Assignment 6

Aufgabe 2:

Human Hemoglobin subunit alpha (HBA_HUMAN) - Quelle

```
CLUSTAL O(1.2.4) multiple sequence alignment

SP|P69905|HBA_HUMAN
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG 60
TR|A0A2R8Y7C0|A0A2R8Y7C0_HUMAN ----------
XKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG 50
TR|G3V1N2|G3V1N2_HUMAN         --------------------------------
MFLSFPTTKTYFPHFDLSHGSAQVKGHG 28

****************************

SP|P69905|HBA_HUMAN
KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 120
TR|A0A2R8Y7C0|A0A2R8Y7C0_HUMAN KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKVSGGPGAIWVEGRDGA-
FLS 109
TR|G3V1N2|G3V1N2_HUMAN
KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 88
                            ****************************************: .    : : .:  * *

SP|P69905|HBA_HUMAN             AVHASLDKFLASVSTVLTSKYR----- 142
TR|A0A2R8Y7C0|A0A2R8Y7C0_HUMAN GQ--RITRVAGGVAQAAAAGLGRTDPL 134
TR|G3V1N2|G3V1N2_HUMAN          AVHASLDKFLASVSTVLTSKYR----- 110
                                .    : :. ..*: . ::
```

Human Hemoglobin subunit beta (HBA_HUMAN) - Quelle

```
CLUSTAL O(1.2.4) multiple sequence alignment

SP|P68871|HBB_HUMAN             MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLL-------------
VVYPWTQRFFESFG 47
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN MVHLTPEEKSAVTALWGKVNVDEVGGCWWSTLGPRGSLSPLGICPLLMLLWATLR---
-- 55
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLL-------------
VVYPWTQRFFESFG 47
TR|F8W6P5|F8W6P5_HUMAN          MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLL-------------
VVYPWTQRFFESFG 47
                                ************************   . *              ::  *:

SP|P68871|HBB_HUMAN
DLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLL 107
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN ------------------------------------------------------------
--
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN
DLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRVS 107
TR|F8W6P5|F8W6P5_HUMAN          DLSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLS---------------
-- 90


SP|P68871|HBB_HUMAN             GNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH 147
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN ---------------------------------------
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN LWDA----------------------------------- 111
TR|F8W6P5|F8W6P5_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

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR

Hemoglobin subunit beta:

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

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMNGPK
VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHHDCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH

1. Globales Alignment mit voreingestellten Parametern - BLOSUM62

```
########################################
# Program: needle
# Rundate: Tue 10 Jul 2018 16:54:10
# Commandline: needle
#    -auto
#    -stdout
#    -asequence emboss_needle-I20180710-165408-0710-25262002-p2m.asequence
#    -bsequence emboss_needle-I20180710-165408-0710-25262002-p2m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 151
# Identity:      65/151 (43.0%)
# Similarity:    89/151 (58.9%)
# Gaps:          11/151 ( 7.3%)
# Score: 278.0
#
#
#=======================================

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

EMBOSS_001        49 LSHGSA-----QVKGHGKKVADALTNAVAHVDDMPNALSALSDLH--AHK     91
                     ||...|     :||.|||||..|.::.:||:|::....:.||:||  ..|
EMBOSS_001        49 LSTPDAVMNGPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHHDCDK     98

EMBOSS_001        92 LRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY    141
                     |.|||.||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..||
EMBOSS_001        99 LHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY    148

EMBOSS_001       142 R    142
                     .
EMBOSS_001       149 H    149


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

Input Parameters

program
needle

version
6.6.0

Matrix
EBLOSUM62

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 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
#    -asequence emboss_needle-I20180710-165943-0526-61799205-p1m.asequence
#    -bsequence emboss_needle-I20180710-165943-0526-61799205-p1m.bsequence
#    -datafile EBLOSUM60
#    -gapopen 10.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_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: 264.0
#
#
#=======================================

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

EMBOSS_001        49 LSHGSA-----QVKGHGKKVADALTNAVAHVDDMPNALSALSDLH--AHK     91
                     ||...|     :||.|||||..|.::.:||:|::....:.||:||  ..|
EMBOSS_001        49 LSTPDAVMNGPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHHDCDK     98

EMBOSS_001        92 LRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY    141
                     |.|||.||:||.:.|:..||.|...||||.|:|:..|.:|.|:..|..||
EMBOSS_001        99 LHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY    148

EMBOSS_001       142 R    142
                     .
EMBOSS_001       149 H    149


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

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
#    -asequence emboss_needle-I20180710-170305-0098-57663728-p1m.asequence
#    -bsequence emboss_needle-I20180710-170305-0098-57663728-p1m.bsequence
#    -datafile EBLOSUM62
#    -gapopen 5.0
#    -gapextend 0.5
#    -endopen 10.0
#    -endextend 0.5
#    -aformat3 pair
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# 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-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D     48
                     || |:|.:|:.|.|.||||  :..|.|.|||.|:.:.:|.|:.:|..| |
EMBOSS_001         1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD     48

EMBOSS_001        49 LS------HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAH--     90
                     ||      :| .:||.|||||..|.::.:||:|::....:.||:|| |
EMBOSS_001        49 LSTPDAVMNG-PKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELH-HDC     96

EMBOSS_001        91 -KLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTS    139
                      ||.|||.||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..
EMBOSS_001        97 DKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAH    146

EMBOSS_001       140 KYR    142
                     ||.
EMBOSS_001       147 KYH    149


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

Input Parameters

program
needle

version
6.6.0

Matrix
EBLOSUM62

Gap open
5.0

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

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

```
########################################
# 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
#    -sprotein1
#    -sprotein2
# Align_format: pair
# Report_file: stdout
########################################

#=======================================
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_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
#
#
#=======================================

EMBOSS_001         3 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSH     51
                     |:|.:|:.|.|.||||  :..|.|.|||.|:.:.:|.|:.:|..| |||.
EMBOSS_001         4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST     51

EMBOSS_001        52 GSA-----QVKGHGKKVADALTNAVAHVDDMPNALSALSDLH--AHKLRV     94
                     ..|     :||.|||||..|.::.:||:|::....:.||:||  ..||.|
EMBOSS_001        52 PDAVMNGPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHHDCDKLHV    101

EMBOSS_001        95 DPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY    141
                     ||.||:||.:.|:..||.|...||||.|.|:..|.:|.|:..|..||
EMBOSS_001       102 DPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY    148


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

Input Parameters

program
water

version
6.6.0

Matrix
EBLOSUM62

Gap open
10.0

Gap extend
0.5

Output Format
pair

Sequence Type
protein

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/