

2. Besorgen Sie sich die Protein-Sequenzen von „Human Hemoglobin subunit alpha“ (HBA_HUMAN) und „Human Hemoglobin subunit beta“ (HBB_HUMAN).

Hemoglobin subunit alpha (HBA_HUMAN)

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

Hemoglobin subunit beta (HBB_HUMAN)

10	20	30	40	50
MVHLTPEEKS	AVTALWGKVN	VDEVGGEALG	RLLVVYPWTQ	RFFESFGDLS
60	70	80	90	100
TPDAVMGNPK	VKAHGKKVLG	AFSDGLAHL	NLKGTFATLS	ELHCDKLHVD
110	120	130	140	
PENFRLLGNV	LVCVLAHHFG	KEFTTPVQAA	YQKVVAGVAN	ALAHKYH

3. Es gibt grob zwei verschiedene Arten eines Alignments für zwei Sequenzen: „Globales Alignment“ und „Lokales Alignment“. Skizzieren Sie in ihrem Bericht kurz zwei zentralen Unterschiede zwischen diesen beiden Arten.

Alignment: ist ein Vergleich zweier oder mehrerer Strings (Sequenz)

Bei einem **globalen Alignment** zwischen 2 Sequenzen werden **all Symbole** berücksichtigt, bei einem **lokalen Alignment** hingegen nur von **einer Teilsequenz** zwischen beiden Sequenzen. Zur Berechnung des optimalen lokalen Alignment zweier Sequenzen müssen die beiden Teilsequenzen gefunden werden, deren optimaler Alignment-Score maximal ist.

Daher werden globale Alignments meist verwendet, wenn die zu untersuchenden Sequenzen ähnlich lang sind starke Sequenzhomologien erwartet werden. Die Anwendung von einem lokalen Alignment ist dagegen die Suche nach gleichen Sequenzmotiven oder Domänen bei Proteinen.

4. Nun sollen verschiedene Arten von Alignments durchgeführt werden.

(1) Globales Alignment mit voreingestellten Parametern

Score: 277

```
#####
# Program: stretcher
# Rundate: Mon  9 Jul 2018 14:47:00
# Commandline: stretcher
#   -auto
#   -stdout
# -asequence emboss_stretcher-I20180709-144657-0209-79819293-plm.asequence
# -bsequence emboss_stretcher-I20180709-144657-0209-79819293-plm.bsequence
# -datafile EBLOSUM62
# -gapopen 12
# -gapextend 2
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 12
# Extend_penalty: 2
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    90/149 (60.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 277
#
#
#=====

EMBOSS_001      1 MV-LSPADKTNVKAAWGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                  |||::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LSH-----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
                  ||.      |::|::|::|::|::|::|::|::|::|::|::|::|::|::|
EMBOSS_001     49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHLNKLKGTfATLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR      142
                  |||.||::|::|::|::|::|::|::|::|::|::|::|::|::|::|
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAAYQKVAGVANALAHKYH      147

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

Ref: https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=emboss_stretcher-I20180709-144657-0209-79819293-plm

Substitution Matrix: BLOSUM62

- BLOSUM62 (BLOcks Substitution Matrix) wird für Sequenzalignment von Proteinen benutzt.

Gap open: 12

Gap extend: 2

(2) Globales Alignment mit einer anderen Substitution MATRIX

Score: 361

```
#####
# Program: stretcher
# Rundate: Mon  9 Jul 2018 16:22:42
# Commandline: stretcher
#   -auto
#   -stdout
#   -asequence emboss_stretcher-I20180709-162239-0429-42485480-p2m.asequence
#   -bsequence emboss_stretcher-I20180709-162239-0429-42485480-p2m.bsequence
#   -datafile EBLOSUM45
#   -gapopen 12
#   -gapextend 2
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM45
# Gap_penalty: 12
# Extend_penalty: 2
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    94/149 (63.1%)
# Gaps:          9/149 ( 6.0%)
# Score: 361
#
#
#=====

EMBOSS_001      1 MVHLTPEEKSAVTALWGKVNVDD--EVGGEALGRLLVVYPWTQRFFESFGD      48
  || |:|.:|.|.|.|||...|.|.|||.|:|:|:|.|.|:|:|..| |
EMBOSS_001      1 MV-LSPADKTNVKAAWGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-D      48

EMBOSS_001     49 LSTPDAMMGNPVKVKAHGKKVLGAFSDGLAHLDDLKGTFTATLSELHCDKLH      98
  ||      .|:|:|.|.|||||.|.|:|:|:|.|:|:|:|:|:|:|..|.
EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93

EMBOSS_001     99 VDPENFRLGNVLCVLAHHFGKEFTPPVQAAYQKVVGAVANALAHKYH      147
  |||.||:|.|.:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR      142

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

Ref: https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=emboss_stretcher-I20180709-162239-0429-42485480-p2m

Score: 361

Substitution Matrix: BLOSUM45

- BLOSUM (BLOcks Substitution Matrix) wird für Sequenzalignment von Proteinen benutzt.

