COMPUTER SCIENCE TRIPOS Part II – 2024 – Paper 8

2 Bioinformatics (pl219)

(*a*) Discuss how to use bioinformatics algorithms to detect the specific pathogenic sequences in the genome of a pathogenic species, by comparing its genome with the genome of an evolutionarily-close non-pathogenic species.

Hint. Most pathogenic bacteria have long DNA sequences containing disease-causing genes that are not present in the genome of similar non-pathogenic species. Consider how to detect extra material, and perhaps inverted repeats (which are usually formed during the insertion of the disease-causing genes.)

[5 marks]

(*b*) Compute the global alignment and the best score of the sequences {CGTGT, TGGCGCC} with the following parameters: match score = +2, mismatch score = −1, gap penalty = −2. Report the final score and alignment(s). [4 marks]

(*c*) Dimerisation occurs when two similar proteins (P) join together to form a dimer (D), and dissociation reverses this process. The Gillespie algorithm may be applied to model dimerisation and dissociation of proteins, with species P and D, and the following rate constants:

- Dimerisation: $2\text{P} \rightarrow \text{D}$ with rate c_1
- Dissociation: $\text{D} \rightarrow 2\text{P}$ with rate c_2

Dimerisation is rare and dimers are unstable, therefore $c_2 \gg c_1$. Explain how to use the Gillespie algorithm to model dimerisation and dissociation reactions.

[7 marks]

(*d*) Discuss advantages and disadvantages of using DNA to store information.

[4 marks]