

## Bioinformatik Übung 6

### Aufgabe 2:

HBA\_Human Sequenz:

```
MVLSPADKTNVKAAWGKVGHAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQ  
VKGHGKKVADALTNAVAHVDDMPNALSALSSDLHAHKLRVDPVNFKLLSHCLLVTLA  
AHLPAEFTPAVHASLDKFLASVSTVLTSKYR
```

HBB\_Human Sequenz:

```
MVHLTPEEKSAVTALWGKVNVDDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAV  
MGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVL  
VCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH
```

### Aufgabe 3:

Bei einem Alignment handelt es sich um den Vergleich zweier Strings. Hierbei wird die Reihenfolge beider Sequenzen beibehalten und auf eine möglichst hohe Homologie geachtet, indem entweder Fehlpaarungen beibehalten oder einem Indel (Insertion/Deletion) zugeordnet werden.

Bei einem globalen Alignment werden die gesamten Strings einander zugeordnet. Dies ist sinnvoll, wenn eine große Sequenzhomologie erwartet wird und die Sequenzen ähnliche Länge aufweisen. Die Berechnung erfolgt auf Grundlage des dynamischen Programmierens.

Bei einem lokalen Alignment werden Substrings, also Fragmente zweier Strings verglichen. Lokale Alignments werden verwendet, um Sequenzhomologien in verschiedenen Genen zu finden, zum Beispiel ähnliche Domänen in Proteinen. Die Berechnung erfolgt auf Grundlage des Smith-Waterman-Algorithmus

### Aufgabe 4:

```

# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20180710-124042-0965-92316069-p2m.asequence
# -bsequence emboss_needle-I20180710-124042-0965-92316069-p2m.bsequence
# -datafile EBLOSUM65
# -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: EBLOSUM65
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    91/149 (61.1%)
# Gaps:          9/149 ( 6.0%)
# Score: 293.5
#
#=====

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

EMBOSS_001     49  LS-----HGSAQVKGGHKKVADALTNVAHVDDMPNALSALSDLHAHKLR    93
   || .|:|:|:|.||| |..|:|:|:|:|:|:|:|:|:|:|:|:|..| |
EMBOSS_001     49  LSTPDAMVGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH    98

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

#-----
#-----

```

Alignment der beiden Untereinheiten mit voreingestellten Parametern. Als Matrix wurde BLOSUM65 benutzt. Diese eignet sich zum Vergleich von Proteinen, die weder starke noch schwache Homologien aufweisen, man kann sie als universelle Matrix auffassen. Sie wurde erstellt, indem Proteinsequenzen mit weniger als 65% Abweichung verglichen wurden.

```

# Program: needle
# Rundate: Tue 10 Jul 2018 12:47:24
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20180710-124721-0222-27205166-p1m.asequence
# -bsequence emboss_needle-I20180710-124721-0222-27205166-p1m.bsequence
# -datafile EBLOSUM90
# -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: EMBOS_001
# 2: EMBOS_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
#
#=====

EMBOS_001      1  MV-LSPADKTNVKAAMGKVGAGHAGEYGAELRMFLSFPTTKTYFPHF-D    48
   || |:|.:|.|.|.||| | ...|.|.|||.|:..:|. |:|.|. |
EMBOS_001      1  NVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD    48

EMBOS_001     49  LSH-----GSAQVKGHGKKVADALTNAAHVDDHMPNALSALDLHAHKLR    93
   ||.   |.:|.|.|||||.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOS_001     49  LSTPDVVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTLSELHCDKLLH    98

EMBOS_001     94  VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR   142
   |||.||:|...|...|.|.|.|||||. |:|.|.|. |:|.|.|.
EMBOS_001     99  VDPENFRLLGNVLCVLAHMFGEFTPPVQAYQKVVAGVANALAHKYH     147

#-----

