

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

Human Hemoglobin subunit alpha

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

Human Hemoglobin subunit beta

### Aufgabe 3

Globale Alignment erzeugt eine Ende-zu-Ende-Ausrichtung der Sequenzen, die ausgerichtet werden sollen, wohingegen lokale Alignment eine oder mehrere Alignments findet, die die am ähnlichsten Regionen innerhalb der Sequenzen beschreiben, die ausgerichtet werden sollen.

## Global Vs Local Alignment

5

### • Global Alignment

- A general **global alignment** technique is the Needleman–Wunsch algorithm, which is based on dynamic programming.
- Attempts to align the maximum of the entire sequence
- Suitable for similar and equal length sequences

### • Local Alignment

- **Local** alignments are more useful for dissimilar sequences that are suspected to contain regions of similarity or similar sequence motifs within their larger sequence context.
- Stretches of sequences with highest density of matches are aligned
- Suitable for partially similar, different length and conserved region containing sequences



## Local Alignment

Target Sequence

5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'

### Query Sequence

Query Sequence 5' TACTCACGGATGAGGTACTTTAGAGGC 3'

## Global Alignment

### Target Sequence

5' ACTACTAGATTACTTACGGATCAGGTACTTTAGAGGCTTGCAACCA 3'

5' ACTACTAGATT---ACGGATC--GTACTTTAGAGGCTAGCAACCA 3'

### Query Sequence

### Aufgabe 4

```
#=====
#
# 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-LSPADKTNVKAAWGKVGGAHAGEYGAEALERMFLSFPTTKTYFPHF-D      48
                  || |:|:|:|:|:|.||| | :..|.|.|||.|:|:|:|.|:|:|:|.|| |
EMBOSS_001      1 MVHLTPPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGD      48

EMBOSS_001      49 LS-----HGSAQVKGHGKKVADALTNAAVAHVDDMPNALSALSDLHAHKLR      93
                  ||      .|:|:|:|.||| |.|.|:|:|:|:|:|:|:|:|:|:|:|.|||.
EMBOSS_001      49 LSTPDAMVGNPKVKKAHGKKVLGAFSDGLAHLNLTGTFATLSELHCDKLH      98

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

Das Ergebnis hier im Vergleich mit dem in der Vorlesung vorhandenen Alignment zeigt, dass die Matches hier geringer und somit die Ähnlichkeit auch niedriger ist.

Eine Substitutionsmatrix beschreibt die Rate, mit der sich ein Zeichen in einer Sequenz im Laufe der Zeit in andere Zeichenzustände ändert.

BLOSUM verwendet einzelne Blöcke (ohne Lücken) innerhalb der Sequenzen von homologen Proteinen, die verglichen wurden. (quelle Wikipedia)

Wobei **BLOSUM62** Vergleichen von Sequenzen mit einer paarweisen Identität von nicht mehr als 62 berechnet.

2)

```
#=====
#
# 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
#
#=====

EMB OSS_001      1 MV-LSPADKTNVKA AWGKV-----GAHAGEYGA EALERM-----F
34
      |||.|.|.|.|.|.|.|||      |      |||.|.      |
EMB OSS_001      1 MVHLTPEEKSAVTALWGKVN VDEVGG-----EALGRLLVVYPWTQRF
42
EMB OSS_001      35 LSFPTTKTYFPHF----DLSHGSAQ-----VKGHGKKV--A--DA
66
      |      |||      |||.||||      |      |.
EMB OSS_001      43 -----FESFGDLS-----TPDAVMGNPKVKAHGKKVLGAFSDG
75
EMB OSS_001      67 LTNAVAHVDDMPN-----ALSALSDLHA HKLRVDPVNFKLLSH---CLLV
108
      |      ||.|      |      |      .|.|.|.|.|.|.|.|.|.|.|.      |.
EMB OSS_001      76 L----AHL D---NLKGTFA--TLSELHCDKLHVDPENFRLLGNVLVCVL-
115
EMB OSS_001      109 TLA AHPA----EFTP AVHASLDKFLASVSTVLTSKYR-----
142
      ||      |||||.|.      |      |.
EMB OSS_001      116 ---AH---HFGKEFTP PVQA-----A-----YQKV VAGVANALAH
144
EMB OSS_001      143 ---      142
EMB OSS_001      145 KYH      147
```

Das Ergebnis hier im Vergleich mit dem in der Vorlesung vorhandenen Alignment zeigt, dass die Matches hier geringer und somit die Ähnlichkeit auch niedriger ist,

Die Rate der Caps so ist hoch

3)

