PROBLEM SET 4 B1 Flows and Complexity

1) Biological background

Brief answers are sufficient:

a) What are **DNA** and **RNA;** what are they primarily used for in the cell?

b) What is the (not brilliantly named) ``**Central Dogma**'' of modern biology? What functions do the molecular machines **RNA polymerase** (RNAp) and the **Ribosome** perform?

c) How many gigabytes of information could one encode in human DNA?

d) What are **proteins** & what do they do in cells? How many different kinds of amino acids are they made of? How are they encoded in DNA? What is meant by a protein's **tertiary** and **quaternary** structure?

e) What are some of the differences between **eukaryotic** and **prokaryotic** cells? Use the difference between passive and active transport to explain why they differ in size.

f) [FOR FUN]: If a single **amino acid** weighs on average 110 Dalton (where one Dalton is the mass of an H atom, 1 amu), how long a strand of protein do you need before the set of all possible sequences would weigh more than the mass of the observable universe (estimated to be about 10^{53} kg)? The average length of proteins in our body is about length 479. How many times the mass of the observable universe would the set of all possible sequences weigh?

g) [FOR FUN] To get a sense of evolutionary time scales, it is useful to draw an analogy with human time scales. Let's say it is just about to turn midnight on New Year's eve, and the earth was first formed on 1 Jan (4.5 billion years ago). Roughly when in the past year would: 1) Life first emerge (3.8 billion years ago)? 2) Eukaryotes first emerge (1.85 billion years ago)? 3) Animal life emerge (500 million years ago)? 4) Dinosaurs go extinct (65 million years ago)? 5) The first anatomically modern Homo Sapiens appear (200,000 years ago)? 6) Recorded human history? 7) First World War? 8) Birth of a typical undergraduate? (question adapted from Jack Miller, Daniel Fisher & PBOC)

2) Random walks and diffusion

Consider a random walk in one dimension along the x-axis. With probability p the walker takes a single step of size δ to the right, and with probability q it takes a single step of size δ to the left, with p+q=1.

a) Work out the probability that the walker takes n+ steps to the right, given that it took a total of n steps.

b) For a symmetric walker where p=q, work out an expression for the mean square displacement $\langle x^2 \rangle$ where $\langle\ \rangle$ denotes an average over many realisations of the random walk. What is the diffusion coefficient for this random walk?

c) Use the handout from book Random Walks in Biology by Howard Berg, which describes how the binomial distribution reduces to a Gaussian in a certain limit, to show under what conditions the expressions worked out in a) and b) reduce to the fundamental solution to the diffusion equation in free space:

$$c(\vec{r},t) = \frac{1}{(4\pi D t)^{d/2}} e^{-\frac{r^2}{4Dt}} \qquad d = 1, 2, 3$$

3.) **Freely jointed chains**

An ideal freely-jointed chain (FJC) consists of N rigid segments of length b, freely hinged where they join. Possible consequences of interference between different parts of the chain are neglected.

(a) Using the links to a random walk, derive an expression for $\langle R^2 \rangle$, the mean square end-to-end distance for a FJC.

(b) The corresponding result for a semi-flexible wormlike chain (WLC), which models a polymer as a continuous filament with a non-zero bending modulus, is

$$\left\langle R^2 \right\rangle = 2\xi_p^2 \left(e^{-L/\xi_p} - 1 + \frac{L}{\xi_p} \right)$$

where L is the contour length and ξ_P is the persistence length. Explain what the concept of persistence length means (a sketch may be useful). What is the relationship between persistence length and the Kuhn length? Evaluate this expression in the limits $L \ll \xi_P$ and $L \gg \xi_P$ (show your working) and comment on both results.

(c) What is $R_E = (\langle R^2 \rangle)^{1/2}$ for the human genome and for the genome of E. coli? (the persistence length of DNA is 50 nm, or about 150 base pairs. How does this compare to the typical size of a nucleus in a human cell or the typical size of an E. coli cell? Calculate the volume taken up by human DNA at close-packing. What does this tell us about the organisation of DNA in eukaryotic cells? Comment on how your result for the R_E of an E. coli genome relates to the image below (made by the legendary Ruth Kavenoff). The bacterial wall was broken (by lysis) to expose the DNA.

