

Advanced Algorithms in Bioinformatics (P4)

Sequence and Structure Analysis

Summer 2007

Review 2

Aufgabe 1: Mutual Information Content 8+9 Punkte

Berechnen Sie den *mutual information content* der beiden folgenden Spalten:

C G

C A Berechnen Sie auch M_{ij} , die Summe der Höhe der

A U beiden 'M's, à la Gorodkin.

A A

Aufgabe 2: Grammatiken 6+6 Punkte

Beweisen oder widerlegen Sie die folgende Aussagen:

- a) Eine reguläre Grammatik kann überhaupt keine Palindrome generieren.
- b) Eine kontext-freie Grammatik ist eine reguläre Grammatik.

Aufgabe 3: Alignments und *traces* 5+5 Punkte

Beschreiben Sie, was eine *trace* ist.

Gegeben zwei Sequenzen, gibt es mehr verschiedene Alignments zwischen diesen Sequenzen, mehr verschiedene *traces* oder gleich viele? Begründen Sie Ihre Antwort.

Aufgabe 4: Motif finding 5+5 Punkte

- a) Let $T[1 \dots 3]$ be a random sequence where each $T[i]$ is chosen uniformly at random from the nucleotide alphabet $\{A, C, G, T\}$, and the choice of $T[i]$ is independent from all other $T[j]$, where $j \neq i$. Compute the probability that $T[1 \dots 3]$ contains $P = AA$ without substitutions.
- b) Assume that n objects are assigned uniformly at random and independently to k buckets. What is the expected number of empty buckets? Give a closed formula in terms of n and k .