

Bericht zum 6. Übungszettel des Moduls BioINF-101:

Aufgabe 1: <https://github.com/rinnerthomas/bioinformatics-BC/tree/master/assignment6>

Aufgabe 2:

Die Protein- bzw. Aminosäuresequenzen der humanen Hämoglobin Untereinheiten wurden der uniprot.org Datenbank entnommen. Dabei wurde für die Alpha-Untereinheit ermittelt:

P69905-1 [UniParc] [FASTA](#) [Add to basket](#)
« Hide

10	20	30	40	50
MVLSPADKTN	VKAAWGKVG	AHAGEYGAEAL	ERMFLSFPTT	KTYFPFHDLS
60	70	80	90	100
HGSAQVKGHG	KKVADALTNA	VAHVDDMPNA	LSALSDLHAH	KLRVDPVNFK
110	120	130	140	
LLSHCLLVTL	AAHLPAEFTP	AVHASLDKFL	ASVSTVLTSK	YR

Length: 142
Mass (Da): 15,258
Last modified: January 23, 2007 - v2
Checksum: ⁱ15E13666573BBBAE
BLAST

Abbildung 1: Aminosäuresequenz der humanen Hämoglobin alpha-Untereinheit (HBA1)¹

Analog für die Beta-Untereinheit:

P68871-1 [UniParc] [FASTA](#) [Add to basket](#)
« Hide

10	20	30	40	50
MVHLTPEERS	AVTALWGKVN	VDEVGGEALG	RLLVVYPWTQ	RFFESFGDLS
60	70	80	90	100
TPDAVMGNPK	VKAHGKKVLG	AFSDGLAHL	NLKGTFTL	ELHCDKLVHD
110	120	130	140	
PENFRLLGNV	LVCVLAHHFG	KEFTPPVQAA	YQKVVAGVAN	ALAHKYH

Length: 147
Mass (Da): 15,998
Last modified: January 23, 2007 - v2
Checksum: ⁱA31F6D621C6556A1
BLAST

Abbildung 2: Aminosäuresequenz der humanen Hämoglobin beta-Untereinheit (HBB1)²

Aufgabe 3:

In einem globalen Alignment wird die gesamte Sequenz bzw. alle Symbole von etwa zwei betrachteten Sequenz zum Alignment verwendet. Dieser Ansatz scheint naheliegend um ein vollständiges Alignment zu erreichen. Globale Alignment können jedoch im biologischen Kontext ungewünschte Schlüsse begünstigen, da im Rahmen der Evolution Genbereiche konserviert bleiben, aber ihre Position im Genom wechseln. In einem globalen Alignment würde man für diese keine Matching-Position finden, daher kann es sich anbieten einen lokales Alignment durchzuführen. Dabei wird jeweils eine bestimmte, kürzere Teilsequenz (wie Proteindomänen) als zu vergleichende Sequenzen ausgewählt. So erhält man zwar nicht einen kompletten Vergleich der Sequenzinformation aber vermeidet das zuvor beschriebene Szenario. Darüber hinaus ist kann die kürzere Laufzeit eines lokalen Alignments vorteilhaft sein.

¹ <https://www.uniprot.org/uniprot/P69905#sequences>

² <https://www.uniprot.org/uniprot/P68871#sequences>

Aufgabe 4:

(1) Globales Alignment mit voreingestellten Parametern

Das globale Alignment wurde erzeugt mit https://www.ebi.ac.uk/Tools/psa/emboss_needle/ und den Voreinstellungen lieferte:

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001          EMBOSS_001      1 MV-LSPADKTNVKAANGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D  48
# 2: EMBOSS_001          EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD  48
# 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
#
#=====
```

Abbildung 3: Globales Alignment von HBA1 und HBB mit unveränderten Voreinstellungen, Score 292.5

(2) Globales Alignment mit einer anderen Substitution MATRIX

(b) Während des Alignments sind 3 Fälle möglich: matches, mismatches und gaps (siehe GAP OPEN Penalty). Um das Alignment mit einem Score zu bewerten, werden den drei möglichen Ereignissen Score-Werte zugewiesen, die wiederum einer Substitution Matrix entnommen werden. Die Werte innerhalb der Matrix basieren auf den Frequenzen mit der die verschiedenen Aminosäuren in lokalen Alignments ausgetauscht wurden. So ergibt sich auch, dass nicht alle Matches bzw. Mismatches gleichwertig sind. Die Zahl 62 in BLOSUM62 bezieht sich darauf, dass zur Berechnung der Matrix Sequenzen verwendet wurden, die einen similarity-Wert von unter oder gleich 62% aufwiesen³.

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001          EMBOSS_001      1 MV-LSPADKTNVKAANGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D  48
# 2: EMBOSS_001          EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD  48
# Matrix: EPAM60
# 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: 295.5
#
#=====
```

