

Übungszettel 6, Ponimaskine

2) Proteinsequenz der „Human Hemoglobin subunit alpha“:

```
MVLSPADKTNVKAAGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  
KKVADALTNVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTF  
AVHASLSDKFLASVSTVLTSKYR
```

Proteinsequenz der „Human Hemoglobin subunit beta“:

```
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPK  
VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG  
KEFTTPPVQAAYQKVAGVANALAHKYH
```

3) Globales Alignment: vergleicht beide Sequenzen vollständig miteinander („end-to-end alignment“), dieses Verfahren eignet sich also für den Vergleich von nahe verwandten Spezies/Proteinen/..., ähnlicher Länge.

Lokales Alignment: es werden einzelne Sequenzbereiche miteinander verglichen, welche einen optimalen score aufzeigen, somit können auch Übereinstimmungen an unterschiedlichen Positionen innerhalb der Sequenz gefunden werden. Dieses Alignment eignet sich für die Auffindung bestimmter Domänen und konservierten Funktionseinheiten.

4)

(1) Globales Alignment mit voreingestellten Parametern:

EBLOSUM62-Matrix: Blocks Substitution Matrix, hohe Nummer (62) also in diesem Fall Verwendung für eher nah verwandte Proteine. Blöcke werden ohne Lücken miteinander verglichen und eine score Matrix erstellt, wobei gleiche bzw. ähnliche AS einen höheren score aufweisen als stark unterschiedliche AS.

Gap penalty: negativer score der beim Einfügen einer Lücke hinzugerechnet wird, ist dieser Wert stark negativ sinkt die Wahrscheinlichkeit/Häufigkeit für eine eingefügte Lücke und es werden z.B. Mismatches beim Alignment bevorzugt (falls diese einen besseren score aufweisen).

```
# 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 48 | 1 MV-LSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-D : . : . . : : . : . : . : . : . . |
| EMBOSS_001 48 | 1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD |
| EMBOSS_001 93 | 49 LS-----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDDLHAHKLR . : . : : . : . : : : : . : . . . |
| EMBOSS_001 98 | 49 LSTPDAMVGNPKVKAHGKKVLGAFSDGLAHLNDNLKGTTFATLSELHCDKLH |
| EMBOSS_001 142 | 94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR . : . : : : . . . : |
| EMBOSS_001 147 | 99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH |

(2) Globales Alignment mit einer anderen Substitution MATRIX (PAM10):

```

Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM10
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 203
# Identity:      61/203 (30.0%)
# Similarity:    61/203 (30.0%)
# Gaps:          117/203 (57.6%)
# Score: 136.0
#=====

EMBOSS_001      1 MV-LSPADKTNVKAAWGKV-----GAHAGEYGAEALERM-----F
34
      |||.|.|.|.|.|||      |||.|.
EMBOSS_001      1 MVHLTPEEKSAVTALWGKVNDEVGG-----EALGRLLVVPWTQRF
42

EMBOSS_001      35 LSFPTTKTYFPHF----DLSHGSAQ-----VKGHGKKV--A--DA
66
      |      |||      ||.||||| | |.
EMBOSS_001      43 -----FESFGDLS-----TPDAVMGNPKVKAHGKKVLGAFSDG
75

EMBOSS_001      67 LTNAVAHVDDMPN-----ALSALSDLHAHKLRVDPVNFKLLSH---CLLV
108
      |      ||.|      |      .|||.|.|.|.|.|.|.|.|.|.
EMBOSS_001      76 L----AHLD---NLKGTFA--TLSELHCDKLHVDPENFRLLGNVLCVL-
115

```

```

EMBOSS_001      109 TLA AHLPA----EFTPAVHASL DKFLASVSTVLTSKYR-----
142
                ||      |||.|.|      |      |.
EMBOSS_001      116 ---AH---HFGKEFTPPVQA-----A-----YQKV VAGVANALAH
144

EMBOSS_001      143 ---      142

EMBOSS_001      145 KYH      147

```

Im Vergleich zu der Vorlesung treten hier viel häufiger Gaps auf und die Übereinstimmung der beiden Sequenzen ist deutlich geringer (nur noch 30%).

(3) Globales Alignment mit einer anderen GAP OPEN penalty (1):

Die gap penalty wurde deutlich runter gesetzt (auf 1), somit werden eingefügte gaps nun wahrscheinlicher, bzw. gegenüber bestimmten mismatches bevorzugt. Man erkennt, dass die Anzahl an Lücken im Vergleich zum ersten Alignment deutlich angestiegen ist (41 Lücken statt vorher nur 9), die generelle Übereinstimmung ist allerdings nur minimal gesunken.

Würde man die gap penalty erhöhen, so würden deutlich weniger gaps auftreten und bestimmte mismatches bevorzugt werden.

```

Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 1.0
# Extend_penalty: 0.5
#
# Length: 165
# Identity:      72/165 (43.6%)
# Similarity:    90/165 (54.5%)
# Gaps:          41/165 (24.8%)
# Score: 347.0
#
#
#=====

EMBOSS_001      1 MV-LSPADKTNVKA AWGKVGAHAGEYGA EALERMFLS--FP-TTKTYFPH
46
                || |:|. |:|.|.|.||| :..|.|.|.|. |: | :| |: :|..
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRL-L-VVYPWTQR-FFES
45

EMBOSS_001      47 F-DLSH-----GSAQVKGHGKKV--A--DALTNAVAHVDDMPNAL--S-
82
                ||||      |:.:|.|.||| | |. | | :| | | :
EMBOSS_001      46 FGDLS-TPDAVMGNPKVKAHGKKVLGAFSDGL----AHL D---N-LKGT F
86

EMBOSS_001      83 A-LSDLHAHKLRVDPVNFKLLSH--CLLVTLA AHLPA-EFTPAVHASLD
127
                | ||:|.|.|.|.|.|. |:|. | | | |. | . |||.|. |:..
EMBOSS_001      87 ATLSELHCDKLHVDPENFRLLGNVLVC--V-LAHHF-GKEFTPPVQAAYQ
132

```

| | | | |
|------------|-----|-----------------|-----|
| EMBOSS_001 | 128 | KFLASVSTVLTSKYR | 142 |
| | | .: . :.. .. . | |
| EMBOSS_001 | 133 | KVVAGVANALAHKYH | 147 |

(4) Lokales Alignment mit voreingestellten Parametern:

Die voreingestellten Parameter des lokalen Alignments ähneln denen des ersten Ansatzes, allerdings gibt es hier keine end gap penalties, da nur Teile der Sequenzen miteinander verglichen werden und nicht ein „end-to-end alignment“.

```

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  PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP
101
EMBOSS_001          97  VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY           141
      .||:|~.:|.|:~.:|||.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOSS_001          102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYOKVVGAVANALAHKY           146

```

Das Ergebnis ähnelt dem des ersten Ansatzes. Im Vergleich zur Vorlesung wurden auch hier weniger Lücken eingeführt, was wahrscheinlich auf einen hier verwendeten negativeren score der gap penalty zurückzuführen ist.