

Aufgabe 2

„Human Hemoglobin subunit alpha“ (HBA_HUMAN)

```
MVLSPADKTN VKAAWGKVG A HAGEYGAEALERMFLSFPTTKTYFPHFDLS
HGSAQVKGHG KKVADALTNA VAHVDDMPNALSALSDLHAH KLRVDPVNFK
LLSHCLLVTLAAHLPAEFTP AVHASLDKFLASVSTVLTSKYR
```

„Human Hemoglobin subunit beta“ (HBB_HUMAN)

```
MVHLTPEEKSAVTALWGKVNVDEVGGEALGRLLVVYPWTQRFFESFGDLS
TPDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLHVD
PENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH
```

Aufgabe 3

Globales Alignment: Zwei komplette Sequenzen werden betrachtet, meist bei ähnlichen Sequenzen mit ähnlicher Länge

Lokales Alignment: Betrachtung von zwei Teilsequenzen, diese werden durch Betrachtung und Vergleichen des Scores gefunden, da dieser im optimalen Fall maximal ist.

Aufgabe 4

1) Globales Alignment mit voreingestellten Parametern

```
EMBOSS_001      1 MV-LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                  || |:|.:|:|.|.|.||| |.:|.|.|||.|:|:|.|.|.|.|.|.|.
EMBOSS_001      1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001     49 LS-----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKL      93
                  || |.:|:|.|.|||||.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOSS_001     49 LSTPDVAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFATLSELHCDKLH      98

EMBOSS_001     94 VDPVNFKLLSHCLLVTLAAHLPAEFTP AVHASLDKFLASVSTVLTSKYR      142
                  |||.||:|.|:|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.|.
EMBOSS_001     99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH      147
```

Aligned_sequences: 2

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

2) Globales Alignment mit einer anderen Matrix (EBLOSUMN)

EMBOSS_001	1	MV-LSPADKTNVKAAWGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D	48
		: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:	
EMBOSS_001	1	MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOSS_001	49	LS-----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLR	93
		. .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:	
EMBOSS_001	49	LSTPDVAVMGNPVKVKAHGKKVLGAFSDGLAHLNLTGTFATLSELHCDKLH	98
EMBOSS_001	94	VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR	142
		. : .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .	
EMBOSS_001	99	VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH	147

Aligned_sequences: 2

Matrix: EBLOSUMN

Gap_penalty: 10.0

Extend_penalty: 0.5

Length: 149

Identity: 65/149 (43.6%)

Similarity: 82/149 (55.0%)

Gaps: 9/149 (6.0%)

Score: 297.5

3) Globales Alignment mit einer anderen GAP OPEN penalty (5)

EMBOSS_001	1	MV-LSPADKTNVKAAWGKVGAGAHAGEYGAEALERMFLSFPTTKTYFPHF-D	48
		: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:	
EMBOSS_001	1	MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD	48
EMBOSS_001	49	LS-----HGSAQVKGHGKKVADALTNVAHVDDMPNALSALSDLHAHKLR	93
		. .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .:	
EMBOSS_001	49	LSTPDVAVMGNPVKVKAHGKKVLGAFSDGLAHLNLTGTFATLSELHCDKLH	98
EMBOSS_001	94	VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTISKYR	142
		. : .: .: .: .: .: .: .: .: .: .: .: .: .: .: .: .	
EMBOSS_001	99	VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH	147

Aligned_sequences: 2

Matrix: EBLOSUM62

Gap_penalty: 5.0

Extend_penalty: 0.5

Length: 149

Identity: 65/149 (43.6%)

Similarity: 90/149 (60.4%)

Gaps: 9/149 (6.0%)

Score: 312.5

4) Lokales Alignment mit voreingestellten Parametern

EMBOSS_001	3	LSPADKTNVKAAWGKVGAHAGEYGAELERMFLSFPTTKTYFPHF-DLS-	50
		: :	
EMBOSS_001	4	LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST	51
EMBOSS_001	51	----HGSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP	96
		: :	
EMBOSS_001	52	PDAVMGNPKVKAHGKKVLGAFSDGLAHLDNLKGTFTATSELHCDKLHVDP	101
EMBOSS_001	97	VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKY	141
		: :	
EMBOSS_001	102	ENFRLLGNVLVLCVLAHHFGKEFTPPVQAAQKVVAGVANALAHKY	146

