

Bericht zum globalen und lokalem Aligment von Hemoglobin subunit alpha und beta:

Human Hemoglobin subunit alpha:

MVLSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAAVHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR

Human Hemoglobin subunit beta:

MVHLTPEEKSAVTALWGKVNVDVGGGALGRLLVVYPWTPQRFFESFGDLSTPDVAMGNPK
VKAHGKKVLGAFSDGLAHLNLTGKFTATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH

Aufgabe 3:

Beim globalen Alignment werden alle Symbole beachtet. Daher bietet sich diese Methode an, um Sequenzen ähnlicher Länge und erwarteter starker Sequenzanalogie zu vergleichen. Beim einem lokalen Alignment werden die Teile einer Sequenz mittels eines globalen Alignments gepaart. Dabei müssen die Teile der Sequenz gefunden werden, die einen maximalen Score bieten. Dieser Ansatz wird zum Beispiel bei dem Vergleich von Domänen in Proteinen verwendet.

Aufgabe 4:

Zum Vergleichen das Alignment aus der Vorlesung:

```

Helix          AAAAAAAAAAAAAAAAAA BBBB BBBB BBBB BBBB BBBBBBBBBB
HBA_HUMAN     -----VLSAPDQKLVNVAAGVKA----- HAGEYGEAEALRMLSPFTTFPTTFPHF
HBB_HUMAN     -----VHLTPEKSAITLAWKY----- NVDEVEGQGLGRLLVYVFWTFQTFPSF
MYB_PHYCA     -----VLEEGEOLWLVNVAAGVKA----- DYVHGQGLTLERFESKPTPEKTFD
GLB3_CHITP    -----LSAQDITSEAPNAPADPKV----- DPGVLVYVFWTFQTFPSF
GLB5_PTPT     PIVDTGVSAPLSAEAKTKIRSNAPVYS----- TYVTSGVDLVKVFSTTPAAQEFKTF
LG82_LUPLU    -----GALTGEQAQVLSWEEFNA----- NTPKHTHFRTFLVLELAPAAKOLF
GLB1_GLYDI    -----GLSAQRCALVIAATWKDLAGDNAGGVKDCILFKFLPGLMAQVAPVG-
Consensus     Ls... v a v kv . . . g . L . f . p F F F

Helix          DDDDDDDDEEEEEEEEEEEEEEEEEEEEEEE FFFFFFFFHHHH
HBA_HUMAN     -DLS-----HGAQGVQGHKGKVALDALTNVAHVH----- DMPNALSALSDLNHHK
HBB_HUMAN     -GLDSTPEAMQMNPKVPAKHKGVLCGASDGLNHL----- N-KGKTGATLSLSELCHKL
MYB_PHYCA     -KHLKTEADMKSEDLKHKGVTLVGTALGALKK----- K-GHNEKEFLPLAQSHAKTK
GLB3_CHITP    -AK-KDLESIKGTAFTPEATHNIRVGFSSKIGEL----- P-NIEADVTNIVASHKPRG
GLB5_PTPT     KGLTTADQLKSDAPRWIHAIRINAVNDVASM----- DPTKMSMKLRDLGSKHAKSP
LG82_LUPLU    -LGT-CTSEPPNPGVLGAHAGVKVLAQVIAQLQVTVGVTADTLKNGLSVSHSGK
GLB1_GLYDI    -SG-----AS-----DPGVAALGAKVLKVLVEGVAVSHL----- GDEGKVAQMKVAVGRHKHSG
Consensus     . t . . . v . Hg kv . a . a l l d . a l l H .

Helix          FFGGGGGGGGGGGGGGGGGGGG KHHHHHHHHHHHHHHHHHHHHHH
HBA_HUMAN     -RVDVPFLLKLLSHCLVTLAAHLPAEITFAVSLDKA----- VVHNTVNTSKYR-----
HBB_HUMAN     -HYVDENFRLLGNLVCVLIHHNHPGCTFPVQAQGLKAVGVAALAKL-----
MYB_PHYCA     -KIPIKYLEIFSEATYVHLHSHRPDGGADAGAMNKALELFPKDLIAKVLKELGYQG
GLB3_CHITP    -VTHDQNLINFRAGFVSVMYHAK----- D-FA-CAEAAQGLTDPFPMI FSKM
GLB5_PTPT     -QYDVPQYFVKVLAIVADTVAG----- -DAGFEKLMSKICILLRAS
LG82_LUPLU    -VADAHFVPKVEALIKTKTEVGVGAKVSEELNSATYDELAIYIKENMDAA-----
GLB1_GLYDI    KHIAQYFVPELGASLLSAMEHRTGGKNNAAKDAWAAAYDISGALISQGS-----
Consensus     v f l . . . . . . . . . . g . aa . k . l sky

```

1) globales Allignment mit voreingestellten Parametern:

Parameter: Program – Needle
 Matrix – EBLOSUM62
 Gap open – 10
 Gap extend – 0.5
 End Gap Open Penalty – 10
 End Gap Extension Penalty – 0.5

