Imperial College London

BSc, MSci and MSc EXAMINATIONS (MATHEMATICS)

May-June 2016

This paper is also taken for the relevant examination for the Associateship of the Royal College of Science

Mathematical Biology of the Cell

Date: **Thursday 26th May 2016**

Time: **09.30 – 11.30**

Time Allowed: **2 Hours**

This paper has Four Questions.

Candidates should use ONE main answer book.

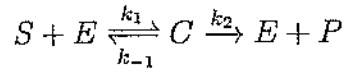Supplementary books may only be used after the relevant main book(s) are full.

Statistical tables will not be provided.

- DO NOT OPEN THIS PAPER UNTIL THE INVIGILATOR TELLS YOU TO.

- Affix one of the labels provided to each answer book that you use, but DO NOT USE THE LABEL WITH YOUR NAME ON IT.

- Credit will be given for all questions attempted, but extra credit will be given for complete or nearly complete answers to each question as per the table below.

Raw Mark	Up to 12	13	14	15	16	17	18	19	20
Extra Credit	0	½	1	1 ½	2	2 ½	3	3 ½	4

- Each question carries equal weight.

- Calculators may not be used.

1. (a) An enzyme E acts on a substrate S and produces product P as described by the following

$$S + E \underset{k_{-1}}{\overset{k_1}{\rightleftharpoons}} C \overset{k_2}{\longrightarrow} E + P$$

Use mass action kinetics to write a complete set of ODEs for S, E, C and P.

(b) Making the quasi-steady state assumption for the complex C ($\frac{dC}{dt} = 0$), and also using the constraint $E + C = E_0$, show that the Michaelis-Menten kinetics is obtained:

$$\frac{dP}{dt} = \frac{vS}{K_m + S}$$

Obtain expressions for v and K_m in terms of k_1, k_{-1}, k_2 and E_0.

Dynamics of a protein kinase cascade. Protein Kinases $X_1, X_2, ..., X_n$ act in a signalling cascade, such that X_1 phosphorylates X_2, which, when phosphorylated gets activated and acts to phosphorylate X_3. X_3 when phosphorylated gets activated and acts to phosphorylate X_4 and so on. There is a common phosphatase that can dephosphorylate all levels of the cascade. Consider the case of zero-order kinetics, when the activated upstream kinase is found in much smaller concentrations than its unphosphorylated target. Michaelis-Menten enzyme kinetics suggests that in zero-order kinetics, the rate of phosphorylation depends on the upstream kinase concentration and not on the concentration of its substrate. However, since levels of phosphorylated substrate is small compare to the phosphatase levels, phosphorylation acts as a first-order reaction. Therefore , the dynamics can be described by a linear set of equations:

$$dX_i/dt = v_{i-1}X_{i-1} - \alpha_i X_i, \quad i \geq 2$$

Where X_i is the phosphorylated level of kinase i and v_i and α_i are phosphorylation and dephosphorylation rates, respectively.

(c) The signal amplitude is defined by $A_i = \int_0^\infty X_i(t)dt$. If the cascade is stimulated by a pulse of X_1 activity with amplitude A_1, what is the amplitude of the final stage of the cascade, A_n? Show that if $v_{i-1} > \alpha_i$ the signal is amplified. You can assume that $X_i(0) = X_i(\infty) = 0$.

(d) The signal duration is defined by $\tau_i = \int_0^\infty tX_i(t)dt/A_i$. Assuming that the input pulse is short ($\tau_1 = 0$), show that

$$\tau_n = \sum_{i=2}^{n} \alpha_i^{-1}.$$

2. (a) What is the main difference between prokaryotic cells (like bacteria) and eukaryotic cells (like yeast and mammalian cells)? It takes 0.01 seconds on average for a protein to diffuse from one side of bacteria E coli to the other side. The diameter of yeast is about 5 times E coli. Assuming similar diffusion coefficients in E coli and yeast how long does it take for a protein to traverse a yeast cell?

