Bioinformatik I, WS99/00
Uebungsblatt 3

Besprechung am 8.12.1999

Lokales Sequenz Alignment

Implementieren Sie den Needleman-Wunsch Sequenz Alignment Algorithmus für lokales Alignment. Ein lokales Alignment ist ein Alignment zwischen Subsequenzen der beiden Sequenzen. Es soll ein optimales solches Alignment bestimmt werden. Für das lokale Alignment wird hier nach folgendem Schema eine Ähnlichkeitsmatrix aufgebaut. a und b seien die zu alignierenden Sequenzen, ihre Längen seien n und m. Setze $D_{0,0} = 0$, $D_{i,0} = 0$ für $1 \leq i \leq n$ und $D_{0,j} = 0$ für $1 \leq j \leq m$. Dann berechne

$$D_{i,j} = \max\{0, D_{i-1,j-1} + s(a_i, b_i), D_{i-1,j} + g, D_{i,j-1} + g\}$$

für $1 < i \leq n, 1 < j \leq m$, wobei s eine Similarity-Matrix ist, die zwischen Match und Mismatch unterscheidet, d.h. z.B. s(x,x)=2 und s(x,y)=-1 für $x \neq y$. g ist eine Konstante für Gapkosten z.B. g=-2.

Um ein optimales Alignment zu bestimmen, beginnt man den Traceback mit einem maximalen Eintrag der Matrix D und verfolgt diesen zurück bis zu einem Eintrag mit Wert 0.