

Übungszettel 6

Aufgabe 2

Human Hemoglobin subunit alpha

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

Human Hemoglobin subunit beta

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

Aufgabe 3

Globales Alignment

Bei einem globalen Alignment werden alle Bestandteile berücksichtigt (Basen oder Aminosäuren). Es wird versucht, die Übereinstimmung der Elemente durch Einfügen von Gaps über die gesamte Sequenz zu maximieren. Diese Methode eignet sich am besten für nah verwandte Sequenzen, die nicht große und häufige Insertionen oder Deletionen enthalten

EMBOSS Needle, BLOSUM Matrix

Lokales Alignment

Das Lokale Alignment hingegen versucht, einzelne Teile der beiden Sequenzen so exakt wie möglich übereinanderzulegen. Es wird das Alignment in Stücken optimiert. Diese Methode ist geeignet, wenn weit entfernte Verwandte bzw. Proteine mit ähnlicher Funktion untersucht werden. So können konservierte Domänen gut untersucht werden.

BLAST, LALIGN



Aufgabe 4

(a)

(1) Globales Alignment mit voreingestellten Parametern

```
#####
# Program: needle
# Rundate: Thu  5 Jul 2018 11:35:24
# Commandline: needle
#   -auto
#   -stdout
#   -asequence emboss_needle-I20180705-113521-0746-31618721-plm.asequence
#   -bsequence emboss_needle-I20180705-113521-0746-31618721-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: EMBOS_001
# 2: EMBOS_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-LSPADKTNVKAANGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                  || |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLR      93
                  || .|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001     49 LSTPDVVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR      142
                  |||:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
```

BLOSUM62 Matrix und Gap Penalty = 10.0

Das Alignment stimmt mit dem aus der Vorlesung überein. Es wurden an den gleichen Stellen Gaps eingefügt. (Gelb markiert die Gaps, die mit dem Vergleichsalignment übereinstimmen)

Auch die PAM Matrix führt zum selben Alignment. (Gelb markiert die Gaps, die mit dem Vergleichsalignment übereinstimmen)

(3) Globales Alignment mit einer anderen GAP OPEN penalty

```
#####  
# Program: needle  
# Rundate: Thu 5 Jul 2018 11:41:34  
# Commandline: needle  
# -auto  
# -stdout  
# -asequence emboss_needle-I20180705-114132-0388-317848-p2m.asequence  
# -bsequence emboss_needle-I20180705-114132-0388-317848-p2m.bsequence  
# -datafile EPAM160  
# -gapopen 100.0  
# -gapextend 0.5  
# -endopen 10.0  
# -endextend 0.5  
# -aformat3 pair  
# -sprtein1  
# -sprtein2  
# Align_format: pair  
# Report_file: stdout  
#####  
  
#=-----=  
#  
# Aligned_sequences: 2  
# 1: EMBOSS_001  
# 2: EMBOSS_001  
# Matrix: EPAM160  
# Gap_penalty: 100.0  
# Extend_penalty: 0.5  
#  
# Length: 147  
# Identity:      44/147 (29.9%)  
# Similarity:    83/147 (56.5%)  
# Gaps:          5/147 ( 3.4%)  
# Score: 123.0  
#  
#  
#=-----=  
  
EMBOSS_001           1 -----MVLSPADKTNVKAAWGVGAHAGEYGAEALERMFSLFPTTKTYFP     45  
                    .....|.....:.....:.....|.~::~:  
EMBOSS_001           1 MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS   50  
  
EMBOSS_001           46 HFDLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSSDLHAHKLRVD     95  
                    .|.~::||:||:|||||..:::::||||:~::||:~::||:~::||  
EMBOSS_001           51 TPDAVMGNPVKVKAHGKKLVGAFSDGLAHLDNLKGTFATLSELHCCKLHVD    100  
  
EMBOSS_001           96 PVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR       142  
                    |.||:~::|...||.|~::|||:~::||:~::||:~::||:~::||:~::||  
EMBOSS_001          101 PENFRLLGNVLVCVLAHHFGKEFTPVPQAAYQKVVGANALAHHKYH       147
```

Die hohe Gap Penalty führt zu einem anderen Alignment. Da eine Lücke in der Sequenz zu einer 10 mal höheren Abwertung der Sequenz führt, wird nur ein Gap ganz am Anfang eingeführt. (Orange markiert die Stellen, die mit dem Vergleichsalignment nicht übereinstimmen). Dieses Alignment hat eine niedrigere Prozentzahl (29 % statt 43%) an identischen und ähnlichen (56% statt 71%) Aminosäuren (an derselben Stelle im Alignment).