(b) Entropy of a closed system is defined as $S = k_B \ln W$, where W is the number of microstates of the system, and k_B is the Boltzman constant. A piece of DNA has N binding cites for proteins and a c fraction of it is occupied, show that the entropy of this system follows (you might need to use Sterling's approximation):

$$S = -k_B N[c \ln c + (1 - c) \ln(1 - c)]$$

For what values of c, entropy is maximised. What is the intuition behind this?

(c) The protein native state is the unique fold it takes in physiological conditions and it is closely related to the function of the protein. What is the entropy of the native state? An unfolded protein can be modelled as a self avoiding walk (SAW). It is conjectured that the number of SAWs of L-steps is $\mu^L L^{\frac{11}{32}}$, where μ is a constant that depends on the dimension of the SAW and the underlying lattice type. What is the entropy of an unfolded protein of length L?

(d) Suppose the energy of the protein native state is E_N and does not depend on L and the energy of the unfolded protein is $E_U = eL$ where e is a constant. By comparing free energies of native and unfolded state derive an expression for T_c the temperature above which the protein native state is unstable (Free energy F is defined as $F = E - TS$). How does T_c changes as L increases (assume $E_N \ll E_U$). How can you explain this effect?

3. (a) The random graph model of Erdos-Renyi (ER) has N nodes, and every pair of nodes are connected with probability p. What is the average number of edges? The degree of a node k is defined by the number of edges that is connected to that node. What is the average degree $\langle k \rangle$ in the ER model? Show that for large N and finite $\langle k \rangle$, the degree distribution $P(k)$ for the ER model is a Poisson Distribution.

Limiting cases of the Barabasi-Albert network model. The power-law scaling of the degree distribution in the Barabasi-Albert (BA) model indicates that growth and preferential attachment play an important role. In this problem we seek to prove both of these are necessary for the emergence of power-law scaling by looking at models that contain only one of these mechanisms.

(b) *Model A* keeps the growing character of the network without preferential attachment. Starting with a small number of nodes m_0, at every time step we add a new node with $m(\leq m_0)$ edges, so the number of nodes at the end of time step t is $m_0 + t$. We assume that the new node connects with equal probability to the nodes already present in the system, i.e. $\Pi(k_i) = 1/(m_0 + t - 1)$, independent of k_i. Use the continuum theory developed for the BA model to show for $t \to \infty$ the degree distribution is:

$$P(k) = \frac{1}{m} e^{1 - \frac{k}{m}}$$

Hint: In continuum theory we wrote an ODE for $k_i(t)$ assuming k_i is a continuous variable. To calculate degree distribution we used the identity
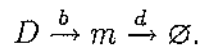
$$P(k) = \partial P(k_i(t) < k)/\partial k.$$

(c) *Model B* starts with N nodes and no edges. At each time step a node is selected randomly and connected with probability $\Pi(k_i) = k_i / \sum_j k_j$ to a node i in the system. Consequently, model B eliminates the growth process, the number of nodes being kept constant. Show that after t time steps, we have $\sum_i k_i = 2t$. Then assuming N is large, use the continuum theory to show that:

$$k_i = \frac{2t}{N},$$

So that, it is independent of t_i. (Hint: the increase in degree of node i has two contributions from random selection of node i (with probabilities $1/N$) and preferential attachment with probability $\Pi(k_i)$.)

4. (a) The simplest model of transcription includes mRNA production and decay:

$$D \xrightarrow{b} m \xrightarrow{d} \varnothing.$$

Write down the master equation describing the evolution of $P_m(t)$.

(b) At steady-state, write down P_{m+1} in terms of P_m and P_{m-1}. Show that P_m follows a Poisson distribution. [Note that P_m for $m < 0$ is by definition zero.]

(c) The generating function is defined as

$$F(z,t) = \sum_{m=0}^{\infty} z^m P_m(t).$$

Use the master equation above to obtain a PDE describing the evolution of $F(z,t)$.

(d) Use the first derivative of the generating function with respect to z at $z = 1$ to obtain an ODE for the mean mRNA. How does this equation compare to the deterministic ODE for the system? Assuming at $t = 0$ there is no mRNA, use the second derivative of the generating function with respect to z at $z = 1$ to find an expression for mRNA noise (coefficient of variation) as a function of time. Show that noise is proportional to the inverse root square of the average mRNA numbers for all times.