

Assignment 6

Aufgabe 2)

Human Hemoglobin subunit alpha α (HBA_HUMAN)

(Quelle: <https://www.uniprot.org/align/A201807058A530B6CA0138AFAA6D2B97CE8C2A924208005W.aln>)

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          KKVDALTNAAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 120
TR|A0A2R8Y7C0|A0A2R8Y7C0_HUMAN KKVDALTNAAHVDDMPNALSALSDLHAHKLRVDPVNFKLVGGPGAIWVEGRDGA-FLS 109
TR|G3V1N2|G3V1N2_HUMAN      KKVDALTNAAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP 88
                               *****: . : : * *

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

Human Hemoglobin subunit beta β (HBB_HUMAN)

(Quelle: <https://www.uniprot.org/align/A2018070583C3DD8CE55183C76102DC5D3A26728B20A4F8O.aln>)

CLUSTAL O(1.2.4) multiple sequence alignment

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

SP|P68871|HBB_HUMAN          DLSTPDVAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTATLSELHCDKLVDPENFRLL 107
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN -----
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN DLSTPDVAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTATLSELHCDKLVDPENFRVS 107
TR|F8W6P5|F8W6P5_HUMAN      DLSTPDVAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTATL----- 90

SP|P68871|HBB_HUMAN          GNVLVCLVAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH 147
TR|A0A0J9YWK4|A0A0J9YWK4_HUMAN -----
TR|A0A2R8Y7R2|A0A2R8Y7R2_HUMAN LWDA----- 111
TR|F8W6P5|F8W6P5_HUMAN      -----
```

Aufgabe 3)

Globales Alignment	Lokales Alignment
<ul style="list-style-type: none">Betrachtung und Vergleich beider Sequenzen komplettÄhnliche Sequenzen (damit auch etwa gleich lang) <p>➔ Needleman-Wunsch Algorithmus</p>	<ul style="list-style-type: none">Suche nach dem am besten übereinstimmenden Substrings in jeder Sequenz und alignment nur dieser TeilstückeSequenzen sind sich nur in Teilen ähnlich <p>➔ Smith-Waterman Algorithmus</p>

Aufgabe 4)

Hemoglobin subunit alpha:

```
>sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens GN=HBA1 PE=1 SV=2
MVLSPADKTNVKAAGWKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR
```

Hemoglobin subunit beta:

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

a)

(1) Globales Alignment mit voreingestellten Parametern:

```
#####
# Program: needle
# Runday: Thu 5 Jul 2018 18:37:15
# Commandline: needle
#
# -auto
# -stdout
# -asequence emboss_needle-I20180705-183711-0506-39475752-p2m.asequence
# -bsequence emboss_needle-I20180705-183711-0506-39475752-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: HBA_HUMAN
# 2: HBA_HUMAN
# 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
#
#=====

HBA_HUMAN      1  MV-LSPADKTNVKAAGWKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D  48
|||.:.|..|..|..|..|..|..|..|..|..|..|..|..|..|..|..|..|
HBA_HUMAN      1  MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD  48

HBA_HUMAN      49  LSHGSA-----QVKGHGKKVADALTNAVAHVDDMPNALSALSDLH--AHK  91
||...|      :||.|||||..|..|..|..|..|..|..|..|..|..|..|..|
HBA_HUMAN      49  LSTPDAVMNGPKVKAHGKKVLGAFSDGLAHLNKLKGTFFATLSELHHDCK  98

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

HBA_HUMAN      142  R      142
HBA_HUMAN      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

(2) Globales Alignment mit einer anderen Substitution *MATRIX*:

MATRIX: BLOSUM60

```
#####
# Program: needle
# Rundate: Thu 5 Jul 2018 18:49:50
# Commandline: needle
#
# -auto
# -stdout
# -asequence emboss_needle-I20180705-184946-0150-9901663-p2m.asequence
# -bsequence emboss_needle-I20180705-184946-0150-9901663-p2m.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: HBA_HUMAN
# 2: HBA_HUMAN
# 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
#
#-----

HBA_HUMAN      1  MV--LSPADKTNVKAANGKVGAGAHAGEYGAELER/MFLSFPTTKTYFPHF-D   48
|| |:|:|:|.|.||| |:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
HBA_HUMAN      1  MVHLTPEEKSAVTALNGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD   48

HBA_HUMAN     49  LSHGSA-----QVKGHGKKVADALTNAVAHVDDMPNALSALSDLH--AHK    91
||...| :|||...|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
HBA_HUMAN     49  LSTPDAMNGPKVKAHGKKVLGAFSDGLAHLNKLKGTFTLSELHHDCK    98

HBA_HUMAN     92  LRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY   141
|.|||.||:|:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
HBA_HUMAN     99  LHVDPENFRLLGNVLCVLAHFGKEFTPPVQAAYQKVVAGVANALAHKY   148

HBA_HUMAN    142  R      142
.
HBA_HUMAN    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

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

(3) Globales Alignment mit einer anderen *GAP OPEN penalty*:

GAP OPEN : 5

