

## Aufgabe 2:

<https://www.uniprot.org/uniprot/P69905#sequences> (8.7.18, 19:02 Uhr)

Protein-Sequenz von „Human Hemoglobin subunit alpha“(HBA\_HUMAN):

```
      10      20      30      40      50
MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHPDLS
      60      70      80      90     100
HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK
     110     120     130     140
LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR
```

<https://www.uniprot.org/uniprot/P68871#sequences> (8.7.18, 19:05 Uhr)

Protein-Sequenz von „Human Hemoglobin subunit beta“(HBB\_HUMAN):

```
      10      20      30      40      50
MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS
      60      70      80      90     100
TPDAVMGNPK VKAHGKKVLG AFSDGLAHLN NLKGTFTALS ELHCDKLHVD
     110     120     130     140
PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
```

## Aufgabe 3:

Beispielsweise möchte man die Protein - SequenzA und - SequenzB miteinander vergleichen.

Wird ein **globales Alignment** durchgeführt, werden alle Symbole (hier: Aminosäuren) mit einbezogen.

Diese Methode wird also zum genauen Vergleich ähnlicher Sequenzen (ähnlich lang und hohe erwartete Ähnlichkeit) verwendet.

Bei einem **lokalen Alignment** hingegen werden Teilsequenzen/Substrings (hier: Teilsequenzen aus A und B) aus den zu vergleichenden Sequenzen (hier: A und B) ermittelt, die sich am meisten ähneln.

Dies dient der Selektion von funktionstragenden/konservierten Sequenzabschnitten in unbekannten Sequenzen.

#### Aufgabe 4 (a):

In der Vorlesung wurden nicht nur die beiden Sequenzen untereinander, sondern auch noch mit weiteren Aminosäure-Sequenzen der Globinfamilie verglichen. Somit weicht das Alignment von den hier aufgeführten Ergebnissen ab.

##### (1) Globales Alignment mit voreingestellten Parametern:

[https://www.ebi.ac.uk/Tools/services/rest/emboss\\_needle/result/emboss\\_needle-I20180708-183728-0426-25423082-p1m/aln](https://www.ebi.ac.uk/Tools/services/rest/emboss_needle/result/emboss_needle-I20180708-183728-0426-25423082-p1m/aln) (8.7.18, 19:40 Uhr)

```
#####
# Program: needle
# Rundate: Sun  8 Jul 2018 18:37:33
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180708-183728-0426-25423082-p1m.asequence
#   -bsequence emboss_needle-I20180708-183728-0426-25423082-p1m.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: 149
# Identity:      65/149 (43.6%)
# Similarity:    90/149 (60.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 292.5
#
#
#=====

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

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDDLHAKLR      93
                  ||  .|:|:|.|.|||||.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOSS_001     49 LSTPDVAVMGNPKVKAHGKKVLGAFSDGLAHLNLDLKGTFATLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSLHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR      142
                  |||.||:|.|:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOSS_001     99 VDPENFRLLGNVLVLCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
```

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

[https://www.ebi.ac.uk/Tools/services/rest/emboss\\_needle/result/emboss\\_needle-I20180708-184600-0497-41743704-p1m/aln](https://www.ebi.ac.uk/Tools/services/rest/emboss_needle/result/emboss_needle-I20180708-184600-0497-41743704-p1m/aln) (8.7.18, 19:48 Uhr)

In **MATRIX** wurde „**BLOSUM62**“ zu „**BLOSUM70**“ geändert.

```
-----
# Program: needle
# Rundate: Sun  8 Jul 2018 18:46:05
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180708-184600-0497-41743704-p1m.asequence
#   -bsequence emboss_needle-I20180708-184600-0497-41743704-p1m.bsequence
#   -datafile EBLOSUM70
#   -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: EBLOSUM70
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    88/149 (59.1%)
# Gaps:          9/149 ( 6.0%)
# Score: 281.5
#
#
#=====

EMBOSS_001      1 MV-LSPADKTNVKAANGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                || |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAAHVDDMPNALSALSDLHAHKLR      93
                ||      .|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001     49 LSTPDVAVMGNPVKVKAHGKKVLGAFSDGLAHLNLIKGTFFATLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR      142
                |||.||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
```

### (3) Globales Alignment mit einer anderen GAP OPEN penalty

[https://www.ebi.ac.uk/Tools/services/rest/emboss\\_needle/result/emboss\\_needle-I20180708-185810-0364-68103356-p2m/aln](https://www.ebi.ac.uk/Tools/services/rest/emboss_needle/result/emboss_needle-I20180708-185810-0364-68103356-p2m/aln) (8.7.18, 19:95 Uhr)

GAP OPEN penalty wurde von „10“ auf „50“ erhöht

```
#####
# Program: needle
# Rundate: Sun  8 Jul 2018 18:58:11
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180708-185810-0364-68103356-p2m.asequence
#   -bsequence emboss_needle-I20180708-185810-0364-68103356-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 50.0
#   -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: 50.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      61/149 (40.9%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 210.0
#
#=====

EMBOSS_001      1 -MVLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF--      47
                  .:.|.:|.:|.:|.|||  :..|.|.|||.|:~::~|.|:~::~|..|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     48 ----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
                  |...|:~::~|.|.||||..|.~::~|:|:~::~.....|:|:~::~|..|..|
EMBOSS_001     49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTATLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR     142
                  |||.||:|.|.:~::~|..|..|..|..|..|..|..|..|..|..|..|..|..|
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH     147
```