Bei der EBLOSUM62 Matrix handelt es sich um die BLOcks SUBstitution Matrix. Dabei sind die höheren Nummern für das Alignment von evolutionär nahen Proteinen geeignet,

während die kleinen Nummern für weit entfernte Proteine verwendet werden. Dabei werden einzelne Blöcke ohne Lücke auf ihre Homologie verglichen.
Gap open und Gap extend sind hierbei auf einen Art Mittelwert voreingestellt.

```
#=====
#
# 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-LSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKYFPHF-D  48
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001      1  MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD  48
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     49  LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSADLAHAKLR  93
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     49  LSTPDVAVGNPKVKAGKGVLAGFSDGLAHLDNLKGTFAFLSELHCDKLH  98
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     94  VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLSKYR  142
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     99  VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH  147
```

Der Vergleich zu dem Alignment aus der Vorlesung zeigt, dass in diesem globalen Alignment weniger Gaps eingefügt wurden (9). Dadurch kommt es öfter zu einem Mismatch. Bei den Mismatches handelt es sich oftmals um ähnliche Aminosäuren.

2) globales Alignment mit der Substitutions Matrix EPAM10

Parameter: Program – Needle
 Matrix – EPAM10
 Gap open – 10
 Gap extend – 0.5
 End Gap Open Penalty – 10
 End Gap Extension Penalty – 0.5

PAM steht für Point accepted Matrix. Darunter wird die Mutation einer Aminosäure in der Sequenz verstanden, welche durch die natürliche Selektion akzeptiert ist. Die restlichen Parameter bleiben im Vergleich zum ersten Alignment erhalten.

```
#
# 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-LSPADKTNVKAANGKV-----GAHAGEYGAEALERM-----F  34
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001      1  MVHLTPEEKSAVTALWGKVNDEVGG-----EALGRLLVVYPWTQRFF  42
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     35  LSFPTTKYFPHF---DLSHGSAQ-----VKGHGKKV--A--DA  66
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     43  -----FESFGDLS-----TPDAVGNPKVKAGKKVLGAFSDG  75
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     67  LTNAVAHVDDMPN-----ALSALSDLHAHLRVDPUNFKLLSH---CLLV  108
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001     76  L---AHLD---NLKGTFA--TLSELHCDKLHVDPENFRLLGNVLVCVL-  115
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001    109  TLAHLPA---EFTPAVHASLDKFLASVSTVLSKYR-----  142
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001    116  ---AH---HFGKEFTPPVQA-----A-----YQKVVAGVANALAH  144
                  || |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|
EMBOSS_001    143  --- 142
EMBOSS_001    145  KYH 147

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

Im Vergleich zu dem Alignment in der Vorlesung hat dieses sehr viel mehr Gaps. Dadurch nimmt der Score ab und die prozentuale Identität bzw. Ähnlichkeit nimmt ab. Es werden weniger Mismatches akzeptiert.

3) Globales ALignment mit Gap Open Penalty von 50

Paraeter: Program – Needle
 Matrix – EBLOSUM62
 Gap open – 50
 Gap extend – 0.5
 End Gap Open Penalty – 10
 End Gap Extension Penalty – 0.5

Die Parameter stimmen mit dem ersten Alignment bis auf die Gap Open penalty überein. Das führt dazu, dass Mismatches gegenüber Gaps bevorzugt sind, da sie, unter diesen Bedingungen, einen besseren Score versprechen. In diesem Alignment gibt es nur 9 Gaps. Jedoch zeigte auch das erste Allignemtn bereits diese Anzahl. Würde die Gap OPene Penalty bei 1 sein, könnte man mit sehr viel mehr Insertion/Deletion rechnen.

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 50.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      61/149 (40.9%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 210.0
#
#=====
EMBOSS_001      1 -MVLSPADKTNVKAAGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-- 47
                  :.:|.:|.:|.:|.:|  :.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD 48
                  |.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001      48 ----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKL 93
                  |.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001      49 LSTPDVAVMGKPKVKAHGKKVLGAFSDGLAHLDNLKGTFTLSELHCDKLH 98
                  |.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001      94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR 142
                  |||.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001      99 VDPENFRLLGNVLCVLAHFGKEFTPPVQAAYQKVVAGVANALAHKYH 147
                  |.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
#=====
#
```

Der Vergleich zu dem Alignment in der Vorlesung stellt sich ähnlich zum ersten Alignment dar. Es gibt in diesem Alignment weniger Gaps und dafür mehr Mismatches. Jedoch ist der Score kleiner als beim ersten Alignment, da die Stellen für die Gaps anders gewählt wurden.

4) lokales Alignment mit voreingestellten Parametern

Paraeter: Program – water
 Matrix – EBLOSUM62
 Gap open – 10
 Gap extend – 0.5

Die Parameter für das locale Alignment gleichen denen im ersten Ansatz. Allerdings gibt es beim lokalen Alignment keine End gap penalties, da es sich um ein Teilalignment der Sequenz handelt.