Abbildung 4: Globales Alignment von HBA1 und HBB unter Verwendung der Substitution Matrix EPAM60, Score 295.5

³ <https://en.wikipedia.org/wiki/BLOSUM#Terminology>

(3) Globales Alignment mit einer GAP OPEN Penalty

(b) Während des Alignments können Gaps in Form von Insertionen bzw. Deletionen eingefügt werden. Würden diese nicht mit einer Strafe belegt werden, könnten auch sehr unterschiedliche Sequenzen durch eine Vielzahl von Indels zu 100% aligned werden, was nicht zielführend ist. Daher wird das Einführen von Indels mit einer Strafe, der Gap-Penalty behaftet. Diese kann konstant sein, kann jedoch auch für spezifische Sequenzbereiche, die beispielsweise hochkonservativ sind, variieren. Wie erwartet und in Abbildung 5 zusehen ist, hat die verdoppelte Gap-Penalty zu einer Reduzierung des Score-Wertes geführt und auch den similarity-Wert leicht verringert.

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001          EMBOSS_001      1 -MVLSPADKTNVKAANGKVGAGHAGEYGAELERMFLSFPTTKTYFPHF-- 47
# 2: EMBOSS_001          EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD 48
# Matrix: EBLOSUM62      EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD 48
# Gap_penalty: 20.0      EMBOSS_001      48 ----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLK 93
# Extend_penalty: 0.5    EMBOSS_001      49 LSTPDVAMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTATLSELHCDKLH 98
#
# Length: 149           EMBOSS_001      94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR 142
# Identity: 61/149 (40.9%) EMBOSS_001      99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH 147
# Similarity: 87/149 (58.4%)
# Gaps: 9/149 ( 6.0%)
# Score: 270.0
#
#=====
```

Abbildung 5: Globales Alignment von HBA1 und HBB mit einer OPEN GAP PENALTY von 20 (gegenüber 10), Score 270

(4) Lokales Alignment mit voreingestellten Parametern:

In Abbildung 6 wurde ein Smith-Waterman Algorithmus zur Erstellung einer lokalen Alignments verwendet.⁴ Da die betrachtete Sequenz ohnehin kurz ist unterscheidet sich der similarity- und Score-Wert nur sehr geringfügig zu Abbildung 3.

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001          EMBOSS_001      3 LSPADKTNVKAANGKVGAGHAGEYGAELERMFLSFPTTKTYFPHF-DLS- 50
# 2: EMBOSS_001          EMBOSS_001      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST 51
# Matrix: EBLOSUM62      EMBOSS_001      51 ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLKRVDP 96
# Gap_penalty: 10.0      EMBOSS_001      52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTATLSELHCDKLHVDP 101
# Extend_penalty: 0.5    EMBOSS_001      97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKY 141
#
# Length: 145           EMBOSS_001      102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH 146
# Identity: 63/145 (43.4%)
# Similarity: 88/145 (60.7%)
# Gaps: 8/145 ( 5.5%)
# Score: 293.5
#
#=====
```

Abbildung 6: Lokales Alignment von HBA1 und HBB mit Voreinstellungen, Score 293.5

⁴ https://www.ebi.ac.uk/Tools/psa/emboss_water/

(a) Vergleich mit Alignment aus der Vorlesung Folie 11 Woche 12:

```

HBA_HUMAN  -----VLSPADKTNVKAAWGKVGA--HAGEYGAEALERMFSLFPTTKTYFPHF
HBB_HUMAN  -----VHLTPPEEKSAVTALWGKV---NVDEVGGEALGRLLVVYPWTQRFFESF
HBA_HUMAN  -DLS-----HGSAQVKGHGKKVADALTNVAHV---D--DMPNALSALSDLHAHKL-
HBB_HUMAN  GDLSTPDVAVMGNPKVKAHGKKVLGAFSDGLAH---D--NLKGTFFATLSELHCDKL-
HBA_HUMAN  -FVDFVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR-----
HBB_HUMAN  -HVDENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH-----

```

Abbildung 7: Alignment aus der Vorlesung

Die Alignments sind aus Abb. 3 und Abbildung 7 sind sich sehr ähnlich, wobei der Start (grün markiert) in Abb. 7 durch eine Deletion und das fehlende Methionin abweicht. Zu Beginn der 2. Zeile enthalten beide Alignments 5 Indels (violett markiert). Es wurden einige Bereiche von mehr als drei aufeinanderfolgenden Aminosäure-Alignments rot markiert. In dem Alignment der Vorlesung existiert (blau markiert) eine Vielzahl von Indels, die nur im Kontext der weiteren Sequenzen notwendig ist und daher in Abb. 1 fehlt.