

26. Oktober 2005

**Abgabe: 2. November 2005**

**Aufgabe 1:** Implementieren Sie den Smith-Waterman-Algorithmus. Erproben Sie mit simulierten und natürlichen Sequenzen in welchen Situationen Ihr Programm andere Ergebnisse liefert als BLAST, siehe <http://www.ncbi.nlm.nih.gov/BLAST/>. Variieren Sie bei BLAST auch den Parameter “Word Size”. (Score-Matrizen für Proteine findet man im Internet z.B. unter [www.es.emblnet.org/Services/ftp/databases/matrices/](http://www.es.emblnet.org/Services/ftp/databases/matrices/).)

**Aufgabe 2:** In manchen Sequenz-Datensätzen gibt es viele sehr kurze Gaps und einige sehr lange. Erläutern Sie, inwiefern dies durch die Gap-Penalty-Funktion

$$\gamma(g) = \min\{d_1 + e_1 \cdot (g - 1), d_2 + e_2 \cdot (g - 1)\}$$

mit  $d_2 < d_1$  und  $e_1 < e_2$  berücksichtigt wird und finden Sie einen Algorithmus, der ein optimales globales Alignment zweier Sequenzen der Längen  $n$  und  $m$  mit dieser Gap-Penalty in Zeit  $O(nm)$  berechnet.

**Aufgabe 3:** Seien  $X$  und  $Y$  Zufallsvariablen mit endlichem Wertebereich  $U \subset \mathbb{R}$ , seien  $a, b \in \mathbb{R}$ , seien  $V$  und  $W$  zufällige Vektoren mit endlichem Wertebereich  $S \subset \mathbb{R}^n$  und sei  $M$  eine  $n \times n$ -Matrix. Zeigen Sie:

- $\mathbb{E}(a \cdot X) = a \cdot \mathbb{E}X$
- $\mathbb{E}(a \cdot X + b \cdot Y) = a \cdot \mathbb{E}X + b \cdot \mathbb{E}Y$
- $\mathbb{E}(a \cdot V + b \cdot W) = a \cdot \mathbb{E}V + b \cdot \mathbb{E}W$
- $\mathbb{E}(M \cdot V) = M \cdot \mathbb{E}V$