

## Aufgabe 2

Protein-Sequenz „Human Hemoglobin subunit alpha“:

MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHFDLSHGSAQVKGHG  
KKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDPVNFKLLSHCLLVTLAAHLPAEFTP AVHASLDK-  
FLASVSTVLTSKYR

Protein-Sequenz „Human Hemoglobin subunit beta“:

MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLSTPDAMGNPK  
VKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVDPENFRLLGNVLVCVLAHHFG KEFTPP-  
VQAAYQKVVAGVANALAHKYH

## Aufgabe 3

„Globales Alignment“ und „Lokales Alignment“ unterscheiden sich darin, dass verschiedene Teile von Sequenzen betrachtet werden. Beim globalen Alignment wird die komplette Sequenz betrachtet und verglichen, wohingegen beim lokalen Alignment lediglich ein Teil der Sequenz betrachtet und verglichen wird. Beim globalen Alignment müssen die Sequenzen eine hohe Sequenzhomologie aufweisen und in etwa die gleiche Länge besitzen, um diese vergleichen zu können. Beim lokalen Alignment muss dies nicht der Fall sein.

### 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-LSPADKTNVKAANGKVGAGHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                  || |:|.:.|:.|.|.||| | :..|.|.|||.|:..:.|:.|:.|..| |
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKHGKGVADALTNAAVHAVDDMPNALSALSDLHAHKLR      93
                  ||      .|:..:| |.|||| |..|.:.:.| |:|:|:.....| |:| |..| |.
EMBOSS_001     49 LSTPDAMVGNPKVKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH      98

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

Abbildung 1: Globales Alignment mit BLOSUM62.

BLOSUM62: Basiert auf Vergleichen von Abschnitten von Sequenzen, die aus Datenbanken stammen. Die Datenbanken enthalten viele Abschnitte ohne Gaps, welche den hoch konservierten Bereichen in Proteinen entsprechen. BLOSUM62 hat eine über 62%ige Übereinstimmung mit den Sequenz Alignments ohne Gaps. Diese Matrix wird vornehmlich für Vergleiche von Proteinsequenzen verwendet.<sup>1</sup>

<sup>1</sup> <http://www.cs.columbia.edu/4761/assignments/assignment1/reference1.pdf> (10.07.2018, 15:04 Uhr).

## (2) Globales Alignment mit einer anderen Substitution MATRIX

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM35
# Gap_penalty: 10.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      65/149 (43.6%)
# Similarity:    97/149 (65.1%)
# Gaps:          9/149 ( 6.0%)
# Score: 398.5
#
#=====
```

EMBOSS_001	1	MV-LSPADKTNVKAANGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D	48
		:   . :   . :   . :       : . .   . :       . : : : :   . : . :   . :	
EMBOSS_001	1	MVHLTPEEKSAVTALWGKVNVDD--EVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOSS_001	49	LS-----HGSAQVKGHGKKVADALTNAAHVDDMPNALSALSDLHAHKLR	93
		:   . . :   . :         . : : : :   : : : : : : :   :     . :   .	
EMBOSS_001	49	LSTPDVAVMGNPVKVKAHGKKVLGAFSDGLAHLDNLKGTFFATLSELHCDKLH	98
EMBOSS_001	94	VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR	142
		. :   : :   . : :     . : . :       . : . : : :   : : : :   . :   .	
EMBOSS_001	99	VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH	147

Abbildung 2: Globales Alignment mit BLOSUM35.

BLOSUM35: BLOSUM35 hat eine über 35%ige Übereinstimmung mit den Sequenz Alignments ohne Gaps.

### (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 -MVLSPADKTNVKAANGKVGGAHAGEYGAEALERMFSLFPTTKTYFPHF--      47
                  :.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     48 ----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDDLHAHKLR      93
                  |...|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|.:|
EMBOSS_001     49 LSTPDAMGPNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTLSELHCDKLH      98

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

Abbildung 3: Globales Alignment mit GAP OPEN 20.

GAP OPEN 20: Gap penalty ist ein Scoring Alignment von zwei Sequenzen, bei welchem durch Einfügen von Gaps mehr Matches erzielt werden können. Je höher das Gap penalty, desto weniger Gaps werden eingefügt, damit der Score nicht zu niedrig wird. Dadurch ist der Score in Abbildung 3 erheblich schlechter als in Abbildung 2.

#### (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 LSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLS-      50
                  |:|:|:|.|.|.|||   :..|.|.|||.|:~::~|.|:~::~|..| |||
EMBOSS_001      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST      51
EMBOSS_001      51 ----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP      96
                  .|:~::~|.|||||.|.::~:|:|:~::~:~::~:|:|:~::~|..|.|||
EMBOSS_001      52 PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVP      101
EMBOSS_001      97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY      141
                  .|:|:|:|.|.|.|||.|||.|.|.::~:~::~:~::~:~::~:~::~:~::~|..|
EMBOSS_001     102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146
```

Abbildung 4: Lokales Alignment mit BLOSUM62.

Beim lokalen Alignment werden nur Teile der Sequenzen betrachtet, wodurch es vergleichsweise zu mehr Matches kommt.

Vergleicht man die Matrizen mit denen aus der Vorlesungsfolie 9, so stellt man fest, dass durch Einfügen von Gaps eine höhere Übereinstimmung erzielt werden kann.