

Assignment 6

Aufgabe 2:

Protein-Sequenzen von (a) „Human Hemoglobin subunit alpha“ (HBA_HUMAN) und (b) „Human Hemoglobin subunit beta“ (HBA_HUMAN):

(a)

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

(b)

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	KEFTPPVQAA	YQKVAVGVAN	ALAHKYH

Aufgabe 3:

Unterschied zwischen „Globalem Alignment“ und „Lokalem Alignment“:

Globales Alignment = Alle Symbole der beiden zu vergleichenden Sequenzen werden betrachtet, wobei eingefügte Gaps an jeder Position gleich stark „bestraft“ werden. Dies ist sinnvoll, sofern die beiden Sequenzen ähnlich lang (z.B. zwei homologe Gensequenzen) sind. Bei starken Unterschieden in der Sequenzlänge werden diese jedoch „zerrissen“ und das Alignment ist wenig aussagekräftig.

Lokales Alignment = Aus den zwei gegebenen Sequenzen werden zwei Teilsequenzen generiert, indem deren maximaler Alignment-Score gesucht wird. Dabei werden, im Gegensatz zum Globalen Alignment, Gaps am Anfang und Ende des lokalen Alignments weniger stark „bestraft“.

Aufgabe 4:

(1) Globales Alignment (Needle, EMBOSS) mit voreingestellten Parametern:

```
#=====
#
# 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-LSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHF-D	48
		: :: . - . :.. . . . :::. . :: .	
EMBOSS_001	1	MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOSS_001	49	LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR	93
		. :: . . :::. : :..... : : . .	
EMBOSS_001	49	LSTPDAMGNPKVKAHGKKVLGAFSDGLAHLNLKGTFATLSELHCCKLH	98
EMBOSS_001	94	VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR	142
		. : : :.. : . :.. .. .	
EMBOSS_001	99	VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH	147

(2) Globales Alignment (Needle, EMBOSS) mit PAM80 und voreingestellten Parametern:

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOS_001
# 2: EMBOS_001
# Matrix: EPAM80
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 318.5
#
#
#=====
```

EMBOS_001	1	MV-LSPADKTNVKAAGWKGVAHAGEYGAEALERMFLSFPTTKTYFPHF-D	48
		: .: .:. . :.. . . . :.... . .:. ..	
EMBOS_001	1	MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOS_001	49	LSH-----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR	93
		. : . . . : : . . . : : . .	
EMBOS_001	49	LSTPDVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTATLSELHCCKLH	98
EMBOS_001	94	VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR	142
		. : . : : : : . . : . .	
EMBOS_001	99	VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH	147

(3) Globales Alignment (Needle, EMBOSS) mit BLOSUM62 und GAP OPEN penalty = 50:

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 50.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      61/149 (40.9%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 210.0
#
#=====
```

EMBOSS_001	1	-MVLSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF--	47
		:. :	
EMBOSS_001	1	MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOSS_001	48	----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR	93
		... :	
EMBOSS_001	49	LSTPDAMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFFATLSELHCDKLH	98
EMBOSS_001	94	VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR	142
		: :	
EMBOSS_001	99	VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH	147

(4) Lokales Alignment (Water, EMBOSS) mit voreingestellten Parametern:

```
#=====
#
# 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	LSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-	50
		: :	
EMBOSS_001	4	LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST	51
EMBOSS_001	51	----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP	96
		. :	
EMBOSS_001	52	PDAMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFFATLSELHCDKLHVDP	101
EMBOSS_001	97	VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKY	141
		. :	
EMBOSS_001	102	ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY	146

Vergleich mit dem Alignment aus der Vorlesung:

