

Übungen zur Vorlesung Grundlagen der Sequenzanalyse

Universität Bielefeld, WS 2006/07

Dr. Sven Rahmann · Dipl.-Bioinf. Katharina Jahn

<http://gi.cebitec.uni-bielefeld.de/teaching/2006winter/sequenzanalyse/>

Blatt 7 vom 30.11.2006

Abgabe am 07.11.2006 vor der Vorlesung um 8:30 in H3

Aufgabe 1 Zeichne schematisch den Alignmentgraphen für globales Alignment mit affinen Gapkosten. Hinweis: Statt eines Knotens (i, j) wird man analog zu den drei Matrizen S , H und V vermutlich jeweils drei Knoten $s_{i,j}$, $h_{i,j}$ und $v_{i,j}$ benötigen. Versuche, die Abhängigkeiten möglichst übersichtlich darzustellen.

Aufgabe 2 Wende Sellers' Algorithmus (ohne Verbesserung) an, um alle Vorkommen des Musters $P = \text{ABBB}$ im Text $T = \text{BBABBAAABABAABAAAB}$ mit bis zu $k = 1$ Fehler zu finden (unter Einheits-Edit-Kosten).

Aufgabe 3 Welche Einträge der Edit-Matrix müssen in der vorigen Aufgabe nicht berechnet werden, wenn man die verbesserte Variante (Ukkonen's cutoff-Algorithmus) verwendet? Markiere in jeder Spalte den letzten essentiellen Index.

Aufgabe 4 Schreibe ein Java-Programm zur Erzeugung von zufälligen Proteinsequenzen mit benutzerdefinierter Länge. Die Häufigkeiten der einzelnen Aminosäuren kann man sich aus UniProt / SwissProt besorgen. Diese Aufgabe wird nicht bewertet, ist aber für weitere Aufgaben (die bewertet werden) nötig!