

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

Institut für Informatik
AG Algorithmische Bioinformatik

Algorithmen und Datenstrukturen in der Bioinformatik

Sechstes Übungsblatt WS 12/13

Abgabe Freitag, 30.11., 15:00 Uhr

Name: _____ Übungsgruppe: A B C

Matrikelnummer: _____

Niveau I

Aufgabe 1: FastA

In dieser Aufgabe soll mit dem **FastA**-Algorithmus ein lokales Alignment der Sequenzen $S_1 = \text{IALEQIAQDI}$ and $S_2 = \text{TIALDIAWEADI}$ gefunden werden. Halten Sie sich dabei an das Skript und führen Sie die folgenden 6 Schritte aus:

- Erstellen Sie eine *Hash-Tabelle* für die Sequenz S_1 mit $ktup = 2$ als Parameter.
- Finden Sie mit Hilfe dieser Hash-Tabelle die *Hot-spots* in Sequenz S_2
- Tragen Sie jetzt diese Hot-spots in einen *Dot-plot* ein und finden Sie die *Diagonal runs*. Hierbei dürfen Diagonale Lücken der Länge 1 (entspricht einem Mismatch) zugelassen werden.
- Rescoring*: Bilden Sie den Score aller ihrer Diagonal runs gemäß der PAM 70 Scoringmatrix.

Aufgabe 2: The chaining problem

- Erzeugen Sie durch *topologische Sortierung* einen *DAG* aus den gefundenen Hot-spots der Aufgabe 1 (Es sollten 5 Runs mit Länge ≥ 2 entstehen).
- Modifizieren Sie diesen Graph, sodass der **DAG shortest paths**-Algorithmus anwendbar ist und lösen Sie anschließend das *chaining problem*.

Aufgabe 3: Verständnis

- a) An welchen Stellen des FastA-Algorithmus wird deutlich, dass er eine *Heuristik* ist? Unter welchen Umständen wird ein optimales lokales Alignment nicht gefunden?
- b) Wie wirken sich große und kleine k – *tup*-Werte auf *Sensitivität* und *Spezifität* des Algorithmus aus?