Multiple sequence alignment (Globin family)

```
Helix      AAAAAAAAAAAAAAAAAA  BBBBBBBBBBBBBBBBCCCCCCCCCCCC
HBA_HUMAN  -----VLSPADKTNVKAANGKVG--HAGEYGAEALERMLSFPTTKTYFPHF
HBB_HUMAN  -----VHLTPEEKSAVTALWGKV---NVDEVGGEALGRLLVVYPWTQRFFESF
MYG_PHYCA  -----VLSEGEWQLVHVWAKVEA--DVAGHGQDILIRLFKSHPETLEKFDRF
GLB3_CHITP -----LSADQISTVQASFDKVKG-----DPVGILYAVFKADPSIMAKPTQF
GLB5_PETMA PIVDTGSVAPLSAAEKTIRSAPVYS--TYETSGVDILVKFFTSTPAAQEFFPKF
LGB2_LUPLU -----GALTESQAALVKSSWEEFNA--NIPKHTHRFFILVLEIAPAAKDLFS-F
GLB1_GLYDI -----GLSAAQRQVIAATWKDIAGADNGAGVGKDKLIKFLSAHPQMAAVFG-F
Consensus   Ls.... v a W kv . . g . L.. f . P . F F

Helix      DDDDDDDDEEEEEEEEEEEEEEEEEEEEEEE FFFFFFFF
HBA_HUMAN  -DLS-----HGSAQVKGHGKKVADALTNAVAHV---D--DMPNALSALSDLHAHKL-
HBB_HUMAN  GDLSTPDVAMGNPKVKAHGKKVLGAFSDGLAHL---D--NLKGTFTATSELHCDKL-
MYG_PHYCA  KHLKTEAEMKASEDLKKGHTVLTALGAILKK---K-GHHEAELKPLAQSHATKH-
GLB3_CHITP AG-KDLESIKGTAPFETHANRIVGFFSKIIGEL--P---NIEADVNTFVASHKPRG-
GLB5_PETMA KGLTTADQLKKSADVRWHAERI INAVNDAVASM--DDTEKMSMKLRDLSGKHAQSF-
LGB2_LUPLU LK-GTSEVPQNNPELQAHAGKVFKLVYEAQIQVTVGVVTDATLKNLGSVHVSKG-
GLB1_GLYDI SG----AS---DPGVAALGAKVLAQIGVAVSHL--GDEGKMVAQMKAVGVHRKGYGN
Consensus   . t . . . v..Hg kv. a a...l d . a l l H .

Helix      FFGGGGGGGGGGGGGGGGGGGGGG HHHHHHHHHHHHHHHHHHHHHHHHH
HBA_HUMAN  -RVDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR-----
HBB_HUMAN  -HVDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVAVANALAHKYH-----
MYG_PHYCA  -KIPIKYLEFISEAIIHVLHSRHPGDFGADAQGAMNKALELFRKDIAAKYKELGYQG
GLB3_CHITP --VTHDQLNNFRAGFVSYMKAHT--DFA-GAEAAGWATLDTFFGMIFSKM-----
GLB5_PETMA -QVDPQYFKVLAADIADTVAAG-----DAGFEKLMSMICILLRSAY-----
LGB2_LUPLU --VADAHFPVVKEAILKTIKEVVGAKWSEELNSAWTIAYDELAIVIKKEMNDAA---
GLB1_GLYDI KHIKAQYFEPLGASLLSAMEHRIGGKMNAAKDAWAAAYADISGALISGLQS----
Consensus   v. f l . . . . . f . aa. k. . l sky
```

Im Unterschied zu den selbstständig durchgeführten Alignments wurden hier mehr als zwei Sequenzen der Globulin-Familie verglichen. Dabei kann eine Consensus-Sequenz erstellt werden, die sich größtenteils in den Sequenzen HBA/B_HUMAN widerspiegelt. Zu Beginn beider Sequenzen sind eine Reihe von Gaps eingefügt, die durch das Alignment mit der GLB5_PETMA-Sequenz verursacht werden. Durch visuellen Vergleich der verschiedenen Alignments scheint für das multiple Sequenz-Alignment das Prinzip des lokalen Alignments verwendet worden zu sein. Dies wird vermutet, da vor allem zu Beginn und Ende der Sequenzen zahlreiche Gaps auftreten, die bestimmte Teilsequenzen umranden.

Erklärung der verwendeten Parameter:

PAM80-Matrix: Es handelt sich um eine Matrix, der ursprünglich das globale Alignment von 100 Proteinen in Gruppen mit mehr als 85% Sequenzidentität zugrunde liegt. Je größer die evolutionären Unterschiede der Sequenzen werden, desto geringer wird die Sequenzähnlichkeit. Dieser evolutionäre Abstand wird hier über die Anzahl der Punktmutationen pro 100 AS definiert: 1 PAM (percent accepted mutations) entspricht einer Punktmutationsrate von 1 AS / 100 AS. Demnach entspricht eine PAM80-Matrix: 80 Punktmutationen pro 100 AS.

Im Gegensatz zu den PAM-Matrizen sind BLOSUM-Matrizen mit hohen Zahlen für den Vergleich sehr ähnlicher Sequenzen geeignet, solche mit kleinen Zahlen hingegen für das Alignment von stärker voneinander abweichenden Sequenzen.

[illegible]

Gap Extend penalty: Höhe der “Bestrafung” für das Einfügen zusätzlicher Gaps (mehr als ein Gap aufeinander folgend) in die aligned Sequenzen.

Das optimale Ergebnis eines Alignments ist ein höchst möglicher Score, weshalb Gap penalties mit negativen Werten verrechnet werden. Matches haben im Normalfall positive Zahlenwerte, die je nach Matrix höhere oder niedrigere Zahlenwerte annehmen können. Mismatches werden mit 0 bzw. kleineren negativen Zahlenwerten verrechnet und verringern so den finalen Score. Meist wird das Einbringen von Gaps jedoch stärker bestraft, um sinnvolle Alignments zu generieren.