

2008 Paper 12 Question 1

Bioinformatics

- (a) Parameters of the positional independence of a transcription factor binding site were estimated by the experimental positional nucleotide frequencies shown in the following table:

$$\begin{pmatrix} & 1 & 2 & 3 & 4 & 5 & 6 \\ T & 0.16 & 0.05 & 0.01 & 0.03 & 0.12 & 0.14 \\ C & 0.08 & 0.04 & 0.01 & 0.03 & 0.05 & 0.11 \\ A & 0.68 & 0.11 & 0.02 & 0.90 & 0.16 & 0.51 \\ G & 0.08 & 0.80 & 0.96 & 0.04 & 0.67 & 0.24 \end{pmatrix}$$

Explain what a *logo* is and determine the parameters of the logo graph. Compute the information content of one column. [8 marks]

- (b) Discuss the complexity of the algorithm for finding a global alignment between two DNA sequences that have a high degree of similarity. Present an example and analyse it using the following scoring parameters: +1 for match, -1 for mismatch, and $d = -1$ for a linear gap penalty. [7 marks]
- (c) In modelling a metabolic process, describe the advantages and disadvantages of using a stochastic approach (for example agents) as opposed to using a set of deterministic differential equations. [5 marks]