

Bericht 1 zum 6. Übungsblatt

Aufgabe 2:

Protein-Sequenz von „Human Hemoglobin subunit alpha“ (HBA_HUMAN):

```
MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGH  
GKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEF  
TPAVHASLDKFLASVSTVLTISKYR
```

Protein-Sequenz von „Human Hemoglobin subunit beta“ (HBB_HUMAN):

```
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAVMGNP  
KVKAHGKKVLGAFSDGLAHLDLNLKGTFATLSELHCDKLVDPENFRLLGNVLVCVLAHH  
FGKEFTPPVQAAYQKVVAGVANALAHKYH
```

Aufgabe 3:

Globales Alignment:

Hierbei werden alle Aminosäuren bzw. Nukleotide in jeder Sequenz gleichzeitig erfasst, wobei die Sequenzen von Anfang bis Ende aligniert werden (end-to-end-Alignment). Globales Alignment ist dabei am sinnvollsten, wenn die zu vergleichenden Sequenzen ähnlich sind und in etwa die gleiche Länge besitzen.

Lokales Alignment:

Im Gegensatz zu Globales Alignment wird Lokales Alignment bei voneinander abweichenden Sequenzen angewendet, bei denen angenommen wird, dass diese vergleichbare Regionen oder ähnliche Sequenzmotive bzw. Teilsequenzen innerhalb der gesamten Sequenz besitzen. Das Alignment kann überall innerhalb der Sequenz anfangen (mehrere Startstellen), wobei mehrere konservierte, lokale Abschnitte sichtbar werden. Hierbei werden nur Teilabschnitte der Sequenzen miteinander verglichen.

Aufgabe 4:

a) + b) Die Ergebnisse der verschiedenen Arten der Alignments und Erklärung der verwendeten Parameter:

Das Alignment aus der Vorlesung zum Vergleich:

```

Helix      AAAAAAAAAAAAAAAAAA      BBBBBBBBBBBBBBBBBCCCCCCCCCCC
HBA_HUMAN  -----VLSPADKTNVKAAGKVGA--HAGEYGAEALERMFLSPPTTKTYFPHF
HBB_HUMAN  -----VHLTPEEKSAVTALWGKV---NVDEVGGEALGRLLVVYPWTQRFFESF
MYG_PHYCA  -----VLSEGEWQLVLHVWAKVEA--DVAGHGQDILIRLFKSHPETLEKFDRF
GLB3_CHITP -----LSADQISTVQASFDKVKG-----DPVGILYAVFKADPSIMAKFTQF
GLB5_PETMA PIVDTGVSAPLSAAEKTIRSAPVYS--TYETSGVDILVKFFTSTPAAQEFFPKF
LGB2_LUPLU -----GALTESQAALVKSSWEEFN--NIPKHTHRFFILVLEIAPAAKDLFS-F
GLB1_GLYDI -----GLSAAQRQVIAATWKDIAGADNGAGVGKDCLIKFLSAHPQMAAVFG-F
Consensus   Ls.... v a W kv . . g . L.. f . P . F F

Helix      DDDDDDEEEEEEEEEEEEEEEEEEEEEEE      FFFFFFFFFFFFFF
HBA_HUMAN  -DLS-----HGSAQVKGHGKKVADALTNVAHV---D--DMPNALSLSDLHAHKL-
HBB_HUMAN  GDLSTPDAMGNPKVKAHGKKVLGAFSDGLAHL---D--NLKGTFTLSELHCDKL-
MYG_PHYCA  KHLKTEAEMKASEDLKKHGVTVLTAALGAILKK---K-GHHEAELKPLAQSHATKH-
GLB3_CHITP AG-KDLESIKGTAPFETHANRIVGFFSKIIGEL--P---NIEADVNTFVASHKPRG-
GLB5_PETMA KGLTTADQLKKSADVRWHAERIINAVNDAVASM--DDTEKMSMKLRDLSGKHAASF-
LGB2_LUPLU LK-GTSEVPQNNPELQAHAGKVFKLVYEAAIQLVTVGVVTDATLKNLGSVHVSKG-
GLB1_GLYDI SG----AS---DPGVAALGAKVLAQIGVAVSHL--GDEGKMVAQMKAVGVVRHKGYGN
Consensus   . t . . . v..Hg kv. a a..l d . a l. l H .

Helix      FFGGGGGGGGGGGGGGGGGGGGG      HHHHHHHHHHHHHHHHHHHHHHHHH
HBA_HUMAN  -RVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR-----
HBB_HUMAN  -HVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVAGVANALAHKYH-----
MYG_PHYCA  -KIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
GLB3_CHITP --VTHDQLNNFRAGFVSYMKAHT--DFA-GAEAAWGATLDTFFGMIFSKM-----
GLB5_PETMA -QVDPQYFKVLAAVIADTVAAAG-----DAGFEKLMSMICILLRSAY-----
LGB2_LUPLU --VADAHFPVVKEAILKTIKEVVGAKWSEELNSAWTIAYDELAIVIKKEMNDAA---
GLB1_GLYDI KHIKAQYFEPLGASLLSAMEHRIGGKMNAAKDAWAAAYADISGALISGLQS-----
Consensus   v. f l . . . . . f . aa. k. . l sky

