

„BIOINFO-Augabe6“

Aufgabe 2) Protein-Sequenzen von:

„Human Hemoglobin subunit alpha“ (HBA_HUMAN) [UniProtKB - P69905 (HBA_HUMAN)]

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG
KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP
AVHASLDKFLASVSTVLTSKYR

„Human Hemoglobin subunit beta“ (HBA_HUMAN) [UniProtKB - P68871 (HBB_HUMAN)]

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPK
VKAHGKKVLGAFSDGLAHLDDLKGTFAITLSELDKLVHDPENFRLLGNVLVCVLAHHFG
KEFTPPVQAAYQKVVAGVANALAHKYH

Aufgabe 3) [<https://www.majordifferences.com/2016/05/difference-between-global-and-local.html>, vom 10.07.18]

„Globales Alignment“:

Bei dieser Form des Alignments wird versucht die gesamte Sequenz anzugleichen („end-to-end alignment“). Es beinhaltet die gesamten Sequenzen, beider Vergleichssequenzen. Zwei Sequenzen sind für das globale Alignment geeignet, wenn sie ungefähr die gleiche Länge haben und sich relativ ähnlich sind. Ein globales Alignment kommt dann zum Einsatz, wenn zwei homologe Gene oder Protein mit ähnlicher Funktion verglichen werden sollen.

„Lokales Alignment“:

Hier sollen lokale Regionen (also solche Sequenzabschnitte mit begrenzter Größe) in zwei Sequenzabschnitten gefunden werden, die sich möglichst ähnlich sehen. Es wird versucht ein Alignment mit je einem „Sub-String“ (kleineren Sequenzabschnitts) von zwei Sequenzen zu erreichen. Es eignen sich hierfür Sequenzen, die sich eher unähnlich sehen. Es kann sich um konservierte Domänen in einer DNA Sequenz oder um eine konservierte Protein-Domäne handeln.

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

=====

MBOSS_001      1 MV-LSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-D    48
                || |:|:|:|.|.||| | :.|.|.|||.|:|:|.|.|.|.|.|.|.|.|.
MBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD    48

MBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLR    93
                || .|:|:|:|.||| |.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
MBOSS_001     49 LSTPDAMGNPKVKAHGKKVLGAFSDGLAHLNLIKGTFFATLSELHCDKLH    98

MBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR    142
                |||.||:|:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
MBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH    147
```

(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

=====

MBOSS_001      1 MV-LSPADKTNVKAANGKV-----GAGHAGEYGAEALERM-----F    34
                || |:|:|:|.|.||| | | | | | | | | | | | | | | | | |
MBOSS_001      1 MVHLTPEEKSAVTALWGKVVNDEVGG-----EALGRLLVVYPWTQRF    42

MBOSS_001     35 LSFPTTKTYFPHF----DLSSHSAQ-----VKGHGKKV--A--DA    66
                | | | | | | | | | | | | | | | | | | | | | | | |
MBOSS_001     43 -----FESFGDLS-----TPDAMGNPKVKAHGKKVLGAFSDG    75

MBOSS_001     67 LTNAAHVDDMPN-----ALSALSDLHAHKLRVDPVNFKLLSH---CLLV    108
                | | | | | | | | | | | | | | | | | | | | | | | |
MBOSS_001     76 L----AHLN---NLKGTFA--TLSELHCDKLHVDPENFRLLGNVLVCVLA    115

MBOSS_001    109 TLAHLPA-----EFTPAVHASLDKFLASVSTVLTSKYR-----    142
                || | | | | | | | | | | | | | | | | | | | | | |
MBOSS_001    116 ---AH---HFGKEFTPPVQA-----A-----YQKVVAGVANALAH    144

MBOSS_001    143 ---    142

MBOSS_001    145 KYH    147
```

(3) Globales Alignment mit einer anderen GAP OPEN penalty

```
#=====
#
# 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-LSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-D | 48 |
| | | : : : : : . . : . . . : : : : : : : : . . | |
| EMBOSS_001 | 1 | MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD | 48 |
| EMBOSS_001 | 49 | LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSOLHAHKLRL | 93 |
| | | : : : : : . . : : : : : : : : . . | |
| EMBOSS_001 | 49 | LSTPDAMGNPKVKAHGKKVLGAFSDGLAHLNKLKGTATLSELHCDKLH | 98 |
| EMBOSS_001 | 94 | VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR | 142 |
| | | : : : : : . . . : : . : : : . . . | |
| EMBOSS_001 | 99 | VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH | 147 |

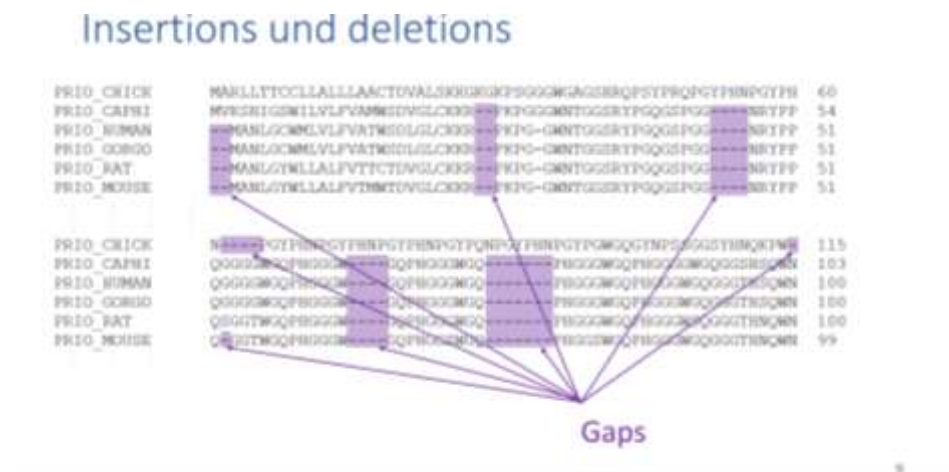
(4) Lokales Alignment mit voreingestellten Parametern

