Hemoglobin subunit alpha/beta

>sp|P69905|HBA_HUMAN Hemoglobin subunit alpha OS=Homo sapiens OX=9606 GN=HBA1 PE=1 SV=2

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP AVHASLDKFLASVSTVLTSKYR

>sp|P68871|HBB_HUMAN Hemoglobin subunit beta OS=Homo sapiens OX=9606 GN=HBB PE=1 SV=2

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNPK VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG KEFTPPVQAAYQKVVAGVANALAHKYH

Lokales vs. Globales Alignment

Ein lokales Alignment vergleicht die Abfrage mit einem Subset der Sequenz, wohingegen bei einem globalen Alignment über die gesamte Länge verglichen wird. Dies führt dazu, dass es in einem globalen Alignment häufiger zu Gaps kommt, wenn Suchsequenz und Vergleichssequenz in der Länge variieren. Folglich ist es häufig sinnvoller ein globales Alignment zum Vergleich homologer Gene und ein lokales Alignment zum Vergleich homologer Domänen heranzuziehen.

(1) Globales Alignment mit voreingestellten Parametern Needle EMBOSS - Blosum62 - Gap Penalty: 10

```
1 MV-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D
EMBOSS 001
                                                                     48
                   || |:|.:|:.|.|.||| :..|.|.||.|:.::|.|:.:|.|
                1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD
EMBOSS 001
                                                                     48
                49 LS----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR
EMBOSS 001
                                                                     93
                         .|:.:||.||||..|.::.:||:|:::....:.||:||...||.
                49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH
EMBOSS 001
                                                                     98
EMBOSS 001
                94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
                                                                   142
                   EMBOSS 001
                99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH
                                                                   147
# Length: 149
```

```
# Identity: 65/149 (43.6%)
# Similarity: 90/149 (60.4%)
# Gaps: 9/149 (6.0%)
# Score: 292.5
```

(2) Globales Alignment mit einer anderen Substitution MATRIX Needle EMBOSS - PAM 50 - Gap Penalty: 10

EMBOSS_001	1 MV-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D	48
EMBOSS_001	1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOSS_001	49 LSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR	93
EMBOSS_001	49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH	98
EMBOSS_001	94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR	142
EMBOSS_001	99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH	147
"		
# Length: 149		
# Identity:	65/149 (43.6%)	
<pre># Similarity:</pre>	81/149 (54.4%)	
# Gaps:	9/149 (6.0%)	
# Score: 290.5		

(3) Globales Alignment mit einer anderen GAPOPEN penalty Needle EMBOSS - Blosum 62 - Gap Penalty: 5

EMBOSS_001	1 MV-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D	48
EMBOSS_001	1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOSS_001	49 LSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR	93
EMBOSS_001	. :.: . .:.: : :::. : . 49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH	98
EMBOSS_001	94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR	142
EMBOSS_001	. : .:. : 	147
<pre># Length: 149 # Identity: # Similarity: # Gaps: # Score: 312.5</pre>	65/149 (43.6%) 90/149 (60.4%) 9/149 (6.0%)	

(4) Lokales Alignment mit voreingestellten Parametern Water EMBOSS - Blosum62 - Gap Penalty: 10

```
EMBOSS 001
                3 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-
                                                               50
                  |:|.:|:.|.|.||| :..|.||||:::::|.|:::|.||
               4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST
EMBOSS 001
                                                               51
EMBOSS 001
               51 ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP
                                                               96
                     52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP
EMBOSS 001
                                                               101
EMBOSS 001
               97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY
                                                          141
                  EMBOSS 001
              102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY
                                                          146
# Length: 145
# Identity:
                 63/145 (43.4%)
# Similarity:
                 88/145 (60.7%)
# Gaps:
                 8/145 ( 5.5%)
# Score: 293.5
```

Die Gap Penalty gibt an, wie stark eingefügt Gaps bestraft werden. Ein geringer Wert hat zur Folge, das mehr Gaps vorkommen um somit einen höheren Score zu erreichen.

Blosum 62 Matrix:

Für die BLOSUM 62 Matrix werden alle Sequenzen mit mehr als 62% Sequenzidentität zusammengefügt, alle verbleibenden verglichenen Sequenzen besitzen untereinander weniger als 62% Identität. In die Matrix eingetragen werden die logarithmischen Wahrscheinlichkeitswerte.

PAM 50 Matrix:

Die Einträge der PAM Matrix sind logarithmischen Wahrscheinlichkeitswerte multipliziert mit einem Faktor.

Vergleicht man die Matrizen mit der in der Vorlesung vorgestellten, so fällt auf, dass mithilfe von eingefügten Gaps eine sehr hohe Ähnlichkeit (künstlich) erzeugt wurde.

Quellen: