

Aufgabe 2

Human Gene Sequenz von "Human Hemoglobin subunit alpha"(HBA)

```
MVLSPADKTN VKAAWGKVGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS  
HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK  
LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR
```

Human Gene Sequenz von "Human Hemoglobin subunit beta"(HBB)

```
MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS  
TPDAVMGNPK VKAHGKKVLG AFSDGLAHLN NLKGTFTLS ELHCDKLHVD  
PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
```

Aufgabe 3

Unterschied zwischen einem globalen Sequenz-Alignment und einem lokalen Sequenz-Alignment

Global Sequenz Alignment

In einem globalen Sequenz-Alignment wird die komplette Sequenz total von Anfang bis zum Ende auf die Gemeinsamkeiten mit der anderen Sequenz analysiert. Die Länge der vergleichbaren Sequenzen sollte idealerweise gleich groß für ein globales Alignment sein. Normalerweise werden die homologen Gene mit der gleichen Funktion (z.B. von einem Maus und dem Menschen) oder zwei Proteine mit der ähnlichen Funktion miteinander verglichen. Das globale Sequenz-Alignment wird auch als Needleman-Wunsch Algorithmus bezeichnet.

Local Sequenz Alignment

In einem lokalen Sequenz-Alignment bindet ein Query Motiv an seinen Target. Hier kann man die höchste Wahrscheinlichkeit an einer Insertion oder Deletion finden, weil die Sequenz an einem bestimmten Ort durchsucht wird. Die anderen Regionen vom Target können dann vernachlässigt werden, da Target größer ist als sein Query.

Aufgabe 4

a.

```
# 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-LSPADKTNVKAANGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
|| |:|:|:|.|.|.|||||. :. |:|.|||||.|:|:|.|.|.|.|.|. |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
|| :|:|:|:|.|.|.|||||. :. |:|.|||||.|:|:|.|.|.|.|.|. |
EMBOSS_001     49 LSTPDAMGNPKVKKAHGKKVLGAFSDGLAHLNLIKGTFFATLSELHCDKLH      98

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

Abb.1. Globales Alignment mit voreingestellten Parametern

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM30
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    102/149 (68.5%)
# Gaps:          9/149 ( 6.0%)
# Score: 401.5
#
#=====
EMBOSS_001      1 MV-LSPADKTNVKAANGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
|| |:|:|:|.|.|.|.|||||. :. |:|.|||||.|:|:|.|.|.|.|.|. |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKVN--VDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
|| :|:|:|:|.|.|.|||||. :. |:|.|||||.|:|:|.|.|.|.|.|. |
EMBOSS_001     49 LSTPDAMGNPKVKKAHGKKVLGAFSDGLAHLNLIKGTFFATLSELHCDKLH      98

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

Abb.2. Globales Alignment mit einer anderen Substitution MATRIX

```

# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 5.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    90/149 (60.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 312.5
#
#
#=====
EMBOSS_001      1 MV-LSPADKTNVKAANGKVGGAHAGEYGAELERMFLSFPTTKTYFPHF-D      48
  || |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
  ||      .|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001     49 LSTPDAMGNPKVKAHGGKVLGAFSDGLAHLNLLKGTATLSELHCDKLH      98

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

```

Abb.3. Globales Alignment mit einer anderen GAP OPEN penalty

```

# 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 LSPADKTNVKAANGKVGGAHAGEYGAELERMFLSFPTTKTYFPHF-DLS-      50
  |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST      51

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

EMBOSS_001     97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKY      141
  .|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001    102 ENFRLLGNVLVLCVLAHFGKEFTPPVQAAYQKVVAGVANALAHKY      146

```

Abb.4. Lokales Alignment mit voreingestellten Parametern

a.)

```
HBA_HUMAN  -----VLSPADKTNVKAANGKVGA--HAGEYGAEALERMFSLFPTTKTYFPHF
HBB_HUMAN  -----VHLTPEEKSAVTALWGKV---NVDEVGGEALGRLLVVYPWTQRFFESF
HBA_HUMAN  -DLS-----HGSAQVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKL-
HBB_HUMAN  GDLSTPDAMGNPKVKAHGKKVLGAFSDGLAHL---D--NLKGTFFATLSELHCDKL-
HBA_HUMAN  -RVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR-----
HBB_HUMAN  -HVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH-----
```

Abb.5. Allignment von Human Hemoglobin subunit alpha und beta aus Vorlesung

Im Vergleich zu den Sequenz von Vorlesung und aus dem Such Tool sehen wir den Unterschied von Anfang, wo es in der zehnte Stelle der Beta Sequenz eine Insertion gibt und deshalb sind allen Basen eine Stelle vor verschoben. Die Anzahl der Deletion ist sogar mehr in die Sequenz aus Vorlesung als der gedruckte.

b.)

Substitutionsmatrix

Gibt die Rate und die Wahrscheinlichkeit an, mit welcher eine Aminosäure in eine andere mutieren/umwandeln kann. Es wird dazu verwendet, um einem bestimmten Alignment einen Score (Wahrscheinlichkeit an die Matches) zuzuordnen und zu bestimmen, wie gut das Alignment ist. Mit Blosum62 wird am häufigsten dazu benutzt, weil sein Score am höchsten ist.

<https://de.wikipedia.org/wiki/Substitutionsmatrix>

Gap Penalty

Mit Gap Penalty Tool führt man die Deletionen ein. Damit wird versucht, die Anzahl an besten Matches zwischen den Basen/Aminosäuren zu erhöhen. Durch die Deletion ändern die verbleibenden Basen/Aminosäuren ihre Position, wodurch neue Matches entstehen und der Score wird dadurch verbessert.