```
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 14
# Extend_penalty: 4
#
# Length: 145
# Identity:      63/145 (43.4%)
# Similarity:    88/145 (60.7%)
# Gaps:          8/145 ( 5.5%)
# Score: 264
#
#=====
```

| | | | |
|------------|-----|---|-----|
| EMBOSS_001 | 3 | LSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-DLSH | 51 |
| | | : : : : . . . : . . . : : : : : : : . . . | |
| EMBOSS_001 | 4 | LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST | 51 |
| EMBOSS_001 | 52 | -----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSOLHAHKLRVDP | 96 |
| | | : : . . . : : : : : : : : . | |
| EMBOSS_001 | 52 | PDAMGNPKVKAHGKKVLGAFSDGLAHLNKLKGTATLSELHCDKLHVDP | 101 |
| EMBOSS_001 | 97 | VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR | 141 |
| | | . : : . : . . . : . . : . : . . . | |
| EMBOSS_001 | 102 | ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH | 146 |

Vergleich mit der Sequenz aus der Vorlesung (Woche-11, S.9, S.10)

-Im Vergleich mit dem Alignment aus der Vorlesung lässt sich sagen, dass das Pairwise Alignment von EBI nicht nur erkennt, ob eine Aminosäure an einer betreffenden Stelle identisch ist oder nicht. Es gibt zusätzlich auch noch die Möglichkeit ähnliche Aminosäuren mit einem „:“ zu versehen und weiter entfernt ähnliche Aminosäuren mit einem „.“ zu versehen. Wie auch bei der Vorlesungsfolie 9 werden ebenso Gaps berücksichtigt und mit einem „-“ anstatt mit einer Aminosäure ersetzt.



(b) je Ergebnis die benutzten Parameter (Substitution Matrix und Gap Penalty) erklärt werden (je 1 Punkt). Z.B. soll kurz (!) erklärt werden, was die BLOSUM62 Matrix enthält.

(1)

-Substitutions Matrix Blosum62: - eigentlich: BLOcks SUBstitutions Matrix

- evidenzbasierte Substitutionsmatrix für Sequenzalignment von Proteinen
- Verwendung von einzelnen Blöcken innerhalb der Sequenzen von homologen Proteinen
- zur Nummer: -hohe Nummer: eher für evolutionär nahe verwandte Proteine geeignet; tiefe Nummern: eher für divergierende Proteine geeignet (62 ist hier relativ hohe Nummer)
- Anfertigung von log-odds Werten: diese geben die Wahrscheinlichkeit an mit der eine Aminosäure im Alignment vorkommt und mit der eine Aminosäure überhaupt vorkommt, auch die Mutationsrate einer Aminosäure kann damit angegeben werden

(<https://de.wikipedia.org/wiki/BLOSUM>, 10.07.18)

-Gap Penalty 10:

- Eine Gap Penalty macht ein flexibleres Alignment möglich als das ohne die Verwendung von Lücken möglich wäre. Zugleich verhindert eine Gap Penalty, dass das Alignment, durch die Verwendung von zu vielen Lücken, seine Aussagekraft verliert. Denn jede Lücke die für das Alignment verwendet wird, wird mit einem Wert versehen, der sich negativ auf die Beurteilung des Gesamtalignments auswirkt. Man tendiert also eher dazu wenige „Gaps“ zu nutzen. Durch den Wert der Gap Penalty kann jedoch bestimmt werden, wie viel eine „Gap“ kostet.

(https://en.wikipedia.org/wiki/Gap_penalty, vom 10.07.18)

(2)

- Substitutions Matrix Pam10:

- PAM: Point Accepted Mutation Matrix
- auch eine Substitutionsmatrix, um die Wahrscheinlichkeit zu berechnen mit der eine Veränderung einer Sequenz auftritt
- notwendig ist hierfür statistische Werte über Sequenzunterschiede
- erstellt, um festzuhalten mit welcher Wahrscheinlichkeit eine Aminosäure zu einer anderen Aminosäure mutiert; auch hier werden odd-logs-Werte verwendet

(https://de.wikipedia.org/wiki/Point_Accepted_Mutation_Matrix, 10.07.18)

-Gap Penalty10: (siehe(1))

(3)

-Substitutions Matrix Blosum62: (siehe (1))

-Gap Penalty10: (siehe(1))

(4)

-Substitutions Matrix Blosum62: (siehe(1))

-Gap Penalty14: (siehe(1))