

Aufgabe 2

Human Hemoglobin subunit alpha (HBA_HUMAN):

```
      10      20      30      40      50
MVLSPADKTN VKAAWGKVGGA HAGEYGAEAL ERMFLSFPTT KTYFPHFDLS
      60      70      80      90     100
HGSAQVKGHG KKVADALTNA VAHVDDMPNA LSALSDLHAH KLRVDPVNFK
     110     120     130     140
LLSHCLLVTL AAHLPAEFTP AVHASLDKFL ASVSTVLTSK YR
```

Human Hemoglobin subunit beta (HBB_HUMAN):

```
      10      20      30      40      50
MVHLTPEEKS AVTALWGKVN VDEVGGEALG RLLVVYPWTQ RFFESFGDLS
      60      70      80      90     100
TPDAVMGNPK VKAHGKKVLG AFSDGLAHLN NLKGTTFATLS ELHCDKLHVD
     110     120     130     140
PENFRLLGNV LVCVLAHHFG KEFTPPVQAA YQKVVAGVAN ALAHKYH
```

Aufgabe 3

Globales Alignment bedeutet, dass die Sequenzen der zu vergleichenden Proteine (oder Gene) in der gesamten Länge verglichen werden, während bei lokalem Alignment nur eine oder mehrere Teilsequenzen verglichen werden. Dabei können entweder die ähnlichsten Regionen oder aber voreingestellte Bereiche miteinander aligned werden.

Aufgabe 4

Global Default Einstellungen (EMBOSS Needle):

```
# 1: HBA_HUMAN          # Length: 149
# 2: HBB_HUMAN          # Identity:    65/149 (43.6%)
# Matrix: EBLOSUM62      # Similarity:  90/149 (60.4%)
# Gap_penalty: 10.0      # Gaps:       9/149 ( 6.0%)
# Extend_penalty: 0.5    # Score: 292.5

HBA_HUMAN      1 MV-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                || |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
HBB_HUMAN      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

HBA_HUMAN     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
                ||      .|:~:|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|
HBB_HUMAN     49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLNLDNLKGTTFATLSELHCDKLH      98

HBA_HUMAN     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR      142
                |||.||:|~|:~|:~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|~|
HBB_HUMAN     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
```

Die Beiden Sequenzen können mit der BLOSUM62 Matrix auf 43,6% Identität gebracht werden, wobei nur 9 Gaps eingefügt werden müssen und sich die Sequenzen an 90 von 149 Stellen ähneln (Matches und "schwache Missmatches"). Insgesamt sieht dieses Alignment besser aus als das in der Vorlesung gezeigte aber der Vergleich ist schwierig, da in der Vorlesung noch viele weitere Sequenzen verglichen wurden und die Erkennung von Übereinstimmungen und Unterschieden dadurch erschwert wird. In der BLOSUM62 Matrix sind sog. log-odds von Aminosäuren aufgetragen. Diese Werte entsprechen grob gesagt dem Logarithmus der Wahrscheinlichkeit einer bestimmten Aminosäureumwandlung (oder auch nicht-Umwandlung). Gewonnen wurden diese Werte aus Sequenzblöcken von homologen Proteinen die, im Falle der BLOSUM62, mindestens 62% Identität aufweisen.

Global EPAM10 Matrix (EMBOSS Needle):

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

HBA_HUMAN      1 MV-LSPADKTNVKAAWGKV-----GAHAGEYGAEALERM-----F      34
                  |||.|.|.|.|.|.||| | |||.|. |
HBB_HUMAN      1 MVHLTPEEKSAVTALWGKVNVDVGG-----EALGRLLVVYPWTQRF      42

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

HBA_HUMAN     67 LTNAVAHVDDMPN-----ALSALSDLHAHKLRVDPVNFKLLSH---CLLV      108
                  | |||. | | .|||.|.|.|.|.|.|.|. |. |
HBB_HUMAN     76 L----AHL D---NLKGTF A--TLSELHCDKLHVDPENFRLLGNVLVCVL-      115

HBA_HUMAN    109 TLA AHLPA----EFTPAVHASLDKFLASVSTVLT SKYR-----      142
                  || |||.|. | | |.
HBB_HUMAN    116 ---AH---HFGKEFTPPVQA-----A-----YQKV VAGVANALAH      144

