

# Übungen zur Vorlesung Sequenzanalyse I

Universität Bielefeld, Wintersemester 2010/2011  
Dipl.-Inform Peter Husemann · Dr. Roland Wittler

<http://wiki.techfak.uni-bielefeld.de/gi/Teaching/2010winter/SequenzAnalyse>

**Blatt 8 vom 03.12.2010**

**Abgabe in einer Woche vor Beginn der Vorlesung.**

## Aufgabe 1 Scorematrizen

(3 Punkte)

Informiere dich eigenständig über die PAM( $t$ ) und BLOSUM( $s$ ) Scorematrizen für Proteinsequenzen.

1. Fasse in eigenen Worten zusammen, wie diese Scorematrizen aus Proteinsequenzen erstellt werden können.
2. Beschreibe im Detail, wofür die Parameter  $t$  und  $s$  stehen.
3. Zähle die Unterschiede von PAM und BLOSSUM auf.

## Aufgabe 2 Substitutions-Scorematrix

(4 Punkte)

Gegeben seien folgende Proteinsequenzen mit dem Alphabet  $\Sigma = \{\text{A, L, M, W}\}$ :

ALLMM  
MALLW  
AALAW  
AALAM  
ALAAM

Bestimme für diese Sequenzen die

1. Observationsmatrix  $M^t$  (Verteile Pseudocounts wenn nötig)
2. Hintergrundverteilung  $\pi$ ,
3. Likelihood-Ratio  $\left(\frac{M^t(x,y)}{\pi_x \pi_y}\right)$ ,

um schließlich daraus die log-Likelihood Scorematrix  $\left(s^t(x,y)\right) := \left(\log_2\left(\frac{M^t(x,y)}{\pi_x \pi_y}\right)\right)$  zu berechnen.

## Aufgabe 3 Edit-Matrix, Edit-Sequenz und Alignment

(2 Punkte)

Die Edit-Matrix, optimale Edit-Sequenzen und Alignments von zwei Sequenzen hängen eng miteinander zusammen.

Schreibe die Edit-Matrix  $D(x,y)$  (mit Einheitskosten) für die Sequenzen  $x = \text{ANANAS}$  und  $y = \text{ANNASHASE}$  auf und zeichne einen optimalen Pfad in die Matrix.

1. Gib zu diesem optimalen Pfad die entsprechende (optimale) Edit-Sequenz an.
2. Gib das dazugehörige Alignment an.

## Aufgabe 4 Die Anzahl globaler Alignments

(3 Punkte)

Erstelle ein Programm in Java, welches

1.  $N(m,n)$  rekursiv berechnet,
2.  $N(n,n)$  approximativ berechnet,
3.  $N'(m,n)$  exakt berechnet und
4.  $N'(m,n)$  approximativ berechnet.

Gib die Ergebnisse für  $n = m \in \{1, 5, 10, 100\}$  grafisch an und vergleiche diese. Diskutiere den Unterschied zwischen  $N(m,n)$  und  $N'(m,n)$ .