```

Bei diesem Alignment wurden die Sequenzen von Hämoglobin A und B miteinander aber auch mit anderen Sequenzen, die zu den Familien der Globuline gehören, aligniert. Auf diese Weise wurden jeweils andere Alignments und damit andere Gaps erhalten.

(1) Globales Alignment mit voreingestellten Parametern (BLOSUM62):

```
#####
# Program: needle
# Rundate: Wed  4 Jul 2018 11:52:42
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180704-115232-0004-68058745-p2m.asequence
#   -bsequence emboss_needle-I20180704-115232-0004-68058745-p2m.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: 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-LSPADKTNVKAAGWKVGAGHAGEYGAELERMFLSFPTTKTYFPHF-D      48
                  || |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

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

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR     142
                  |||.||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH     147
```

Mit BLOSUM (BLOcks Substitution Matrix) werden Substitutionsmatrices bezeichnet, die für das Alignment von Aminosäure-Sequenzen verwendet werden. Eine Substitutions-Matrix ordnet jedem möglichen Alignment eines Aminosäurepaares einen „score“ zu. Dabei wird eine BLOSUM-Matrix mit einer hohen Zahl (in diesem Fall BLOSUM62) zum Vergleich von nahe verwandten Sequenzen angewendet, wohingegen BLOSUM mit einer niedrigen Zahl zum Vergleich entfernt verwandter Sequenzen dient. Die Algorithmen für globales und lokales Alignment verwenden für Proteinsequenzen häufig BLOSUM. (Quelle: <https://de.wikipedia.org/wiki/BLOSUM#Verwendung>)

Mit der „Gap Penalty“ erhält das Einführen und Verlängern von Lücken innerhalb einer Sequenz eine unterschiedliche Bewertung beim Vergleich zwischen zwei oder mehr

Sequenzen. Hierbei werden „Gap Penalties“ verwendet, um „Alignment Scores“ basierend auf der Länge und der Anzahl von Gaps zu verbessern. Durch das Einführen von „Gaps“ können die Sequenzen mithilfe von Alignment Algorithmen eine höhere Übereinstimmung erhalten. Jedoch müssen die „Gaps“ auch minimiert werden, um eine valide Aussage machen zu können. (Quelle: https://en.wikipedia.org/wiki/Gap_penalty)

In diesem Fall wird durch eine Gap Penalty von 10.0 für jedes Gap 10 Punkte abgezogen.

Bei diesem Alignment gibt es weniger Gaps als bei dem Alignment aus der Vorlesung. Es gibt mehrere Mismatches, wobei es sich bei einigen von ihnen um den Austausch einer ähnlichen Aminosäure handelt.

(2) Globales Alignment mit einer anderen Substitutionsmatrix (PAM250):

```
#####
# Program: needle
# Rundate: Tue 10 Jul 2018 18:46:26
# Commandline: needle
# -auto
# -stdout
# -asequence emboss_needle-I20180710-184624-0950-91595832-plm.asequence
# -bsequence emboss_needle-I20180710-184624-0950-91595832-plm.bsequence
# -datafile EPAM250
# -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: EMBOS_001
# 2: EMBOS_001
# Matrix: EPAM250
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      63/149 (42.3%)
# Similarity:    110/149 (73.8%)
# Gaps:          9/149 ( 6.0%)
# Score: 346.5
#
#
#=====

EMBOSS_001      1 -MVLSPADKTNVKAAGWKVGAHAGEYGAEEALERMFLSFPTTKTYFPHF-D      48
                  :.:|.:|.:|.:|.|.||| :.:|.:|.|.|.:|.:|.:|.:| |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

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

EMBOSS_001      94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR      142
                  |||.||.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001      99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
```