Aligned_sequences: 2

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

PRIO_CHICK	MARLLTTCCLLALLLAACTDVALSKKGKGPSSGGWAGSHRQPSYPRQPGYPHNPGYPH	60
PRIO_CAPHI	MVKSHIGSWILVLFVAMWSDVGLCKKR--PKPGGGWNTGGSRYPGQGSPPG----NRYPP	54
PRIO_HUMAN	--MANLGCWMLVLFVATWSDLG LCKKR--PKPG-GWNTGGSRYPGQGSPPG----NRYPP	51
PRIO_GORGO	--MANLGCWMLVLFVATWSDLG LCKKR--PKPG-GWNTGGSRYPGQGSPPG----NRYPP	51
PRIO_RAT	--MANLGYWLLALFVTTC T D VGLCKKR--PKPG-GWNTGGSRYPGQGSPPG----NRYPP	51
PRIO_MOUSE	--MANLGYWLLALFVTMTWTDVGLCKKR--PKPG-GWNTGGSRYPGQGSPPG----NRYPP	51
PRIO_CHICK	N----PGYPHNPGYPHNPGYPHNPGYPQNPGYPHNPGYPGWGQGYNPSSGGSYHNQKPW-	115
PRIO_CAPHI	QGGGGWGQPHGGGW----GQPHGGGWGQ-----PHGGGWGQPHGGGGWGQGGGSHSQWN	103
PRIO_HUMAN	QGGGGWGQPHGGGW----GQPHGGGWGQ-----PHGGGWGQPHGGGGWGQGGGTHSQWN	100
PRIO_GORGO	QGGGGWGQPHGGGW----GQPHGGGWGQ-----PHGGGWGQPHGGGGWGQGGGTHSQWN	100
PRIO_RAT	QSGGTWGP HGGGW----GQPHGGGWGQ-----PHGGGWGQPHGGGSQGGGTHSQWN	100
PRIO_MOUSE	Q-GGTWGP HGGGW----GQPHGGGWGQ-----PHGGGWGQPHGGGGWGQGGGTHNQWN	99

Alignment aus der Vorlesung (1.7.2018, Folie 9)

Auf der Folie wurden mehrere Sequenzen miteinander verglichen, nicht wie in den vorherigen Aufgaben jeweils 2. Zu sehen ist, dass in der Vorlesungsfolie eindeutig mehr Abschnitte der Sequenzen übereinstimmen und sich diese dementsprechend ähnlicher sind (mit Ausnahme der ersten Sequenz, PRIO_CHICK). Diese Ähnlichkeiten sind vermutlich dadurch zustande gekommen, indem viele Gaps eingefügt wurden, welche in den erstellten Alignments kaum vorhanden sind.

Substitutionsmatrizen: Bewertet individuelle Substitutionen, wird genutzt, um einem Alignment einen Score zuzuordnen und das Alignment so bewertet.

Häufig verwendete Substitutionsmatrizen sind BLOSUM und PAM.

BLOSUM (Blocks Substitution Matrix): basiert auf homologen Proteinsequenzen, deren Alignments evolutionär weit entfernt sind.

BLOSUM62: Alle Sequenzen in der Matrix werden mit mehr als 62% Sequenzidentität zu einer Sequenz zusammengefasst.

PAM (Point-Accepted Mutations): Die Mutationswahrscheinlichkeit ist unabhängig von der Position in der Sequenz. Die Matrix berechnet sich aus der absoluten Häufigkeit aller Aminosäuren über alle Sequenzen, welche auf die Gesamtlänge aller Sequenzen normiert werden.

Gap Penalty: Wird verwendet, um die Anzahl der Gaps minimal zu halten und den Alignment Score anzupassen basierend auf der Zahl und Länge der Gaps, um noch ein nützliches Alignment zu erhalten.

Quelle:

https://www.informatik.hu-berlin.de/de/forschung/gebiete/wbi/teaching/archive/ws0506/bioinformatik-bpi/05_pam.pdf
(10.7.2018, 11:43)