HBA_HUMAN    143 ---      142
HBB_HUMAN    145 KYH      147
```

Ich konnte leider nicht rausfinden wie die EPAM10 Matrix aussieht aber da die Anzahl der Gaps in diesem Alignment viel höher ist obwohl die OPEN GAP penalty gleichbleibend 10 ist nehme ich an, dass die "Strafe" für ein Missmatch in dieser Matrix deutlich höher ist. Außerdem gibt es keinen Unterschied mehr zwischen Identity und Similarity, was natürlich Zufall sein könnte aber ich vermute eher, dass in dieser Matrix alle Missmatches den gleichen Wert haben und daher kein Unterschied gemacht werden kann zwischen "starken" und "schwachen" Missmatches.

Insgesamt ist hier das Alignment schlechter als vorher mit der BLOSUM62 Matrix, da eine Sequenz hier häufig für größere Bereiche nur aus Gaps besteht, sodass das Alignment letztendlich knapp ein Drittel länger ist als die ursprünglichen Sequenzen. Die tatsächlich überlappenden Bereiche sind dafür sehr gut aligned.

Global GAP OPEN penalty = 1 (EMBOSS Needle):

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

HBA_HUMAN      1 MV-LSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLS--FP-TTKTYFPH      46
  || |:|.:|:|.|.||||| :..|.|.||||.:| | :| |:| :|..
HBB_HUMAN      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRL-L-VVYPWTQR-FFES      45

HBA_HUMAN     47 F-DLSH-----GSAQVKGHGKKV--A--DALTNAVAHVDDMPNAL--S-      82
  |||| |.:|.|.||||| | |. | |.:| | | :
HBB_HUMAN     46 FGDLS-TPDAVMGNPKVKAHGKKVLGAFSDGL----AHLN---N-LKGTFF      86

HBA_HUMAN     83 A-LSDLHAHKLRVDPVNFKLLSH---CLLVTLAAHLPA-EFTPAVHASLD      127
  ||:|.|.|.|.|.|.||:|.:| | | |.|. . ||||.|.:|..
HBB_HUMAN     87 ATLSELHCDKLVDPENFRLLGNVLVC--V-LAHHF-GKEFTPPVQAAYQ      132

HBA_HUMAN    128 KFLASVSTVLTSKYR      142
  |.:|.|.:|.|.|.|.
HBB_HUMAN    133 KVVAGVANALAHKYH      147
```

Durch eine Senkung der OPEN GAP penalty konnte ein besserer Score erreicht werden, da durch die geringere Strafe eine höhere Anzahl an Matches gebildet werden konnte. Interessanterweise ist aber der Anteil der Matches auf die Stellen des Alignments hochgerechnet exakt gleich geblieben, was zeigt, dass ein höherer Score nicht zwangsläufig auf ein besseres Alignment hinweist, wenn die verwendeten Parameter geändert wurden.

Lokal Default Einstellungen (EMBOSS Matcher):

```
# 1: HBA_HUMAN          # Length: 145
# 2: HBB_HUMAN          # Identity:      63/145 (43.4%)
# Matrix: EBLOSUM62     # Similarity:   88/145 (60.7%)
# Gap_penalty: 14       # Gaps:         8/145 ( 5.5%)
# Extend_penalty: 4     # Score: 264#

HBA_HUMAN      3 LSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-DLSH      51
  |:|.:|:|.|.|.||||| :..|.|.||||.:|.:|.:|.:|.:| | | |
HBB_HUMAN      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST      51

HBA_HUMAN     52 -----GSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLRVDP      96
  |.:|.|.|||||.|.|.:.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
HBB_HUMAN     52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLNLLKGTFTATLSELHCDKLVDP      101

HBA_HUMAN     97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY      141
  .||:|.|.:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
HBB_HUMAN    102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146
```

Zumindest mit den Voreinstellungen scheint ein lokales Alignment der beiden Sequenzen keinen großen Unterschied zu einem globalen Alignment zu haben. Das ist aber durchaus verständlich, da die beiden Sequenzen recht kurz sind und sich außerdem über die gesamte Länge hinweg (bis auf die 5 Aminosäuren nach Position 52) recht stark ähneln. Daher wird für das Lokale Alignment nur ein sehr kleines Stück am Anfang (3 AS) und Ende (1 AS) der Sequenzen entfernt.