```
#####
# Program: needle
# Rundate: Thu  5 Jul 2018 18:50:48
# Commandline: needle
#
# -auto
# -stdout
# -asequence emboss_needle-I20180705-185043-0474-50408505-p2m.asequence
# -bsequence emboss_needle-I20180705-185043-0474-50408505-p2m.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: HBA_HUMAN
# 2: HBA_HUMAN
# 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
#
#=====

HBA_HUMAN      1  MV-LSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-D   48
                  || |:|:|:|.|.||| |..|.|||.|:..|.|:|:|..| |
HBA_HUMAN      1  MVHLTPEEKSAVTALMGKV--NVDEVGGEALGRLLVVYPHTQRFFESFGD   48

HBA_HUMAN     49  LS-----HGSAQVKVGHGKKVADALTNAVAHVDDMPNALSALSDLHAH--   90
                  || :|:|:|:|:|.|.|:..|:|:|:|:|:|:|:|:| |
HBA_HUMAN     49  LSTPDVAVNG-PKVKAHGKKVLGAFSDGLAHLNKLKGTFTLSELH-HDC   96

HBA_HUMAN     91  -KLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTS   139
                  ||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|..
HBA_HUMAN     97  DKLVHDPENFRLLGNVLVCVLAHHFGKEFTPPVQARYQKVAVGAVANALAH   146

HBA_HUMAN    140  KYR      142
                  ||.
HBA_HUMAN    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

Da das **gap open penalty** kleiner (5.0) und somit Negativpunkte für gaps für den Gesamtscore niedriger ist, ist der Score im Vergleich zu (1) größer (**gap open penalty** bei (1): 10.0)

(4) Lokales Alignment mit voreingestellten Parametern:

```
#####
# Program: water
# Rundate: Thu 5 Jul 2018 18:43:50
# Commandline: water
# -auto
# -stdout
# -asequence emboss_water-I20180705-184348-0898-80429575-p1m.asequence
# -bsequence emboss_water-I20180705-184348-0898-80429575-p1m.bsequence
# -datafile EBLOSUM62
# -gapopen 10.0
# -gapextend 0.5
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#-----
#
# Aligned_sequences: 2
# 1: HBA_HUMAN
# 2: HBA_HUMAN
# 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
#
#-----

HBA_HUMAN      3 LSPADKTNKAAAMGKVGAHAGEYGAELERMFLSPTTKTYPFHF-DLSH 51
                |:.|:.|:.|:.||| |.:.|:.|:.|:.|:.|:.|:.|:.|:.|:.|
HBA_HUMAN      4 LTPEKSAVTALMGKV- -NDEVGGEALGRLLVVYPWTRFFESFGDLST 51
                | | | | | | | | | | | | | | | | | | | | | | | | |
HBA_HUMAN      52 GSA-----QVKGHGKKVADALTNAVAHVDDMPNALSALSDLH- -AHLKRV 94
                ..| | | | | | | | | | | | | | | | | | | | | | | | |
HBA_HUMAN      52 PDAVNMGPKVKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHDDCKLHV 101
                | | | | | | | | | | | | | | | | | | | | | | | | |
HBA_HUMAN      95 DPVNFKLLSHCLLVTLAHLPAEFTPAWASLDKFLASVSTVLTISKY 141
                | | | | | | | | | | | | | | | | | | | | | | | | |
HBA_HUMAN     102 DPNFRLLGMVLCVLAHFGKEFTPPVQAQYKQVAGVANALAHYK 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 sowohl beim lokalem als auch beim globalem Alignment nahezu identisch.

Da das globale Alignment beide Sequenzen vergleicht, das lokale aber nur die sich ähnelnde Teilstücke kann man bei diesem Score sagen, dass die zu vergleichenden Sequenzen sich sehr ähnlich sind.

4b) (Quelle: https://www.ebi.ac.uk/Tools/psa/emboss_stretcher/help/index-protein.html)

Gap open: „Strafpunkte“ für eine einzelne Gap-Spalte im Alignment (der Score ist das Negative davon)

Gap extend: „Strafpunkte“ für weitere darauffolgende Gap-Spalten (z.B. gewertet wird hier bei drei aufeinanderfolgende Gap-Spalten nur $10+0.5+0.5 = 11$ statt $3*10 = 30$)

Matrix: BLOSUM60/62 (BLOcks Substitution Matrix)

Verschiedene Matrixe werden je nach dem Ähnlichkeitsgrad bzw. dem Verwandtschaftsgrad der vergleichenden Sequenzen. Je höher die Zahl einer BLOSUM Matrix desto besser ist sie für den Vergleich von nah verwandten Sequenzen geeignet.

Die BLOSUM-62 Matrix ist eine der geeignetsten Matrixen für die Aufklärung von schwachen Proteinähnlichkeiten .

(Quelle: https://en.wikipedia.org/wiki/Substitution_matrix)