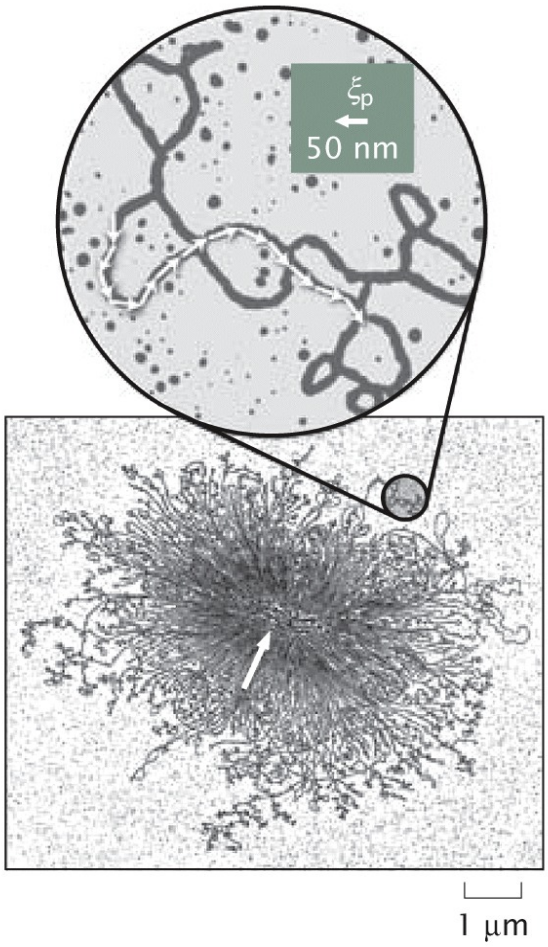


Figure 8.5 Physical Biology of the Cell, 2ed. (© Garland Science 2013)

(d) Write out an explicit expression for P(**R**,N), the end-to-end probability distribution function for a FJC and and compare it to the probability distribution of a particle that undergoes a random walk. In particular, what is the relationship between: i) the time t for the particle and the number of segments N for the FJC and ii) the particle diffusion coefficient and its analogue in the FJC?

(e) Write down an expression for the partition function of the FJC in the case that a force f is applied between the ends of the chain, and in the z direction. Show that in the low-force limit the chain behaves as a linear spring. How does the spring constant vary with temperature? Explain the physical origin of this temperature dependence.
Now show that in the large-force limit:

$$F = \frac{k_B T}{b} \frac{1}{(1 - z/L)}$$

and comment on your results.

(f) In the class we also derived a linear spring constant directly from the (unperturbed) probability distribution function P(**R**,N). Compare this expression to the low-force and high force limits you just derived, and comment on your results. In particular why does the high-force limit differ, i.e. what goes wrong with the derivation from P(**R**,N) ?

(f) Calculate the cyclization probability (probability that ends fuse together) of a FJC by working out the probability that the two ends of the chain are within a distance $\delta << R_E$. Comment on your results.
[HINT: You can assume that P(**R**,N) is a constant for $|\mathbf{R}| \leq \delta$]

4**) Noise in gene expression**

(a) A gene is transcribed to mRNA molecules, with a concentration R, that are in turn translated to proteins, with a concentration P. The concentrations can be described by the Langevin equations:

$$\frac{dR}{dt} = k_R - \gamma_R R + \eta_R(t)$$

$$\frac{dP}{dt} = k_p R - \gamma_P P + \eta_P(t)$$

$$<\eta_i(t)>= 0; \quad <\eta_i(t)\eta_j(t-t')>= q_i\delta(t')\delta_{ij}$$

where the γ_R and γ_P are the degradation rates, and k_R and k_P are production rates and η_R and η_P are the noise terms. for the RNA and proteins respectively. What are the typical time-scales for the degradation rates of RNA and proteins and why are they different?
(FOR FUN: can you think of an equivalent particle system that would obey the same system of Langevin equations?)