- BLOSUM-Matrizen mit hohen Zahlen sind zum Vergleichen eng verwandter Sequenzen vorgesehen, während solche mit niedrigen Zahlen zum Vergleichen entfernter verwandter Sequenzen ausgelegt sind.
- z.B. BLOSUM-80 wird für Sequenzalignment von Proteinen mit ähnlicher Sequenzen, BLOSUM-45 hingegen für Proteinen mit vielen Mismatches.
- Die Similarity ist um 3% und Score um 84 gestiegen.

Gap open: 12

Gap extend: 2

(3) Globales Alignment mit einer anderen GAP OPEN penalty

```
#####
# Program: stretcher
# Rundate: Mon  9 Jul 2018 16:27:10
# Commandline: stretcher
#   -auto
#   -stdout
#   -asequence emboss_stretcher-I20180709-162709-0539-76143305-plm.asequence
#   -bsequence emboss_stretcher-I20180709-162709-0539-76143305-plm.bsequence
#   -datafile EBLOSUM62
#   -gapopen 1
#   -gapextend 2
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 1
# Extend_penalty: 2
#
# Length: 171
# Identity:      72/171 (42.1%)
# Similarity:    90/171 (52.6%)
# Gaps:          53/171 (31.0%)
# Score: 339
#
#
#=====

EMBOSS_001      1 MVHLTPEEKs-AVTALWGKVNvDEVG--G-EAL-GRL-L-VVYPWTQR-F      42
  || |:|.:|: |.|.|||||.. ..| | ||| |:| |:| |.:|
EMBOSS_001      1 MV-LSPADKTN-VKAAWGKVGA-HAGEYGAEALE-RMFLS--FP-TTKTY      43

EMBOSS_001     43 F-ESFGDLSTPDVAMGNPKVKAHGKKVLGAFSDGL--A--HLD---N-LK      83
  | . | ||| . |.:|.|.|||| | |.| | |:| | |
EMBOSS_001     44 FPH-F-DLS--H---GSAQVKGHGKKV--A--DALTNAVAHVDDMPNAL-      81

EMBOSS_001     84 GTFATLSELHCDKLHVDP-ENFRLLGNVLVC--V-LA-HHF-GKEFTPPV      127
  : | ||:|.|.|||| | |:|.: | | || | . . |||.|
EMBOSS_001     82 -S-A-LSDLHAHKLRVDPV-NFKLLSH---CLLVTLAAH-LPA-EFTPAV      122

EMBOSS_001    128 QAAYQKV VAGVANAL-AHKYH      147
  .|:..|.:|.:|.:| : ||.
EMBOSS_001    123 HASLDKFLASVSTVLTS-KYR      142

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

Ref: https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=emboss_stretcher-I20180709-162709-0539-76143305-plm

Score: 339

Substitution Matrix: BLOSUM62

Gap open: 1

- Gap extend: 2

```
#####  
# Program: water  
# Rundate: Mon   9 Jul 2018 16:43:27  
# Commandline: water  
# -auto  
# -stdout  
# -asequence emboss_water-I20180709-164324-0815-95261709-plm.asequence  
# -bsequence emboss_water-I20180709-164324-0815-95261709-plm.bsequence  
# -datafile EBLOSUM62  
# -gapopen 10.0  
# -gapextend 0.5  
# -aformat3 pair  
# -sprtein1  
# -sprtein2  
# Align_format: pair  
# Report_file: stdout  
#####  
  
#=====
```

#
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 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFSLFPTTKTYFPHF-DLS-           50  
      |:|.:|:.|.|.||| | :..|.|.||||.|:~::~|.|:~::|.~| |||  
EMBOSS_001          4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST       51  
  
EMBOSS_001         51 ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP        96  
      .|:~::~|.||| |||.~::~||:|:~::~::~.||:| |||.~|||  
EMBOSS_001         52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDDLKGTFATLSELDKLAHVDP       101  
  
EMBOSS_001         97 VNFKLLSHCLLVTLAAHLPAEFTTPAVHASLDKFLASVSTVLTISKY          141  
      .||:| |.:|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|.~|  
EMBOSS_001        102 ENFRLLGNVLVCVLAHHFGKEFTPVPQAAYQKVAGVANALA HKY            146  
  
#-----  
#-----
```

Ref: https://www.ebi.ac.uk/Tools/services/web/toolresult.ebi?jobId=emboss_water-I20180709-164324-0815-95261709-p1m

Score: 293.5

- Bei dem Lokalen Alignment wurden die Teilen von beiden Sequenzen verglichen. Die Score ist 16.5 gestiegen.