```
#=====
#
# Aligned_sequences: 2
# 1: EMBOSS_001
# 2: EMBOSS_001
# Matrix: EBLOSUM62
# Gap_penalty: 25.0
# Extend_penalty: 0.5
#
# Length: 149
# Identity:      61/149 (40.9%)
# Similarity:    87/149 (58.4%)
# Gaps:          9/149 ( 6.0%)
# Score: 260.0
#
#
#=====

EMBOSS_001           1 -MVLSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF--
47

                        :.:|.:|:.|.|.||||   :..|.|.|||.|:~::~|.|:~::|.|.
EMBOSS_001           1 MVHLTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTVRFFESFGD
48

EMBOSS_001           48 ----DLSHGSAQVKGHGKKVADALTNAVAHVDDMPNALSALS~LHAHKLR
93

                        |...|:~::~|.|||||.|.~::~|:|:~::.....|:|:|..|.|.
EMBOSS_001           49 LSTPDAVMGNPKVKAHGKKV~LGAFSDGLAHL~DNLKGTFATLSELHCDKLH
98

EMBOSS_001           94 VDPVNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLTSKYR
142

                        |||.||:|.|:~::~|.|.|.~::~|.|.~::~|.|.~::~|.|.|.
EMBOSS_001           99 VDPENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKYH
147
```

Das Ergebnis hier im Vergleich mit dem in der Vorlesung vorhandenen Alignment zeigt, dass die Matches hier geringer und somit die Ähnlichkeit auch niedriger ist.  
Mit wenigem Zahl an Caps

## Cap Opening 25:

Gap Opening hat den Nachteil, dass es eine Lücke in der Ausrichtung öffnet, wodurch die Lücken für zunehmende Lückenöffnungswerte weniger häufig werden, und somit wurde im Vergleich zu 1 vom Score abgezogen

4)

```
#####  
# Program: matcher  
# Rundate: Tue 10 Jul 2018 19:13:57  
# Commandline: matcher  
# -auto  
# -stdout  
# -asequence emboss_matcher-E20180710-191356-0079-46001268-plm.asequence  
# -bsequence emboss_matcher-E20180710-191356-0079-46001268-plm.bsequence  
# -datafile EBLOSUM62
```

```

# -gapopen 14
# -gapextend 4
# -alternatives 1
# -aformat3 pair
# -sprotein1
# -sprotein2
# Align_format: pair
# Report_file: stdout
#####

#=====
#
# 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 LSPADKTNVKAAWGKVGAHAGEYGAEALERMFLSFPTTKTYFPHF-DLSH
51
      |:|.:|:|.|.||| |.:|.|.|||.|:~.:|.|:~.:|.|||.
EMBOSS_001      4 LTPEEKSAVTALWGKV--NVDEVGGEALGRLLVVYPWTQRFFESFGDLST
51
EMBOSS_001      52 -----GSAQVKGHGKKVADALTNAVAHVDDMPNALSALSDLHAHKLRVDP
96
      |:~.:|.|.|||||.|.~.:~.:|:|:~.:~.:~.:~.:|:|.|.||||
EMBOSS_001      52 PDAVMGNPKVKAHGKKVLGAFSDGLAHL DNLKGT FATLSELHCDKLHVDP
101
EMBOSS_001      97 VNFKLLSHCLLVTLAAHLPAEFTPAVHASLDKFLASVSTVLT SKY      141
      .|:|:|:~.:|:~.:|:~.:|:~.:|:~.:|:~.:|:~.:|:~.:|:~.:|
EMBOSS_001     102 ENFRLLGNVLVCVLAHHFGKEFTPPVQAAYQKVVAGVANALAHKY      146

```