

Klausur zur Vorlesung „Softwarewerkzeuge der Bioinformatik“ WS11/12



Montag 20.02.2012

Dauer: 10:00 bis 12.00 Uhr = 120 Minuten

Erlaubte Hilfsmittel: keine

Maximale Punktzahl: 100. Ab 50 Punkten gilt die Klausur als bestanden

Name:

Vorname:

MatrikelNr:

Kennzeichnen Sie jedes zusätzliche mit der Klausur abgegebene Blatt mit Ihrem Namen und Vornamen.

Nr.	1	2	3	4	5	6	7	8	9	10	11	Σ
mögliche Punkte	6	6	6	12	10	8	8	14	10	10	10	100
Ihre Punkte												

Bereich I Sequenzanalyse

Aufgabe 1 (6 Punkte)

Definieren Sie die Begriffe *Sequenzidentität* und *Sequenzähnlichkeit*.

Sequenzidentität:

Sequenzähnlichkeit:

Wie hängt der Begriff *Homologie* mit einem oder beiden dieser Begriffe zusammen?

Aufgabe 2 (6 Punkte)

Vergleichen Sie die Stärken und Schwächen von Algorithmen für das paarweise Sequenzalignment:

dynamische Programmierung kontra heuristischer BLAST-Algorithmus

Aufgabe 3 (6 Punkte)

Was ist der grundsätzliche Unterschied von BLAST und PSI-BLAST?

Aufgabe 5 (10 Punkte)

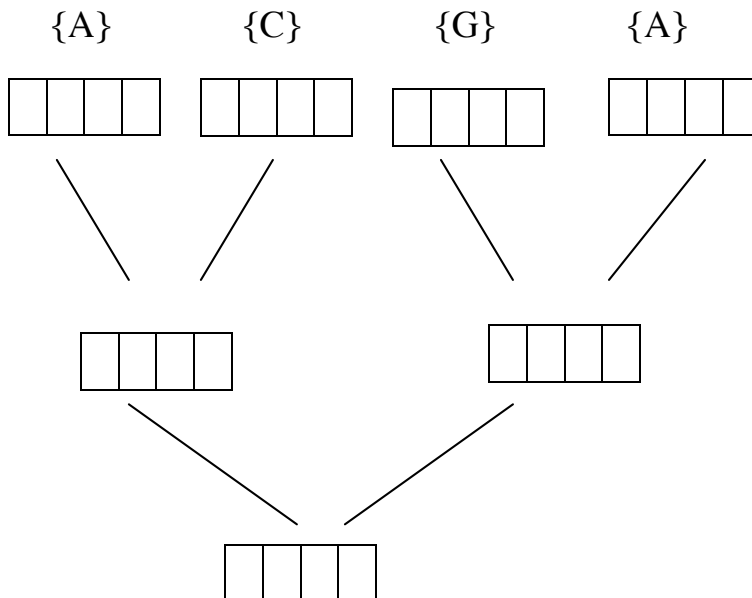
(a) Der Sankoff-Algorithmus für die Erstellung von Phylogenien verwendet folgende Formel zur Berechnung der evolutionären Kosten für einen Vorfahren-Knoten a :

$$S_a(i) = \min_j [c_{ij} + S_l(j)] + \min_k [c_{ik} + S_r(k)]$$

Benutzen Sie diese Formel und die symmetrische Kostenmatrix

	A	C	G	T
A	0	3	1	3
C	3	0	3	1
G	1	3	0	3
T	3	1	3	0

um die fehlenden Werte in dem Baum (unten) zu ergänzen.
(7 Punkte)



(b) Spekulieren Sie, weshalb man die evolutionären Kosten für Basenaustausche symmetrisch ansetzt (A → C erhält dieselben Kosten wie C → A) (3 Punkte)

Bereich II Proteinstruktur

Aufgabe 6 (8 Punkte)

(a) Was versteht man unter Protein-Threading? (4 Punkte)

(b) Wozu benutzt man Profil-Profil-Alignment? (4 Punkte)

Aufgabe 7 (8 Punkte)

(a) Skizzieren Sie die Twilight-Zone, die den Zusammenhang zwischen Proteinsequenz und -struktur beschreibt. Achsen beschriften! (6 Punkte)

(b) Wie verläuft die Trennlinie, wenn Sie statt Sequenzidentität nur Sequenz-ähnlichkeit fordern? (2 Punkte)

Aufgabe 8 (14 Punkte)

Das Bild zeigt einen Ausschnitt des Komplexes von Galanthamin, einem Inhibitormolekül (hellgrün gezeichnet) mit dem Protein Acetylcholinesterase, wobei der Inhibitor in der Bindungstasche des Proteins gebunden ist.

(a) Was könnten die dicken, roten Bälle sein? (2 Punkte)?