Die Point Accepted Mutation Matrix (kurz: PAM-Matrix) ist wie die BLOSUM-Matrix ebenfalls eine Substitutions-Matrix, die verwendet wird, um Sequenz-Alignments einen Score zuzuordnen und die Wahrscheinlichkeit eines Aminosäureaustausches zu bestimmen. Unter PAM wird hierbei der Austausch einer einzelnen Aminosäure in der Primärstruktur eines Proteins mit einer anderen Aminosäure verstanden. In einer PAM-Matrix stellt jede Spalte und Zeile eine der 20 Standard-Aminosäuren dar (Wahrscheinlichkeit eines Austausches einer Aminosäure einer Zeile mit einer Aminosäure aus einer Spalte durch einen oder mehrerer PAMs). Die PAM250 Matrix würde der BLOSUM45 entsprechen. (Quelle: https://en.wikipedia.org/wiki/Point_accepted_mutation)

Bei diesem Alignment gibt es weniger Gaps als bei dem Alignment aus der Vorlesung. Es gibt viel mehr Mismatches (werden eher toleriert) und der score nimmt zu. Bei den Mismatches handelt es sich zum größten Teil um ähnliche Aminosäuren.

(3) Globales Alignment mit einer anderen GAP OPEN Penalty (1.0) (BLOSUM62):

```
#####
# Program: needle
# Rundate: Tue 10 Jul 2018 19:04:15
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180710-190408-0654-24534045-plm.asequence
#   -bsequence emboss_needle-I20180710-190408-0654-24534045-plm.bsequence
#   -datafile EBLOSUM62
#   -gapopen 1.0
#   -gapextend 0.5
#   -endopen 10.0
#   -endextend 0.5
#   -aformat3 pair
#   -sprotein1
#   -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 1.0
# Extend_penalty: 0.5
#
# Length: 165
# Identity:      72/165 (43.6%)
# Similarity:    90/165 (54.5%)
# Gaps:          41/165 (24.8%)
# Score: 347.0
#
#=====

EMBOSS_001      1 MV-LSPADKTNVKAAWGKVGAGAHAGEYGAEALERMFLS--FP-TTKTYFPH      46
                  || |:|.:|.:|.|.|||| :..|.||||.:| | :| |:| :|..
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLL-L-VVYPWTQR-FFES      45

EMBOSS_001      47 F-DLSH-----GSAQVKGHGKKV--A--DALTNAVAHVDDMPNAL--S-      82
                  | ||| :|.:|.|.|||| | |. | |.:| | | :
EMBOSS_001      46 FGDLS-TPDAVMGNPKVKAHGKKVLGAFSDGL---AHLD---N-LKGTFF      86

EMBOSS_001      83 A-LSDLHAHKLRVDPVNFKLLSH---CLLVTLAAHLPA-EFTPAVHASLD      127
                  | ||:|.|.|.|.|||||:|.:| | | |.|. . |||||.|.:..
EMBOSS_001      87 ATLSELHCDKLHVDPENFRLLGNVLVC--V-LAHHF-GKEFTPPVQAAAYQ      132

EMBOSS_001      128 KFLASVSTVLTSKYR      142
                   |.:|.:|.:|.|.|.
EMBOSS_001      133 KVVAGVANALAHKYH      147
```

Hier wurde die Gap Penalty auf 1.0 verringert, was heißt, dass für das Einführen von Gaps jeweils 1.0 Punkt angezogen werden. Dies ist ein geringerer Wert als 10.0 (die Gap Penalty, die für die restlichen Alignments verwendet wurde), was zu einer höheren Toleranz gegenüber Gaps als gegenüber Mismatches führt. Dies müsste zu einer höheren Anzahl an Gaps führen, als im Alignment von davor und dies ist hier auch der

Fall. Jedoch führt das Einführen von zu vielen Gaps zu einer geringeren Ähnlichkeit der Sequenzen zueinander.

```
#####  
# Program: water  
# Rundate: Thu 5 Jul 2018 11:22:55  
# Commandline: water  
# -auto  
# -stdout  
# -asequence emboss_water-I20180705-112252-0324-98468452-plm.asequence  
# -bsequence emboss_water-I20180705-112252-0324-98468452-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 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFSLFPTTKTYFPFH-DLS-           50  
                  |::|::|.||.||||   ..|.|||.||:::|.||:|..| |||  
EMBOSS_001      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST         51  
  
EMBOSS_001     51 ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP       96  
                  .|::|.|.|||||.||::|:::|.||:|..|..| |||  
EMBOSS_001     52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDP        101  
  
EMBOSS_001     97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY             141  
                  .||:|.|.:|:..|||.||.|||||.||:~|.:|:~|..|..|  
EMBOSS_001    102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY              146
```