

## Assignment 6

### Aufgabe 2:

#### Human Hemoglobin subunit alpha (HBA\_HUMAN) - [Quelle](#)

```
CLUSTAL O(1.2.4) multiple sequence alignment

SP|P69905|HBA_HUMAN
MVLSPADKTMVKAAMGKVGAGAGEYGARALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG 60
TR|A0A2R8Y7C0|A0A2R8Y7C0_HUMAN -----
XKAAMGKVGAGAGEYGAEALERMFSLFPTTKTYFPHFDLSHGSAQVKGHG 50
TR|G3VIN2|G3VIN2_HUMAN -----
MFLSFPTTKTYFPHFDLSHGSAQVKGHG 20

*****

SP|P69905|HBA_HUMAN
KEVADALTNVAHVVDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 120
TR|A0A2R8Y7C0|A0A2R8Y7C0_HUMAN KEVADALTNVAHVVDMPNALSALSDLHAHKLRVDPVNFKVSGGPGRINWVEGRDGA-
FLS 109
TR|G3VIN2|G3VIN2_HUMAN
KEVADALTNVAHVVDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 88
*****! : : .! * *

SP|P69905|HBA_HUMAN AVHASLDKFLASVSTVLTSKYR----- 142
TR|A0A2R8Y7C0|A0A2R8Y7C0_HUMAN GQ--RITRVAGGVAQAAAAGLGRTDPL 134
TR|G3VIN2|G3VIN2_HUMAN AVHASLDKFLASVSTVLTSKYR----- 110
. : !. ..*! . !!
```

#### Human Hemoglobin subunit beta (HBB\_HUMAN) - [Quelle](#)

```
CLUSTAL O(1.2.4) multiple sequence alignment

SP|P68871|HBB_HUMAN MVHLTFEEKSAVTALWGCKVNVDDEVGGEALGRLL-----
VVYPWTQRFPFESFG 47
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN MVHLTFEEKSAVTALWGCKVNVDDEVGGCWNSTLGPBGSLSPIGICPLMLMLWATLR---
-- 55
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN MVHLTFEEKSAVTALWGCKVNVDDEVGGEALGRLL-----
VVYPWTQRFPFESFG 47
TR|P8W6P5|P8W6P5_HUMAN MVHLTFEEKSAVTALWGCKVNVDDEVGGEALGRLL-----
VVYPWTQRFPFESFG 47
***** . * : : * :

SP|P68871|HBB_HUMAN
DLSTPDVVMGNPKVKAHGKKVLGAFSDGLAHLDNLEKGTFTATLSLHCDKLEHVDPEMFRLL 107
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN -----
--
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN
DLSTPDVVMGNPKVKAHGKKVLGAFSDGLAHLDNLEKGTFTATLSLHCDKLEHVDPEMFRVS 107
TR|P8W6P5|P8W6P5_HUMAN DLSTPDVVMGNPKVKAHGKKVLGAFSDGLAHLDNLEKGTFTATLS-----
-- 90

SP|P68871|HBB_HUMAN GNVLVCVLAHFGKEFTTPVQAAAYQKVVAGVANALAHKYH 147
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN -----
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN LWDA----- 111
TR|P8W6P5|P8W6P5_HUMAN -----
```

### Aufgabe 3:

Globales Alignment	Lokales Alignment
Alignment der Sequenzen von Anfang bis Ende.	Alignment führt zu Identifikation von einer oder mehreren besonders übereinstimmenden Regionen innerhalb der Sequenzen.
Needleman-Wunsch Algorithmus	Smith-Waterman Algorithmus

### Aufgabe 4:

#### Hemoglobin subunit alpha:

>sp|P69905|HBA\_HUMAN Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2

MVLSPADKTNVKAAWGKVGAAHAGEYGAEALERMFSLFPTTKTYFPHFDLSHGSAQVKGHG  
KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP  
AVHASLDKFLASVSTVLTSKYR

#### Hemoglobin subunit beta:

>sp|P68871|HBA\_HUMAN Hemoglobin subunit beta OS=Homo sapiens GN=HBB PE=1 SV=2

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMNGPK  
VKAHGKKVLGAFSDGLAHLNLIKGTFTLSELHHDCDKLHVDPENFRLLGNVLVCVLAHHFG  
KEFTPPVQAAYQKVVAGVANALAHKYH