(b) Umkreisen Sie in dem Bild je ein/eine
- Glutaminsäure
- Tryptophan
- Phenylalanin
des Proteins und beschriften Sie deren Seitenketten. (6 Punkte)

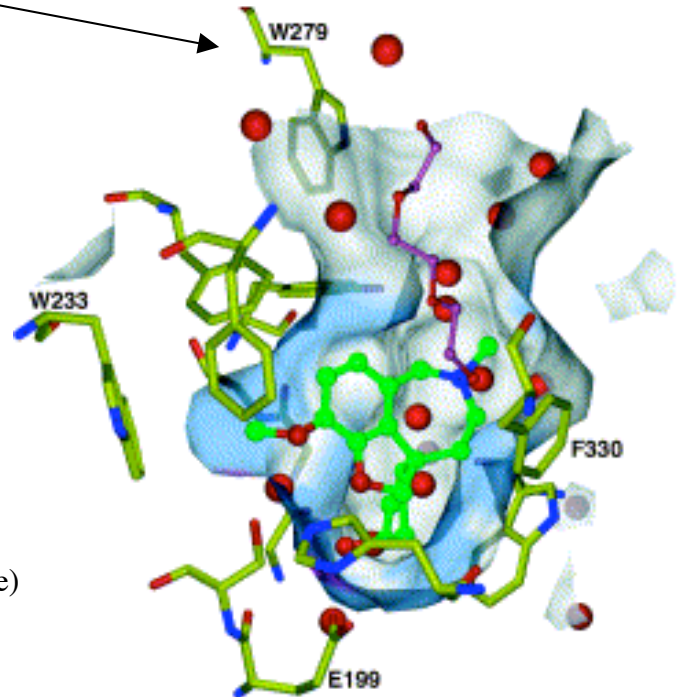
(c) Was für Wechselwirkungen können folgende Aminosäuren ausbilden? (6 Punkte)

- Glutaminsäure:

- Tryptophan:

- Phenylalanin:

(d) Wenn Sie in dem Bild zusätzlich das versteckte Histidin kennzeichnen können, erhalten Sie dafür 2 Bonuspunkte.



Bereich III Netzwerke

Aufgabe 9 Genexpression/funktionelle Annotation (10 Punkte)

(a) Was ist die Gen-Ontologie (GO)? (3 Punkte)

(b) Der hypergeometrische Test benutzt folgende Formel:

$$\sum_{i=k_{\pi}}^{\min(n, K_{\pi})} \frac{\binom{K_{\pi}}{i} \binom{N-K_{\pi}}{n-i}}{\binom{N}{n}}$$

Sei $N = 6$ die Anzahl an Genen im Genom,

$n = 4$ die Anzahl an Genen in der Testmenge.

$K_{\pi} = 4$ die Anzahl an Genen im Genom mit der Annotation π .

$k_{\pi} = 4$ die Anzahl an Genen in der Testmenge mit der Annotation π .

Berechnen Sie mit obiger Formel den p-Wert. (5 Punkte)

Beantworten Sie damit die Frage, ob die Annotation π in der Testmenge von 4 Genen signifikant angereichert ist? (2 Punkte)

Aufgabe 10 Diffusionsgleichung (10 Punkte)

(a) Die Kontinuitätsgleichung lautet:

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -\nabla \cdot \vec{j}(\vec{r}, t) = -\text{div } \vec{j}(\vec{r}, t)$$

Was wird hier mit ρ bezeichnet und was mit j ? (2 Punkte)

ρ :

j :

(b) Der Diffusionsstrom aufgrund eines Dichteunterschiedes (Gradienten) lautet gemäß dem Fick'schen Gesetz:

$$\vec{j}(\vec{r}, t) = -D \nabla \rho(\vec{r}, t) = -D \text{ grad } \rho(\vec{r}, t)$$

Wie kann man damit die Diffusionsgleichung

$$\frac{\partial \rho(\vec{r}, t)}{\partial t} = -\nabla \cdot (-D \nabla \rho(\vec{r}, t)) = D \Delta \rho(\vec{r}, t)$$

herleiten? (3 Punkte)

(c) Beschreiben Sie einen Algorithmus, wie man den zeitlichen Verlauf eines Diffusionsprozesses auf einem räumlichen Gitter simulieren kann (5 Punkte).

Aufgabe 11 Michaelis-Menten Modell (10 Punkte)

(a) Wie lautet die Gleichung für den Umsatz V als Funktion der Substratkonzentration S und der Michaelis-Menten-Konstanten K_M ? (3 Punkte).

(b) Bei welcher Substratkonzentration S erhält man die Hälfte des maximalen Umsatzes V_{\max} ? (2 Punkte)

(c) Angenommen, die Substratkonzentration S sei zeitlich konstant und zum Zeitpunkt $t = 0$ sei die Konzentration des Enzym-Substrat Komplexes $ES = 0$.

Zeichnen Sie einen Plot, in dem auf der x-Achse die Zeit aufgetragen ist und auf der y-Achse die Menge des gebildeten Produkts.

Tragen Sie im Plot ein, welche Menge an Produkt zum einen gemäß der Michaelis-Menten-Gleichung und zum anderen in einer einzigen stochastischen Simulation erhalten wird. (3 Punkte)

(d) Wie verläuft der Plot, wenn Sie den Mittelwert aus 1000 stochastischen Simulationen eintragen? (2 Punkte)