

Aufgabe 2

HBA

<https://www.uniprot.org/uniprot/P69905>

```
MVLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  
KKVADALTNVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTF  
AVHASLSDKFLASVSTVLTSKYR
```

HBB

<https://www.uniprot.org/uniprot/P68871>

```
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDVAVMGNPK  
VKAHGKKVLGAFSDGLAHLDNLKGTFTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  
KEFTTPVQAAVQKVVAGVANALAHKYH
```

Aufgabe 3

Die zwei Arten des Alignments (global, lokal) weisen zunächst eine unterschiedliche algorithmische Herangehensweise auf. Globale Alignments werden meist mit dem Needleman-Algorithmus ermittelt und lokale mit dem Waterman-Algorithmus.

Die beiden Alignments werden für unterschiedliche Zwecke verwendet. Globale Alignments für Sequenzen, welche über ihre komplette Länge eine große Ähnlichkeit aufweisen und optimalerweise gleich lang sind. Daher werden diese Alignments für Segmente verwendet, welche biologisch hoch konserviert sind bzw. zur Prüfung dessen.

Lokale Alignments suchen in einer Sequenz nach einer Teilsequenz, welche am besten zu einer weiteren gegeben (Teil-)Sequenz passt. Aus den Teilstücken wird das Alignment berechnet.

Aufgabe 4

(1)

```
#=====
#
# 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-LSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
      || |:|.:|.:|.|.||| :..|.|.|||.|:~.:|.:~.:|..| |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLR      93
      || .|:~.:|.|.|||||.|.~.:~.:|:~.:~.:~.:~.:|:~.:|..|.
EMBOSS_001     49 LSTPDVAVMGNPKVKAHGKKVLGAFSDGLAHLNLRKGTATLSELHCDKLH      98

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

Algorithmus: NEEDLE

Matrix: EBLOSUM62

End gap open: 10.0

BLOSUM62:

BLOSUM (Blocks Substitution Matrix) generell ist eine Datenbasis, welche aus verschiedenen Proteinfamilien mit verschiedenen evolutionären Abständen erstellt wurde. Mit dieser Matrix wird ein Verwandtschaftsmaß je Aminosäurepaar festgestellt und aufsummiert. Hierbei zählen physiko-chemische Eigenschaften und die natürliche Mutation/Veränderung von Aminosäuren mit hinein.

Die Zahl nach dem Namen der Matrix beschreibt, wie sehr die Matrix für ähnliche Sequenzen verwendet werden kann, da sich hierbei die jeweiligen Scores verändern. Kleine Zahlen werden für Sequenzen mit größeren evolutionären Unterschieden verwendet und große Zahlen für geringe evolutionäre Unterschiede.

BLOSUM62 ist der heutige Standard.

(2)

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM90
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    83/149 (55.7%)
# Gaps:          9/149 ( 6.0%)
# Score: 311.5
#
#=====

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

EMBOSS_001     49 LSH-----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
                  ||.      |.:|.|.|||||.|.:.:.|.|.:.:.:.|.|:|..||.
EMBOSS_001     49 LSTPDVAMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCCKLH      98

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

Algorithmus: NEEDLE

Matrix: EBLOSUM90

End gap open: 10.0

BLOSUM (Blocks Substitution Matrix) generell ist eine Datenbasis, welche aus verschiedenen Proteinfamilien mit verschiedenen evolutionären Abständen erstellt wurde. Mit dieser Matrix wird ein Verwandtschaftsmaß je Aminosäurepaar festgestellt und aufsummiert. Hierbei zählen physiko-chemische Eigenschaften und die natürliche Mutation/Veränderung von Aminosäuren mit hinein.

BLOSUM90 wird für Sequenzen mit sehr geringen Unterschieden verwendet.

(3)

```
#####
# Program: needle
# Rundate: Sun  8 Jul 2018 15:22:33
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180708-152231-0121-53139733-plm.asequence
#   -bsequence emboss_needle-I20180708-152231-0121-53139733-plm.bsequence
#   -datafile EBLOSUM62
#   -gapopen 10.0
#   -gapextend 0.5
#   -endopen 50.0
#   -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-LSPADKTNVKAAWGKVGGAHAGEYGAELERMFLSFPTTKTYFPHF-D      48
|| |:|.:|.:|.|.|||| :..|.|.|.|.|:..:|.|.:|.|.| |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAAVAVDDMPNALSALSDDLHAHKLR      93
||      .|:..|.|.|||||.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOSS_001     49 LSTPDVAVMGNPKVKAHGKKVLGAFSDGLAHLNKLKGTATLSELHCDKLH      98

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

Algorithmus: NEEDLE

Matrix: EBLOSUM62

End gap open: 50.0

Die "end gap open penalty" wird als ganzzahliger Wert nach der Ermittlung des gesamten Scores von selbigem abgezogen, wenn am N-oder C-Terminus der Sequenzen eine Lücke vorhanden ist. Dies ist bei den gegebenen Sequenzen nicht der Fall.

(4)

```
#=====
#
# 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  LSPADKTNVKAANGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-      50
   |:|:|:|.|.|.||||   :..|.|.||||.|:~::~|. |:~::~|. | ||
EMBOSS_001      4  LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST      51

EMBOSS_001     51  ----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDDLHAHKLRVDP      96
   .|:~::~|. | || |||.|.|.~::~| |:|:~::~|. |:|:~::~|. | ||
EMBOSS_001     52  PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTLSELHCDKLHVDP      101

EMBOSS_001     97  VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY      141
   -| |:|:|:|. |:~::~|. | ..| |||.|. |:~::~|. |:~::~|. | ..|
EMBOSS_001    102  ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146
```

Algorithmus: WATER

Matrix: EBLOSUM62

End gap open: 10.0

Der Waterman-Algorithmus wird vor allem für das lokale Alignment von Teilsequenzen verwendet und ist eine Modifikation des Needleman-Algorithmus. Im vorliegenden Fall wurde die globale Sequenz des HBA und HBB als Teilsequenz betrachtet und aligniert. Die Verwendung des Algorithmus ergab keine signifikante Veränderung des Scores oder Identität [%] der Sequenz, im Vergleich zu den anderen Alignment-Varianten.

YAFDLGYTCMFPVLLGGGELHIVQKETYTAPDEIAHYIKEHGITYIKLTPSLFHTIVNTASFAFDANFESLRLIVLGGEKIIPIDVIAFRKMYGHTE-FINHYGPTEATIGA
 -AFDVSAGDFARALLTGGQLIVCPNEVKMDPASLYAIIKKYDITIFEATPALVIPLMEYI-YEQKLDISQLQILIVGSDSCSMEDFKTLVSRFGSTIRIVNSYGVTEACIDS
 IAFDASSWEIYAPLLNGGTVCIDYYTTIDIKALEAVFKQHHIRGAMLPALQKCLVSA-----PTMISSLEILFAAGDRLSSQDAILARRAVGSGV-Y-NAYGPTENTVLS

Alignment von Folie 9 (?)

Sämtliche berechneten Alignments, welche in (1) – (4) gezeigt wurden, zeigen eine ungefähr gleiche Identität von 43.6 %. Im Vergleich zum Alignment der Vorlesung fällt auf, dass hierbei eine geringe Identität gegeben ist: Es sind 112 Aminosäuren gegeben, wovon 19 identisch sind. Dies ergibt eine Identität von ca. 17.0 %. Daher ist das Alignment aus der Vorlesung wesentlich schlechter.