(4) Lokales Alignment mit voreingestellten Parametern

```

The best non-identical alignments are:      ls-w bits E(1) %_id %_sim alen
EMBOSS_001                                ( 147) 375 106.6 1.7e-28 0.434 0.745 145
+-                                         42 15.9      0.29 0.333 0.762 21
+-                                         37 14.5      0.58 0.355 0.677 31
+-                                         32 13.2      0.89 0.500 0.833 12
+-                                         29 12.4      0.98 0.294 0.647 17
+-                                         29 12.4      0.98 0.600 0.800 10
+-                                         28 12.1      0.99 0.417 0.667 12
+-                                         28 12.1      0.99 0.375 0.625 16
+-                                         27 11.8      1 0.462 0.538 13
+-                                         26 11.6      1 0.286 0.667 21

>>>EMBOSS_001, 142 aa vs lalign-I20180705-114451-0687-19980705-p2m.bsequence library

>>EMBOSS_001                                (147 aa)
  Waterman-Eggert score: 375; 106.6 bits; E(1) < 1.7e-28
  43.4% identity (74.5% similar) in 145 aa overlap (3-141:4-146)

      10      20      30      40      50
EMBOSS LSPADKTNVKAAGWKVGAHAGEYGAEALERMFSLFPTTKTYFPHF-DLS-----HGSAQV
      ::  ::  :  :  ::  ::  :  ::  ::  :  ::  :  ::  :  ::  :  ::  :  ::
EMBOSS LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPKV
      10      20      30      40      50      60

      60      70      80      90      100     110
EMBOSS KGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPA
      :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  :::::  ::
EMBOSS KAHGKKVLGAFSDGLAHLNLLKGTFTLSELHCDKLHVDPENFRLLGNVLCVLAHFGK
      70      80      90      100     110     120

      120     130     140
EMBOSS EFTPAVHASLDKFLASVSTVLTSKY
      ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::::  ::
EMBOSS EFTPPVQAAYQKVVAGVANALAHKY
      130     140

```

Das Lokale Alignment stimmt wieder mit dem Vergleichsalignment überein. In diesem Falle gibt es keinen Unterschied zwischen Globalem und Lokalem Alignment mit Standardparametern.