## 1. Globales Alignment mit voreingestellten Parametern - BLOSUM62

```

#####
# Programs: needle
# Run dates: Tue 10 Jul 2018 16:54:10
# Command line: needle
#   -auto
#   -stdout
#   -sequence subseqs_needle-226189710-165408-9716-25262002-p2m.assembly
#   -background subseqs_needle-226189710-165408-9716-25262002-p2m.background
#   -datafile ERL0SUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -format pair
#   -xprotein1
#   -xprotein2
# Align_format: pair
# Report_file: stdout
#####

#-----
#
# Aligned_sequences: 2
# 1: ERL0SUM62
# 2: ERL0SUM62
# Matrix: ERL0SUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 151
# Identity: 65/151 (43.04%)
# Similarity: 89/151 (58.94%)
# Gaps: 11/151 ( 7.35%)
# Score: 278.0
#
#-----

ERL0SUM62_1 1 MVLEPADITGVNVAHWKVGAGHAGETGAGALSINFLSPITPTTYTPEP-D 48
|||*||*||*||*||*||*||*||*||*||*||*||*||*||*||*||*||
ERL0SUM62_2 1 MVHLEPRRPSAVTALWKKV--NVRNGGKALGRLLVVVPWTPGRFPSPGD 48

ERL0SUM62_1 49 LSHGSA-----QVQCHCKKVVNDNWKAAFTDIMPNALSLSDSH--SHK 91
||...||*||*||*||*||*||*||*||*||*||*||*||*||*||*||
ERL0SUM62_2 49 LSTPDVWNGKTVVKAHKKVLSAFEDGLAELOSLRSTPATLSLHSDCK 98

ERL0SUM62_1 92 LHVDSVRFELLGNCLLVTLAAHLFAETTPVWASLNDPLASVGVTLTKKY 141
|||*||*||*||*||*||*||*||*||*||*||*||*||*||*||*||
ERL0SUM62_2 99 LHVDSGRFELLGNYLVGVLAHHFGCFETPTPPQANLQKVVVIGVAAALAKY 148

ERL0SUM62_1 142 R 142
.
ERL0SUM62_2 149 H 149

#-----
#-----

```

### Input Parameters

program

needle

version

### 6.6.0

Matrix

EBLOSUM162

Gap open

10.0

Gap extend

0.5

### End Gap Penalty

false

End Gap Open Penalty

10.0

End Gap Extension Penalty

0.5

### Output Format

050117

Sequence Type

protein

Die BLOSUM62 (BLOcks Substitution Matrix) wird für das Sequenzalignment von Proteinen verwendet.

## 2. Globales Alignment mit einer anderen Substitution MATRIX - BLOSUM60

```
#####
# Program: needle
# Rundate: Tue 10 Jul 2018 16:59:44
# Commandline: needle
#   -auto
#   -stdout
#   -anosequence emboss_needle-I20130710-165943-0526-61799205-pls.anosequence
#   -besequence emboss_needle-I20130710-165943-0526-61799205-pls.besequence
#   -datafile EBLOSUM60
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat1 pair
#   -aprotein1
#   -aprotein2
# Align_format: pair
# Report_file: stdout
#####

#-----
#
# Aligned_sequences: 2
# 1: EMBOS8_001
# 2: EMBOS8_001
# Matrix: EBLOSUM60
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 151
# Identity:      65/151 (43.0%)
# Similarity:    90/151 (59.6%)
# Gaps:          11/151 ( 7.3%)
# Score: 264.9
#
#-----

EMBOS8_001      1  MV-LSPADKTHVKAMGKVGANAGEYGAELERKPLSEPTKTYFFHF-D      48
|| |||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
EMBOS8_001      1  MVRLTFEERSEKVTALMKV--HYDRNDEALGRILAFVFWQRFPSGQ      48

EMBOS8_001     45  LSHGSA-----QVGHGKREVDALTYDAVAVFDKHALSALSLH--ARE      91
||---| :||-||| ||..||..||:|:.....|:| |..|
EMBOS8_001     45  LSTPDAVEKQPEVSAHKEKVLGAFEDGLAHLDNLGTFATLSELHDCDE      98

EMBOS8_001     92  LRVPVNHFELLSHCLLVTLAHLPAEFTFAVEASLQKFLASVTVLTSKY      141
|.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.|||.
EMBOS8_001     95  LHVDPENFELLGNVLVGVLAHHFQKEFTFPQANYQKVVACVANALAHKY      148

EMBOS8_001     142 R      142
      .
EMBOS8_001     143 H      143

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

### Input Parameters

program  
needle

version  
6.6.0

Matrix  
EBLOSUM60

Gap open  
10.0

Gap extend  
0.5

End Gap Penalty  
false

End Gap Open Penalty  
10.0

End Gap Extension Penalty  
0.5

Output Format  
pair

Sequence Type  
protein

Die BLOSUM60 (BLOcks Substitution Matrix) wird ebenfalls für das Sequenzalignment von Proteinen verwendet.

Je höher die Zahl einer BLOSUM Matrix desto eher ist sie für Sequenzen mit einem großen zugrundeliegenden Verwandtschaftsgrad geeignet. Im Umkehrschluss bedeutet das, dass BLOSUM Matrizen mit einer niedrigen Zahl für Sequenzen mit einem geringen zugrundeliegenden Verwandtschaftsgrad geeignet sind.

Durch das Ändern der Matrix verändern sich auch die Werte für den Score und weitere Kenngrößen.



### 3. Globales Alignment mit einer anderen GAP OPEN penalty - BLOSUM62

