

Übungszettel 6

Alignment zweier Proteinsequenzen Human Hemoglobin subunits

Aufgabe 2

Hemoglobin subunit alpha

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

<https://www.uniprot.org/uniprot/P69905>

Hemoglobin subunit beta

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

<https://www.uniprot.org/uniprot/P68871#sequences>

Aufgabe 3

Beim globalen Alignment werden beide Gesamtsequenzen vollständig zueinander ausgerichtet. Beim lokalen Alignment werden nur Sequenzbereiche betrachtet. Daher können beim lokalen Alignment auch Bereiche an unterschiedlichen Positionen innerhalb der Sequenzen entdeckt und zueinander ausgerichtet werden. Beim globalen Alignment würden solche Übereinstimmungen nicht entdeckt werden.

Aufgabe 4

(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-LSPADKTNVKAAWGKVGHAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                  || |:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTPQRFESFGD      48

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

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

#-----
```

Substitution Matrix: BLOSUM62 (Blocks Substitution Matrix), enthält unterschiedliche Scores der einzelnen Aminosäuren (AS) zueinander. So haben nahverwandte AS einen höheren Score als stark unterschiedliche, die einen geringen Score aufweisen.

Gap penalty: ist der (negative) Score der anfällt, wenn eine Insertion/Deletion in Betracht gezogen wird. Bei höheren Gap penalties wird das Einfügen einer Deletion/Insertion statistisch geringer.

(2) Globales Alignment mit einer anderen Substitution MATRIX

```
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EPAM10
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 203
# Identity:      61/203 (30.0%)
# Similarity:    61/203 (30.0%)
# Gaps:          117/203 (57.6%)
# Score: 136.0
#
#=====

EMBOSS_001      1 MV-LSPADKTNVKAAWGKV-----GAHAGEYGAEALERM-----F      34
                  || |:|..|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKVNDEVGG-----EALGRLLVVYPWTPQRF      42

EMBOSS_001     35 LSFPTTKTYFPHF----DLSHGSAQ-----VKGHGKKV--A--DA      66
                  |   |||               ||.||||| | |.
EMBOSS_001     43 -----FESFGDLS-----TPDAVMGNPKVKAHGKKVLGAFSDG      75
```

```

EMBOSS_001      67 LTNAVAHVDDMPN-----ALSALSDLHAHKLRVDPVNFKLLSH---CLLV      108
                  |      ||.|      |      |      .||.|..|||.|||.|||.||..      |.|
EMBOSS_001      76 L-----AHLN---NLKGTFA--TLSELHCDKLHVDPENFRLLGNVLVCVL-      115

EMBOSS_001     109 TLA AHLPA---EFTPAVHASLDKFLASVSTVLT SKYR-----      142
                  ||      ||||.|.|      |      |.
EMBOSS_001     116 ---AH---HFGKEFTPPVQA-----A-----YQKVVAGVANALAH      144

EMBOSS_001     143 ---      142
EMBOSS_001     145 KYH      147

```

#-----

Substitution Matrix: PAM10 (Point Accepted Mutation Matrix), enthält die Möglichkeit zur Betrachtung von Punktmutationen innerhalb des Alignments von 10%

Gap penalty: ist der (negative) Score der anfällt, wenn eine Insertion/Deletion in Betracht gezogen wird. Bei höheren Gap penalties wird das Einfügen einer Deletion/Insertion statistisch geringer.

(3) Globales Alignment mit einer anderen GAP OPEN penalty

```

# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 20.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      61/149 (40.9%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 270.0
#
#=====

EMBOSS_001      1 -MVLSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHF--      47
                  :.:|.:|.:|.:|.|.||||      :.:|.:|.|.||||.:|.:|.:|.:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     48 ----DLSHGSAQVKGHGKKVADALTNVAHVDDMPNALALSALSDLHAHKLR      93
                  |...|.:|.|.|||||.||.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001     49 LSTPDVAVMGNPVKVKAHGGKVLGAFSDDLALDNLKGTFTLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLT SKYR      142
                  |||.||.:|.|.:|.:|.|.|||||.||.:|.:|.:|.:|.:|.:|
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147

```

#-----

Substitution Matrix: BLOSUM62 (Blocks Substitution Matrix), enthält unterschiedliche Scores der einzelnen Aminosäuren (AS) zueinander. So haben nahverwandte AS einen höheren Score als stark unterschiedliche, die einen geringen Score aufweisen.

Gap penalty: ist der (negative) Score der anfällt, wenn eine Insertion/Deletion in Betracht gezogen wird. Bei höheren Gap penalties wird das Einfügen einer Deletion/Insertion statistisch geringer.

(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	LSPADKTNVKAAWGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-	50
		: .: .:. . . .:.:.:. . .:. . .	
EMBOSS_001	4	LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST	51
EMBOSS_001	51	----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDDLHAHKLRVDP	96
		. : .:.:.:. . . .:.:.:.	
EMBOSS_001	52	PDAVMGNPKVKAHGKKVLGAFSDGLAHL DNLKGT FATLSELHCDKLHVDP	101
EMBOSS_001	97	VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY	141
		. : . .:. .:.:. . .:. . .	
EMBOSS_001	102	ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY	146

#-----

Substitution Matrix: BLOSUM62 (Blocks Substitution Matrix), enthält unterschiedliche Scores der einzelnen Aminosäuren (AS) zueinander. So haben nahverwandte AS einen höheren Score als stark unterschiedliche, die einen geringen Score aufweisen.

Gap penalty: ist der (negative) Score der anfällt, wenn eine Insertion/Deletion in Betracht gezogen wird. Bei höheren Gap penalties wird das Einfügen einer Deletion/Insertion statistisch geringer.