Alignment zum Vergleich

```

Helix      AAAAAAAAAAAAAAAAAA  BBBBBBBBBBBBBBBBBCCCCCCCCCCCC
HBA_HUMAN  -----VLSPADKTNVKAAGWKVGA--HAGEYGAEALERMFSLFPTTKTYFPHF
HBB_HUMAN  -----VHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESF
MYG_PHYCA  -----VLSGEWQLVLHVWAKVEA--DVAGHGQDILIRLFKSHPETLEKFDRE
GLB3_CHITP -----LSADQISTVQASFDKVKG-----DPVCILYAVFKADPSIMAKFTQF
GLB5_PETMA PIVDTGSAPLSAAEKTKIRSAWAPVYS--TYETSGVDILVKFFSTPAAQEFFPKF
LGB2_LUPLU -----GALTESQAALVKSSWEEFNA--NIPKHTHRFFILVLEIAPAAKDLFS-F
GLB1_GLYDI -----GLSAAQRCVIAATWCKDIAGADNGAGVGKDCLIKFLSAHPQMAAVFG-F
Consensus  Ls... v a W kv . . g . L . f . P . F F

Helix      DDDDDDDDEEEEEEEEEEEEEEEEEEE FFFFFFFF
HBA_HUMAN  -DLS-----HGSAQVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKT-
HBB_HUMAN  GDLSTPDAMGNPKVKAHGKKVLGAFSDGLAHL---D--NFKGTFTLSELHCDKL-
MYG_PHYCA  KHLKTRAEKASEDLKKHGVTVLTALGAILKK---K-GHHEAELKPLAQSHATKH-
GLB3_CHITP AG-KDLESIKGTAPFETHANRIVGFSSKILIGEL--P---NIEADVNTFVASHKPRG-
GLB5_PETMA KGLTTADQLKKSADVRWHAERI INAVNDAVASM--DDTEKMSMKLRDLGSKHAKSF-
LGB2_LUPLU LK-GTSEVPQNNPELQAHAGKVFKLVYEAAIQLQVTVGVVTDATLKNLQSVHVSKG-
GLB1_GLYDI SG----AS----DPGVAALGAKVLAQIGVAVSHI--GDEGKMVAQMKAVGVVRHKGYGN
Consensus  . t . . . v . Hg kv . a a . . l d . a l . l H .

Helix      FFGCGGGGGGGGGGGGGGGGGGGG HHHHHHHHHHHHHHHHHHHHHHHHHHHHH
HBA_HUMAN  -RVDPNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR-----
HBB_HUMAN  -HYDPENFRLLGNVLCVLAHFGKFTPPVQAAYQKVVAGVANALAHKYH-----
MYG_PHYCA  -KIPKYLEFISEATTHVLHSRHPGDFGADAQGAMNKALELFRKIDIAAKYKELGYQG
GLB3_CHITP --VTHDQLNNFRAGFVSYMKAHT--DFA-CASAAGCATLDTFFGMIFSKM-----
GLB5_PETMA -QVDPQYFKVLAAVIADTVAG-----DAGFEKIMSMICILLRSAY-----
LGB2_LUPLU -VADAHFPVVKRAILKTIKEVVGAKWSEELNSAWTIAYDELAIVIKKEMNDAA---
GLB1_GLYDI KHIAQYFEPPLGASLLSAMEHRIGKMNAAAKDAWAAAYADISGALISGLQS-----
Consensus  v . f l . . . . . f . aa . k . . l sky

```

(b) Parameter

BLOSUM

Die BLOcks Substitution Matrix ist eine Bewertungsmatrix, die bei Protein-Alignments verwendet wird. Sie basiert auf dem blockweisen Vergleich homologer Proteine (es werden nur Teile ohne Gaps verglichen). Aufgrund des häufigen oder seltenen Ersetzen von einzelnen Aminosäuren durch andere an den gleichen Stellen in homologen Proteinen, wird eine Scoring Matrix errechnet.

Es gibt unterschiedliche BLOSUM-Matrices abhängig davon, ob evolutionär verwandte oder entfernte Proteine untersucht werden (BLOSUM80 bspw. für nahe Sequenzen und BLOSUM30 für entfernte Proteine).

PAM Matrix

Eine Alternative ist die PAM Matrix (Point Accepted Mutation Matrix). Sie basiert auf den einzelnen Wahrscheinlichkeiten, dass Aminosäuren durch Aminosäuren mit ähnlichen Eigenschaften ersetzt sind. So kann eine missense Mutation in der DNA zu einem Lys → Arg Austausch führen. Ist das Protein trotzdem einsatzfähig, wäre dies eine akzeptierte Mutation. Die Wahrscheinlichkeiten werden über phylogenetische Bäume errechnet.

GAP Penalty

Wird verwendet, um die Alignments zu bewerten. Pro eingefügtem ‚-‘ Gap also Loch zwischen den Aminosäuren wird eine ‚Bestrafung‘, also ein Score hochgezählt. Bei einem Alignment werden unterschiedliche Möglichkeiten ausprobiert und die mit dem niedrigsten Score gilt als die beste.

Ein niedriger Gap Penalty ist geeignet für Sequenzen, die nicht nah verwandt sind und wahrscheinlich größere Insertionen bzw. Mutationen an vielen Stellen enthalten. Eine hohe Gap Penalty führt dazu, dass die Sequenzen so angeordnet werden, dass es möglichst wenig Löcher gibt. Dies kann zu einem weniger gutem Alignment führen.