

Human Hemoglobin subunit alpha HBA_HUMAN

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

Human Hemoglobin subunit beta HBB_HUMAN

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

Globales Alignment

- Kompletter Vergleich der zu untersuchenden Sequenzen
- Hauptsächlich bei starken Sequenzhomologien und ähnlicher Länge verwendet

Lokales Alignment

- Ein oder mehrere Alignments, die eine oder mehrere Regionen beschreiben die sich am ähnlichsten sind in den beiden zu vergleichenden Sequenzen
- Alignment einer Teilsequenz (substring)
- Anwendungsbeispiel: wenn die gegebene Sequenz eine Teilsequenz der Referenz sein könnte oder nur nach ähnlichen Motiven in der Referenz gesucht wird

Global Alignment default

```
#####
# Program: needle
# Rundate: Sun  8 Jul 2018 16:37:48
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180708-163746-0077-52237448-plm.asequence
#   -bsequence emboss_needle-I20180708-163746-0077-52237448-plm.bsequence
#   -datafile EBLOSUM62
#   -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: HBA_HUMAN
# 2: HBB_HUMAN
# 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
#
#
#=====
```

HBA_HUMAN	1	MV-LSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-
48		: .: : . . . :..: : :
HBB_HUMAN	1	MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD
48		
HBA_HUMAN	49	LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDDLHAHKLR
93		. : : .
HBB_HUMAN	49	LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLDDLKGTFTATLSELHCDKLH
98		
HBA_HUMAN	94	VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
142		. : : .
HBB_HUMAN	99	VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH
147		

#-----

#-----

Global Alignment mit anderer Substitution MATRIX

```
#####  
# Program: needle  
# Rundate: Sun 8 Jul 2018 16:38:48  
# Commandline: needle  
# -auto  
# -stdout  
# -asequence emboss_needle-I20180708-163840-0721-65808114-plm.asequence  
# -bsequence emboss_needle-I20180708-163840-0721-65808114-plm.bsequence  
# -datafile EPAM10  
# -gapopen 10.0  
# -gapextend 0.5  
# -endopen 10.0  
# -endextend 0.5  
# -aformat3 pair  
# -sprotein1  
# -sprotein2  
# Align_format: pair  
# Report_file: stdout  
#####  
  
#=====
```

```

# 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
#
#
#=====

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----AHLN---NLKGTFA--TLSELHCDKLHVDPENFRLLGNVLVCVL-
115

HBA_HUMAN      109 TLAHLPA---EFTPAVHASLDKFLASVSTVLTSKYR-----
142
      ||      |||.|.      |      |.
HBB_HUMAN      116 ---AH---HFGKEFTPPVQA-----A-----YQKVVGAVANALAH
144

HBA_HUMAN      143 ---      142
HBB_HUMAN      145 KYH      147

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

```

Global Alignment mit anderer GAP OPEN penatly

```

#####
# Program: needle
# Rundate: Sun  8 Jul 2018 16:39:25
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180708-163924-0142-36419351-p1m.asequence
#   -bsequence emboss_needle-I20180708-163924-0142-36419351-p1m.bsequence
#   -datafile EBLOSUM62
#   -gapopen 20.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2

```



```
#####

#=====
#
# Aligned_sequences: 2
# 1: HBA_HUMAN
# 2: HBB_HUMAN
# 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
#
#=====

HBA_HUMAN      3 LSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-
50
      |:|.:|.:|.|.||||  :..|.|.||||.|:~::~|.|:~::~|..| |||
HBB_HUMAN      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST
51

HBA_HUMAN      51 ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALS~DLHAHKLRVDP
96
      .|:~::~||.|||||.~|.~::~||:|:~::~.....:|~::~||.~::~|||
HBB_HUMAN      52 PDAVMGNPKVKAHGKKVLGAFSDGLAHL~DNLKGT~FATLSELHCDK~LHVDP
101

HBA_HUMAN      97 VNFKLLSHCLLVTLAAHLPAEFTTPAVHASLDKFLASVSTVLT~SKY      141
      .||:|~::~.:|:~::~||.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|
HBB_HUMAN      102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146

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

BLOSUM62 Matrix

- Aminosäuren sind in der Tabelle basierend auf den chemischen Eigenschaften der Seitenketten gruppiert
- Jeder Wert der Matrix wird berechnet, indem die Häufigkeit des Aminosäurepaars in der BLOCKS Datenbank (62% level) durch die Wahrscheinlichkeit, dass zwei Aminosäuren zufällig „aligned“ werden geteilt wird

PAM10 Matrix

- PAM = point accepted mutation
- Jeder Eintrag in der Matrix zeigt die Wahrscheinlichkeit mit der eine Aminosäure einer Reihe durch eine andere Aminosäure der gleichen Spalte durch eine oder mehrere Punktmutationen im Verlauf der Evolution (festgelegter Intervall) ersetzt wurde, statt durch Zufall zusammenzupassen. Dabei beziehen sich verschiedene PAM Matrices auf verschiedene Intervalle der Evolution der gegebenen Protein Sequenzen
- PAM10: Annahme, dass 10 Mutation pro 100 Aminosäuren möglich sind (nur sehr ähnliche Sequenzen werden dann hohe Alignment Scores bekommen)

Gap Open penalty

- Alignment scores können basierend auf der Anzahl und Länge von Lücken angepasst werden
- Eine Lücke einzufügen kann dazu führen, dass der Score verbessert wird, als bei einem lückenlosen Alignment, wobei zu viele Lücken auch zu keinem sinnvollen Ergebnis führen
- Lücken können bei Mutationen wie Insertionen oder Deletionen eingefügt werden
- Gap open penalty: Wert für das Einsetzen einer Lücke
 - High gap open penalty: bei Insertionen/Deletionen werden keine Lücken eingesetzt, bei einer hohen Punktmutationsrate gibt es also einen niedrigen Alignment score
 - Low gap open penalty: es werden mehr Insertionen/Deletionen geduldet, Alignment score kann auch bei höherer Punktmutationsrate gut sein