

Das vierte Alignement wurde als lokales Alignement durchgeführt (siehe Aufgabe 3). Die Substitutionsmatrix war BLOSUM62 mit default Einstellungen (Gap Penalty 10)