Answer THREE questions.

The numbers in square brackets in the right-hand margin indicate the provisional allocation of marks for each sub-section of a question.

$$k_B = 1.38 \times 10^{-23} m^2 kg\ s^{-2} T^{-1}$$

$$\ln n! \approx n \ln n - n + \frac{1}{2} \ln(2\pi n)$$

1. Cooperative binding in Hemoglobin

(a) Give a full description of the structure of the hemoglobin molecule, indicating the nature of its secondary structure and the nature and location of its heme groups. Indicate where the oxygen molecules bind and by what mechanism they can interact with each other. [5]

(b) Explain the meaning of "cooperativity" in the binding of oxygen to hemoglobin. You may wish to explain this by including diagrams of the binding probability as a function of oxygen partial pressure. What is the difference between the Langmuir isotherm and Hill Function used to describe the binding? How is cooperative binding useful in carrying oxygen from the lungs to the muscles of higher animals. [5]

(c) A 2-site abstraction of hemoglobin, *H*, binds two oxygen molecules, *O*, at the same time according to the reaction,

$$O + O + H \longleftrightarrow O_2H$$

with equilibrium constant K. Show that the probability of the bound state is given by the expression

$$p_{bound} = \frac{([O]/K)^2}{1 + ([O]/K)^2}$$

where *[O]* denotes the concentration. [5]

(d) The Shannon entropy $S = -\sum_i p_i \ln p_i$ measures the "missing information" of a set of microscopic states of probability p_i. Calculate the Shannon entropy of a Hemoglobin molecule as a function of oxygen concentration *[O]*. Sketch how this entropy varies with *[O]*, paying attention to the limits of *[O]* both going to zero and infinity. [5]

2. Nucleosome binding.

(a) Describe the operating principles of "chromosome conformation capture" methods (3C, 5C, Hi-C etc). How can information about the internal organization of chromosomes be determined from Hi-C data? How would you expect this to be different between the interphase and metaphase points of the cell cycle? [5]

(b) Describe the role of nucleosomes in eukaryotic chromosomes. What are the key structural features of nucleosomes that favour their binding to the chromosome? How many nucleosomes are present in a typical chromosome? How specific is the binding and how does this allow transcription and replication to take place? [5]

(c) Show that the probability of a random walk of N steps, randomly pointing to the left or right, returning to the origin is given by:

$$p_0 = \frac{1}{2^N} \frac{N!}{(N/2)!(N/2)!}$$

Hence, using standard approximations, obtain a simplified expression for p_0 in the limit $N >> 1$. [5]

(d) Hi-C data are often presented as probability distributions of crosslinks between pairs of genetic loci separated by a distance N base pairs. Under what circumstances does this distribution follow the form of your answer to part (c)? Show how you could distinguish between two models of chromosome structure, one built as loops extending from a linear protein scaffold, and one in which the DNA forms an unconstrained polymer coil. [5]

3. Separation by Diffusion

(a) Describe two ways how a centrifuge can be used to separate macromolecules in solution. For each method, are the particles are separated according to their size, mass or density? Explain how thermally generated diffusion can affect the result of a centrifugation experiment. Is a centrifuge more effective at separating large macromolecular complexes, such as ribosomes, or smaller proteins, such as enzymes? [5]

(b) Defining your symbols, write down an equation for the net flow across an area A inside a solution of variable concentration *c(x)* due to diffusion and the presence of a thermal driving force $F = -dU/dx$ in the low-Reynolds number limit (viscous forces only). In thermal equilibrium, when the potential *U(x)* follows a Boltzmann distribution, show how this leads to the famous Einstein relation. State what additional assumptions are needed to give the Stokes-Einstein formula for the diffusion constant, *D*, of a spherical particle of radius *R*: [6]

$$D = \frac{k_B T}{6\pi\eta R}$$

(c) Show that the Gaussian function:

$$c(x,t) = \frac{c_0}{\sqrt{4\pi D t}} \exp\left(\frac{-x^2}{4Dt}\right)$$

is a solution of the 1D diffusion equation, [5]

$$\frac{\partial c(x,t)}{\partial t} = D\frac{\partial^2 c(x,t)}{\partial x^2}$$

(d) We wish to use diffusion in a long, narrow, water-filled tube of a microfluidics device to separate a solution containing equal concentrations of two proteins of size $R_1 = 2nm$ and $R_2 = 4nm$. At time $t=0$, a drop of the solution is placed at one end of the tube. Determine how long we have to wait for the faster-diffusing protein to become *10x* more concentrated than the slower one at a position $L = 1mm$ along the tube. Viscosity of water = 10^{-3} *kg / ms*. [4]

4. 1-state and 2-state motors

(a) Give an account of the principles of operation of biological motor systems. Provide details (with sketches) of three distinct examples of energy-consuming biological motors. What features are common to all these motors and what differentiates them? [6]

(b) In physical models of biological motor systems what is the distinction between single-state, 2-state and N-state models? How is the energy supply typically coupled to the different steps? Give a detailed example of an energy diagram describing a 2-state motor. [4]

(c) A 1-state model of a linear motor consists of binding of a single molecular complex to a series of identical receptor sites along a track. The complex is only allowed to move between adjacent sites. Assuming thermodynamic equilibrium between the complex and the binding sites, show that the distribution of waiting times at each site is given by the probability distribution,

$$p(t) = \frac{1}{\tau_0} e^{-t/\tau_0}$$

To what quantity does the constant τ_0 correspond? [4]

(d) This model is now generalized to one with 2-states of the complex, A and B, with different conformations and energies, whose sequence is linked to the order of the receptors along the track:

$$[n]_A \longrightarrow [n]_B \longrightarrow [n+1]_A \longrightarrow [n+1]_B \longrightarrow etc$$

If the two states each have probability distributions given by the expression above in part (c) with constants τ_A and τ_B, derive the combined probability distribution of the complex at each position along the track. Sketch this distribution and that for p(t) above. Describe how 1-state and 2-state motors can be distinguished experimentally. [6]

5. Lac repressor energetics

(a) Describe how X-ray crystallography can be used to determine the atomic level structure of proteins: how proteins can be made into crystals; how they can be measured with X-rays and why X-rays are necessary. What are the main features of protein structures that are revealed by X-ray crystallography? [5]

(b) The Lac repressor binds to two sites O1 and O2, separated by 401 base pairs (bp), located just in front of the beta-galactosidase gene of *E. coli*. Describe how the structure of the lac repressor protein (and the way that it binds to DNA) is capable of switching the expression of the gene. What does the distance of 401bp suggest about the structure of the protein and/or the DNA? [5]

(c) If DNA is modeled as a continuum elastic cylinder of solid material of radius a and Young's modulus E, show that the elastic energy required to bend it into a circular loop of radius R is given by: [5]

$$\epsilon_{Loop} = \pi^2 \frac{Ea^4}{4R}$$

(d) From the 3D random walk model, it is known that the probability of a polymer of $N >> 1$ segments returning to its starting point is proportional to $N^{-3/2}$. From this and the equation above, derive an expression for the free energy ΔG of the formation of a loop of DNA and deduce its most likely length, L_0. Estimate a value of L_0 for DNA using $a=1nm$, $E=3x10^8$ Pa and the persistence length of DNA $\approx 50nm$. [5]