Derive the steady-state concentration of protein, P and comment on how it varies with degradation rates and production rates. The scientific literature on this problem often defines an average burst size $b = k_P/\gamma_R$. What does this dimensionless number describe?

(b)In a classic paper, E. M. Ozbudak, I. Kurtser, A.D. Grossman and A. van Oudenaarden, Nature Genet. **31**, 69 (2002), showed that the noise strength, measured by the variance σ_P^2 in protein concentration P, takes the form: $\sigma_P^2 = <P>(1 + b)$ where the Fano factor F= 1+b measures the deviation from Poissonian behaviour. In the Appendix of this paper they sketch a derivation of this result. (The paper is available on the course website and well worth reading in full.) Go through their derivation equation by equation and explain what assumptions they make in each step. (note that there is a small error – for you to spot …). In particular, show explicitly how to derive each step up to the equation for the steady-state value of the fluctuations in RNA concentration, and explain how the strength of the noise was fixed through a fluctuation-dissipation relation.

To increase physical insight into their results, first briefly describe what a Poisson process is and how its fluctuations scale. Explain qualitatively why the fluctuations here are non-Poissonian and give a semi-quantitative argument for the origin of the factor b. [FOR FUN: Can you think of other systems that should exhibit non-Poissonian noise?]

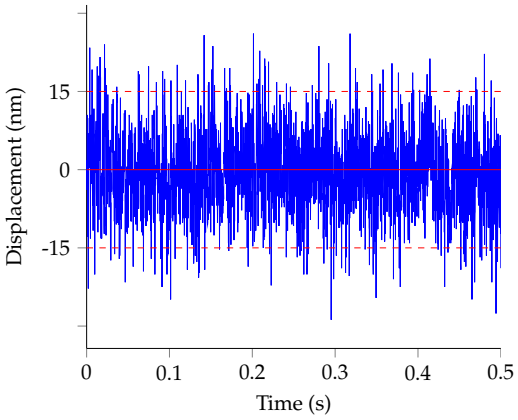
Describe the difference between intrinsic and extrinsic noise in gene expression. How would you measure this in a cell? What kind of noise are Ozbudak et al., describing?

(c) **More challenging:** Use Fourier transform techniques to derive the full expression for the fluctuations in protein concentration from the Langevin equations above, that is explicitly fill in all the missing steps in the derivation by Ozbudak et al.. Hint: you may need the identity:

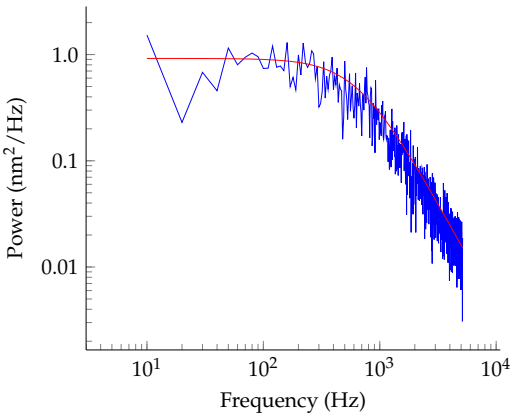
$$\int_{-\infty}^{\infty} \frac{1}{(2\pi)\left(a^2 + x^2\right)}\, dx = \frac{1}{2a}$$

5.) Statistical Mechanics of Optical tweezers.

Briefly describe how an optical tweezer set up works.



