

Algorithmen und Datenstrukturen in der Bioinformatik

Drittes Übungsblatt WS 14/15

Abgabe Donnerstag 12:00

Niveau I

Aufgabe 1: Needleman-Wunsch Algorithmus für Protein-Sequenzen

Berechnen Sie für die Sequenzen $S = \text{LINK}$ (horizontal) und $T = \text{ALIGN}$ (vertikal) ein globales Alignment nach dem Algorithmus von Needleman-Wunsch.

- Suchen Sie im Internet nach der PAM120-Matrix (z.B. <http://www.behav.org/gene/bio-info/pam120.jpg>) und nehmen Sie zusätzlich eine Gap Penalty von -4 an.
- Berechnen Sie die DP-Matrix, und
- tragen Sie dabei jeweils *alle* Traceback-Kanten ein (also eventuell mehrere pro Zelle).
- Heben Sie *die beim Traceback benutzten* Kanten besonders hervor, und
- notieren Sie das resultierende *globale Alignment*.
- Recherchieren Sie, was die Zahlen in PAM120 und BLOSUM50 bedeuten.

Aufgabe 2: Varianten von Alignments

Finden Sie Beispiele für *global*, *overlap* und *local* alignments.

Niveau II

Aufgabe 3: Dynamic Programming

Der *längste gemeinsame Teilstring (LGT)* einer Menge von Strings sei definiert als die längste zusammenhängende Buchstabenfolge, die in jedem String aus der Menge auftaucht.

Beispiel: $S_1 = \text{MANGO}$ und $S_2 = \text{STANGENBOHNE}$ $S_3 = \text{MANGOLD}$ Der LGT dieser 3 Strings ist ANG mit der Länge 3.

- Lösen Sie das LGT-Problem für zwei Strings mit *dynamischer Programmierung*
- Wie lässt sich Ihr Algorithmus auf k Strings ausweiten?
- Welche Laufzeit hat Ihr Algorithmus?