

Prof. Dr. Knut Reinert
Rene Rahn
Kathrin Trappe
Kathleen Gallo
Thomas Krannich

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmen und Datenstrukturen in der Bioinformatik

Viertes Übungsblatt WS 12/13

Abgabe Freitag, 16.11., 15:00 Uhr

Name: _____ Übungsgruppe: A B C

Matrikelnummer: _____

Niveau I

Aufgabe 1: Internet-Übung

Berechnen Sie mit Hilfe eines Needleman Wunsch Web Applets die paarweise Ähnlichkeit der menschlichen Haemoglobin- α - und Haemoglobin- β -Untereinheit und der Haemoglobin- β -Untereinheit der Maus. Die Sequenzen finden sich bei UniProt unter folgenden IDs:

HBA_HUMAN HBB_HUMAN HBA_MOUSE

Erklären Sie kurz Ihre Ergebnisse.

Aufgabe 2: Sequence alignment

Gegeben seien die Sequenzen $S1=ATAC$ und $S2=AA$.

a)

Machen Sie ein globales Alignment mit folgenden affinen Gapkosten:
 $match=2$, $mismatch=-4$, $gap\ opening(d)=-6$, $gap\ extension(e)=-1$.

Verwenden Sie hierzu den GOTOH-Algorithmus aus der Vorlesung und notieren Sie ein optimales Alignment. Zur Intuition, argumentieren Sie, warum ein direktes matchen von beiden A's aus S2 mit den beiden A's aus S1 kein optimales Alingnment ergibt.

b)

Konstruieren Sie einen Algorithmus, der ein *overlap alignment* zweier Sequenzen erstellt.

Niveau II

Aufgabe 3: Suchen mit *lcp*-Tabelle

Beweisen Sie Theorem 2 aus dem Skript über exakte Stringsuche (`exact.pdf`):

Mithilfe der *lcp*-Werte braucht die Suche eines Patterns in einem Suffixarray höchstens $\mathcal{O}(m + \log n)$ Vergleiche und besitzt diese Laufzeit.

Hinweis: Nutzen Sie die Beobachtung, dass weder l noch r während der Suche jemals kleiner werden. Finden Sie eine Schranke für die Anzahl redundanter Vergleiche pro Iteration der Binärsuche.