

Übungszettel 6 BIOINF101 – Aufgabe 6:

Aufgabe 2

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

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

Protein-Sequenzen von „Human 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	KEFTPPVQAA	YQKVVAGVAN	ALAHKYH

Aufgabe 3

Globale Alignments werden für Sequenzen genutzt die eine Ähnlichkeit hinsichtlich ihrer Länge aufweisen und zudem starke Sequenzhomologie haben, wobei alle Symbole berücksichtigt werden. Ein lokales Alignment von zwei Sequenzen ist ein globales Alignment von einer Teilsequenz (Substring) von S und einer Teilsequenz T. (Quelle: https://de.wikipedia.org/wiki/Sequenzalignment#Globales_Alignment).

Aufgabe 4

Im Folgendem wurden Alignments von den beiden Sequenzen Human *Hemoglobin subunit alpha* und Human *Hemoglobin subunit beta* erstellt. Hierbei wurden einige Parameter variiert wie z. B die Substitutionsmatrix (2) oder Gap open penalty (3).

(1) Globales Alignment 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-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDDLHAKLR      93
      ||      .|:~::~|.|.|||||.|.|.~::~|:|:~::~~::~~::~|:|:~::~|.|.|.
EMBOSS_001     49 LSTPDVAMGNPKVKAHGKKVLGAFSDGLAHLNLTGTFATLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR      142
      |||.|.|:|:|.~::~|:~::~|.|.~::~|:~::~|.|.~::~|:~::~|.~::~|.~::~|.
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147

#-----
#_____
```

Die *BLOSUM62*-Matrix wird für evolutionär nahe verwandte Proteine verwendet, in dem einzelne Blöcke ohne Lücken innerhalb der Sequenz von homologen Proteinen verglichen werden (Quelle: <https://de.wikipedia.org/wiki/BLOSUM>).

Beim Gap penalty werden Positionen in einer der Zeichenketten ausgelassen, in Form von gaps/Lücken.

(2) Globales Alignment miteinanderen Substitution MATRIX EPAM160

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM160
# Gap_penalty: 10.0
```

```
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    106/149 (71.1%)
# Gaps:          9/149 ( 6.0%)
# Score: 274.5
#
#
#=====
EMBOSS_001      1 MV-LSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
      || |:|.:.|:|.|.|.||| | :..|. |:| |||. |:::|. |.:.|..| |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTPQRFESFGD      48
EMBOSS_001     49 LSH-----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
      ||.      |:|. |:| ||| |||. |:::| |:| :|:|:|:|:| |:| |..| |:
EMBOSS_001     49 LSTPDAVMGNPKVKAHGKKVLGAFSDGLAHL DNLKGT FATLSELHCDKLH      98
EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR      142
      |||. |:| |:|. |:|. |. |. |. | | |:| |:|:|. |:| |:|:|. |:| |:
EMBOSS_001     99 VDPENFRL LGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
#-----
#-----
```

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 25.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      61/149 (40.9%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 260.0
#
#
#=====

EMBOSS_001      1 -MVLSPADKTNVKAAWGKVGGAHAGEYGAEALERMFSLFPTTKTYFPHF--      47
                  :.:|.:|.:|.|.||||   :..|.|.|||.|:~::~|.|:~::~|..|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001      48 ----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR      93
                  |...|:~::~|.|.|||||.|.|:~::~|:~::~|:~::~|:~::~|:~::~|..|
EMBOSS_001      49 LSTPDAVMGNPVKVKAHGKKVLGAFSDGLAHL DNLKGT FATLSELHCDKLH      98
```

EMBOSS_001	94	VDPVNFKLLSHCLLVTLAAHLP AEFTPAVHASL DKFLASVSTVLTSKYR	142
		. : . : : . . : . : . . .	
EMBOSS_001	99	VDPENFRLLGNVLCVLAH HFGKEFTPPVQAAYQKV VAGVANALAHKYH	147

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

Mit einem Gap penalty von 25.0 wurde ein geringer Score (260) erzielt statt 274,5 mit einem Gap penalty von 10.0.

(4) Lokales Alignment 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	LSPADKTNVKA AWGKVG AHAGEYGA EALERMFLSFPTTKTYFPHF-DLS-	50
		: . : . . . : . . . : : : . : : .	
EMBOSS_001	4	LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST	51
EMBOSS_001	51	----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALS DLHAHKLRVDP	96
		. : : . . . : : : : : : . .	
EMBOSS_001	52	PDAVMGNPKVKAHGKKVLGAFSDGLAHL DNLKGT FATLSELHCDKLHVDP	101
EMBOSS_001	97	VNFKLLSHCLLVTLAAHLP AEFTPAVHASL DKFLASVSTVLTSKY	141
		. : . : . : : . . : . : . .	
EMBOSS_001	102	ENFRLLGNVLCVLAH HFGKEFTPPVQAAYQKV VAGVANALAHKY	146

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

Mit dem lokalen Alignment wurde die größte Übereinstimmung gefunden/ größter Score (293,5).