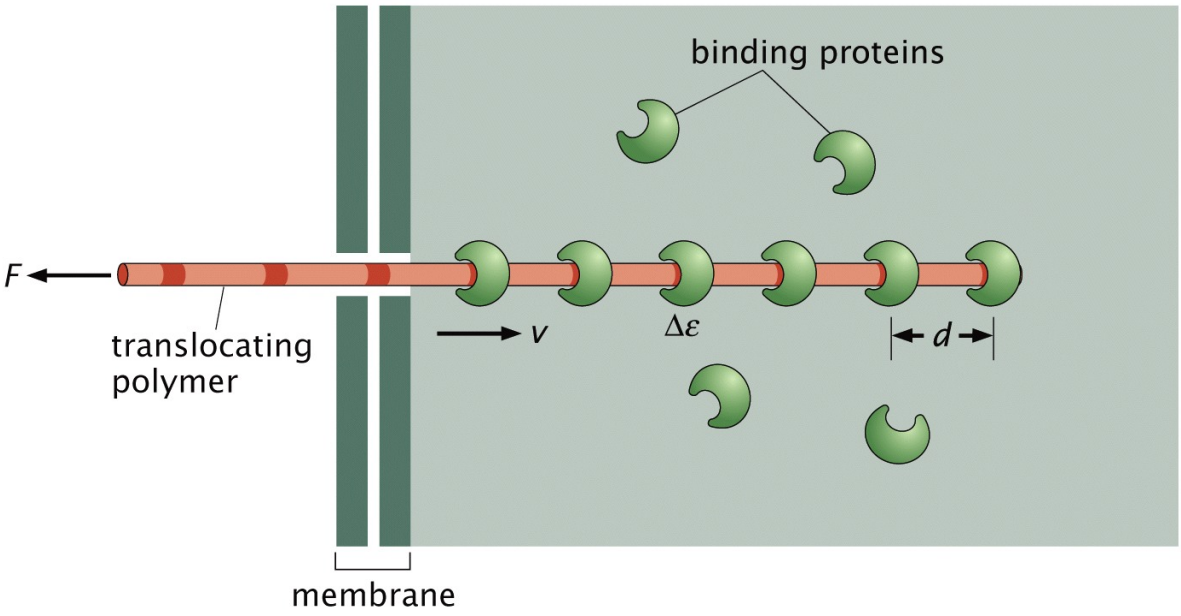
(a) Brownian motion



(b) Power spectrum of trapped bead

The motion of the bead can be described by a Langevin equation. Write this equation out for a bead of mass m in a trap of stiffness K, with a friction coefficient γ (you may assume that you are in the large friction limit so that the mass m can be neglected). The displacement x(t) is measured as a function of time (left plot - taken from optical tweezers practical), and Fourier transformed to x(ω). This, in turn, allows us to obtain the power spectrum (right plot) which is proportional to $|x(\omega)|^2$. The friction coefficient γ can be obtained independently by measuring the diffusion coefficient of the bead (give the relationship). Show how to measure the trap stiffness K by writing out $|x(\omega)|^2$ in terms of the "corner frequency" $f_c = K/2\pi\gamma$, and comparing to the right plot.

6) **Molecular Ratchet**



The translocating polymer of length L above has N = L/d binding sites (coloured red) spaced a distance d apart. On the inside of the membrane there are binding proteins that irreversibly bind to the translocating polymer, which locally (on lengths << than the persistence length, which in turn is much larger than the diameter of the proteins) can be treated as a rigid rod. There is also a force F acting against the translocation of the polymer.

(a) To analyse the molecular ratchet above, we introduce a probability p(x,t)dx of finding the last binding site to have crossed the pore at position (x, x+dx) at time t. Here x measures the distance from the pore. Show that the flux of protein binding sites (number coming through per unit time) J_x (t) = v(t)/d is then given by

$$J_x(t) = -\frac{DF}{k_B T} p(x,t) - D\frac{\partial p(x,t)}{\partial x}$$

where the first term describes the drift, and the second term the diffusion. This equation is subject to the boundary condition p(d,t)=0, because if the protein binding site reaches a distance d from the pore, then another binding site emerges from the pore and is now the new "last" site to emerge. The old "last" site effectively disappears at this point. Solve the differential equation above in steady state (for a constant J = v d).

(b) Next, use the normalisation condition

$$\int_0^d p(x)dx = 1$$

to show that the velocity as a function of force can be written as:

$$v = \frac{D}{d}\frac{w^2}{e^w - w - 1}$$

where w=Fd/k_B T is the dimensionless force. What happens for large forces? Take the limit F —> 0, and give a physical explanation for this form. How does it compare to diffusion without a ratchet mechanism?