#### (4) Lokales Alignment mit voreingestellten Parametern:

[https://www.ebi.ac.uk/Tools/services/rest/emboss\\_water/result/emboss\\_water-I20180708-190306-0080-44401789-p1m/aln](https://www.ebi.ac.uk/Tools/services/rest/emboss_water/result/emboss_water-I20180708-190306-0080-44401789-p1m/aln) (8.7.18, 20:04 Uhr)

```
#####
# Program: water
# Rundate: Sun  8 Jul 2018 19:03:08
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20180708-190306-0080-44401789-p1m.asequence
#   -bsequence emboss_water-I20180708-190306-0080-44401789-p1m.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: 145
# Identity:      63/145 (43.4%)
# Similarity:    88/145 (60.7%)
# Gaps:          8/145 ( 5.5%)
# Score: 293.5
#
#=====

EMBOSS_001      3 LSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-      50
                  |:|:|:|:|.|.||||  :..|.|.||||.|:~::~|.|:~::~|.|. | ||
EMBOSS_001      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST      51

EMBOSS_001     51 ----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLRVDP      96
                  .|:~::~|.|||||.|.|:~::~|:|:~::~.....|:|:|..|||.|||
EMBOSS_001     52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTATLSELHCDKLVHVP     101

EMBOSS_001     97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKY      141
                  .|:|:|:|.|.|:~::~|.|.|||||.|.|:~::~|.|:~::~|.|. | ||
EMBOSS_001    102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146
```

#### **Aufgabe 4 (b):**

**(1):**

Es wurden die voreingestellten Parameter für die MATRIX: BLOSUM62 und die GAP OPEN penalty: 10 verwendet.

*BLOSUM:*

„In bioinformatics, the BLOSUM (BLOcks SUBstitution Matrix) matrix is a substitution matrix used for sequence alignment of proteins. BLOSUM matrices are used to score alignments between evolutionarily divergent protein sequences. They are based on local alignments.”

*BLOSUM 62:*

“The BLOSUM62 matrix with the amino acids in the table grouped according to the chemistry of the side chain [...]. Each value in the matrix is calculated by dividing the frequency of occurrence of the amino acid pair in the BLOCKS database, clustered at the 62% level, divided by the probability that the same two amino acids might align by chance. The ratio is then converted to a logarithm and expressed as a log odds score, as for PAM. BLOSUM matrices are usually scaled in half-bit units. A score of zero indicates that the frequency with which a given two amino acids were found aligned in the database was as expected by chance, while a positive score indicates that the alignment was found more often than by chance, and negative score indicates that the alignment was found less often than by chance.”

Es wird also erwartet, dass die Sequenzen durchschnittlich zu 62% identisch sind.

(Quelle: <https://en.wikipedia.org/wiki/BLOSUM>, 8.7.18, 20:21 Uhr)

BLOSUM mit hoher hinten angestellter Zahl -> für ähnlichere Sequenzen

BLOSUM mit niedriger hinten angestellter Zahl -> weniger ähnliche Sequenzen

BLOSUM62 = Mittelklasse

*GAP penalty:*

“A Gap penalty is a method of scoring alignments of two or more sequences. When aligning sequences, introducing a gaps in the sequences can allow an alignment algorithm to match more terms than a gap-less alignment can. However, minimizing gaps in an alignment is important to create a useful alignment. Too many gaps can cause an alignment to become meaningless. Gap penalties are used to adjust alignment scores based on the number and length of gaps.”

(Quelle: [https://en.wikipedia.org/wiki/Gap\\_penalty](https://en.wikipedia.org/wiki/Gap_penalty), 8.7.18, 20:40 Uhr)

GOP („gap open penalty“) = Kosten für das Beginnen einer Lücke

Je höher also die GAP penalty, desto eher werden Lücken (Gaps) im Alignment vermieden bzw, desto schlechter wird der Score beim Einsetzen von Lücken. Bei einem GAP OPEN von „10“, werden also 10 Score-Punkte je GAP abgezogen.

**(2):**

Der Parameter „BLOSUM“ wurde hochgestellt auf „BLOSUM70“. Es wurde als davon ausgegangen, dass sich die verglichenen Sequenzen stärker ähneln müssten. Abweichungen werden also stärker gewichtet und verringern den Score.

Es wird also erwartet, dass die Sequenzen durchschnittlich zu 70% identisch sind.

Der Parameter „GAP OPEN penalty“ blieb unverändert.

**(3):**

Der Parameter „GAP OPEN penalty“ wurde auf „50“ hochgestellt. Somit wurden die verwendeten GAPs negativer gewichtet und setzten den Score herab.

Eine verwendete GAP setzte den Score um 50 Punkte hinab.

Der Parameter „BLOSUM“ blieb unverändert.

**(4):**

Die verwendeten Parameter wurden wie in (1) verwendet.

Der Score hat sich dennoch erhöht (verbessert), da lokal aligned (siehe oben) wurde. Es wurden also nur sehr ähnliche Sequenzbereiche einbezogen und in die Score-Rechnung integriert.