```

Alignment unter Benutzung der BLOSUM90 Matrix. Diese Matrix wurde erstellt, indem Sequenzen mit weniger als 90% Homologie verglichen wurden. BLOSUM90 ist sinnvoll, um Sequenzen mit hoher Homologie zu vergleichen.

```

#####
# Program: needle
# Rundate: Tue 10 Jul 2018 12:51:09
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20180710-125106-0733-76339261-p1m.asequence
# -bsequence emboss_needle-I20180710-125106-0733-76339261-p1m.bsequence
# -datafile EBLOSUM90
# -gapopen 100.0
# -gapextend 0.5
# -endopen 10.0
# -endextend 0.5
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EBLOSUM90
# Gap_penalty: 100.0
# Extend_penalty: 0.5
#
# Length: 148
# Identity:      52/148 (35.1%)
# Similarity:    67/148 (45.3%)
# Gaps:          7/148 ( 4.7%)
# Score: 134.5
#
#
#=====

EMBOS_001      1  MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF--- 47
                  .....|.|.|.|.:...|.:.|.|.
EMBOS_001      1  -MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDL 49

EMBOS_001     48  ---DLSHGSAQVKVGHGKKVADALTNVAHVDDNPALSALSOLHAHKLRV 94
                  |...|.:.|.|.|.|.|.|.:.|.|.:.|.|.|.|.|.|.
EMBOS_001     50  STPDVHNGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTLSELCOKLHV 99

EMBOS_001     95  DPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR 142
                  |||.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOS_001    100  DPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH 147

```

Alignment mit BLOSUM90 und einer Gap Penalty von 100. Eine Gap Penalty von 100 bedeutet, dass pro Indel 100 Punkte vom Score abgezogen werden, wodurch die Anzahl der Gaps reduziert wird.

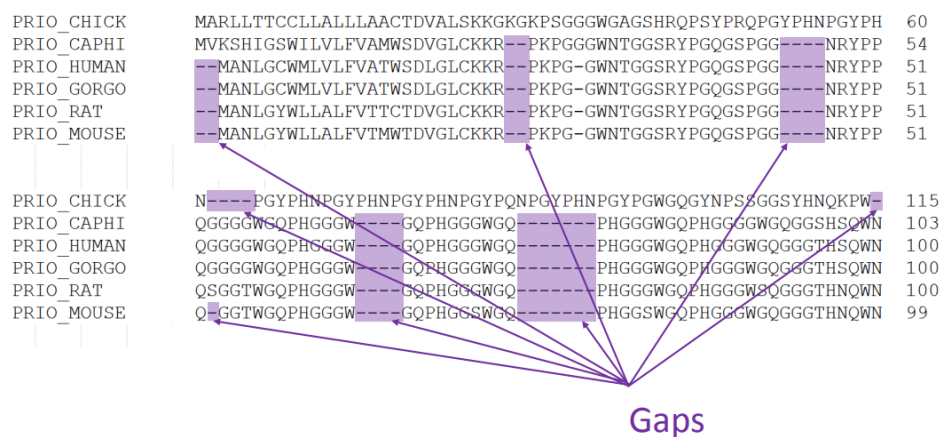
```
#####
# Program: water
# Rundate: Tue 10 Jul 2018 12:56:08
# Commandline: water
#
# -auto
# -stdout
# -asequence emboss_water-I20180710-125605-0661-3785908-p1m.asequence
# -bsequence emboss_water-I20180710-125605-0661-3785908-p1m.bsequence
# -datafile EBLOSUM62
# -gapopen 10.0
# -gapextend 0.5
# -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 LSPADKTNVKAANGKVGAGAGEYGAELERMFLSFPTTKTYFPHP-DLS-      50
EMBOSS_001      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST      51
EMBOSS_001     51 ----HGSAQVKKGKGVADALTNVAHVDDMPNALSALSDLHAHKLRVDP      96
EMBOSS_001     52 PDAVMGNPKVKAHGKKVLAFAFSDGLAHLDNLKGTATLSELHCDKLHVDP     101
EMBOSS_001     97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY      141
EMBOSS_001    102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146

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

Lokales Alignment der Sequenzen mit Hilfe von BLOSUM62. Blosum 62 wurde mit Hilfe von Sequenzen erstellt, die weniger als 62% Homologie aufwiesen.



Vergleicht man das Alignemnt mit den in der Vorlesung vorgestellten Alignments, so fällt auf, dass in der Vorlesung durch Gaps ein hoher Score erreicht wurde.<sup>[1]</sup>

<sup>[1]</sup>: Mit freundlicher Genehmigung von Phillip Nawrath, Berlin, 10.07.2018, 14:12 Uhr.