```
#####
# Program: needle
# RunDate: Tue 10 Jul 2018 17:03:08
# Commandline: needle
#   -auto
#   -stdout
#   -sequence emboss needle-I20180710-
#   -sequence emboss_needle-I20180710-
#   -database BRCOSUN62
#   -gapopen 5.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -format3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####
```

### Input Parameters

program  
needle

version  
6.6.0

Matrix  
EBLOSUM62

Gap open  
5.0

Gap extended  
0.5

End Gap Penalty	false
-----------------	-------

End Gap Open Penalty	10.0
----------------------	------

End Gap Extension Penalty  
0.5

Output Format  
pair

Sequence Type  
protein

```

#
# Aligned sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 5.0
# Extend_penalty: 0.5
#
# Length: 153
# Identity: 66/153 (43.1%)
# Similarity: 91/153 (59.5%)
# Gaps: 15/153 (9.8%)
# Score: 307.0
#
#
#-----
EMBOSS_001      1  MV-LSPADKTVTKAAMGKV
                        |||.|.|.|.|.|||
EMBOSS_001      1  NWHLTPEEKSAVTALMGKV
#
EMBOSS_001      (9  LS-----EGSNQVKGCS
                        |||.|.|.|.|||
EMBOSS_001      (9  LSTPEAVNGK-PEKVGKNG
#
EMBOSS_001      91  -KLAVDFVWPFELLSGLLV
                        |||.|.|.|.|.|.|.|.
EMBOSS_001      97  DKLTPVDFPWFELLSGLTVG
#
EMBOSS_001      140  KTH      142
                        |||.
EMBOSS_001      147  KTH      149
#
#-----
#

```

Je geringer die gap open penalty ist, desto mehr gaps werden beim alignment zugelassen. Der Score ist durch diese Input-Parameter Änderung gestiegen.

#### 4. Lokales Alignment mit voreingestellten Parametern

Input Parameters
program
water
version
6.6.0
Matrix
EBLOSUM62
Gap open
10.0
Gap extend
0.5
Output Format
pair
Sequence Type
protein

```

#####
# Program: water
# Rundate: Tue 10 Jul 2018 17:05:52
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20180710-170551-0138-86734015-p2m.asequence
#   -bsequence emboss_water-I20180710-170551-0138-86734015-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -aformat3 pair
#   -aprotein1
#   -aprotein2
# Align_format: pair
# Report_file: stdout
#####

#-----
#
# Aligned sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 147
# Identity:      63/147 (42.9%)
# Similarity:    87/147 (59.2%)
# Gaps:          10/147 ( 6.8%)
# Score: 279.0
#
#-----

EMBOS_001      3  I SPADKTVVEAAMIEKVGHAAGSYVGATALRNPLSPPTTETTFPHF--EIGH      51
      |||.:.|:.|.|.||| |...|.|.|||.|.:.:.|.:.|.:.| |||.
EMBOS_001      4  LTPREKSAVTALMGKV--NVEEVGGEALGHLLVVYTWQMPFEPGGLSV      51
      |||.:.|:.|.|.||| |...|.|.|||.|.:.:.|.:.|.:.| |||.
EMBOS_001      52  CEA-----QVKCHKEKVALATKAVAEVDNPNALSALSULH--AHKLRV      94
      ..| | | | | | | | | | | | | | | | | | | | | | | | | | | |
EMBOS_001      52  PDVYNGSPVKAMGKKVLGATSGGLALDNLKGTATATLSELMHECCNLGV      101
      |||.:.|:.|.|.||| |...|.|.|||.|.:.:.|.:.|.:.| |||.
EMBOS_001      95  EDVNEKLLLSHCLLVLLAAHLFAEFTFAVEASLDKFLASVSTVLISKY      141
      |||.:.|:.|.|.||| |...|.|.|||.|.:.:.|.:.|.:.| |||.
EMBOS_001     102  EDENPRLLEGVLVCLVAHHFCKEFTFPVQAAYQKVVAGVAMALAHRY      148
      |||.:.|:.|.|.||| |...|.|.|||.|.:.:.|.:.|.:.| |||.

#-----
#

```

Der Score ist praktisch gleich geblieben.

Tools:

UniProt sequence database:

<https://www.uniprot.org/uniprot/>

Pairwise sequence alignment (z.B. vom EBI):

<https://www.ebi.ac.uk/Tools/psa/>