

# Übungsblatt 6

## Aminosäuresequenz der „Human Hemoglobin subunit alpha“:

MVLSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHGKKVADALTNAVAH  
VDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR

## Aminosäuresequenz der „Human Hemoglobin subunit beta“:

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKVKAHGKKVLGAFSD  
GLAHLNKLKGTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH

## Globales Alignment vs Lokales Alignment:

Bei einem globalen Alignment werden alle Symbole, in diesem Fall alle Aminosäureabkürzungen, der zwei zu vergleichenden Sequenzen betrachtet. Im Gegensatz dazu wird bei einem lokalen Alignment nur eine oder mehrere Teilsequenzen berücksichtigt. Einsatz findet das globale Alignment bei Sequenzen, die ähnlich lang sind und eine starke Ähnlichkeit erwartet wird. Die benutzte Methode des globalen Alignments ist der Needleman-Wunsch-Algorithmus.

Ein lokales Alignment wird durchgeführt, wenn nach ähnlichen/gleichen Sequenzmotiven oder beispielsweise ähnlichen Domänen bei Proteinen gesucht wird. Beim lokalen Alignment wird klassischerweise der Smith-Waterman-Algorithmus verwendet.

### 1. Globales Alignment mit voreingestellten Parametern (Seite 3):

Es wurde eine BLOSUM62-Matrix verwendet. Die BLOcks Substitution Matrix basiert auf einem Vergleich von ganzen Blöcken der Sequenzen. Dabei werden diese Blöcke ohne Lücken verglichen. BLOSUM mit hohen Zahlen werden für evolutionär nahe verwandte Proteine verwendet, BLOSUM mit geringen Zahlen eher für stark divergierende Proteine.

Für die Berechnung der BLOSUM X-Matrixen wurden verwandte Proteinsequenzen der BLOCKS-Datenbank verglichen, die zu X % identisch waren, wodurch dann die relativen Mutationsraten der Aminosäuren bestimmt werden konnten.

Als Gap Penalty wurde 10 gewählt.

Es wurde ein Score von 292,5 errechnet mit einer Identität von 43.6 % und einer Ähnlichkeit von 60.4%. Das ist nicht verwunderlich, da es sich nicht um homologe Proteine verschiedener Species handelt, sondern um verschiedene Untereinheiten eines größeren Komplexes.

## **2. Globales Alignment mit der PAM200-Matrix (Seite 4):**

Es wurde eine PAM200-Matrix verwendet. Diese Matrix ist eine mit PAM-Werten gefüllte Substitutionsmatrix. Ein PAM ist dabei ein Maß für den Verwandtschaftsgrad von verschiedenen Species in phylogenetischen Stammbäumen, wobei die relativen evolutionären Entfernungen zwischen benachbarten Verzweigungspunkten in der Anzahl der Aminosäureunterschiede pro 100 Aminosäuren des Proteins ausgedrückt werden. [1]

Als Gap Penalty wurde 10 gewählt.

Es wurde ein Score von 409.5 errechnet mit einer Identität von 43.6 % und einer Ähnlichkeit von 72.5%.

[1] <https://www.spektrum.de/lexikon/biochemie/pam/4542>

## **3. Globales Alignment mit einer GAP OPEN penalty von 25 (Seite 5):**

Es wurde eine BLOSUM62-Matrix mit einer GAP OPEN penalty von 25 verwendet. Es wurde ein Score von 260.0 errechnet mit einer Identität von 40.9 % und einer Ähnlichkeit von 58.4%. Die Anzahl der Gaps blieb mit 9/149 allerdings identisch zum 1. Alignment mit einer Strafe von 10.

## **4. Lokales Alignment mit voreingestellten Parametern (Seite 6):**

Es wurde eine BLOSUM62-Matrix mit einer GAP OPEN penalty von 25 verwendet. Es wurde ein Score von 293.5 errechnet mit einer Identität von 43.4 % und einer Ähnlichkeit von 60.7%.

Im Vergleich mit dem Alignment der Vorlesung fällt auf, dass es deutlich mehr Unterschiede/Aminosäurenabweichungen gibt. Das ist allerdings nicht verwunderlich, da es sich um zwei verschiedene Untereinheiten eines Komplexes handelt und nicht um homologe/ähnliche Proteine aus verschiedenen Organismen.

1/1

1/1

```
#####
# Program: needle
# Rundate: Tue 10 Jul 2018 14:31:35
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180710-143133-0627-50097100-p2m.asequence
#   -bsequence emboss_needle-I20180710-143133-0627-50097100-p2m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 25.0
#   -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: 25.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      61/149 (40.9%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 260.0
#
#
#=====

EMBOSS_001      1 -MVLSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHF--      47
                  :.:|.:|.:|.:|.|||  :..|.|.|||.|:..:|.|:..|..|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     48 ----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
                  |...|.:|.|.|||||.|.:.:.:|:|:|:.....:|:|..|..|
EMBOSS_001     49 LSTPDAVMGNPVKVKAHGKKVLGAFSDGLAHLDNLKGTFTATLSELHCDKLH      98

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

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

```
#####
# Program: water
# Rundate: Tue 10 Jul 2018 14:33:13
# Commandline: water
#   -auto
#   -stdout
#   -asequence emboss_water-I20180710-143310-0867-78433583-p2m.asequence
#   -bsequence emboss_water-I20180710-143310-0867-78433583-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: 145
# Identity:      63/145 (43.4%)
# Similarity:    88/145 (60.7%)
# Gaps:          8/145 ( 5.5%)
# Score: 293.5
#
#
#=====

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

EMBOSS_001     51  ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP      96
      .|:..|||.|||||..|:..:|:|:|:.....:|:|..|||.|||
EMBOSS_001     52  PDAVMGNPKVKAHGKKVLGAFSDGLAHLNKGTFATLSELHCDKLHVDP      101

EMBOSS_001     97  VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY      141
      .||:|.|.:|.|.|.|||.|||.|.|:..|:|.|:..|..||
EMBOSS_001    102  ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146

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